Pharmacogenomics of Sickle Cell Disease Therapeutics:

Pain and Drug Metabolism Associated Gene Variants and Hydroxyurea-induced Post-Transcriptional Expression of miRNAs

By: Khuthala Mnika

MNKKHU004

Submitted to the University of Cape Town

In fulfilment of the requirements of PhD degree in Human Genetics

University of Cape Town, Faculty of Health Sciences, Department of Pathology, Division of Human Genetics

Supervisor: Prof Ambroise Wonkam

Co-Supervisors: Prof Collet Dandara, Dr Gaston Mazandu, Dr Shaheen Mowla and Dr Emile Chimusa

Date of Submission:13 December 2019

Acknowledgement

I would like to express my utmost gratitude:

To my Father in heaven (God) for the wisdom, strength and direction.

To my sister, Andiswa Mnika for her unconditional love and support since I was 11 years old. Enkosi Mbhokodo, awukhange uwavumele amaxhalanga anditye

To my sibling and my nephew, Lumka Mnika and Indiphile Mnika, for their encouragement and confidence in me.

To my friends, Zinhle Cindi and Lerato Mpye, for the undying comfort and support throughout the ups and downs of this journey.

To those who generously funded this research: The National Research Foundation, University of Cape Town funding and Sickle Africa Data Coordinating Center (SADaCC).

To Dr Gift Pule, for the endless support and motivation towards this journey.

To my co-supervisors, Prof Dandara, Dr Mowla, Dr Mazandu and Ass/Prof Chimusa, for the encouragement and support.

To my supervisor, Prof Ambroise Wonkam, for believing and having faith in me, also the opportunities, encouragement and endless support. Je vous remercie, Asante, Enkosi Prof.

To GeneMap team, for their love and support.

To my colleagues in the Divisions of Human Genetics and Haematology; international collaborators, for their input and support of this work.

To Dr Gloudi Agenbag, for all the unselfish help in the laboratory and for always going the extra mile to help me.

To Dr Victoria Nembaware, for the mentorship, support and advice.

Declaration

I, Khuthala Mnika, hereby declare that this dissertation/thesis is based in my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been or is being submitted for another degree in this or any other university.

I give permission to the university to reproduce, for the purpose of research, either the whole or portion of the content, in any manner whatsoever.

Signed by candidate

Signature: _____

Date :13 December 2019

"I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s)":

List of publication included in this thesis

1. **Mnika K,** Pule GD, Dandara C, Wonkam A. An Expert Review of Pharmacogenomics of Sickle Cell Disease Therapeutics: Not Yet Ready for Global Precision Medicine. OMICS, 2016 Oct; 20 (10):565-574. Epub 2016 Sep 16. (Status: *published*).
2. Wonkam A, **Mnika K**, Ngo Bitoungui VJ,Chemegni BC, Chimusa ER, Dandara C, and Kengne AP. Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. Br J Haematol. 2018 Jan; 180 (1): 134–146. (Status: *published*).
3. **Minka K**, Mazandu GK, Chimusa ER, Dandara C, Wonkam. Analyses 299 Pharmacogenes in Sickle Cell Disease Patients in Africa. *Frontiers in Pharmacology* (2019) (Status: *under review*).
4. **Mnika K**, Mazandu GK, Pule G, Jonas M, Chimusa ER, Hanchard N, and Wonkam A. Using miRNA expression level to analyse the effect of Hydroxyurea in SCD patient at Groote Schuur Hospital. Frontier in Genetics. 2019 May; 509 (10); 1-6. (Status: *published*).

List of publications candidate contributed to (excluded)

1. Pule GD, **Mnika K**, Jourbert M, Mowla S, Novitzky N, Wonkam A. Increasing burden of adult patients living with Sickle Cell Disease in Cape Town. S. Afr. med. j. 2017 Feb; 107 (2). (Status: *published*).

2. Pule GD, **Mnika K**, Chimusa E, Mhandire K, Kampira E, Dandara C, Wonkam A. Beta globin gene haplotypes and tag variants for malaria resistance among black

southern African populations. Global Health, Epidemiology and Genomics. 2017; 2, (17); 1-11. (Status: *published*).

3. Wonkam, A., Chimusa, E. R., **Mnika, K**., Pule, G. D., Ngo Bitoungui, V. J., Mulder, N., Shriner, D., Rotimi, C. N., & Adeyemo, A. (2020). Genetic modifiers of long-term survival in sickle cell anemia. Clinical and Translational Medicine, 10(4), e152. https://doi.org/10.1002/ctm2.152. (Status: *published*).

4. Sickle Cell Disease Ontology Working GroupThe Sickle Cell Disease Ontology: enabling universal sickle cell-based knowledge representation. Database, Volume 2019, 2019, baz118, https://doi.org/10.1093/database/baz118. (Status: published).

Preface

Sickle Cell Disease (SCD) is a global health problem with a burden comparable to that of communicable diseases and other major global diseases such as diabetes and hypertension. Although over 75% of the SCD burden is in sub-Saharan Africa, there are limited studies in this region investigating the genetic polymorphisms that affect the severity, hematologic phenotypes, disease course and mechanism of action of the most widely used and the Food and Drug Administration (FDA) approved medication for treatment of SCD, namely hydroxyurea (HU).

Research Concept and Funding

With the supervision of Prof. Ambroise Wonkam, the candidate crafted the research concept and proposal. This work was supported through competitively secured independent student funding (National Research Foundation and University of Cape Town (UCT) funding) and research funding obtained by Prof. Ambroise Wonkam as a principal investigator for this project. With advice and guidance from the supervisors, the candidate's contributions to the project included: conceiving and designing the experiments; recruiting and sampling of biological material from participants; execution of all experiments; analysis of experimental results; drafting manuscripts and incorporating revisions from co-authors and journal reviewers. The study and all experiments were conceived by the candidate in collaboration with the principal supervisor Professor Ambroise Wonkam and co-supervisors, Dr Shaheen Mowla, Prof Collet Dandara, Dr Gaston Mazandu and Dr Emile Chimusa.

Data collection

Recruitment of participants, obtaining approvals from the Faculty of Health Sciences Ethics Board, obtaining informed consent and sampling of biological material in this project were performed by the candidate. Extraction of participants' clinical data from

hospital records and patient files, DNA and RNA isolation were performed by the candidate. The candidate also managed patient samples and data.

Experimentation and Analysis

All experiments including molecular analysis by genotyping was performed in full by the candidate. Sequencing of microRNA was outsourced to Baylor College of Medicine, United States, using the NanoString Platform (NanoString Technologies, Inc., Seattle, WA, United States). The primary analysis of all data was performed by the candidate and when additional statistical analysis was required, a collaborator was included by the primary supervisor and their contribution to the work is clearly indicated in each publication.

Publications

Synthesis of all the work and drafting of all manuscripts for publication included in this thesis were executed in full by the candidate, after which revisions from all co-authors were similarly incorporated before submission to the journal by the primary supervisor. After review, all reviewer comments were addressed by the candidate under supervision from the primary supervisor. Peer-reviewed publications by multiple international researchers were therefore included to improve the quality of the final thesis. Detailed contributions by the candidate for each publication included in this work is defined in each chapter.

The format of this thesis by publication will hopefully contribute significantly to the scholarship of this field. In addition, inclusion of this published work also aimed to: i) set a path that would strengthen the academic career of the candidate
ii)increase her candidacy for continued financial support from bodies that supporting this work: from the National Research Funding, Genetic Medicine of African populations and the University of Cape Town; iii) qualify her participation and attendance to academic

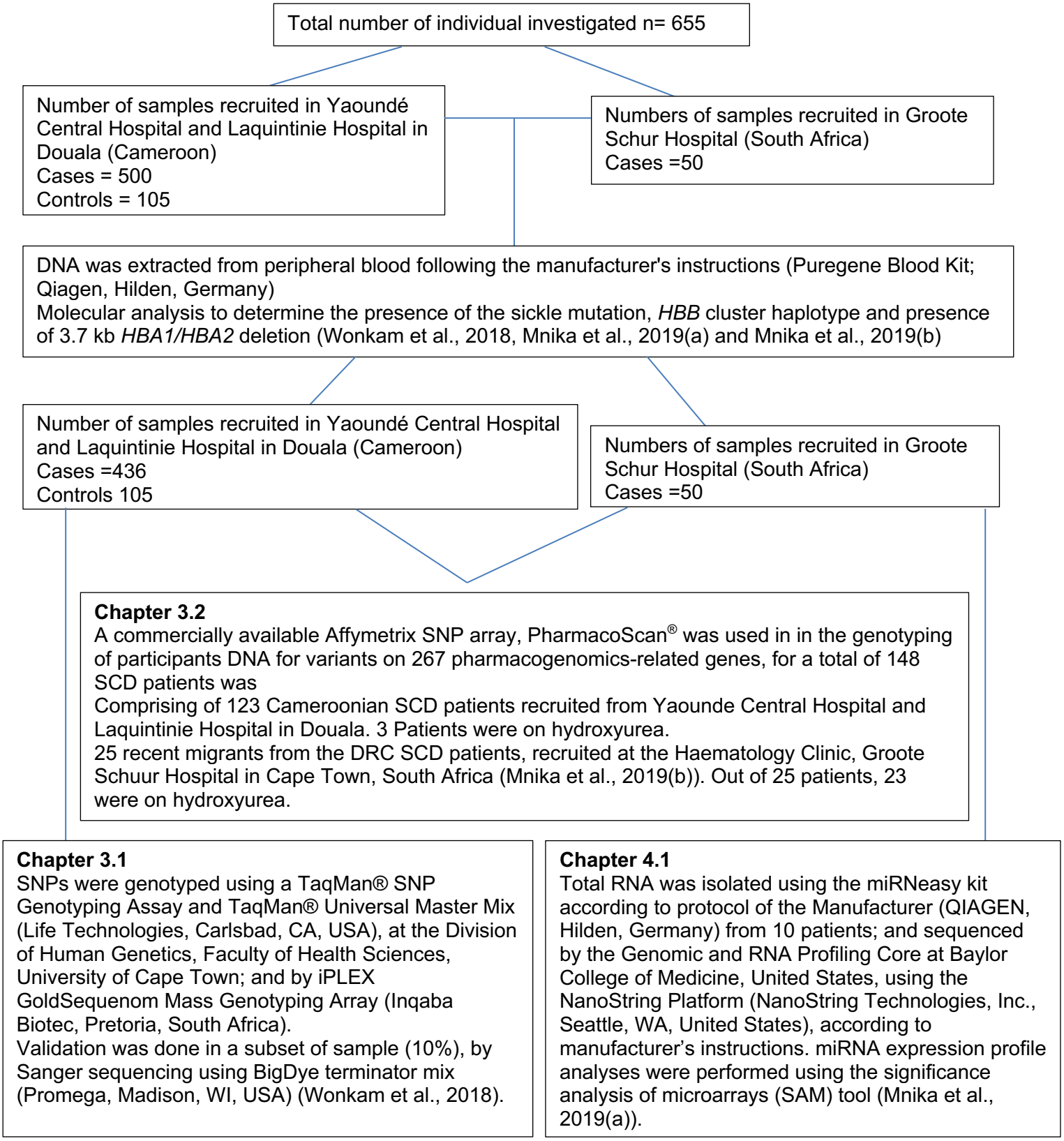
conferences and training workshops. The candidate successfully attended several specialised national and international conferences in Cape Town, London, Mauritius, Cairo, Paris, Corvallis USA, Nigeria and Tunisia) and lastly; iv) the published work will form important parts of a consistent body of research that fits the Faculty of Health Sciences policies. These policies promote publishing of thesis research work as much as possible in order to disseminate knowledge generated and further improve the profile of the institution and of the candidate.

