

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give <i>P</i> values as exact values whenever suitable.</i>
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Raw data of SNP genotyping were collected on the 2.3M H3Africa SNP array at Illumina® FastTrack™ Microarray services (Illumina, San Diego, USA) and Illumina Human Omnichip 2.3 platform (Illumina Inc., San Diego, CA, USA) genotyping platforms. Summary statistics of USA-based cohorts for meta-analysis obtained from collaborators were generated using METAL (https://genome.sph.umich.edu/wiki/METAL_Documentation). High-throughput sequencing of edited cells: Illumina Miseq Sequencer System. HPLC data collection: Prominence HPLC System (Shimadzu Corporation). RT-qPCR expression data: QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems). Digital PCR: QIAcuity Digital PCR System (QIAGEN). mRNA sequencing data: NovaSeq X+ (Illumina). Flow cytometry: Attune NxT Flow Cytometer
Data analysis	GeneMAP hbf-gwas-scripts (https://zenodo.org/records/14607341) IAAP-CLI (v1.1) PLINK (v1.9 and v2) KING (v2.2.4) smartpca (EIGENSOFT v7.2.1) conform-gt (https://faculty.washington.edu/browning/conform-gt.html) Eagle (v.2.4.2) TOPMed Imputation Service: phasing, Eagle v.2.4; imputation, Minimac v4; reference panel release freeze 8 Mihigan Imputation Service: phasing, Eagle v.2.4; imputation Minimac v4; reference panels, CAAPA and 1000g-phase-3-v5 (hg19) Sanger Imputation Service: phasing, Eagle v2; imputation, PBWT (https://github.com/VertebrateResequencing/pbwt); reference panel, African

Genome Resource (AGR)
 BEAGLE (v5)
 Minimac (v3 and v4)
 FASTQC (v0.11.9)
 BWA-MEM (v0.7)
 SAMtools (v1.19)
 GATK (v4.2.5.0)
 DeepVariant (v1.3.0)
 GLNexus (1.4.3)
 SAIGE (v0.38)
 METAL (https://genome.sph.umich.edu/wiki/METAL_Documentation)
 LiftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>)
 SusieR (<https://stephenslab.github.io/susieR/>)
 LocusZoom (<http://locuszoom.org/>)
 UCSC Genome Browser (<https://genome.ucsc.edu/>)
 JASPAR (<https://jaspar.elixir.no/>)
 ENCODE (<https://www.encodeproject.org/>)
 VISTA Enhancer (<https://enhancer.lbl.gov/vista/>)
 ENSEMBL (<https://www.ensembl.org/index.html?redirect=no>)
 FUMA (v1.5.4)
 haploview (v4.2)
 RHE-mc (<https://github.com/sriramlab/RHE-mc>)
 R (<https://cran.r-project.org/>)
 ggplot2 (<https://ggplot2.tidyverse.org/>)
 R Limma package (v3.42.2)
 STAR (v2.7)
 RSEM (v1.31)
 CRIS.py (<https://github.com/patrickc01/CRIS.py>)
 QuantStudio 12K Flex Software V1.5 (Applied Biosystems).
 QIAcuity (<https://www.qiagen.com/us/products/instruments-and-automation/pcr-instruments/qiacuity-digital-pcr-system>).
 Attune Cytometric Software, FlowJo (<https://docs.flowjo.com/flowjo/workspaces-and-samples/flowjo-and-your-cytometer/ws-cytometer-attune/>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw data from Cameroon used in this study have been deposited in the dbGaP database under the accession code phs003748.v1.p1 []. Raw data from Tanzania are available from the EGA database under the accession code EGAS00001000990 [<https://ega-archive.org/search/EGAS00001000990>]. The processed supplementary data are provided in Source Data and Supplementary Data files.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex information was utilized as a covariate in the study. Sex information was self-reported and genotyped. Samples with a discordance in reported and genotyped sex were excluded. Ethical approval covered individual-level data. A total of 827 samples were analyzed from Cameroon after quality control, 50.8% were females. From Tanzania, 884 samples were analyzed, 52.8% were females. Only samples for which there was concordance between reported and genotyped sex were considered

Reporting on race, ethnicity, or other socially relevant groupings

No social categories were used in the study.

