




Clinical and genetic ancestry profile of a large multi-centre sickle cell disease cohort in Brazil

Anna B. F. Carneiro-Proietti,¹ Shannon Kelly,^{2,3}  Carolina Miranda Teixeira,¹ Ester C. Sabino,⁴ Cecilia S. Alencar,^{4,*} Ligia Capuani,⁵ Tassila P. Salomon Silva,¹ Aderson Araujo,⁶ Paula Loureiro,⁶ Cláudia Máximo,⁷ Clarisse Lobo,⁷ Miriam V. Flor-Park,⁸ Daniela O. W. Rodrigues,¹ Rosimere A. Mota,¹ Thelma T. Gonzalez,² Carolyn Hoppe,³ João E. Ferreira,⁴ Mina Ozahata,⁴ Grier P. Page,⁹ Yuelong Guo,⁹ Liliana R. Preiss,⁹ Donald Brambilla,⁹ Michael P. Busch²  and Brian Custer,²  for the International Component of the NHLBI Recipient Epidemiology and Donor Evaluation Study (REDS-III)

¹Hemominas Foundation, Belo Horizonte, Brazil,

²Blood Systems Research Institute, San Francisco,

³BCHO – UCSF Benioff Children's Hospital

Oakland, Oakland, CA, USA, ⁴University of São

Paulo, ⁵Pró-Sangue Foundation, São Paulo,

⁶Hemope Foundation, Recife, ⁷Hemorio

Foundation, Rio de Janeiro, ⁸Instituto da

Criança, Hospital das Clínicas da Faculdade de

Medicina da Universidade de São Paulo, São

Paulo, Brazil, and ⁹RTI – Research Triangle

Institute International, Triangle Park, NC, USA

*Present address: Cecilia S. Alencar,

Laboratório de Investigação Médica LIM03,

Hospital das Clínicas da FMUSP, São Paulo,

Brazil

Received 11 December 2017; accepted for publication 14 April 2018

Correspondence: Dr Shannon Kelly,

Epidemiology, Blood Systems Research

Institute, 270 Masonic Ave, San Francisco, CA

94118, USA.

E-mail: skelly1@bloodsystems.org

Sickle cell disease (SCD) is the most prevalent clinically significant inherited blood disorder, with more than 13 million affected people worldwide (Weatherall, 2010; Piel *et al.*, 2017). SCD is caused by the sickle mutation (Glu6Val, β^S) in the beta

Summary

Approximately 3500 children with sickle cell disease (SCD) are born in Brazil each year, but the burden of SCD morbidity is not fully characterised. A large, multi-centre cohort was established to characterise clinical outcomes in the Brazilian SCD population and create the infrastructure to perform genotype-phenotype association studies. Eligible patients were randomly selected from participating sites and recruited at routine visits. A biorepository of blood samples was created and comprehensive demographic and clinical outcome data were entered in a centralized electronic database. Peripheral blood genome-wide single nucleotide polymorphism (SNP) genotyping was performed using a customized Transfusion Medicine (TM) Array. A total of 2795 participants at six Brazilian sites were enrolled between 2013 and 2015. The cohort included slight predominance of children <18 years (55.9%) and females (53.0%). Haemoglobin (Hb) SS was the most common SCD genotype (70.7%), followed by HbSC (23%), S β 0 (3.0%) and S β + (2.9%). SNP data from the TM Array were analysed to evaluate the genetic ancestry of the cohort and revealed significant admixture among the population. Demographics and clinical complications, stratified by age and SCD genotype, are summarized and future studies in this cohort are discussed.

Keywords: sickle cell disease, clinical aspects, single nucleotide polymorphisms.

globin gene (*HBB*), inherited in either the homozygous (HbSS) or compound heterozygous conditions, in which the sickle mutation is inherited with β -thalassaemia or additional abnormal haemoglobins, such as HbC, HbD and others. The

pathophysiology of SCD is complex. It involves a combination of vaso-occlusion, haemolysis, endothelial dysfunction and inflammation, and patients with SCD suffer severe complications of virtually every organ system (Kato *et al*, 2017). There is significant heterogeneity in clinical manifestations among patients with SCD, even within each SCD genotype (Saraf *et al*, 2014) and the ability to predict specific complications in individual patients and therefore tailor patient-specific screening and therapy is not currently possible.

In Brazil, the mutations that cause SCD were introduced during the colonial era after 1500 by the slave trade of people of African origin (Curtin, 1969). The gene mutations causing alpha and beta thalassaemia were introduced by European populations (Salzano & Bortolini, 2002; Wagner *et al*, 2010), which also mixed with the populations already present in Brazil, thus creating a genetically admixed population of SCD patients with varying heritage. Brazil currently has between 25 000 and 30 000 people with SCD and about 7 million carriers, leading to the birth of approximately 3500 infants with SCD each year, though the prevalence varies markedly across the country (Cançado & Jesus, 2007).

Despite clinical guidelines and advances in SCD treatment (see Clinical Context below) in Brazil, there is no national database or organized national network to comprehensively evaluate and characterize the impact of SCD in Brazil. The Recipient Epidemiology and Donor Evaluation (REDS)-III Brazil SCD Cohort was established to define the prevalence of clinical complications with a focus on transfusion outcomes in a large sample of the Brazilian SCD population and to create the infrastructure to perform targeted studies within the cohort.

Clinical Context

Brazil has a mixed healthcare system of public and private health insurance and service provision. Universal access to care is provided by the public healthcare system, the Unified Health System (SUS), and the vast majority of patients with SCD receive care through this. The Brazilian Government, through the National Agency of Sanitary Vigilance (ANVISA) and the Blood Coordination of the Ministry of Health (MoH), implemented a series of new regulations from 2001 to 2017, requiring neonatal screening (starting in the late 1980s and now available in all 27 states) and comprehensive care of SCD (ANVISA, 2002, Brazilian Ministry of Health, 2005, 2012, 2015, 2016). Practice guidelines include:

1. Standardized use of folic acid (1–5 mg/day) and prophylactic antibiotics for children from 3 months to 5 years of age using oral penicillin V, monthly injected benzathine or erythromycin for patients with documented penicillin allergy.
2. Hydroxycarbamide (HC) is recommended for the following indications: three or more acute vaso-occlusive pain episodes (VOE) requiring hospital medical care or proven productive incapacity (school/work); >1 acute chest

syndrome (ACS) event; ≥ 1 ACS that required ≥ 2 blood transfusions, ≥ 1 ACS that required hospitalization in an intensive care unit; chronic hypoxaemia defined as oxygen saturation persistently lower than 94% measured at two consecutive routine clinical visits; other evidence of chronic organ damage (priapism, bone necrosis, proliferative retinopathy, among others) (Brazilian Ministry of Health, 2012).

3. Vaccinations against pneumococcus, diphtheria, tetanus, pertussis, poliomyelitis, measles, mumps, rubella, Bacillus Calmette–Guérin (BCG), *Haemophilus influenzae*, hepatitis B, varicella and hepatitis A as well as annual influenza vaccination are recommended. At the time of cohort inception, yellow fever vaccine was not recommended in all haemocentres, but was added to recommendations during the outbreak in 2018.
4. Transfusion recommendations for SCD include use of leucoreduced and phenotyped (Rh, Kidd, Kell and Duffy) red blood cells as well as detailed protocols for specific situations, such as ACS and stroke (Brazilian Ministry of Health, 2012).
5. Pain management includes recommendations on the precise diagnosis and management of the crisis, including hydration (oral and/or parenteral), psychological support, removal of the triggering cause, determination of oxygen saturation, red blood cell transfusion if VOE is associated with a >20% reduction of steady state haemoglobin, and use of pain medication (including narcotics or anti-inflammatory drugs among others) (Brazilian Ministry of Health, 2012).
6. Recently, the MoH issued recommendations for SUS to provide haematopoietic stem cell transplant for SCD patients with defined eligibility criteria and a human leucocyte antigen (HLA) matched sibling donor (Brazilian Ministry of Health, 2015, 2016).

Methods

REDS-III overview

This study is part of the international component of the multi-centre USA National Institutes of Health, National Heart Lung and Blood Institute (NHLBI) REDS-III programme that conducts research focused on the safety and adequacy of the blood supply and impact of blood transfusion in recipients in the USA, Brazil, China and South Africa (Kleinman *et al*, 2014). Multiple different research projects have been implemented through REDS-III, including studies to characterize blood product epidemiology and utilisation (Karafin *et al*, 2017), the risk of human immunodeficiency virus (HIV) and other transfusion-transmitted infections primarily in international sites (Sabino *et al*, 2016), identify donor characteristics that impact storage of blood (Kanas *et al*, 2017), and characterize transfusion outcomes in recipients of blood (Bloch *et al*, 2015). The Brazil programme is

the only programme within REDS-III to establish a SCD cohort in order to characterize health outcomes (including transfusion outcomes) in patients with SCD. The study was funded by NHLBI and Research Triangle Institute (RTI) International was the data-coordinating centre. The Brazilian National Ethical Committee for Research, local Ethical Committees at each participating centre and the Institutional Review Board at University of California, San Francisco (UCSF) all reviewed and approved the study.