This thesis is presented as three distinct sections: 1) clinical and genetic factors associated with pain and hospitalization, 2) pharmacogenomics of SCD therapeutics and 3) mechanisms of HU-induced γ-globin expression (in vivo and in vitro models). The candidate has met all requirements and approval of UCT's Doctoral Degrees Board, under Rules GP6.7 as follows:

1. The candidate's proposal to include publications in the current thesis was approved by the UCT Faculty of Health Sciences Doctoral Degrees Board.
2. The thesis contains an adequate summary, introduction; a chapter on the Aims and Objectives; a comprehensive Academic Discussion of the results as a whole, forming the basis of the Conclusions and Perspectives drawn from this research.
3. Each chapter is a peer-reviewed publication which is preceded by a synopsis of how the publication ties to the aims and objectives of the project, as well as to the thesis as a whole.
4. All included publications were written and published during the candidate's tenure as a PhD student since 2016.

The candidate,
Khuthala Mnika

Overall workflow



Abstract

Sickle cell disease (SCD) is a common blood disease caused by a single nucleotide substitution (c.20T>A, p.Glu6Val) in the beta globin gene on chromosome 11. The prevalence of the disease is high throughout large areas in sub-Saharan Africa, the Mediterranean basin, the Middle East, and India due to the level of protection that the sickle cell trait, provides against severe malaria. Approximately 300,000 infants are born per year with sickle cell anemia, which is defined as homozygosity for the sickle hemoglobin (HbS). The majority (nearly 75%) of these births occur in sub-Saharan Africa, particularly in two countries: Nigeria, and the Democratic Republic of the Congo where there are poorly resourced healthcare systems. Early diagnosis, penicillin prophylaxis, blood transfusions, hydroxyurea, and hematopoietic stem-cell transplantation can dramatically improve survival and quality of life for patients with SCD. However, our understanding of the role of genetic and clinical factors in explaining the complex phenotypic diversity of this disease is still limited.

Early prediction of the severity, and patients' responses to specific therapeutics of SCD could lead to more precise treatment and management. Beyond well-known modifiers of disease severity, such as fetal hemoglobin (HbF) levels and α-thalassemia, other genetic variants might influence specific sub-phenotypes. New treatments and management strategies accounting for these genetic and nongenetic factors could substantially and rapidly improve the quality of life and reduce health care costs for patients with SCD. Patients with SCD are subjected to long term administration of drugs and there is a limited data on pharmacogenomics of SCD therapeutics. Vaso-occlusive crisis (VOC) are the main clinical events of SCD and are associated with recurrent and long-term use of antalgics/opioids and HU. This project aimed to investigate the clinical and genetic predictors of painful vaso-occlusive crisis (VOC) among SCD Cameroon patients by exploring pharmacokinetic determinants of treatment responses as well as post-

transcriptional signatures triggered by hydroxyurea treatment, particularly, miRNA expression.

SCD patients were recruited from Yaounde Central Hospital and Laquintinie Hospital in Douala (Wonkam et al., 2018, Mnika et al., 2019 (b)), and recent migrants SCD patients from the DRC, recruited at the Haematology Clinic, Groote Schuur Hospital in Cape Town, South Africa (Mnika et al., 2019 (a) and Mnika et al., 2019 (b)). Socio-demographic and clinical data were collected by means of a structured questionnaire. Patients' medical records were reviewed to extract their clinical features over the past 3 years. Specifically, the occurrences of VOC, hematological parameters, hospital outpatient visits, hospitalisation, overt strokes, blood transfusions, and administration of hydroxyurea were recorded. Height, weight, body mass index (BMI), systolic and diastolic blood pressures (SBP and DBP) were measured. Detailed descriptions of patients and sampling methods used in the Cameroonian patients have been reported previously (Wonkam et al., 2018 Mnika et al., 2019 (a) and Mnika et al., 2019 (b)). For the purpose of comparing frequencies of variants, ethnically matched Cameroonian controls were randomly recruited from apparently healthy blood donors in Yaounde for participation in the study. All blood samples were collected for genomic characterisation and analysis. DNA was extracted from peripheral blood, following instructions on the available commercial kit [QIAamp DNA Blood Maxi Kit ® (Qiagen, United States)]. Genotyping (TaqMan and MassArray) was performed for 40 variants in 17 pain-related genes, three fetal haemoglobin (HbF)-promoting loci, two kidney dysfunction-related genes, and *HBA1*/*HBA2* genes for 436 patients. A subset of these samples was also genotyped to analyse 32 core and 267 extended pharmacogenes using commercially available PharmacoScan® platform for characterisation of pharmacokinetic determinant of response. We also compared the pharmacogenes variants from these African groups, to data extracted from the 1000 genomes Project. Moreover, association studies were carried out on pharmacogenes variants with SCD clinical variability. Additionally, protein-

protein interaction (PPI) network and enriched biological processes and pathways were investigated.

For association studies, statistical models using regression frameworks to analyse 40 variants were performed in R®. For miRNA expression, total RNA was isolated using the miRNeasy kit according to protocol of the Manufacturer (QIAGEN, Hilden, Germany); and sequenced by the Genomic and RNA Profiling Core at Baylor College of Medicine, United States, using the NanoString Platform (NanoString Technologies, Inc., Seattle, WA, United States), according to manufacturer's instructions. Genes with statistically significant changes in expression were analysed using the significance analyses of microarrays (SAM) tools.

Female sex, body mass index, Hb/HbF, blood transfusions, leucocytosis and consultation or hospitalisation rates significantly correlated with VOC. Three pain-related gene variants correlated with VOC (*CACNA2D3*-rs6777055, P = 0·025; *DRD2*-rs4274224, P = 0·037; *KCNS1*-rs734784, P= 0·01). Five pain-related gene variants correlated with hospitalization/consultation rates (*COMT*-rs6269, P = 0·027; *FAAH*-rs4141964, P = 0·003; *OPRM1*- rs1799971, P = 0·031; *ADRB2-rs*1042713; P < 0·001; *UGT2B7*-rs7438135, P = 0·037). The 3·7 kb *HBA1/HBA2* deletion correlated with increased VOC (P = 0·002). HbF-promoting loci variants correlated with decreased hospitalisation (*BCL11A*-rs4671393, P = 0·026; *HBS1L-MYB*-rs28384513*,* P = 0·01). *APOL1* G1/G2 correlated with increased hospitalisation (P = 0·048).

A commercial genotyping array platform (PharmacoScan®) with 4627 markers located in 1191 genes was used to investigate 299 pharmacogenes (32 ADME core and 267 extended pharmacogenes). Based on the PharmacoScan analyses, no statistically significant differences in allele frequencies were detected between SCD cases and controls from Cameroon. A principal component analysis (PCA) revealed that Cameroonians' data clustered with other Africans, but this population is significantly

distinct from American, European and Asian populations data. Variant allele frequencies in 21/32 core pharmacogenes were significantly different between the two SCD groups (Cameroon vs. Congo). No correlation between clinical variability and variants in the core genes was detected for both populations under study.

An association study of the core and extended PharmacoScan variants to VOC identified statistically significant associations between two single nucleotide polymorphisms (SNPs) to VOC after correction of multiple testing. These two SNPs mapped to 50 genes, with two SNPs located in core pharmacogenes (*SLCO4A1*- rs118042746, p=1.21e-07; *UGT1A1*0, *UGT1A8*- rs10176426, p=1.22e-07). Functional enrichment analyses revealed that these 50 genes are involved in three biological processes and four pathways relevant to SCD pathophysiology, including xenobiotic glucuronidation (GO:0052697, p = 2.3e-03), and drug metabolism - other enzymes (p = 2.1e-02). Further analyses of the 50 genes, identified key genes in human protein-protein networks: *NTSR1, LRMDA, SMAD SMAD4 and CDH2.* These four genes also interacted with three core pharmacogenes associated with VOC: *UGT1A8, UGT1A10* and *SLCO4A1.*

We found 22/798 miRNAs to be differentially expressed under HU treatment, with the majority (13/22) being functionally associated with HbF-regulatory genes, including *BCL11A* (miR-148b-3p, miR-32-5p, miR-340-5p, miR-29c-3p), *MYB* (miR-105-5p), *KLF-3* (miR-106b-5), and *SP1 (*miR-29b-3p, miR-625-5p, miR-324-5p, miR-125a-5p, miR-99b-5p, miR-374b-5p, miR-145-5p).