Population characteristics

The age of the participants was used as a covariate. Analysis was restricted to participants aged five years and older to match the age distribution in the USA-based cohort that was used for meta-analysis. For 827 samples from Cameroon after quality control, median age was 15 years (ranging from 5 to 66 years). From 884 samples from Tanzania after quality control, median age was 13 (ranging from 5 to 44 years).

Recruitment

The study participants of the primary cohort (Cameroon) were recruited from five cities in Cameroon (Yaoundé, Douala, Bafoussam, Bertoua, and Maroua) between May 2016 and July 2018. Structured questionnaires were used to collect socio-demographic and clinical events from adult sickle cell disease (SCD) patients or parents/guardians of younger patients. Clinical features for the past 3 years were extracted from patient medical record. Only patients older than 5 years of age,

who had not received a blood transfusion in the past 6 weeks were included. None was currently treated with hydroxycarbamide (hydroxyurea) or opioids. The sampling strategy was not restricted to hospital-based patients to avoid the bias that might result from including only the sickest patients. To accomplish this, two SCD patient groups in Cameroon were engaged for collaboration and additional patients were recruited during their monthly meetings. No incentives were provided for participation in the study. Raw data from Tanzania has been previously published (<https://doi.org/10.1371/journal.pone.0111464>). The data was accessed in EGA on approval using the accession EGAS00001000990

Ethics oversight

Approval was obtained from the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee, Cape Town, South Africa (HREC/REF: R015/2018), and National Ethical Committee of the Ministry of Public Health of Cameroon (No 193/CNE/SE/15).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. All available samples from the previous genetic epidemiological study were used.
Data exclusions	Samples with genotyping call rate less 90% were excluded. One individual from each pair of closely related individuals, up to second degree relationship, was excluded. Samples with outlying heterozygosity and missing data rate greater than 10% were excluded. Population outlying samples were excluded (outlier pruning). SNPs with minor allele frequency less than 1%, missing data more than 5%, and failed the Hardy-Weinberg equilibrium test at p-value 0.000005 were excluded. SNPs with imputation accuracy less than 60% and multi-allelic SNPs and INDELs were excluded.
Replication	HUDEP2 and CD34+ HSPCs experiments were repeated three times independently, as indicated in figure legends.
Randomization	Randomization was not relevant in the association test since it was a quantitative trait analysis and did not involve groups.
Blinding	Blinding was irrelevant since all procedures were based on objective quantitative methods.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-Human CD235a FITC, clone GA-R2 (HIR2), BD Pharmingen™ 559943 Anti-Human CD49d PE, clone 9F10, BioLegend 304304 Anti-Human Band3 APC, clone custom, New York Blood Center Gift from X. An Anti-Human Fetal Hemoglobin, clone HBF-1, ThermoFisher MHFH05
Validation	Anti-Human CD235a FITC, clone GA-R2 (HIR2), BD Pharmingen™ 559943 (1:100 dilution for FACS) Metais et al, Blood Adv, 2019 Anti-Human CD49d PE, clone 9F10, BioLegend 304304 (1:20 dilution for FACS) Validation: Metais et al, Blood Adv, 2019 Anti-Human Band3 APC, clone custom, New York Blood Center Gift from X. An (1:100 dilution for FACS) Validation: Metais et al, Blood

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.
Authentication	Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.
Mycoplasma contamination	Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cultured cells were resuspended in PBS with 0.1% BSA. Cells were fixed and penetrated for intracellular F cell staining.
Instrument	Attune NxT Flow Cytometer
Software	Attune Cytometric Software, FlowJo (https://docs.flowjo.com/flowjo/workspaces-and-samples/flowjo-and-your-cytometer/ws-cytometer-attune/)
Cell population abundance	A minimum of 10,000 live cells was collected for each sample for analysis.
Gating strategy	FSC-A/SSC-A for mononuclear cells, followed by FSC-H/SSC-W for singlets. For erythroid maturation, single cells that were positive for CD235a+ were then gated into quadrants to compare Band3 and CD49d expression For F-cell quantification, single cells that were positive for CD235a+ were then gated for HbF expression.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.