The REDS-III SCD cohort study is a collaboration between American investigators at Blood Systems Research Institute, San Francisco, California, and Brazilian investigators at four haemocentres (blood donation and transfusion centres which provide outpatient care to haematology patients). Six patient care sites are included: (i) Fundação Hemominas in the three cities of Belo Horizonte (HBH), Juiz de Fora (JFO) and Montes Claros (MOC) in the state of Minas Gerais, (ii) Fundação Hemorio in Rio de Janeiro in the state of Rio de Janeiro, (iii) Fundação Hemope in Recife in the state of Pernambuco and, (iv) Instituto da Criança, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HCFMUSP) in São Paulo in the state of São Paulo.

Patient recruitment and study procedures

A list of the 9676 active patients (patients with at least one clinical encounter in the last 3 years) at the six participating sites was created and a random sample from this list was selected by RTI. Approximately 50% of the active patient population (4956 of 9676) was enlisted as the sample frame in order to achieve a recruitment goal of 3000 participants. Randomization was conducted in strata proportional to the age (children defined as aged <18 years or adults defined as aged ≥18 years), gender and SCD genotype distributions of the patients at each haemocentre to ensure the study population was reflective of the SCD population at each site. Patients randomly selected as eligible were recruited at scheduled clinic visits. An interview, medical record abstraction and blood collection were performed for all consenting participants. The interview captured comprehensive demographics and vital signs were measured. Race was defined according to the standards of the Brazilian census, which refers to skin colour instead of heritage. Medical records were reviewed by haematologists, or research nurses under the supervision of haematologists, to abstract clinical data using standardised definitions of the phenotypic manifestations of SCD (Ballas *et al*, 2010). Abstraction included any documentation in the patient's record of specific clinical events, such as stroke, avascular necrosis, priapism, leg ulcers, bacteraemia and other SCD complications which had occurred during patients' lifetime. Detailed information related to hospitalisations, transfusions, laboratory testing and screening examinations that occurred in the year prior to enrolment was also abstracted (summary of collected data in Table SI). Transfusion history and blood bank data, including red blood cell phenotype and

antibody data, were extracted from each haemocentre's records. A comprehensive electronic database was created to centralise all clinical, laboratory and transfusion information. In the current phase of the cohort, approximately annual follow-up visits are scheduled through 2018 to capture interim clinical and transfusion events. This paper presents the baseline data collected during enrolment visits in 2013–2015.

Biospecimen collection and single nucleotide polymorphism genotyping

A biospecimen repository of blood samples from all enrolled patients was established to complement the database and support translational studies relevant to SCD pathogenesis and transfusion complications. Blood was collected in EDTA tubes, separated into cellular and plasma components and aliquoted for storage in a Biobank held at the University of São Paulo. DNA was extracted from cellular aliquots using alcohol precipitation and quantified using real-time polymerase chain reaction (PCR). DNA was normalized to 10 ng/μl with a minimum volume of 60 μl.

Peripheral blood genomewide single nucleotide polymorphism (SNP) genotyping was performed using a customized Axiom™ array, entitled the Transfusion Medicine Array (TM Array; Affymetrix, Santa Clara, CA, USA). The TM Array was developed by REDS-III to facilitate targeted blood donor and recipient studies and includes 875 188 SNPs, of which 549 000 were designed to whole genome coverage down to a minor allele frequency (MAF) of 5% in European, African and East Asian descended populations. In addition, SNPs down to a MAF of 1% were included in 141 iron-related genes, 3856 platelet genes, 1285 red blood cell genes, 238 cytokines, 155 tumour growth factor genes and 48 genes associated with SCD-related phenotypes. The TM Array also includes ~1000 copy number polymorphisms (CNPs) across alpha globin, beta globin and RH loci to allow the detection of copy number changes (Page *et al*, 2016).

The DNA samples were genotyped using the TM Array in the UCSF Genomics Core Facility. SNPs were called using the Axiom Analysis Suite Software (Affymetrix Thermo Fisher Scientific, Waltham, MA, USA), HLA alleles were called with Axiom HLA Analysis software (Affymetrix) and CNPs were called with Axiom CNV Summary Tools with further analysis with PennCNV (<http://penncnv.openbioinformatics.org/en/latest/>). 838 722 SNPs were considered high quality by the software. We examined for genotype mismatches between duplicate samples and excluded 6925 SNPs. Additionally, 3546 markers that were not successfully called in >97% of samples were also excluded resulting in a final data set of 831 797 SNPs.

Confirmation of SCD genotype

Genotyping for haemoglobin mutations for all participants was confirmed by allele-specific pyrosequencing (Qiagen, Hilden,

Germany) (Ronaghi, 2001). If pyrosequencing diagnosis matched the original haemocentre diagnosis of SS or SC, no further testing was performed. If pyrosequencing showed either SS or SC that did not match the diagnosis recorded in medical records or heterozygous S was detected, Sanger sequencing of *HBB* exons 1 and 2 was performed. This allowed confirmation of SS or SC diagnosis and defined the specific β^0 or β^+ mutations for most S β thalassaemia subjects. Samples from 26 participants with an unresolved sickle cell genotype after sequencing were sent to the Hemoglobinopathy Reference Laboratory at UCSF Benioff Children's Hospital Oakland (CHO) for definitive diagnosis. Sequencing at CHO included *HBB* exon 3 in addition to introns and promoter regions, and SCD genotype was resolved in all 26 participants. For data analyses of clinical complications, genotypes SS, S β^0 , SD and S/Quebec-CHORI were combined to create a single sickle cell anaemia (SCA) category.

Statistical methods

Statistical analyses were performed by RTI, using SAS 9.4 (SAS Institute, Inc., Cary, NC, USA). Continuous variables are reported as medians (interquartile range) or means (standard deviation), depending on the distribution, while categorical variables are summarized as number (%). Differences between haemocentres and sickle cell type groups were assessed using analysis of variance (ANOVA) or Kruskal–Wallis tests as appropriate, for continuous variables; and chi square test or Fishers exact test as appropriate for categorical variables.

The TM Array SNP data were analysed to evaluate the genetic ancestry background of the cohort compared to self-reported skin colour using the categories established by the Brazilian Census. A random set of 100 000 common markers (minor allele frequency 5%) was extracted from the REDS-III Brazil SCD cohort and the 1000 Genomes Project (1000G) phase three release May 2013 (Auton, 2015). Principle component analyses (PCA) were conducted using PLINK version 1.90a (Purcell *et al*, 2007). Principle components were extracted and plotted using the ggplot2 (v2.2.1) package in R (v3.2.5) (<https://CRAN.R-project.org/package=ggplot2>). *T*-tests were conducted to compare the mean of principle components between subgroups of the cohort by self-reported skin colour. For both clinical and PCA comparisons, a *P* value of <0.01 was defined as statistically significant, considering multiple outcomes were compared.

Results

Demographics

Of the 4956 randomly selected eligible patients, 3029 were recruited in the defined enrolment period from 2013 to 2015 and 2795 participants (92.3% of recruited patients) were enrolled in the cohort. Of 3029 recruited patients, 234 were not included in the cohort either because the participant

decided he/she was not interested ($n = 218$) or the patient was excluded when further evaluation revealed the patient did not have SCD ($n = 16$).

In the enrolled cohort, there were slightly more children <18 years old (55.9%) than adults and slightly more females (53.0%) than males. The majority of the patients reported being mixed skin colour (58.7%) or black (26.8%), with significant variation in self-reported skin colour among the participating centres. Only 34.2% of the adults ≥ 18 years of age reported current employment and most patients (93.5%) reported a monthly household income of \leq \$3000 Brazilian reals (approximately \$1351 US dollars in 2014). Most adult patients (82.2%) did not have more than secondary (high) school-level education (6.5% had attended post-secondary technical school (public schools, equivalent to a high school level, but with additional hours to learn a specific skill, such as electronics, computation, clinical laboratory, etc., to work as a technician in industry, laboratories, hospitals and services) and 11.3% had attended college (Table I).

Most patients had HbSS (1976 of 2795, 70.7%), but the genotypes varied among the six clinical centres, with the highest prevalence of SS in Hemope in the State of Pernambuco (83.5%, Table I). Patients with HbSC accounted for 23% of the cohort, but again with uneven distribution among the regions studied, with the three sites in the state of Minas Gerais having the highest prevalence (28.8–33.6%). The prevalences of S β^+ (2.9%) and S β^0 (3.0%) were lower among the patients enrolled and other heterozygous forms of SCD such as SD and other variants were rare (0.4%). Demographics of the study population enrolled at each haemocentre did not significantly differ from those of the overall active SCD population at the centre (data not shown). Comprehensive demographics are summarized in Table I, with socioeconomic data restricted to adults aged ≥ 18 years.