The present thesis started by highlighting the scarcity of studies investigating variable responses to pain in SCD patients and then proceeded to addressing this research gap. To our knowledge this is the first body of from Africa to provide evidence supporting the possible development of a genetic risk model for pain in SCD. This is also the first body of work to report an association between these two SNPs and VOC in core and extended pharmacogenes. Our data reveals that the commercial pharmacogenes arrays investigated might need additional evidence for appropriateness among Africans.

Therefore, it advocates the need to invest in research exploring population-specific arrays, drug design, targeting, and efficacy, for improved clinical management of patients of African descent. Previous studies have investigated various mechanisms to understand the genomic variations affecting responses to HU, but full understanding of the variable HU-mediated HbF production among individuals affected by SCD remains elusive. The present study showed that mechanisms of HbF production in response to HU, could particularly be mediated through miRNA regulation. The data reveals some alternative perspectives and routes towards identifying new therapeutic targets and approaches for SCD. However, this study needs to be replicated in larger samples in multiple African populations.

Table of Contents

Chapter 1: Introduction

Sickle cell disease (SCD) is caused by a single point mutation (c.20A>T, p. Glu6Val) in the beta-globin gene, that leads to polymerisation of haemoglobin S (HbS) and the sickling of erythrocytes. HbS is associated with inflammation, haemolysis, microvascular obstruction, and multiple organ damage, specifically in the cardiovascular system, leading to acute and chronic complications in SCD such as stroke, acute chest syndrome, and kidney dysfunctions (Piel et al., 2017). The pathophysiology of SCD is influenced by both environmental and genetic factors. The sickle cell mutation is estimated to have appeared in Africa about 7000 years ago (Shriner and Rotimi, 2018) and its high prevalence in Africa is driven by the protection against malaria conferred on mutation carriers. It is estimated that globally, more than 300 000 babies are born each year with SCD, with nearly 75% of these births in sub-Saharan Africa (Piel et al., 2013). Most SCD patients are subjected to long-term treatment, mainly using hydroxyurea (HU), antibiotics, and opioids (Mnika et al., 2016). High variability in individual responses to these therapies has been reported (Mnika et al., 2016; Wonkam et al., 2018). While environmental factors affecting responses to treatment have been reported, little is known about pharmacogenomic drivers of treatment responses in SCD (Yahouédéhou et al., 2018).

Despite the high incidence of SCD, there is currently no effective public health program in any SSA country focused on SCD (Rahimy et al., 2009; Wonkam, et al., 2014). As a consequence, up to 90% of infants with SCD in SSA are believed to die by the age of 5 years (Flemings et al., 1989). While there have been recent efforts in selected African countries to implement new-born screening (Rahimy et al., 2009; Tshilolo et al., 2009; McGann et al., 2013), to use HU more frequently (Makubi, Soka & Makani, 2012), and to initiate genetic studies (Cox et al., 2014; Mtatiro et al., 2014; Wonkam, Mba, et al., 2014; Wonkam, Ngo Bitoungui, et al., 2014a,b; Wonkam, Rumaney, et al., 2014; Mmbando et al., 2015; Pule, Ngo Bitoungui, et al., 2015), there is still a lack of integration and coordination of these emerging research efforts.

The United States of America (USA) has reduced SCD-related premature deaths by 70% through the use of comprehensive clinical care programs (Yanni et al., 2009; Colombatti et al., 2012), which is in sharp contrast to SSA countries. Evidence from West Africa indicates that the institution of interventions, such as new-born screening and penicillin prophylaxis, can reduce the devastating disease burden in SSA (Rahimy et al., 2003) and to decrease the morbidity and mortality associated with pneumococcal septicemia (Gaston et al., 1986). There is, therefore, a need for research to develop effective therapies across the life span of SCD patients in all parts of the world (Hamideh & Alvarez, 2013; Chaturvedi & Debaun, 2016), including the incorporation of personalized medicine and pharmacogenomics.

Environmental and multiple genetic factors influence many pathophysiological aspects of SCD that contribute to a highly variable clinical expression in individual patients. Fetal hemoglobin (HbF) has emerged as a central disease modifier, and genetic variants at three principal loci, B-Cell CLL/Lymphoma 11A (*BCL11A), HBS1L-MYB*, and Hemoglobin Subunit Beta (*HBB*) cluster, account for 10–20% of HbF variation among SCD patients in the USA, Brazil, and the United Kingdom (Lettre et al., 2008; Thein & Menzel, 2009). These studies have been replicated in patients living with SCD in Tanzania and Cameroon (Mtatiro et al., 2014; Wonkam, Ngo Bitoungui, et al., 2014a; Pule, Ngo Bitoungui, et al., 2015). Interestingly, the expression of these modifiers is responsive to therapeutic manipulation (Xu et al., 2011; Bukar & Abjah, 2013; Canver et al., 2015).

Up till 2018 (Ansari et al., 2018), HU was the only Food and Drug Administration (FDA)-approved treatment for SCD in adults and children (Bockaert, 1999; Shenoy, 2011). HU is a ribonucleotide reductase inhibitor that increases HbF levels, which is a known ameliorator of the disease. However, patients have variable responses to HU due to genetic variations (Charache et al., 1995a; Steinberg, 1997; Bockaert, 1999; Zimmerman

et al., 2004). Nevertheless, the common medications used by SCD patients are analgesic. HU can also reduce pain in SCD which is classified as acute, chronic, and mixed pain, which varies in severity (Steinberg et al., 2010; Ballas et al., 2012). The analgesics are commonly used to manage pain in SCD patients. Genetic differences are known to influence inter-individual variability in pain perception, experience, and responses to anti-inflammatory and opioid drugs (Chou et al., 2006). Homozygous individuals for the 118A>G polymorphism in the opioid receptor mu 1 (*OPRM1)* (a major site of action for most opioid analgesics) experience more pain and need more morphine to subdue pain (Rakvåg et al., 2005). Single-nucleotide polymorphisms (SNPs) in the catechol-O-methyltransferase (*COMT)* gene affect pain sensitivity, with low *COMT* activity being associated increasing pain sensitivity due to increased levels of norepinephrine and epinephrine (Slade et al., 2007).

Mnika et al., 2016, identified a limited number of studies that investigated the genetic/genomic basis of variable responses to pain (e.g., variants in *OPRM1, HMOX-1, GCH1, VEGFA COMT* genes), and the pharmacogenomics of analgesics and opioids (e.g., variants in *OPRM1, STAT6*, *ABCB1,* and *COMT* genes) in SCD. The progress towards identifying the key genomic variants, mainly in *BCL11A, HBS1L-MYB*, or *SAR1*, which contribute to HU treatment response. However, the complete picture of pharmacogenomic determinants of the above therapeutic phenotypes remains elusive. Strikingly, no study has been conducted in sub-Saharan Africa on SCD pharmacogenomics, where majority of the patients live. This alerts the broader global research community to existing disparities in optimal and ethical targeting of research and innovation investments for SCD specifically in pharmacology research and broadly in precision medicine (Mnika et al., 2016).

Clinical complications of SCD including multiple organ damage, specifically in the cardiovascular systems (stroke, acute chest syndrome, and kidney dysfunctions), and susceptibility of SCD patients to infections, require that most patients be subjected to long

term administration of various other types of medications, in addition to HU or opioid therapy. Therefore, there is a need to explore genes involved in absorption, distribution, metabolism and excretion (ADME) genes in SCD, as it is expected that this could influence future therapeutic regimes. It has been indicated that significant differences in the minor allele frequencies (MAF) of about 1/3 of pain-associated variants among SCD patients versus controls from the same population are likely due to positive selection of possible protective variants among patients (Wonkam et al., 2018). This indicates the need to also compare ADME gene variants with non-sickle controls from the same population background. This may provide valuable information for pharmacogenomic interpretations and for understanding the relationship between pharmacogenomics and disease genes/variants.

Through a human erythroid stem cell model, our research group has previously shown that critical regulators of HU-induced γ-globin expression (*MYB, BCL11A* and *KLF-1*), interacted post-transcriptionally with specific miRNAs (Pule, et al., 2015). We have demonstrated a mechanism of HbF production through HU-induced targeted miRNA inhibition of *MYB.* The *in vivo* effects of HU on global transcriptomics still needs to be investigated in African patients. The role of miRNA-mediated post-transcriptional regulation of HbF provides potential targets for new treatments of SCD, that may minimize alterations to the cellular transcriptome. However, the full library of important miRNAs that are involved in HbF- medicated action remain to be determined. In our group, we previously reported the countries of origin of SCD patients in Groote Schuur Hospital (GSH) (Pule et al., 2017). The novel aspect of the study was assessment of the genetic background of four key modifiers of HbF for 34 SCDD patients: variants at the *BCL11A* erythroid specific enhancer, β-globin haplotypes, α-thalassaemia 3.7 kb gene deletion and several other known HbF-promoting polymorphisms. The study also revealed that most of the patients in Cape Town were undergoing HU treatment. It also offered the opportunity to investigate the effect of HU on the transcriptome, specifically miRNA expression in these patients.

In this thesis, we investigate genetic polymorphisms responsible for drug metabolism in a cohort of African patients with severe clinical events such as stroke and vaso-occlusive crises (VOC). This will improve our understanding of the genomics and, subsequently, the clinical management of African patients. Furthermore, we challenge the hypothesis of using miRNAs as agents for SCD therapeutics in African populations by performing global analyses of microRNA expression in peripheral blood of SCD patients (chapter 3.3.1). We also provide valuable insights about the mechanism of action of HU treatment (this data is not included in this thesis) by investigating miRNA-mediated post-transcriptional mechanisms of HbF induction through *ex-vivo* HU treatment using hematopoietic stem cells (commercial CD34) in the African context. Investigation of ADME variants and mechanisms of HU carry important clinical significance as a better understanding of the HU mechanism of effect could reveal novel miRNA agents in SCD. This could improve population-specific drug design, targeting, and efficacy, as well as the clinical management of patients. Most importantly, *ex-vivo* and *in-vivo* disease modeling will build on existing knowledge of drug mechanisms and elucidate potential alternative therapeutic approaches and targets.