TM Array data

Comparison of the TM Array data from the enrolled participants to the 1000G revealed the REDS-III Brazil SCD cohort is a heterogeneous population (Fig 1). Varying levels of heterogeneity were observed within each self-reported skin colour (as a surrogate for race based on Brazilian census categories). However, patients who self-reported as white were significantly more similar to 1000G Caucasians than other members of the Brazil SCD populations ($P < 0.0001$); similarly, self-reported black patients were significantly more similar to 1000G Africans ($P < 0.0001$) as supported by the first principle component. Patients reporting other skin colours were distributed between the white and black groups. The self-reported Indigenous category in the SCD cohort did not correspond with Native American populations, but the 1000G samples do not include representation of any sub-Amazonian Native American populations. Patients with S β^+ /S β^0 were more likely to report white skin colour than other genotypes (23.8% compared to 9.9%, $P < 0.0001$).

Table 1. Sickle cell type and selected sociodemographic characteristics of 2795 patients at the time of enrolment into the REDS-III Brazil SCD cohort study.

Characteristic	Clinical Site								Total (N = 2795), n (%)	P-value
	Hemominas									
	Belo Horizonte (N = 783), n (%)	Juiz de Fora (N = 274), n (%)	Montes Claros (N = 363), n (%)	Hemope Recife (N = 550), n (%)	Hemorio Rio de Janeiro (N = 720), n (%)	HCFMUSP São Paulo (N = 105), n (%)				
SCD genotype										
SS	475 (60.7)	180 (65.7)	228 (62.8)	459 (83.5)	557 (77.4)	77 (73.3)	1976 (70.7)	<0.0001		
SC	263 (33.6)	79 (28.8)	120 (33.1)	49 (8.9)	118 (16.4)	15 (14.3)	644 (23.0)			
Sβ0	25 (3.2)	10 (3.6)	9 (2.5)	14 (2.5)	18 (2.5)	7 (6.7)	83 (3.0)			
Sβ+	18 (2.3)	4 (1.5)	3 (0.8)	26 (4.7)	24 (3.3)	6 (5.7)	81 (2.9)			
Other§	2 (0.3)	1 (0.4)	3 (0.8)	2 (0.4)	3 (0.4)	0	11 (0.4)			
Sex										
Female	439 (56.1)	148 (54.0)	184 (50.7)	289 (52.5)	376 (52.2)	45 (42.9)	1481 (53.0)	0.14		
Male	344 (43.9)	126 (46.0)	179 (49.3)	261 (47.5)	344 (47.8)	60 (57.1)	1314 (47.0)			
Race										
Black	224 (28.6)	131 (47.8)	34 (9.4)	81 (14.8)	252 (35.0)	26 (24.8)	748 (26.8)	<0.0001		
Indigenous	8 (1.0)	0	1 (0.3)	1 (0.2)	0	4 (3.8)	14 (0.5)			
Mixed	438 (55.9)	113 (41.2)	266 (73.3)	398 (72.5)	385 (53.5)	38 (36.2)	1638 (58.7)			
White	65 (8.3)	21 (7.7)	38 (10.5)	78 (10.8)	78 (10.8)	37 (35.2)	299 (10.7)			
Unknown/not reported‡	48 (6.1)	9 (3.3)	24 (6.6)	10 (1.8)	5 (0.7)	0	96 (3.4)			
Age group (years)										
0-4	17 (2.2)	38 (13.9)	43 (11.8)	33 (6.0)	94 (13.1)	12 (11.4)	237 (8.5)	<0.0001*		
5-9	129 (16.5)	42 (15.3)	59 (16.3)	98 (17.8)	140 (19.4)	28 (26.7)	496 (17.7)			
10-17	273 (34.9%)	71 (25.9%)	113 (31.1%)	136 (24.7%)	171 (23.8%)	65 (61.9%)	829 (29.7%)			
18-29	166 (21.2%)	58 (21.2%)	94 (25.9%)	140 (25.5%)	137 (19.0%)	-	595 (21.3%)			
30-39	106 (13.5)	34 (12.4%)	35 (9.6%)	91 (16.5%)	88 (12.2)	-	353 (12.6)			
40-49	61 (7.8)	17 (6.2)	11 (3.0)	37 (6.7)	49 (6.8)	-	175 (6.3)			
50-59	20 (2.6)	12 (4.4)	7 (1.9)	11 (2.0)	28 (3.9)	-	78 (2.8)			
60+	11 (1.4)	2 (0.7)	1 (0.3)	4 (0.7)	15 (2.1)	-	33 (1.2)			
Patient category										
Adult	364 (46.5)	123 (44.9)	148 (40.8)	283 (51.5)	315 (43.8)	-	1233 (44.1)	0.02†		
Paediatric	419 (53.5)	151 (55.1)	215 (59.2)	267 (48.5)	405 (56.3)	105 (100)	1562 (55.9)			
Education¶										
Never attended	2 (0.6)	0	1 (0.7)	4 (1.4)	3 (1.0)	-	10 (0.8)	<0.0001†		
Primary school	129 (35.7)	67 (54.5)	64 (43.2)	100 (35.3)	81 (25.7)	-	441 (35.9)			
Secondary school	175 (48.5)	44 (35.8)	61 (41.2)	124 (43.8)	156 (49.5)	-	560 (45.5)			
Technical school	20 (5.5)	4 (3.3)	10 (6.8)	17 (6.0)	29 (9.2)	-	80 (6.5%)			
College/post graduate	35 (9.7)	8 (6.5)	12 (8.1)	38 (13.4)	46 (14.6)	-	139 (11.3)			

Table 1. (Continued)

Characteristic	Clinical Site							P-value
	Hemominas							
	Belo Horizonte (N = 783), n (%)	Juiz de Fora (N = 274), n (%)	Montes Claros (N = 363), n (%)	Hemope Recife (N = 550), n (%)	Hemorio Rio de Janeiro (N = 720), n (%)	HCFMUSP São Paulo (N = 105), n (%)	Total (N = 2795), n (%)	
Marital status¶								
Single	212 (58.7)	75 (61.0)	97 (65.5)	145 (51.2)	169 (53.7)	—	698 (56.7)	0.0002†
Living together	36 (10.0)	13 (10.6)	20 (13.5)	65 (23.0)	37 (11.7)	—	171 (13.9)	
Married	90 (24.9)	31 (25.2)	29 (19.6)	53 (18.7)	92 (29.2)	—	295 (24.0)	
Separated/divorced/widowed	23 (6.4)	4 (3.2)	2 (1.4)	20 (7.1)	17 (5.4)	—	66 (5.3)	
Employment¶¶								
Currently working	160 (44.3)	38 (30.9)	47 (31.8)	54 (19.1)	122 (38.7)	—	421 (34.2)	<0.0001†
Household income								
<R\$700 per month	64 (18.4)	21 (17.1)	21 (14.3)	43 (16.9)	38 (12.5)	—	187 (15.9)	<0.0001†
R\$701–1400	159 (45.7)	61 (49.6)	97 (66.0)	166 (65.1)	142 (46.6)	—	625 (53.1)	
R\$1401–3000	102 (29.3)	35 (28.5)	27 (18.4)	39 (15.3)	86 (28.2)	—	289 (24.5)	
R\$3000+	23 (6.6)	6 (4.9)	2 (1.4)	7 (2.7)	39 (12.8)	—	77 (6.5)	
Unknown/not reported	13 (3.6)	0	1 (0.7)	28 (9.9)	10 (3.2)	—	52 (4.2)	
Smoking history¶¶¶								
Ever smoke	116 (32.2)	37 (30.8)	37 (25.0)	64 (22.6)	79 (25.1)	—	333 (27.1)	0.06†
Currently smoke	26 (7.2)	10 (8.1)	8 (5.4)	18 (6.4)	18 (5.7)	—	80 (6.5)	0.80†
Alcohol history¶¶¶								
Ever drank	273 (75.8)	63 (51.2)	116 (78.4)	213 (75.3)	174 (55.2)	—	839 (68.3)	<0.0001†
Currently drink	105 (29.2)	23 (18.7)	46 (31.3)	56 (20.1)	54 (17.1)	—	284 (23.2)	0.0002†

HCFMUSP, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo; R\$, Brazilian reals; SCD, sickle cell disease.

All P values for comparisons of haemocentre specific differences.

*Same P-value excluding and including HCFMUSP.

†Excluding HCFMUSP.

‡2 refused to answer, 4 did not complete the patient interview.

§Includes 6 SD, 3 S/HPPH, 1 S/K-Woolwich, 1 S/Quebec-CHORI

Brazilian school categories include primary (elementary), secondary (high school) and technical school (post secondary school that provides education regarding specific skills such as electronics, clinical laboratory etc.)

¶¶Questions asked of adult patient category only.

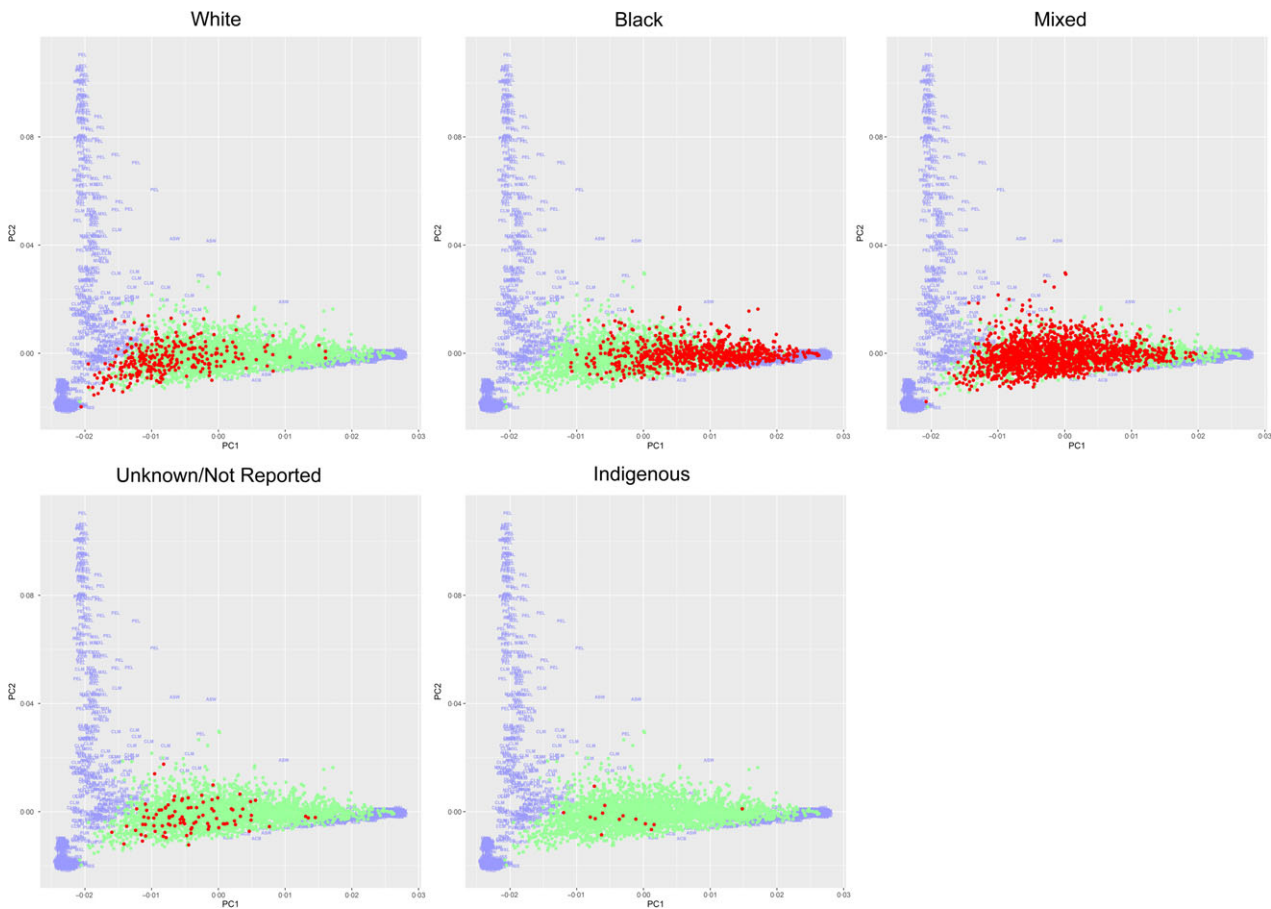


Fig 1. Principle component analysis comparison of TM-Array data of REDS-III SCD Cohort to 1000G Reference Population by Self-Reported Race. Plot of top principle components of the Brazilian sickle cell disease (SCD) cohort are co-graphed with the 1000 Genome Project (1000G) reference population. 1000G participants are coloured purple, with subpopulations indicated by the three-lettered coding. REDS-III Brazilian SCD patients are shown as dots. The race in focus, as identified in the header, is coloured red; while the remaining REDS-III SCD patients are coloured green. 1000G three-lettered coding: GWD, Gambian in Western Division, The Gambia – Mandinka; MSL, Mende in Sierra Leone; ESN, Esan in Nigeria; CLM, Colombian in Medellin, Colombia; PEL, Peruvian in Lima, Peru; TSI, Toscani in Italy; IBS, Iberian populations in Spain; MXL, Mexican Ancestry in Los Angeles, California; GWJ, Gambian in Western Division, The Gambia – Jola; GBR, British in England and Scotland; CEU, Utah residents (CEPH) with Northern and Western European ancestry; YRI, Yoruba in Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya; GWF, Gambian in Western Division, The Gambia – Fula; ASW, African Ancestry in Southwest US; ACB, African Caribbean in Barbados; PUR, Puerto Rican in Puerto Rico; FIN, Finnish in Finland; GWW, Gambian in Western Division, The Gambia – Wolof.

Disease modifying therapies, clinical complications and steady state laboratory values

Summary of the treatments, surgeries and clinical complications of the cohort participants, stratified by children/adults and SCD genotype are shown in Table II. To evaluate differences among the six participating centres, comparisons were restricted to SCA patients for all outcomes and summarized in Table SII. Treatment with HC was reported in 458 children (29.3% of 1104) and 447 adults (36.3% of 1044) at the time of enrolment, with a significantly higher percentage of SCA patients treated with HC than other SCD types (Table II). The proportion of patients treated with HC varied significantly across the centres, both in children (range from 15.4% of SCA children in Montes Claros to 57.1% in São Paulo) and

adults (range from 23.4% of SCA adults in Montes Claros to 53.7% in Belo Horizonte, Table SII). Most cohort patients (75.0% of children and 89.2% of adults) had been transfused at least once, with significantly more SCA patients demonstrating a history of transfusion (children 88.0%, adults 95.9%) than SC (children 41.3%, adults 66.0%, Table II). 10.6% of children and 9.2% of adults were treated with chronic transfusion therapy (approximately monthly red blood cell transfusion to prevent severe complications of SCD), almost all of which were SCA patients (Table II).

Mirroring the results of chronic transfusion therapy, the use of iron chelators was similar in children (6.1%) and adults (7.2%) and was significantly higher in patients with SCA (8.5% children, 9.1% adults). The majority of the chelated patients (174/185, 94.0%) were treated with Desferasirox®.

Table II. Comparison of treatment and clinical complications between paediatric and adult patients by genotype at enrolment in the REDS-III Brazil SCD cohort study.