Rationale for the Study

SCD is a multisystem disease which is associated with pain episodes (chronic and acute illness) and organ damage, and commonly occurs in Sub-Saharan African countries. Acute episodes of pain, commonly referred to as sickle cell pain crises or VOC, are the primary phenotype expressions associated with SCD and the cause of hospitalization of approximately 95% of cases. Furthermore, the frequency of VOC, along with acute chest syndrome (ACS), is the most common predictor of death in patients with SCD. In addition to VOC and ACS, there is a number of other clinical manifestations, including hepatic and renal involvement, cerebrovascular accident, and multi-organ failure resulting in death. However, a complete picture on pharmacogenomic determinants governing observed variable therapeutic outcomes remains elusive. Strikingly, no study has been conducted

in Sub-Saharan Africa where the majority of SCD patients SCD live. There is a need to evaluate the in vivo impact of efficacious concentrations of HU on the erythroblast transcriptome and/or proteome, as well as the erythroid-specific micronome of SCD patients (before treatment and at maximum tolerated dosage (MTD). When administered at MTD, HU increases fetal HbF to levels ranging from 10% to 40% and a global analysis of these epigenetic mechanisms could highlight multiple components of this complex system, which may possibly yield alternative (e.g., miRNA-based) therapeutic approaches to hemoglobinopathies. This has allowed us to invest in SCD precision medicine and pharmacology research. In this study, we aim to gain a better understanding of genetic variants affecting the predisposition to specific complications such as stroke, acute chest syndrome, and polymorphisms affecting susceptibility to pain, as well as the pharmacogenomics of commonly prescribed treatments such as HU, malaria prophylaxis and pain medication for future precision medicine in SCD.

Chapter 2: Aims and Objectives

This project investigated the clinical and genetic predictors of painful VOC in SCD in Cameroon patients by exploring pharmacokinetic determinants of treatment response. In addition, this work evaluated post-transcriptional signatures triggered by actions of HU treatment, particularly, on miRNA expression. The aims were accomplished through a set of objectives as listed below;

Objectives

1. Literature search on pharmacogenomics of SCD patients
2. Recruitment of a suitable cohort to study
3. Investigation of the clinical and genetic predictors of painful VOC in SCD among Cameroonians.

3.1 Genetic characterisation for variants that are associated with pain and drug metabolisms among SCD patients.

3.2 Evaluation of the association of these variants with health care utilisation (hospitalisations or consultations), considered as direct proxies of VOC.

3.3 Analysis of core (32) ad extended (276) genes using PharmacoScan.

4. Hydroxyurea-induced miRNA profile in SCD patients through studying global miRNA expression.

Chapter 3: Literature review

Pharmacogenomics of Sickle Cell Disease Therapeutics in African Population.

Synopsis: This chapter presents the current scientific understanding of literature background of the work presented in publications which explored the most commonly SCD/pain-associated genes in Sub-Saharan Africa. This chapter includes a peer reviewed review article on the effectiveness of pharmacogenomics/genetics of pain management in SCD, with specific focus on

Mnika K, Pule GD, Dandara C, Wonkam A. An Expert Review of Pharmacogenomics of Sickle Cell Disease Therapeutics: Not Yet Ready for Global Precision Medicine. OMICS, 2016 Oct; 20 (10):565-574. Epub 2016 Sep 16.

Abstract

Sickle cell disease (SCD) is a blood disease caused by a single nucleotide substitution (T > A) in the beta globin gene on chromosome 11. The single point mutation (Glu6Val) promotes polymerization of hemoglobin S (HbS) and causes sickling of erythrocytes. Vaso-occlusive painful crises are associated with recurrent and long-term use of analgesics/opioids and hydroxyurea (HU) by people living with SCD. The present analysis offers a state-of-the-art expert review of the effectiveness of pharmacogenomics/genetics of pain management in SCD, with specific focus on HU and opioids. The literature search used the following keywords: SCD, pharmacogenomics, pharmacogenetics, pain, antalgics, opioids, morphine, and HU. The literature was scanned until March 2016, with specific inclusion of targeted landmark and background articles on SCD. Surprisingly, our review identified only a limited number of studies that addressed the genetic/genomic basis of variable responses to pain (e.g., variants in *OPRM1*, *HMOX-1*, *GCH1*, *VEGFA* *COMT* genes), and pharmacogenomics of antalgics and opioids (e.g., variants in *OPRM1*, *STAT6*, *ABCB1*, and *COMT* genes) in SCD. There has been greater progress made

toward identifying the key genomic variants, mainly in *BCL11A*, *HBS1L-MYB*, and *SAR1*, which contribute to response to HU treatment. However, the complete picture on pharmacogenomics determinants of the above therapeutic phenotypes remains elusive. Strikingly, no study has been conducted in Sub-Saharan Africa where the majority of the patients with SCD live. This alerts the broader global life sciences community toward the existing disparities in optimal and ethical targeting of research and innovation investments for SCD specifically and precision medicine and pharmacology research broadly

Nature of Publication: Original Full Journal Article
Journal/Publisher: OMICS, Journal of Integrative Biology, Mary Anne Liebert Inc.; Peer reviewed

Candidate's contribution: Performed experiments pertaining to the present analysis offers a state-of-the-art expert review of the effectiveness of pharmacogenomics/genetics of pain management in SCD, with specific focus on HU and opioids. Contributed to writing manuscripts on the pharmacogenomics/genetics of pain management in SCD.

Co-Author contribution:

KM and AW: Performed the literature search, analysed data, wrote and revised the manuscript

PGD: Revised manuscript

CD: Contributed to data analysis, revised manuscript

AW: Conceived and designed the experiments, analysed data, wrote and revised and approved manuscript

OMICS A Journal of Integrative Biology
Volume 20, Number 10, 2016

DOI: 10.1089/omi.2016.0105

An Expert Review of Pharmacogenomics of Sickle Cell Disease Therapeutics: Not Yet Ready for Global Precision Medicine

Khuthala Mnika,[1] Gift D. Pule,[1] Collet Dandara,[1] and Ambroise Wonkam[1,2]

Abstract

Sickle cell disease (SCD) is a blood disease caused by a single nucleotide substitution (T > A) in the beta globin gene on chromosome 11. The single point mutation (Glu6Val) promotes polymerization of hemoglobin S (HbS) and causes sickling of erythrocytes. Vaso-occlusive painful crises are associated with recurrent and long-term use of analgesics/opioids and hydroxyurea (HU) by people living with SCD. The present analysis offers a state-of-the-art expert review of the effectiveness of pharmacogenomics/genetics of pain management in SCD, with specific focus on HU and opioids. The literature search used the following keywords: SCD, pharmacogenomics, pharmacogenetics, pain, antalgics, opioids, morphine, and HU. The literature was scanned until March 2016, with specific inclusion of targeted landmark and background articles on SCD. Surprisingly, our review identified only a limited number of studies that addressed the genetic/genomic basis of variable responses to pain (e.g., variants in *OPRM1*, *HMOX-1*, *GCH1*, *VEGFA COMT* genes), and pharmacogenomics of antalgics and opioids (e.g., variants in *OPRM1*, *STAT6*, *ABCB1*, and *COMT* genes) in SCD. There has been greater progress made toward identifying the key genomic variants, mainly in *BCL11A*, *HBS1L-MYB*, or *SAR1*, which contribute to response to HU treatment. However, the complete picture on pharmacogenomic determinants of the above therapeutic phenotypes remains elusive. Strikingly, no study has been conducted in sub-Saharan Africa where majority of the patients with SCD live. This alerts the broader global life sciences community toward the existing disparities in optimal and ethical targeting of research and innovation investments for SCD specifically and precision medicine and pharmacology research broadly.

Introduction

Sickle Cell Disease (SCD) is a multisystem disease, which is associated with episode pain (chronic and acute illness) and organ damage, and commonly occurs in sub-Saharan African countries. SCD is a genetic blood disease caused by a single nucleotide substitution (T > A) in the beta globin gene on chromosome 11 (Brousseau et al., 2007). The resulting HbS leads to polymerization and precipitation of hemoglobin during deoxygenation or dehydration. This results in sickling of red blood cells, abnormal adhesion of leukocytes and platelets, inflammation, hemolysis, and hypercoagulation, which could lead to vaso-occlusive crisis and hypoxia and ultimately organ damage (Bartolucci and Galacteros, 2012).

There is a strong association between the frequency of the HbS mutation and endemicity of malaria (Charache et al., 1995; Williams et al., 2005). It is estimated that 305,800 babies are born each year with SCD worldwide with nearly 75% of the births occurring in sub-Saharan Africa (SSA) (Piel et al., 2013). However, as a result of migration, there is a reported increasing burden of SCD in other countries where it was not initially prevalent, such as South Africa (Wonkam et al., 2012), Ireland (Gibbons et al., 2015), Italy (Colombatti et al., 2013), Germany (Kunz et al., 2015; Zur, 2016), England (Pizzo et al., 2015), and France (Dzierzynski et al., 2016), with, for example, 1300–2600 affected newborns annually in France. SCD is now an accepted worldwide health problem and comparable with other major global noncommunicable diseases such as diabetes and hypertension (Weatherall and Clegg, 2008).

Despite the high incidence, there is currently no effective public health program in any SSA country focused on SCD (Rahimy et al., 2009; Tekola-Ayele and Rotimi, 2015;

[1]Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, Republic of South Africa.
[2]Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, Republic of South Africa.

Wonkam et al., 2014b). As a consequence, up to 90% of infants with SCD in SSA are believed to die by the age of 5 years (Grosse et al., 2011; Makani et al., 2013). While there have been recent efforts in selected African countries to implement newborn screening (McGann et al., 2013; Rahimy et al., 2009; Tshilolo et al., 2009; Tubman et al., 2016), to use hydroxyurea (HU) more frequently (Makubi et al., 2012; Olabode and Shokunbi, 2006; Ware, 2013), and to initiate genetic studies (Cox et al., 2014; Mmbando et al., 2015; Mtatiro et al., 2014; Pule et al., 2015; Rumaney et al., 2014; Wonkam et al., 2014a, 2014b, 2014c), there is still a lack of integration and coordination of these emerging research efforts.

In sharp contrast to SSA, comprehensive clinical care programs have reduced SCD-related premature childhood deaths by 70% in high-income nations such as the United State of America (Vichinsky, 1991; Yanni et al., 2009). This evidence from the West indicates that the institution of interventions such as newborn screening and penicillin prophylaxis can reduce the horrendous disease burden in SSA (Rahimy et al., 2003). Therefore, there is a major need for research to help develop effective therapies across the life span of SCD patients in all parts of the world (Chaturvedi and DeBaun, 2016; Hamideh and Alvarez, 2013), including the incorporation of personalized medicine and pharmacogenomics.

Indeed, environmental and multiple genetic factors influence many pathophysiological aspects of SCD that contribute to a highly variable clinical expression in individual patients. Fetal hemoglobin (HbF) has emerged as a central disease modifier and genetic variants at three principal loci, *BCL11A*, *HBS1L-MYB*, and *HBB* cluster, which account for 10–20% of HbF variation among SCD patients in USA, Brazil, and the United Kingdom (Lettre et al., 2008; Thein and Menzel, 2009). These studies have been replicated in patients living with SCD in Tanzania and Cameroon (Makani et al., 2011; Mtatiro et al., 2014; Pule et al., 2015; Wonkam et al., 2014a). Interestingly, the expression of these modifiers is amenable to therapeutic manipulation (Bukar et al., 2013; Canver et al., 2015; Xu et al., 2011), leading to new hope for treatment routes for SCD (Orkin, 2016).

HU is the only Food and Drug Administration (FDA)-approved treatment of SCD in adults and children (Shenoy, 2011). HU is a ribonucleotide reductase inhibitor that increases the fetal hemoglobin level, a known ameliorator of the disease. Patients respond differently to HU due to genetic variations (Bockaert and Pin, 1999; Charache et al., 1995; Steinberg et al., 1997; Zimmerman et al., 2004).

Nevertheless, the common medications used by SCD patients are antalgics to manage pain. Pain in SCD is classified as acute, chronic, and mixed pain, which varies in severity (Ballas, 2015; Ballas et al., 2012; Steinberg et al., 2010). Genetic differences are suggested to be the reason for inter-individual variability in pain perception and experience and variable responses to anti-inflammatory (Chou et al., 2006) and opioid drugs (Chou et al., 2006). Individuals who are homozygous for 118A>G polymorphism in the *OPRM1* (a major site of action for most opioid analgesics) have more pain and need more morphine to subdue the pain (Klepstad et al., 2004). Single-nucleotide polymorphisms (SNPs) in the *COMT* gene affected pain sensitivity and with low *COM*T activity lead to increased levels of norepinephrine and epinephrine, which resulted in more pain sensitivity (Slade et al., 2007).

The aim of the present analysis was to provide an expert literature review of the effectiveness of pharmacogenomics/genetics for pain management in SCD, with specific focus on pharmacogenetics/pharmacogenomics of pain, HU, and opioids.

Methods

A comprehensive literature search was conducted by the authors covering the subject until March 2016, with specific addition of landmark and background articles on SCD published articles. We used the PubMed® (National Library of Medicine), Medline®, and Google Scholar®. Keywords included individual use or a combination of the following: ''Pharmacogenomics,'' ''Pharmacogenetics,'' ''Hydroxyurea,'' ''Sickle Cell Disease,'' ''Pain,'' ''Painkillers,'' and ''Morphine'' and ''Opioids.'' Additionally, specific expert authors' names that are active in the field of SCD and its therapeutics were also used to complement the literature searches.

Selection criteria

The inclusion criteria were confined to articles written in English, with major emphasis being focused on research articles and review articles describing pharmacogenomics of pain, particularly on SCD patients, and effectiveness of pharmacogenomics of drug therapies for HU and pain management. Prior knowledge of research groups working on HU, pain episodes, and SCD in Africa globally further facilitated the identification and selection of research articles. Only available full-length articles, in English, with the use of ''HU,'' ''Painkillers,'' ''Morphine,'' and ''Opioids'' were selected. In cases where multiple studies reported a similar pathway, the most recent report with the most detailed associations' studies was included. The main search was conducted, separately, by an MSc student and a PhD student (First and Second authors) in Human Genetics working on SCD (to maximize the inclusion of potentially relevant articles) and reviewed successively by a medical geneticist and a human geneticist, with expertise in SCD and pharmacogenomics (Fourth and Third authors), respectively.

A total of 316 articles were consulted after the search from Google Scholar (of which 47 were from PubMed); exclusion criteria were performed based on the article title and its relevance to the scope of the review; additional study performed on the same cohort for the same experiment; and studies that were not clearly stated were excluded. Subsequently, 158 articles were fully retrieved and their abstract and result sections perused for further elimination, of which a final total of 125 articles were selected for inclusion in the review (Fig. 1 and Supplementary Table S1).

Data collection

Data were collected using an extraction form to summarize the following information: type of study, year of publication, patients' sample, study country, title, and author names (Supplementary Table S1).

Results

Pharmacogenomics of pain susceptibility in SCD

Acute pain acts as a protective mechanism in response to tissue injury (Ballas and Lusardi, 2005; Bergman, 2005) and

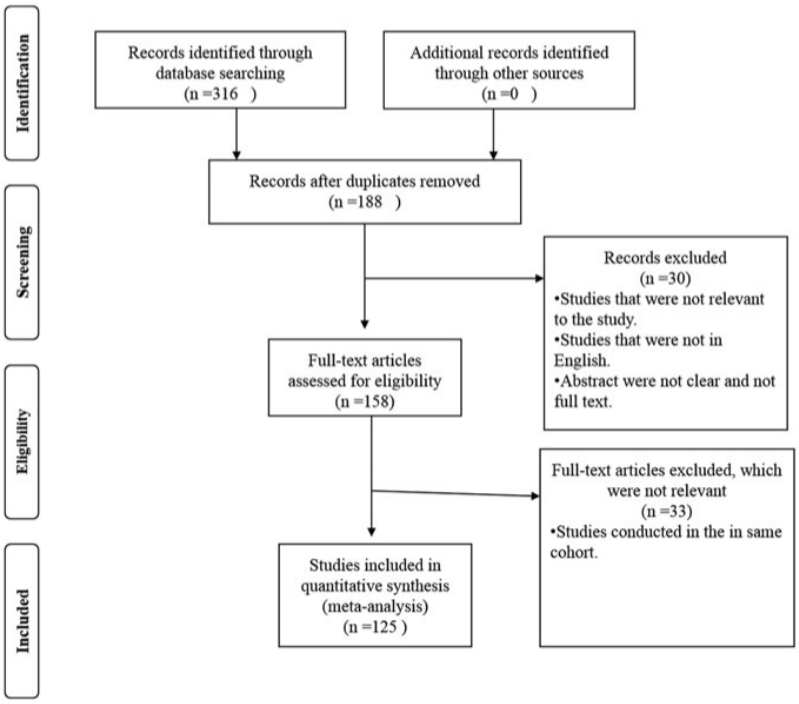


FIG. 1. Flowchart of the literature review employed in the present expert review.

can worsen and prolong to a chronic state, which results in mixed pain. Chronic pain persists longer than acute pain (Todd, 2005; Todd et al., 2006). Chronic pain can result in psychopathology disorders such as depression, anxiety, and personality disorder (Dersh et al., 2002), which is called chronic pain syndrome (Knorring, 1989). Chronic Pain in SCD has a direct impact on the quality of life of patients (Kanter and Kruse-Jarres, 2013; Platt et al., 1991; Rees et al., 2010).

A few studies have been conducted to establish the difference in pain perception and response to opioids (Stamer and Stuber, 2007). It was found that genomic variations influence both perception and vulnerability to chronic pain (Mogil, 2004; Mogil and Devor, 2004; Stamer and Stuber, 2007). Furthermore, SNPs of specific genes were associated with variable degrees of pain perception (Diatchenko et al., 2005), leading to the hypothesis that some variants were located in genes related to the inflammatory process of vaso-occlusive painful crises, resulting in nerve and tissue damage and thus the development of secondary pain (Mogil, 2004). Table 1 summarizes the selected genes that have been associated with pain susceptibility in SCD.

HMOX-1 codes for heme oxygenase-1, which is a rate-limiting step in the catalysis of heme. It exhibits a GT dinucleotide repeat in the promoter region, and long repeat lengths (>25 repeats) are associated with decreased activity and inducibility, and therefore higher rates of SCD patient hospitalization, but not directly associated with pain (Bean et al., 2013). It is reported that among African-Americans, a polymorphism in the GTP cyclohydrolase (*GCH1*) on chromosome 14 (rs8007267) is significantly associated with pain crises (Belfer et al., 2014). *GCH1* catalyzes the rate-limiting step for tetrahydrobiopterin synthesis, thus variation in its gene is likely to have pathophysiological roles in pain. Acute pain has been a subject of some studies with the most relevant to SCD referring to an SNP (rs614803) located in a region about 8 kb from the COMM domain-containing protein COMMD7. This polymorphism is significantly associated with painful crises.

COMMD7 modulates many proteins and is associated with NF-kappa-B complex, suppressing its transcriptional activity (Galarneau et al., 2013). Investigations among Egyptians reported the *GSTM1* null allele to be significantly associated with increased risk of severe vaso-occlusive crises (Shiba et al., 2014). *GSTM1* is located on chromosome 1 and catalyzes the addition of glutathione on molecules to increase the antioxidant status, while the *GSTM1* null refers to deletion of this gene. A higher incidence of pain was observed among SCD patients who were carriers of the methylenetetrahydrofolate reductase (*MTHFR*; C677T) polymorphism as well as Factor V Leiden *(FVL*; G191A) polymorphism (Nishank et al., 2013). The vascular endothelial growth factor gene (*VEGFA*) has several mutations of which three, rs2010963, rs833068, and rs3025020, have been associated with vaso-occlusive crisis when inherited in a homozygous state (Al-Habboubi et al., 2012).