Treatment/clinical complication	Genotypes of adult patients										P-value
	Genotypes of paediatric patients					Genotypes of adult patients					
	SCA* (N = 1114), n (%)	SC (N = 397), n (%)	Sβ+† (N = 51), n (%)	Total (N = 1562), n (%)	P-value	SCA* (N = 952), n (%)	SC (N = 247), n (%)	Sβ+† (N = 34), n (%)	Total (N = 1233), n (%)	P-value	
RBC transfusion											
Ever	980 (88.0%)	164 (41.3%)	28 (54.9)	1172 (75%)	<0.0001	913 (95.9%)	163 (66.0%)	24 (70.6%)	1100 (89.2%)	<0.0001	
Past year	427 (38.3)	21 (5.3)	8 (15.7)	456 (29.2)	<0.0001	344 (36.1)	12 (4.9)	4 (11.8)	360 (29.2)	<0.0001	
Chronic	164 (14.7)	2 (0.5)	0	166 (10.6)	<0.0001	110 (11.6)	3 (1.2)	0	113 (9.2)	<0.0001	
Medications‡											
Hydroxycarbamide	432 (38.8)	16 (4.0)	10 (19.6)	458 (29.3)	<0.0001	402 (42.2)	36 (14.6)	9 (26.5)	447 (36.3)	<0.0001	
Prophylactic antibiotics (0–6 years old)§	213 (72.4)	74 (74.0)	12 (80.0)	299 (73.1)	0.7	–	–	–	–	–	
Analgesics for 30+ consecutive days	134 (12.0)	22 (5.5)	6 (11.8)	162 (10.4)	0.0012	132 (13.9)	27 (10.9)	11 (32.4)	170 (13.8)	0.004	
Antidepressants	21 (1.9)	5 (1.3)	1 (2.0)	27 (1.7)	0.7	63 (6.6)	19 (7.7)	3 (8.8)	85 (6.9)	0.8	
Antihypertensives	22 (2.0)	1 (0.3)	0	23 (1.5)	0.03	125 (13.1)	32 (13.0)	4 (11.8)	161 (13.1)	0.9	
Any Chelator	95 (8.5)	1 (0.3)	0	96 (6.1)	<0.0001	87 (9.1)	2 (0.8)	0	89 (7.2)	<0.0001	
Deferasirox	92 (8.3)	1 (0.3)	0	93 (6.0)	<0.0001	79 (8.3)	2 (0.8)	0	81 (6.6)	<0.0001	
Desferoxamine	4 (0.4)	0	0	4 (0.3)	0.6	10 (1.1)	0	0	10 (0.8)	0.4	
Deferiprone	1 (0.1)	0	0	1 (0.1)	–	0	0	0	0	–	
ACE inhibitor	36 (3.2)	2 (0.5)	0	38 (2.4)	0.005	62 (6.5)	8 (3.2)	0	70 (5.7)	0.04	
Surgery											
Splenectomy	217 (19.5)	11 (2.8)	9 (17.6)	237 (15.2)	<0.0001	70 (7.4)	7 (2.8)	4 (11.8)	81 (6.6)	0.02	
Cholecystectomy	154 (13.8)	14 (3.5)	4 (7.8)	172 (11.0)	<0.0001	378 (39.7)	55 (22.3)	3 (8.8)	436 (35.4)	<0.0001	
Hip replacement	3 (0.3)	0	0	3 (0.2)	0.6	32 (3.4)	10 (4.0)	0	42 (3.4)	0.5	
Complications											
Hospital admission in past year	409 (36.7)	71 (17.9)	15 (29.4)	495 (31.7)	<0.0001	311 (32.7)	41 (16.6)	8 (23.5)	360 (29.2)	<0.0001	
VOE hospitalisation in past year	246 (22.1%)	48 (12.1)	13 (25.5)	307 (19.7%)	<0.0001	244 (25.6%)	25 (10.1)	8 (23.5)	277 (22.5%)	<0.0001	
ACS hospitalisation in past year	216 (19.4%)	22 (5.5)	7 (13.7)	245 (15.7%)	<0.0001	88 (9.2)	6 (2.4)	2 (5.9)	96 (7.8)	0.002	
Stroke – infarct	98 (8.8)	0	0	98 (6.3)	<0.0001	104 (10.9)	10 (4.1)	0	114 (9.2%)	0.0007	
Stroke – haemorrhagic	9 (0.8)	0	0	9 (0.6)	0.2	8 (0.8)	1 (0.4)	0	9 (0.4)	0.7	
Avascular necrosis	27 (2.4)	17 (4.3)	2 (3.9)	46 (3.0)	0.16	140 (14.8)	48 (19.8)	6 (17.6)	194 (15.8)	0.15	
Leg ulcers	10 (0.9)	1 (0.3)	0	11 (0.7)	0.5	207 (21.8)	10 (4.1)	2 (5.9)	219 (17.8)	<0.0001	
Splenic sequestration	328 (30.0)	45 (11.5)	10 (19.6)	383 (24.9)	<0.0001	71 (7.8)	11 (4.5)	2 (6.1)	84 (7.1)	0.2	

Table II. (Continued)

Treatment/clinical complication	Genotypes of paediatric patients			Genotypes of adult patients			P-value	
	SCA* (N = 1114), n (%)	SC (N = 397), n (%)	Sβ+† (N = 51), n (%)	Total (N = 1562), n (%)	SCA* (N = 952), n (%)	SC (N = 247), n (%)		Sβ+† (N = 34), n (%)
Bacteraemia	67 (6.1)	11 (2.8)	3 (5.9)	81 (5.2)	70 (7.4)	7 (2.9)	2 (5.9)	79 (6.5)
Chronic renal failure	1 (0.1)	0	0	1 (0.1)	32 (3.4)	5 (2.0)	0	37 (3.0)
Priapism (males only)	48 (8.5)	5 (2.6)	1 (3.7)	54 (6.9)	109 (27.8)	4 (4.2)	4 (23.5)	117 (23.2)

ACE, angiotensin-converting-enzyme; ACS, acute chest syndrome; RBC, red blood cell; SCA, sickle cell anaemia; SCD, sickle cell disease; VOE, vaso-occlusive pain episode.

*Includes SS, Sβ0, SD, S/Quebec-CHORI.

†Less severe genotype variants such as S/HFFH and S/K-Woolwich were included with Sβ+ patients.

‡Medications specifically included on study abstraction forms.

§Denominators for percentages are restricted to participants ≤6 years of age.

Many complications were relatively rare in children but became common by adulthood. For example, avascular necrosis (AVN) and leg ulcers were reported in only 3% and 0.7%, respectively, of children but reached a prevalence of 15.8% and 17.8% in the adult population. Similarly, treatment with antidepressant drugs and antihypertensive treatment were not common in children (1.7% and 1.5%, respectively) but were used in 6.9% and 13.1% of adults. Less significant differences between the paediatric and adult population was seen in the prevalence of bacteraemia/sepsis, which occurred in 5.2% of children and 6.5% of adults. Differences in the prevalence of selected SCD complications and treatments in age categories of the REDS-III patients are shown in Fig 2.

The use of chronic pain medication (defined as 30+ consecutive days of treatment with any analgesic in the year prior to enrolment) was significantly higher in SCA patients, both in children and adults, but was present also in SC and Sβ+ patients (10.4% of all children and 13.8% of all adults), and was more commonly used in Hemorio than other centres. Prophylactic antibiotics were used by the majority (85.8%) of the children aged 0–5 years in the cohort and showed some variation among the centres, with Montes Claros having the highest level of use (93.5% of the children, Table II).

The most commonly reported surgery in the paediatric population was splenectomy (217 children, 15.2%). The proportion of adults with splenectomy was only 6.6%, probably reflecting a combination of changes in clinical practice, with splenectomy being less common in the past, and also some survival bias of the cohort, with splenectomised adults potentially less likely to be alive and included in the cohort. In adults, history of cholecystectomy was the most commonly recorded surgery (378 adults, 39.7%). Both surgeries were most common in SCA patients.

Almost one-third of both children and adults had been hospitalised in the year prior to enrolment. Hospital admissions varied by centre, with São Paulo having the highest (42.9%) proportion of children with a hospitalisation in the year prior to enrolment, and Recife and Juiz de Fora the highest for adults (46.0% and 36.6%, respectively). Montes Claros had the lowest rates for hospital admissions, both in children and adults (25.4% and 17.8%, respectively). The most common indications for hospitalization based on discharge diagnoses abstracted from the medical records (up to three recorded for each hospitalization) in both children and adults were VOE (61.3% of paediatric and 80.1% of adult hospitalisations) followed by ACS (40% of paediatric and 16.3% of adult hospitalisations). Clinically overt ischaemic stroke had occurred in 6.3% of all children (8.8% of SCA) and 9.2% of all adults (10.9% of SCA). Haemorrhagic stroke was reported to a lesser extent (less than 1.0%) in both children and adults with SCA (Table II).

As expected clinically, SCA patients had lower steady state haemoglobin and higher steady state white blood cell and platelet counts as well as higher levels of markers of haemolysis