Morphine metabolism

Morphine is a member of the opioid family and is mostly used because it is globally available and shows successful clinical efficacy (Adegbola, 2009). Morphine is derived from codeine through the action of *CYP2D6*-catalyzed demethylation. Through the actions of UDP-glucuronosyltransferases, 2B7 and 1A1 (*UGT2B7* and *UGT1A1*), morphine is converted to morphine-3-glucuronide (M3G) and morphine-6-

Table 1. Genomic Variants that Influence Pain in Sickle Cell Disease

Gene	*SNPS*	*Chromosomes locus*	*Association*	*References*
OPRM1	rs1799971	6:154039662	Pain	Joly et al. (2012); Jhun et al. (2015)
HMOX-1	A(GT) VNTR	Chromosome 22	Vaso-occlusive crises	Bean et al. (2013)
GCH1	rs8007267	14:54912273	Pain	Belfer et al. (2014)
COMMD7	rs614803	18:79389574	Painful crisis	Galarneau et al. (2013)
GSTM1	GSTM1 null allele	Chromosome 1	Severe vaso-occlusive crisis	Shiba et al. (2014)
MTHFR	rs1801133 (C677T)	1:11796321	Pain	Nishank et al. (2013)
FVL	rs6025 (G1691A; R506Q)	1:169549811	Pain	Nishank et al. (2013)
VEGFA	rs833068 (G398A)	6:43774790	Vaso-occlusive crisis	Al-Habboubi et al. (2012)
VEGFA	rs2010963	6:43770613	Vaso-occlusive crisis	Al-Habboubi et al. (2012)
VEGFA	rs3025020	6:43781373	Vaso-occlusive crisis	Al-Habboubi et al. (2012)
CYP2D6	rs1065852	22:42130692	Pain and drug metabolism	Joly et al. (2012); Jhun et al. (2015)
COMT	rs4633	22:19962712	Pain	Joly et al. (2012); Jhun et al. (2015)
	rs6269	22:19962429	Pain	
	rs737865	22:19942598	Pain	
CPY3A	rs1057868	7:75985688	Pain	Joly et al. (2012); Jhun et al. (2015)
UGTB7	rs1799971	6:154039662	Pain	Joly et al. (2012); Jhun et al. (2015)
ABCB1	rs1045642	7:87509329	Pain	Jhun et al. (2015)

glucuronide (M6G) through glucuronic acid conjugation. Ultimately, the glucuronidated morphine is effluxed by transporters such as *ABCB1*, *ABCC2*, *ABCC3*, and *SLC01B1*. M6G is responsible for analgesia contribution by binding to μ-opioid receptor; there are arguments about the role of M6G in analgesia that results from morphine (Höllt, 2002; Murthy et al., 2002; Osborne et al., 1990; Smith et al., 1990). M3G has small pull force for opioid receptors (Smith et al., 1990) and it might be responsible for the excitatory effect of morphine (Smith et al., 1990). Blood plasma concentration of morphine and its metabolites is a function of morphine dose and renal clearance, which might be affected by genetic variations, and it is therefore anticipated that variants in the above genes could be associated with variable response to the drug treatment in patients living with SCD. This is supported by evidence from a population study, indicating that the allele variation in genes that are involved in morphine mechanism might regulate the response of opioid analgesic (Lotsch and Geisslinger, 2006).

Genetic variations and morphine metabolism

Patients respond differently to drugs due to variations in genes coding for metabolizing enzymes (Table 2). *UGT2B7* (rs7438135), *OPRM1* (rs1799971), and *ABCB1* (rs1045642) influence the pharmacokinetic and pharmacodynamic measurements and affect the clinical effectiveness of morphine (Adegbola, 2009). *COMT* (rs4633) is not directly involved in the metabolism, but can improve the productivity of morphine. This can occur by influencing μ-opioid receptors and its concentration in different areas of the brain by affecting the neuronal activity; with reduction in *COMT* activity then resulting in sensitivity to pain and morphine (Bockaert and Pin, 1999; Bohn et al., 1999; Kraus et al., 2001; Loh et al., 1998; Matthes et al., 1996; Meineke et al., 2002; Rakvåg et al., 2005; Weinshilboum and Raymond, 1977; Zubieta et al., 2003). Individuals who have the lowest *COMT* activity (met/met variant) have higher sensory and higher effective rates of pain, as well as a more effective state, as the met/met variant reduces the ability to activate the μ-opioid receptor system (Zubieta et al., 2003). This also causes upregulation of the opioid receptors and low concentrations of morphine are required to produce sufficient analgesia to ease the pain (Rakvåg et al., 2005).

ABCB1, which is also known as the *MDR1* transporter gene (Weinshilboum and Raymond, 1977), contributes to the variability in morphine metabolism to produce analgesia by moving the efflux of morphine and M6G across the blood–brain barrier (Darbari et al., 2008). *OPRM1* is the major site

Table 2. Genetic Variants Associated with Morphine Metabolism

Gene	*SNPS*	*Chromosome locus*	*Effect of variant allele*	*References*
UGT2B7	rs7438135	4:69095621	Drug metabolism	Höllt (2002); Duguay et al. (2004)
OPRM1	rs1799971	6:154039662	Pain and mediates analgesic effect of morphine.	Lötsch et al. (2002); Zubieta et al. (2003); Jhun et al. (2015)
ARRB2	rs1045280	17:4719343	Drug metabolism	Ross et al. (2005)
STAT6	rs167769	12:57109992	Drug metabolism	Ross et al. (2005); Jhun et al. (2015)
	rs841718	12:57099213	Drug metabolism	
	rs3024971	12:57099944	Drug metabolism	
COMT	rs4633	22:19962712	Drug metabolism and pain	Zubieta et al. (2003); Jhun et al. (2015)
ABCB1	rs1045642	7:87509329	Responsible for analgesia	Meineke et al. (2002); Jhun et al. (2015)

of action for most opioid analgesics, including morphine (Adegbola, 2009; Beyer et al., 2004; Lotsch and Geisslinger, 2006). This gene is responsible for both pain response and opioid addiction (Adegbola, 2009; Compton et al., 2003). Each individual has different responses to morphine due to polymorphisms in *OPRM1*, which affect the functioning and expression of the binding site (Adegbola, 2009; Chou et al., 2006; Klepstad et al., 2004; Lotsch and Geisslinger, 2006; Mantione et al., 2005; Stamer and Stuber, 2007); and *OPRM1* has two SNPS; A118G and C17T, with A118 being the one that is a commonly identified SNP (Adegbola, 2009; Bond et al., 1998). There is therefore an urgent need to explore the knowledge on pharmacogenomics on morphine metabolism among the population of people affected by SCD.

Pharmacogenomics of HU

HU is the only available treatment for induction of HbF in patients living with SCD that has been approved by both the FDA in 1998 and by the European Medicines Agency in 2007. It was also mentioned as an effective treatment for both adult and children with SCD by the National Institutes of Health (Officer of Medical Applications of Research) (NIH-OMAR) and the Agency of Healthcare Research and Quality (AHRQ) (Herrick, 2000; Loh et al., 1998; Weatherall et al., 2005). HU is an oral, S-phase-specific cytotoxic, antimetabolic, and antineoplastic drug treatment. It is a strong inhibitor of a universal enzyme called ribonucleotide reductase (Elford, 1968; Modell and Darlison, 2008). In 1984, the first clinical application of HU in hemoglobinopathies successfully demonstrated a swift and vivid increase in HbF concentration within immature red blood cells called reticulocytes (Platt et al., 1984).

Besides increasing HbF, HU also plays an important clinical role by increasing the concentration of hemoglobin and simultaneously decreasing white blood cells, absolute neutrophil count, absolute reticulocyte count, and platelets (Charache et al., 1992; de Montalembert et al., 2006; Kinney et al., 1999; Thornburg et al., 2009; Zimmerman et al., 2004). Treatment of HU is associated with a decrease in the frequency of pain episodes, acute chest syndrome, hospitalization, and the need for a blood transfusion (Charache et al., 1995).

The reduction of the clinical phenotype results in increase of efficiency in survival rates and life expectancy among SCD patients (Nagel et al., 1985; Voskaridou et al., 2010; Zago et al., 2000). It may also provide protection against cerebrovascular disease (Zimmerman et al., 2007), long-term drug safety, capacity to prevent organ damage, and reduced morbidity and mortality in school-age children (Kinney et al., 1999), toddlers (Hankins et al., 2005; Thornburg et al., 2009), and infants (Alvarez et al., 2012). HU also helps with related complications of SCD such as stroke prevention, priapism, and pulmonary hypertension (DeBaun, 2014). Maximum tolerated dose for various phenotypes was observed to be different for patients using HU, showing that patients respond differently to HU (Charache et al., 1992; Heeney and Ware, 2008; Ware et al., 2011).

Genetic variation in HU treatment response

Induced HbF levels range from 10% to greater than 30% (Kinney et al., 1999; Zimmerman et al., 2004) among patients with SCD, highlighting the variation in response to HU. This is due to pharmacogenomic interactions (Steinberg et al., 2003). Previous studies have shown that haplotypes in the *HBB* gene cluster that are associated with SCD could possibly affect the clinical response to HU, likely refereed by their genetically determined effect on the HbF level (Adekile, 2011; Friedrisch et al., 2008). *XMNL-HHBG2* (rs7482144) is associated with high level of HbF in response to HU drug treatment in both SCD and β-thalassemia individuals

Table 3. Genomic Variants Associated with Hydroxyurea-Induced HbF Level

Gene	*SNPs*	*Chromosome: locus*	*References*
HBB	rs7482144	11:5254939	Friedrisch et al. (2008); Adekile (2011)
BCL11A	rs1427407	2:60490908	Ware et al. (2011); Ware (2013); Friedrisch et al. (2008); Adekile (2011)
	rs4671393	2:60491212	
	rs7606173	3:60493111	
	rs7557939	2:60494212	
	rs1186868	2:61764103	
ARG1/2	rs2295644	14:67599842	Friedrisch et al. (2008); Adekile (2011)
	rs17599586	6:131583579	
	rs28384513	6:135055071	
HBS1L-MYB	rs9399137	6:135097880	Friedrisch et al. (2008); Adekile (2011)
SAR1	rs2310991	3:142444839	Kumkhaek et al. (2008); Zhu et al. (2014)
	rs4282891	10:70171890	
	rs76901216	10:70170313	
SALL2	rs61743453	14:21523209	Sheehan et al. (2013)
FLT1	rs2182008	13:28412924	Ma et al. (2007)
	rs8002446	13:28423263	
	rs9319428	13:28399484	
	rs3751395	13:28384818	
	rs2387634	13:28416291	
TOX	rs826729	8:58826354	Ma et al. (2007)
	rs765587	8:58878344	
	rs9693712	8:59034864	
	rs172652	8:59045582	
	rs380620	8:59069973	
	rs2693430	8:58812489	
	rs12155519	8:58936271	
ARG2	rs10483801	14:67650289	Ma et al. (2007)
	rs10483802	14:67650704	
NOS1	rs816361	12:117217326	Ma et al. (2007)
	rs7977109	12:117292535	
	rs7309163	12:117291469	
NOS2A	rs1137933	17:27778906	Ma et al. (2007)
	rs944725	17:27782545	
MAP3K5	rs9376230	6:136781227	Ma et al. (2007)
	rs9483947	6:136784262	
PDE7B	rs11154849	6:136032167	Ma et al. (2007)
	rs9376173	6:136038308	
	rs1480642	6:136178390	
	rs487278	6:136180690	
HAO2	rs10494225	1:119375480	Ma et al. (2007)
KLF10	rs3191333	8:102649991	Borg et al. (2012)

SNP, single-nucleotide polymorphism.

(Alebouyeh et al., 2004; Dixit et al., 2005; Yavarian et al., 2004). Research provides some evidences that the effect of HU on HbF level could act through other HbF-promoting loci such as *BCL11A* (Ware et al., 2011). *BCL11A* is central to the fetal switch. It is coexpressed with *SOX-6* as well as directly interacting and co-occupying the β-globin loci. It also has an association with the Mi-2/nucleosome remodeling and deacetylase (NuRD) complex for long-range reconformation of the β-globin cluster for the transcriptional silencing of γ-globin (Xu et al., 2010).

Besides *BCL11A*, from DNA structural alteration to sequence modification, the secretion-associated and ras-related protein (SAR-1) has been shown to play a significant role in γ-globin regulation (Zhu et al., 2014) and three SNPs in the *SAR-1a* promoter sequence have been associated with HbF level in the peripheral blood of SCD patients on HU (Kumkhaek et al., 2008). In addition, in relation to HU responses, it was reported that 17 SNPs are associated with HbF and 20 SNPs with response to HU (Solovieff et al., 2010). It was shown that the absence of *KLF10* (rs3191333) was found to be significantly associated with induction of HbF level in β-thalassemia intermedia compared with the majority patients with β-thalassemia and healthy individuals (Borg et al., 2012). Additional variants, which have been less consistently associated with HU-induced HbF level, are summarized in Table 3.

Discussion

There are emerging data summarized in the present article that indicate that genetic differences in SCD individuals influence the sensitivity to pain (Table 1). There are also a few studies indicating variability in analgesic response that is produced by morphine treatment with some considerable overlap with variants in genes also associated with pain sensitivity (Table 2). However, there are limited data on genetic interindividual variants and responses to morphine treatment in SCD that is directly associated with morphine metabolism. Surprisingly, there are very few data on pharmacogenetics of antalgics and analgesics and specifically opioids used in managing SCD and no data from SSA. Understanding the pharmacogenomics of pain medication in SCD could potentially improve personalized medicine and explore new routes for therapeutic intervention.

Besides antalgics and analgesics, HU drug treatment, which is prescribed for SCD patients, has produced successful results in both children and adults by decreasing pain, blood transfusions, and hospitalization. There is more consistent evidence of the association between several SNPs and HbF levels in response to HU treatment (Table 3). Again, none of the studies were conducted in SSA where the disease burden is highest, further supporting a call for action if the wide use of HU is to be implemented in Africa. Fortunately, there are emerging clinical data from multiple sites on the implementation of HU in Africa in an effort to close this gap, and they have taken the opportunity to perform association studies that could hopefully provide new insight into pharmacogenomics of HU in SCD.

Expert Commentary

Vaso-occlusive painful crises are the main clinical events of SCD and are associated with recurrent and long-term use of antalgics/opioids and HU. The present article has provided evidence of the scarcity of studies investigating the variable response to pain in SCD patients. More consistent studies have addressed the various mechanisms to understand genomic variation affecting the response to HU, but the full understanding of the variable HU-mediated HbF production among individuals affected by SCD remains elusive. Therefore, more research is needed to understand their various mechanisms and pharmacogenomics of both painkiller/opioids and HU to improve the management of people living with SCD.

Five-Year View

The global burden of SCD is anticipated to increase due to an increase in the life expectancy of people living with SCD in the West as well as in Africa. This is due to the emerging implementation of newborn screening and the use of HU treatment and the global migration that is associated with the increase in SCD incidence in countries where this condition was not initially prevalent. The improvement in the treatment of SCD will continue to contribute toward an increase in the global burden of the disease as well as dependency on chronic medications. Therefore, it is expected that most patients living with SCD will have access to pain and HU treatment worldwide, including in SSA. Thus, it could be anticipated in the coming years to observe more global interest in the field of pharmacogenetics of SCD, especially for antalgics and HU.

It is expected that future studies will also give potential explanations regarding the regulatory mechanism level of these drugs and associated gene expression, which could reveal additional pathways to explore novel therapeutic interventions that could maximize benefits while avoiding side effects. As the level of science advances, it is also suspected that there will be more medications that will be developed that are outside HbF induction, for example, to induce stress hematopoiesis, endothelial nitric oxide release, the reduction of leucocyte counts, the reduction of red blood cell adhesion to the endothelium, the reduction in inflammation processes, or medication aiming to reduce blood viscosity to name a few.

There are also topics of SCD pharmacogenomics in need of more research that have not been discussed in the current article, such as those related to recurrent blood transfusions and associated immunogenic issues, and chronic use of antibio-prophylaxy with penicillin in SCD. More research on pharmacogenomics in various aspects of treatment of SCD will result, hopefully, in a complete profile and possible algorithm that could be usable for a successful personalized medicine in SCD.

Key Issues

- Vaso-occlusive painful crises are associated with the recurrent pain and long-term use of antalgics/opioid by people living with SCD. Surprisingly, the present article has provided evidence of limited number of studies to understand the variable responses to pain and pharmacogenomics of antalgics and opioids in people living with SCD.
- There has been great progress made toward understanding and identifying key genomic variants in *BCL11A*, *HBS1L-MYB*, or *SAR1* that predispose the response to the HU treatment; however, the complete picture remains elusive.
- The global burden of SCD is anticipated to increase due to increase of the life expectancy of patients in the

West, emerging implementation of newborn screening and the use of HU treatment in Africa, and the global migrations. Therefore, it could be anticipated to see, in the coming years, more global interest in the field of pharmacogenetics of SCD.

- Strikingly, no study has been conducted in SSA where majority of the patients with SCD live. This alerts the broader global life sciences community toward the existing disparities in optimal and ethical targeting of research and innovation investments for SCD specifically and precision medicine and pharmacology research broadly.

Acknowledgments

The student's bursary was funded by the National Research Foundation (NRF) and the National Health Laboratory Services (NHLS), South Africa, to K.M.; A.W. is funded by the NIH, USA, and Grant number 1U01HG007459-01. The funders had no role in study design, data collection, and analysis; decision to publish; or preparation of the manuscript.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Adegbola MA. (2009). Can heterogeneity of chronic sickle-cell disease pain be explained by genomics? A literature review. Biol Res Nurs 11, 81–97.

Adekile AD. (2011). Limitations of Hb F as a phenotypic modifier in sickle cell disease: Study of Kuwaiti Arab patients. Hemoglobin 35, 607–617.

Alebouyeh M, Moussavi F, Haddad-Deylami H, and Vossough P. (2004). Hydroxyurea in the treatment of major β-thalassemia and importance of genetic screening. Ann Hematol 83, 430–433.

Al-Habboubi HH, Mahdi N, Abu-Hijleh TM, Abu-Hijleh FM, Sater MS, and Almawi WY. (2012). The relation of vascular endothelial growth factor (VEGF) gene polymorphisms on VEGF levels and the risk of vasoocclusive crisis in sickle cell disease. Eur J Haematol 89, 403–409.

Alvarez O, Miller ST, Wang WC, et al. (2012). Effect of hydroxyurea treatment on renal function parameters: Results from the multi-center placebo-controlled BABY HUG clinical trial for infants with sickle cell anemia. Pediatr Blood Cancer 59, 668–674.

Ballas SK. (2015). *Sickle Cell Pain*. Lippincott: Williams and Wilkins.

Ballas SK, and Lusardi M. (2005). Hospital readmission for adult acute sickle cell painful episodes: Frequency, etiology, and prognostic significance. Am J Hematol 79, 17–25.

Ballas SK, Gupta K, and Adams-Graves P. (2012). Sickle cell pain: A critical reappraisal. Blood 120, 3647–3656.

Bartolucci P, and Galacteros F. (2012). Clinical management of adult sickle-cell disease. Curr Opin Hematol 19, 149–155.

Bean CJ, Boulet SL, Yang G, et al. (2013). Acute chest syndrome is associated with single nucleotide polymorphism-defined beta globin cluster haplotype in children with sickle cell anaemia. Br J Haematol 163, 268–276.

Belfer I, Youngblood V, Darbari DS, et al. (2014). A GCH1 haplotype confers sex-specific susceptibility to pain crises and altered endothelial function in adults with sickle cell anemia. Am J Hematol 89, 187–193.

Bergman S. (2005). Psychosocial aspects of chronic widespread pain and fibromyalgia. Disabil Rehabil 27, 675–683.

Beyer A, Koch T, Schröder H, Schulz S, and Höllt V. (2004). Effect of the A118G polymorphism on binding affinity, potency and agonist-mediated endocytosis, desensitization, and resensitization of the human mu-opioid receptor. J Neurochem 89, 553–560.

Bockaert J, and Pin JP. (1999). Molecular tinkering of G protein-coupled receptors: An evolutionary success. EMBO J 18, 1723–1729.

Bohn LM, Lefkowitz RJ, Gainetdinov RR, Peppel K, Caron MG, and Lin FT. (1999). Enhanced morphine analgesia in mice lacking beta-arrestin 2. Science 286, 2495–2498.