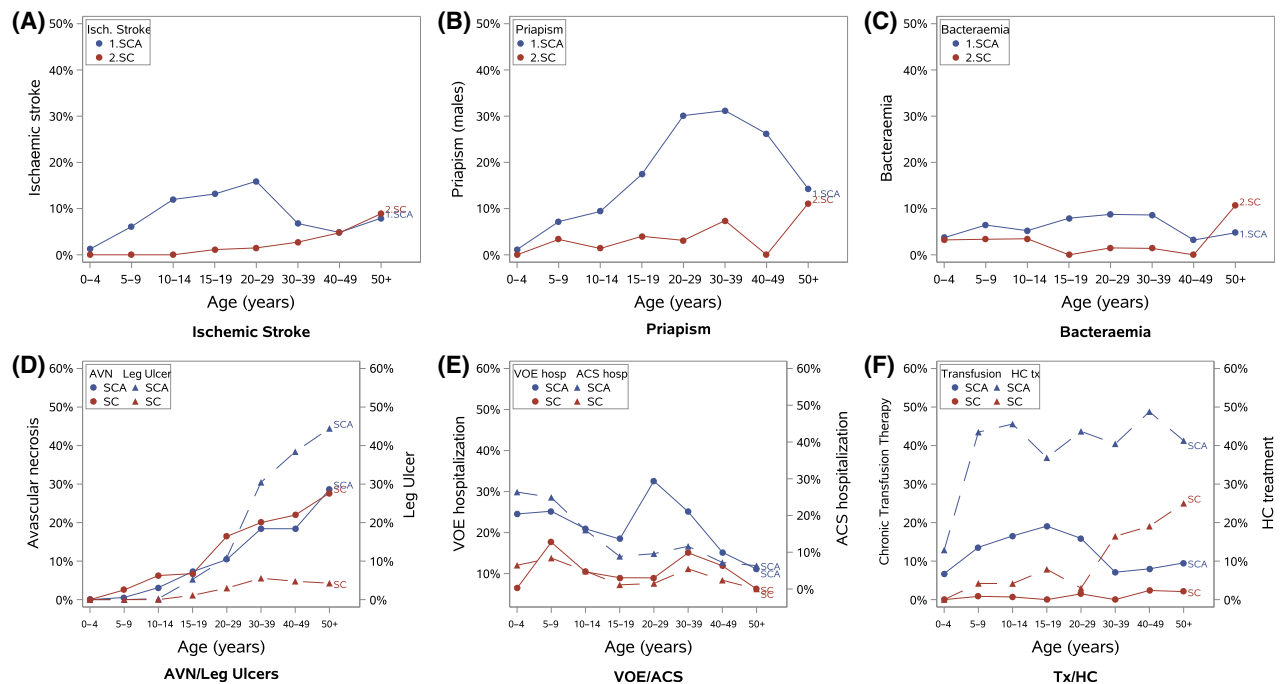


Fig 2. Prevalence of select SCD clinical complications and treatments within age groups of the REDS-III Brazil SCD cohort. The proportion of REDS-III patients within defined age group categories with a history of select sickle cell disease (SCD) clinical complications is shown – ischaemic stroke (a), priapism (b), bacteraemia (c), avascular necrosis (AVN) and leg ulcers (d). (e) The proportion of patients with a hospitalization for vaso-occlusive pain episode (VOE) or acute chest syndrome (ACS) in the year prior to enrolment into the cohort. (f) The prevalence of patients treated with chronic transfusion therapy and hydroxycarbamide (HC tx) at the time of enrolment into the cohort. In all graphs, sickle cell anemia (SCA) patients (SS, Sβ⁰, SD) are shown in blue and HbSC patients are shown in red. Because of the significant heterogeneity of clinical outcomes in only 81 Sβ⁺ patients, the Sβ⁺ patients are not included in the figures. Specific phenotypes are indicated with circle/solid line or triangle/dashed line as defined in the key in each panel. [Colour figure can be viewed at wileyonlinelibrary.com]

compared to SC or Sβ⁺ patients (steady state was defined as laboratory values taken at routine visits, with no acute event or transfusion in the previous 2 months, Table III).

Discussion

The REDS-III Brazil SCD cohort was established to characterize clinical outcomes with a focus on transfusion outcomes in the Brazilian SCD population and create the infrastructure to perform genotype-phenotype association studies. This report of the baseline demographics and clinical complications of the REDS-III participants highlights some similarities to other international cohorts of SCD. For example, the demographic profile of the REDS-III cohort reveals the SCD population in Brazil has higher rates of unemployment, lower monthly household income and lower education level when compared to the general population in the country (<https://brasilemsintese.ibge.gov.br/educacao.html>; <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2013-agencia-de-noticias/releases/15693-pnad-continua-taxa-de-desocupacao-cai-em-11-das-27-ufs-no-2-trimestre-de-2017.html>), underscoring the challenges faced by this vulnerable population coping with a chronic disease. The socioeconomic and educational status of the Brazilian population varies according to the region of the country, especially when the Northeastern

and Southeastern parts of the country are considered (https://ww2.ibge.gov.br/home/estatistica/indicadores/trabalho-erendimento/pnad-continua/default_renda_percapita.shtm). We observed differences in the SCD population among the studied centres which were comparable to differences observed in the general population, i.e., the patients in the Northeast region of Brazil and northern part of Minas Gerais State reported lower socioeconomic status indicators.

The observed unemployment rate in the cohort (65.7%) is considerably higher than the general population of Brazil, which was 6.3–9% according to 2013–2015 Brazilian Institute of Geography and Statistics data (<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2013-agencia-de-noticias/releases/15693-pnad-continua-taxa-de-desocupacao-cai-em-11-das-27-ufs-no-2-trimestre-de-2017.html>). Similarly, the monthly household income of ≤3000 Brazilian reals, reported in most patients in this cohort, is less than the household monthly income of 5072 Brazilian reals in the general population during the same time period (https://ww2.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pnad-continua/default_renda_percapita.shtm). These challenges are common to SCD patients in other parts of the world, however the unemployment rates in this Brazilian SCD cohort are even higher than the ~30–50% unemployment reported in various US and Jamaican studies (Sanger *et al*, 2016; Asnani *et al*, 2017).

Table III. Comparison of steady state* laboratory values between paediatric and adult patients by genotype at enrolment in the REDS-III Brazil SCD cohort study.

Analyte	Children <18 years				Adults ≥18 years				P-value
	SC		Sβ+‡		SCA†		SC		
	No HC	Median (IQR) Range	No HC	Median (IQR) Range	HC	No HC	Median (IQR) Range	Sβ+‡ Median (IQR) Range	
Haemoglobin (g/l)	n = 397 84 (76–91) 54–126	n = 638 81 (74–89) 52–138	n = 361 110 (104–118) 72–145	n = 43 99 (88–108) 60–150	n = 361 88 (78–99) 48–138	n = 485 80 (72–90) 35–139	n = 223 116 (106–126) 58–157	n = 32 104 (92–122) 69–159	<0.0001
Leucocyte count (×10 ⁹ /l)	n = 393 10.4 (8–13.6) 3.4–28.8	n = 633 13.2 (10.8–16.4) 3.2–42.4	n = 360 8.3 (6.6–10.6) 1.1–25.4	n = 43 8.9 (6.9–12) 3.3–17.3	n = 361 8.2 (6.5–11) 2.21–7	n = 485 11.4 (8.9–13.7) 4.2–28.4	n = 223 8.2 (5.8–10.4) 2.6–18.2	n = 32 7.1 (5.7–10.8) 2.7–20.6	<0.0001
Platelet count (×10 ⁹ /l)	n = 395 420 (322–526) 52.2–986	n = 633 442 (356–551) 54–999	n = 361 258 (188.0–355) 80.5–693	n = 42 338 (213–413) 58.2–563	n = 361 348 (286–436) 83–977	n = 484 404 (303–496) 72–938	n = 222 269 (148–357) 55.2–578	n = 32 180 (157.0–341.5) 104–713	<0.0001
Reticulocytes (%)	n = 369 8.3 (5.6–12.4) 0.6–38	n = 566 11 (7.6–15) 0.6–43	n = 320 3.4 (2.4–5.0) 0.4–30	n = 41 3.5 (2.0–6.0) 1.0–11	n = 335 8.2 (5.1–12) 0.1–26.2	n = 402 11.2 (7.8–15) 0.3–31.6	n = 200 3.8 (2.6–5.3) 0.2–16	n = 28 3.4 (2.0–8.0) 0.8–17.7	<0.0001
Creatinine (μmol/l)	n = 420 44.2 (35.4–53.0) 8.8–97.2	n = 688 44.2 (35.4–53.0) 17.7–141.4	n = 393 53.0 (44.2–61.9) 17.7–79.6	n = 51 35.4 (35.4–53.0) 17.7–97.2	n = 385 70.7 (53.0–79.6) 17.7–274.0	n = 567 61.9 (53.0–79.6) 17.7–919.4	n = 247 79.6 (70.7–97.2) 35.4–203.3	n = 34 70.7 (61.9–97.2) 44.2–123.8	<0.0001
LDH (μ/l)	n = 420 536 (372–753) 120–2488	n = 686 603 (450–861) 58–3481	n = 392 294 (217–420) 49–2289	n = 25 313 (208–501) 84–988	n = 385 430 (303–639) 52–2945	n = 565 594 (398–853) 111–2583	n = 247 243 (178–378) 82–1124	n = 34 269 (184–378) 78–767	<0.0001
Total bilirubin (μmol/l)	n = 421 15.4 (10.3–25.7) 1.7–171	n = 689 20.5 (13.7–32.5) 1.7–444.2	n = 393 10.3 (6.8–13.7) 1.7–49.6	n = 51 8.6 (5.1–11.9) 1.7–562	n = 385 17.1 (10.3–25.7) 1.7–225.7	n = 567 25.7 (17.1–37.6) 3.4–278.7	n = 247 11.9 (8.6–17.1) 1.7–106.0	n = 34 8.6 (6.8–13.7) 1.7–259.9	<0.0001
Haemoglobin F (%)	n = 315 11 (5.3–18) 0–48	n = 363 8 (3.6–14) 0.8–47	n = 152 3 (1.4–5) 0.1–31.0	n = 24 5 (2–11.3) 0.7–34	n = 230 14.6 (6.7–23.1) 0.3–47	n = 224 4.6 (2–11) 0.1–34	n = 73 3.0 (1.4–5) 0.1–21.0	n = 11 8.8 (2.3–15.2) 0.1–16.9	<0.0001

HC, hydroxycarbamide; IQR, interquartile range; LDH, lactate dehydrogenase; SCA, sickle cell anaemia; SCD, sickle cell disease.

*Steady state was defined as laboratory values taken at routine visits, with no acute event or transfusion in last 2 months.

†Includes SS, Sβ0, SD, S/Quebec-CHORI, excludes 2 SS patients with diagnosis SS status post stem cell transplant. SCA group stratified by patients treated or not with HC at the time of laboratory assessment.

‡Less severe genotype variants such as S/HFFH and S/K-Woolwich were included with Sβ+ patients.

Additional similarities to other large SCD cohorts include the increasing prevalence of most clinical complications with age (Farber *et al*, 1985; Makani *et al*, 2011), demonstrating the chronic cumulative organ damage experienced by SCD patients over their lifetime, and the more prevalent occurrence of most clinical phenotypes in patients with SCA compared to other genotypes (Ohene-Frempong *et al*, 1998). A notable exception to this trend was the prevalence of avascular necrosis, which occurred more commonly in SC than other genotypes. Use of antihypertensives and antidepressant medications did not differ between the genotype groups.

Although heterogeneity in the clinical presentation of the S β + genotype is well reported, depending on the specific β + mutation and quantity of normal HbA (Christakis *et al*, 1991), S β + is generally associated with milder disease severity. However, the S β + patients in the REDS-III Cohort appear to exhibit a more severe phenotype, particularly in complications related to pain (treatment with chronic pain medication, vaso-occlusive pain hospitalisations), which were observed as frequently in S β + patients as SCA patients. An analysis of complications in S β + participants within categories of specific β + mutations is underway to further characterize these enrolment findings.

Other differences observed in the cohort may be related to regional and clinical care differences among the patient populations studied within Brazil. The most common genotype in the REDS-III patients was SS (70.7%), but the frequency of genotypes varied by centre, with the prevalence of SC much higher in the three sites in the state of Minas Gerais, consistent with the overall SCD population in the state. We observed variation in the rates of many treatments, including episodic and chronic transfusions among the centres that may reflect better access to care for the patients in larger metropolitan centres of the Southeast part of the country. We also cannot exclude variation among the participating centres on the utilization of the national protocols of SCD treatment instituted by the Brazilian Ministry of Health. In addition, the possibility of bias in the data exist due to missing information from clinical encounters that occurred outside the participating sites. While the centres in São Paulo, Rio de Janeiro and Recife have hospitals attached to them and patients at these centres receive both outpatient and inpatient care at the same location, the three Hemominas centres have more fragmented inpatient care and therefore data to related to hospitalisations may be less complete, especially for adult patients. The lower prevalence of complications, such as ischaemic stroke, in the oldest age group (Fig 2) may reflect survival bias to some extent, but missing data from adults' remote medical history is also likely.

An additional source of bias could result from including patients more frequently seen at the haemocentres in the study. Although a random sampling of each haemocentres' population was made to select eligible patients and attempt to obtain a representative spectrum of disease severity in the cohort, patients with milder clinical disease may not attend

routine care appointments as frequently and may therefore be underrepresented.

Despite these limitations, this is one of the largest contemporary SCD cohorts in the world, and a significant strength of the cohort will be to define SCD outcomes in the current era of increased HC and chronic transfusion therapy. Approximately a third of the REDS-III cohort was treated with HC and the impact of HC in the cohort is currently being evaluated. With respect to transfusion, even with variation in the use of chronic transfusions among centres, the proportion of the REDS-III cohort treated with chronic transfusions at the time of enrolment (approximately 10%) is similar to a recent survey of 31 USA institutions collectively treating 12 644 SCD patients that reported 1274 (10%) were chronically transfused (Kelly *et al*, 2016).

An additional strength of the cohort is the extensive genotyping performed which will allow broad genotype-phenotype association studies. Whole genome SNP typing was planned as part of REDS-III with the primary goal to identify genetic determinants of red blood cell alloimmunization. Additional genotype-phenotype associations studies are underway to investigate genetic determinants of stroke, fetal haemoglobin, frequent pain and other key SCD phenotypes. These genome wide association studies will be conducted in one of the most genetically admixed populations in the world due to the combination of the Native South American, African, Caucasian and other immigrant populations in Brazil. The most commonly reported race in the REDS-III SCD patients was "mixed" and, regardless of self-reported race, significant admixture was seen when comparing the cohort to the 1000G reference populations. The admixture in Brazil is significantly more complex than SCD populations in North America (Solovieff *et al*, 2011) therefore this initial analysis of the population structure is critical to provide context for genotype-phenotype association studies currently underway.

The REDS-III Brazil SCD cohort was established to define the clinical impact of SCD in the Brazilian population in the current era. These data provide the context to targeted analyses currently underway to characterise blood utilisation and predictors of transfusion adverse events, identify changes in SCD pathophysiology that occur over time with chronic transfusion therapy, define the prevalence of eligible candidates for haematopoietic stem cell transplant based on recent MoH guidelines, identify predictors of severe pain and characterize the prevalence and risk factors for infectious diseases, such as HIV, Hepatitis and, more recently, Zika Virus and yellow fever virus in this vulnerable population. In addition, genotype-phenotype analyses to identify the genetic determinants of key SCD complications are planned. After enrolment commenced, REDS-III was selected to participate in the NHLBI-funded Trans-Omics for Precision Medicine (TOPMed) study. The TOPMed programme provides whole genome sequencing (WGS) data to well-phenotyped cohorts in order to link WGS data to clinical outcomes and

ultimately promote discoveries about the fundamental mechanisms that underlie heart, lung, blood and sleep disorders.

In conclusion, we have successfully established a large cohort representative of the range of SCD genotypes found in Brazil and defined the baseline demographic, clinical and laboratory characteristics at the time of enrolment. Given Brazil's position as a high middle-income country, this cohort provides a unique opportunity to provide broad insights into SCD and predictors of outcomes for patients relevant to low-, middle- and high-income countries with large numbers of SCD patients. We anticipate these results will be critical for defining the impact of SCD in the Brazilian population and ultimately will help guide management and research for SCD patients worldwide.

Acknowledgements

We first wish to give a heart-felt thank you to all of the patients who agreed to participate in this research. We gratefully acknowledge all of the research staff at each haemocentre in Brazil who have enrolled patients into the study and completed all study procedures. At each site the following specific people are recognized for their commitment and contribution to this project: Fundação Pró-Sangue (São Paulo) – Alfredo Mendrone Jr., Cesar de Almeida Neto; ITACI – Instituto de Tratamento do Câncer Infantil (São Paulo) – Roberta Calucci, Erivanda Bezerra; Hemominas – Belo Horizonte (Minas Gerais) – Franciane Mendes de Oliveira, Valquíria Reis, Nayara Durte, Barbara Malta; Hemominas; Montes Claros (Minas Gerais) – José Wilson Sales, Maria Aparecida Souza, Rodrigo Ferreira; Fundação Hemope – Recife (Pernambuco) – Maria do Carmo Valgueir; Regina Gomes, Airly Goes Maciel, Rebeca Talamatu Dantas; Hemorio – (Rio de

Janeiro) – Flavia Herculano, Ana Claudia Pereira, Ana Carla Alvarenga, Adriana Grilo, Fabiana Canedo; Instituto de Matemática e Estatística da Universidade de São Paulo – USP (São Paulo) – Pedro Losco Takecian, Mina Cintho Ozahata, Rodrigo Muller de Carvalho. US Investigators: RTI – Research Triangle Institute, International – Christopher McClure; National Institutes of Health, National Heart, Lung, and Blood Institute – Simone A. Glynn.

Author contributions

The research was designed by ABCF, SK, BC, DB, TTG, MB and ES. Patient enrolment and data collection/interpretation at each site was led by ABCF, CMT, TPSS, AA, PL, CM, CL, MVFP, DOWR and RAM. Study databases were built and maintained by LC, JEF and MC. Haemoglobin sequencing was performed and analysed by CSA and CH. The clinical data was analysed by LC, LP and DB. SNP array data was analysed and interpreted by YG and GP. Primary responsibility for writing the manuscript was shared by ABFC, SK and BC. All co-authors reviewed and made revisions to the final manuscript.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Overview of the demographic and clinical data collected for REDS-III Brazil SCD cohort members.

Table S2. Treatments and clinical complications of REDS-III SCA (SS/SBO/SD/SQuebec-CHORI) Children by Hemocenter.

References

- ANVISA (2002) Manual de Diagnóstico e Tratamento de Doenças Falciformes. Agência Nacional de Vigilância Sanitária, Brazil. <http://bvsmms.saude.gov.br/bvs/publicacoes/anvisa/diagnostico.pdf>
- Asnani, M.R., Knight Madden, J., Reid, M., Greene, L.G. & Lyew-Ayee, P. (2017) Socio-environmental exposures and health outcomes among persons with sickle cell disease. *PLoS ONE*, **12**, e0175260.
- Auton (2015) A global reference for human genetic variation. 1000 Genomes Project Consortium. *Nature*, **527**, 68–74.
- Ballas, S.K., Lieff, S., Benjamin, L.J., Dampier, C.D., Heeney, M.M., Hoppe, C., Johnson, C.S., Rogers, Z.R., Smith-Whitley, K., Wang, W.C. & Telen, M.J. (2010) Definitions of the phenotypic manifestations of sickle cell disease. *American Journal of Hematology*, **85**, 6–13.
- Bloch, E.M., Crookes, R.L., Hull, J., Fawcus, S., Gangaram, R., Anthony, J., Ingram, C., Ngcobo, S., Croxford, J., Creel, D.V. & Murphy, E.L. (2015) International Component of the NHLBI Recipient Epidemiology and Donor Evaluation Study-III (REDS-III). *Transfusion*, **55**, 1675–1684.
- Brazilian Ministry of Health (2005) Ordinance No. 1,391/GM of August 16, 2005. Article 1 To establish, within the scope of the Unified Health System – SUS, as guidelines for the National Policy for Comprehensive Care for Persons with Sickle Cell Disease and other Hemoglobinopathies. http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2005/prt1391_16_08_2005.html
- Brazilian Ministry of Health (2012) Basic treatments of sickle cell disease. http://bvsmms.saude.gov.br/bvs/publicacoes/doenca_falciforme_condutas_basicas.pdf
- Brazilian Ministry of Health (2015) Stem cell transplant in sickle cell disease. Ordinance No. 30, 2015 <http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?journal=1&pagina=49&data=01/07/2015>
- Brazilian Ministry of Health (2016) Clinical protocol for treatment of sickle cell disease. http://conitec.gov.br/imagens/Consultas/Relatorios/2016/Relatorio_PCDT_DoencaFalciforme_CP_2016_v2.pdf
- Cançado, R.D. & Jesus, J.A. (2007) Sickle cell in Brazil. *Brazilian Journal of Hematology and Hemotherapy*, **29**, 204–206.
- Christakis, J., Vavatsi, N., Hassapopoulou, H., Angeloudi, M., Papadopoulou, M., Loukopoulos, D., Morris, J.S., Serjeant, B.E. & Serjeant, G.R. (1991) A comparison of sickle cell syndromes in northern Greece. *British Journal of Haematology*, **77**, 386.
- Curtin, P.D. (1969) The Atlantic Slave Trade: A Census. The University of Wisconsin Press, Milwaukee, WI.
- Farber, M.D., Koshy, M. & Kinney, T.R. (1985) The Cooperative Study of Sickle Cell Disease. Demographic and socioeconomic characteristics of patients and families with sickle cell disease. *Journal of Chronic Diseases*, **38**, 495.
- Kanias, T., Lanteri, M.C., Page, G.P., Guo, Y., Endres, S.M., Stone, M., Keating, S., Mast, A.E., Cable, R.G., Triulzi, D.J., Kiss, J.E., Murphy, E.L., Kleinman, S., Busch, M.P., Gladwin, M.T. & for the National Heart Lung and Blood

- Institute Recipient Epidemiology and Donor Evaluation Study-III (REDS-III) (2017) Ethnicity, sex, and age are determinants of red blood cell storage and stress hemolysis: results of the REDS-III RBC-Omics study. *Blood Advances*, **1**, 1132–1141.
- Karafin, M.S., Bruhn, R., Westlake, M., Sullivan, M.T., Bialkowski, W., Idgren, G., Roubinian, N.H., Hauser, R.G., Kor, D.J., Fleischmann, D., Gottschall, J.L., Murphy, E.L., Triulzi, D.J. & National Heart Lung and Blood Institute Recipient Epidemiology and Donor Evaluation Study-III (REDS-III) (2017) Demographic and epidemiologic characterization of transfusion recipients from four US regions: evidence from the REDS-III recipient database. *Transfusion*, **57**, 2903–2913.
- Kato, G.J., Steinberg, M.H. & Gladwin, M.T. (2017) Intravascular hemolysis and the pathophysiology of sickle cell disease. *The Journal of Clinical Investigation*, **127**, 750–760.
- Kelly, S., Quirolo, K., Marsh, A., Neumayr, L., Garcia, A. & Custer, B. (2016) Erythrocytapheresis for chronic transfusion therapy in sickle cell disease: survey of current practices and review of the literature. *Transfusion*, **56**, 2877–2888.
- Kleinman, S., Busch, M.P., Murphy, E.L., Shan, H., Ness, P., Glynn, S.A. & for The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III) (2014) The National Heart, Lung and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III): a research program striving to improve blood donor and recipient outcomes. *Transfusion*, **54**, 942–955.
- Makani, J., Cox, S.E., Soka, D., Komba, A.N., Oruo, J., Mwamtemi, H., Magesa, P., Rwezaula, S., Meda, E., Mgaya, J., Lowe, B., Muturi, D., Roberts, D.J., Williams, T.N., Pallangyo, K., Kitundu, J., Fegan, G., Kirkham, F.J., Marsh, K. & Newton, C.R. (2011) Mortality in sickle cell anemia in Africa: a prospective cohort study in Tanzania. *PLoS ONE*, **6**, e14699.
- Ohene-Frempong, K., Weiner, S.J., Sleeper, L.A., Miller, S.T., Embury, S., Moohr, J.W., Wethers, D.L., Pegelow, C.H. & Gill, F.M. (1998) Cerebrovascular accidents in sickle cell disease: rates and risk factors. *Blood*, **91**, 288–294.
- Page, G.P., Guo, Y., Seielstad, M., Keating, B., Westhoff, C.M., Hoppe, C., Bordbar, A., Pals-son, B.O., Custer, B., Lu, Y. & Busch, M.P. (2016) Development and evaluation of a transfusion medicine genome-wide SNP Array. *Transfusion*, **56**, 151A.
- Piel, F.B., Steinberg, M.H. & Rees, D.C. (2017) Sickle cell disease. *New England Journal of Medicine*, **376**, 1561–1573.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. & Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.
- Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research*, **11**, 3–11.
- Sabino, E.C., Loureiro, P., Lopes, M.E., Capuani, L., McClure, C., Chowdhury, D., Di-Lorenzo-Oliveira, C., Oliveira, L.C., Linnen, J.M., Lee, T.H., Gonçalez, T., Brambilla, D., Kleinman, S., Busch, M.P. & Custer B for the REDS-III Study (2016) Transfusion transmitted dengue and associated clinical symptoms during the 2012 epidemic in Brazil. *Journal of Infectious Diseases*, **213**, 694–702.
- Salzano, F.M. & Bortolini, M.C. (2002) The Evolution and Genetics of Latin American Populations. Cambridge University Press, Cambridge.
- Sanger, M., Jordan, L., Pruthi, S., Day, M., Covert, B., Merriweather, B., Rodeghier, M., DeBaun, M. & Kassim, A. (2016) Cognitive deficits are associated with unemployment in adults with sickle cell anemia. *Journal of Clinical and Experimental Neuropsychology*, **38**, 661–671.
- Saraf, S.L., Molokie, R.E., Nouriaie, M., Sable, C.A., Luchtman-Jones, L., Ensing, G.J., Campbell, A.D., Rana, S.R., Niu, X.M., Machado, R.F., Gladwin, M.T. & Gordeuk, V.R. (2014) Differences in the clinical and genotypic presentation of sickle cell disease around the world. *Paediatric Respiratory Reviews*, **15**, 4–12.
- Solovieff, N., Hartley, S.W., Baldwin, C.T., Klings, E.S., Gladwin, M.T., Taylor, J.G., Kato, G.J., Farrer, L.A., Steinberg, M.H. & Sebastiani, P. (2011) Ancestry of Americans, African with sickle cell disease. *Blood Cells, Molecules, and Diseases*, **47**, 41–45.
- Wagner, S.C., de Castro, S.M., Gonzalez, T.P., Santin, A.P., Filippin, L., Zaleski, C.F., Azevedo, L.A., Amorin, B., Callegari-Jacques, S.M. & Hutz, M.H. (2010) Prevalence of common α -thalassaemia determinants in south Brazil: importance for the diagnosis of microcytic anemia. *Genetics and Molecular Biology*, **33**, 641–645.
- Weatherall, D.J. (2010) The inherited diseases of haemoglobin are an emerging global health burden. *Blood*, **115**, 4331–4336.