Bond C, Laforge KS, Tian M, et al. (1998). Single-nucleotide polymorphism in the human mu opioid receptor gene alters ß-endorphin binding and activity: Possible implications for opiate addiction. Proc Natl Acad Sci USA 95, 9608–9613.

Borg J, Phylactides M, Bartsakoulia M, et al. (2012). KLF10 gene expression is associated with high fetal hemoglobin levels and with response to hydroxyurea treatment in β-hemoglobinnopathy patients. Pharmacogenomics 13, 1487–1500.

Bukar AA, Abjah UAM, Kagu MB, et al. (2013). Seroprevalence of parvovirus B19 and its clinical effect among anaemic SCA patients in Northeastern Nigeria. Am J Sci Ind Res 4, 195–200.

Brousseau DC, McCarver DG, Drendel AL, et al. (2007). The effect of CYP2D6 polymorphisms on the response to pain treatment for paediatric sickle cell pain crisis. J Pediatr 150, 623–626.

Canver MC, Smith EC, Sher F, et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527, 192–197.

Charache S, Dover GJ, Moore RD. et al. (1992). Hydroxyurea: Effects on hemoglobin F production in patients with sickle cell anemia. Blood 79, 2555–2565.

Charache S, Terrin ML, Moore RD, et al. (1995). Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. N Engl J Med 332, 1317–1322.

Chaturvedi S, and DeBaun MR. (2016). Evolution of sickle cell disease from a life-threatening disease of children to a chronic disease of adults: The last 40 years. Am J Hematol 91, 5–14.

Chou W, Wang C, Liu P, Liu C, Tseng C, and Jawan, B. (2006). Human opioid receptor A118G polymorphism affects intravenous patient-controlled analgesia morphine consumption after total abdominal hysterectomy. Anesthesiology 105, 334–337.

Colombatti R, Perrotta S, Samperi P, et al. (2013). Organizing national responses for rare blood disorders: The Italian experience with sickle cell disease in childhood. Orphanet J Rare Dis 8, 1.

Compton P, Geschwind DH, and Alarcon M. (2003). Association between human μ-opioid receptor gene polymorphism, pain tolerance, and opioid addiction. Am J Med Genet B Neuropsychiatr Genet 121, 76–82.

Cox SE, Makani J, Soka D, et al. (2014). Haptoglobin, alpha-thalassaemia and glucose-6-phosphate dehydrogenase polymorphisms and risk of abnormal transcranial Doppler among patients with sickle cell anaemia in Tanzania. Br J Haematol 165, 699–706.

Darbari DS, Minniti CP, Rana S, and van den Anker J. (2008). Pharmacogenetics of morphine: Potential implications in sickle cell disease. Am J Hematol 83, 233–236.

de Montalembert M, Brousse V, Elie C, et al. (2006). Long-term hydroxyurea treatment in children with sickle cell disease: Tolerance and clinical outcomes. Haematologica 91, 125–128.

DeBaun MR. (2014). Hydroxyurea therapy contributes to infertility in adult men with sickle cell disease: A review. Expert Rev Hematol 7, 767–773.

Dersh J, Polatin PB, and Gatchel RJ. (2002). Chronic pain and psychopathology: Research findings and theoretical considerations. Psychosom Med 64, 773–786.

Diatchenko L, Slade GD, Nackley AG, et al. (2005). Genetic basis for individual variations in pain perception and the development of a chronic pain condition. Hum Mol Gen 14, 135–143.

Dixit A, Chatterjee T, Mishra P, et al. (2005). Hydroxyurea in thalassemia intermedia—A promising therapy. Ann Hematol 84, 441–446.

Duguay Y, Baar C, Skorpen F, and Guillemette C. (2004). A novel functional polymorphism in the uridine diphosphate-glucuronosyltransferase 2B7 promoter with significant impact on promoter activity. Clin Pharmacol Ther 75, 223–233.

Dzierzynski N, Stojanovic KS, Georgin-Lavialle S, and Lionnet F. (2016). Enjeux et difficultés de la relation entre soignants et patients drépanocytaires au cours de la crise douloureuse aiguë. Rev Med Interne 37, 111–116.

Elford HL. (1968). Effect of hydroxyurea on ribonucleotide reductase. Biochem Biophys Res Commun 33, 129–135.

Friedrisch JR, Prá D, Maluf SW, et al. (2008). DNA damage in blood leukocytes of individuals with sickle cell disease treated with hydroxyurea. Mutat Res Genet Toxicol Environ Mutagen 649, 213–220.

Galarneau G, Coady S, Garrett ME, et al. (2013). Gene-centric association study of acute chest syndrome and painful crisis in sickle cell disease patients. Blood 122, 434–442.

Gibbons C, Geoghegan R, Conroy H, et al. (2015). Sickle cell disease: Time for a targeted neonatal screening programme. Ir Med J 8, 43–45.

Grosse SD, Odame I, Atrash HK, Amendah DD, Piel FB, and Williams TN. (2011). Sickle cell disease in Africa: A neglected cause of early childhood mortality. Am J Prev Med 41, 398–405.

Hamideh D, and Alvarez O. (2013). Sickle cell disease related mortality in the United States (1999–2009). Pediatr Blood Cancer 60, 1482–1486.

Hankins JS, Ware RE, Rogers ZR, et al. (2005). Long-term hydroxyurea therapy for infants with sickle cell anemia: The HUSOFT extension study. Blood 106, 2269–2275.

Heeney MM, and Ware RE. (2008). Hydroxyurea for children with sickle cell disease. Pediatr Clin North Am 55, 483–501.

Herrick JB. (2000). Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. Yale J Biol Med 74, 179–184.

Höllt V. (2002). A polymorphism (A118G) in the μ-opioid receptor gene affects the response to morphine-6-glucuronide in humans. Pharmacogenet Genomics 12, 1–2.

Jhun, EH, Yao Y, He Y, et al. (2015). Prevalence of pain-related single nucleotide polymorphisms in patients of African origin with sickle cell disease. Pharmacogenomics 16, 1795–1806.

Joly P, Gagnieu, MC, Bardel, C, et al. (2012). Genotypic screening of the main opiate-related polymorphisms in a cohort of 139 sickle cell disease patients. Am J Hematol 87, 534–536.

Kanter J, and Kruse-Jarres R. (2013). Management of sickle cell disease from childhood through adulthood. Blood Rev 27, 279–287.

Kinney TR, Helms RW, O'Branski EE, et al. (1999). Safety of hydroxyurea in children with sickle cell anemia: Results of the HUG-KIDS study, a phase I/II trial. Pediatric Hydroxyurea Group. Blood 94, 1550–1554.

Klepstad P, Rakvåg T, Kaasa S, et al. (2004). The 118 A> G polymorphism in the human μ-opioid receptor gene may increase morphine requirements in patients with pain caused by malignant disease. Acta Anaesthesiol Scand 48, 1232–1239.

Knorring LV. (1989). The pathogenesis of chronic pain syndromes. Nord Psykiatr Tidsskr 43, 35–41.

Kraus J, Borner C, Giannini E, et al. (2001). Regulation of mu-opioid receptor gene transcription by interleukin-4 and influence of an allelic variation within a STAT6 transcription factor binding site. J Biol Chem 276, 43901–43908.

Kumkhaek C, James G, Taylor VI, et al. (2008). Fetal haemoglobin response to hydroxycarbamide treatment and sar1a promoter polymorphisms in sickle cell anaemia. Br J Haematol 141, 254–259.

Kunz JB, Awad S, Happich M, et al. (2015). Significant prevalence of sickle cell disease in Southwest Germany: Results from a birth cohort study indicate the necessity for newborn screening. Ann Hematol 95, 397–402.

Lettre G, Sankaran VG, Bezerra MA, et al. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci U S A 105, 11869–11874.

Loh HH, Liu H, Cavalli A, Yang W, Chen Y, and Wei L. (1998). μ Opioid receptor knockout in mice: Effects on ligand-induced analgesia and morphine lethality. Mol Brain Res 54, 321–326.

Lotsch J, Sharke C, Grosch S, et al. (2002). The polymorphism A118G of the human mu-opioid receptor gene decreases the pupil constrictory effect of morphine-6-glucuronide but not that of morphine. Pharmacogenetics 12, 3–9.

Lotsch J, and Geisslinger G. (2006). Current evidence for a genetic modulation of the response to analgesics. Pain 121, 1–5.

Ma Q, Wyszynski DF, Farrell JJ, et al. (2007). Fetal hemoglobin in sickle cell anaemia: Genetic determinants of response to hydroxyurea. Pharmacogenomics J 7, 386–394.

Makani J, Menzel S, Nkya S, et al. (2011). Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. Blood 117, 1390–1392.

Makani J, Ofori-Acquah SF, Nnodu O, Wonkam A, and Ohene-Frempong K. (2013). Sickle cell disease: New opportunities and challenges in Africa. Sci World J 2013.

Makubi A, Soka D, and Makani J. (2012). Moyamoya disease, a rare cause of recurrent strokes in an African sickle cell child: Does hydroxyurea have a role in this context? Int J Child Health Nutr 1, 82–85.

Mantione KJ, Goumon Y, Esch T, and Stefano GB. (2005). Morphine 6beta glucuronide: Fortuitous morphine metabolite or preferred peripheral regulatory opiate? Med Sci Monitr 11, 43–46.

Matthes HW, Maldonado R, Simonin F, et al. (1996). Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ-opioid-receptor gene. Nature 383, 819–823.

McGann PT, Ferris MG, Ramamurthy U, et al. (2013). A prospective newborn screening and treatment program for sickle cell anemia in Luanda, Angola. Am J Hematol 88, 984–989.

Meineke I, Freudenthaler S, Hofmann U, et al. (2002). Pharmacokinetic modelling of morphine, morphine-3-glucuronide

and morphine-6-glucuronide in plasma and cerebrospinal fluid of neurosurgical patients after short-term infusion of morphine. Br J Clin Pharmacol 54, 592–603.
Mmbando BP, Mgaya J, Cox SE, et al. (2015). Negative epistasis between sickle and foetal haemoglobin suggests a reduction in protection against malaria. PLoS One 10, e0125929.
Modell B, and Darlison M. (2008). Global epidemiology of haemoglobin disorders and derived service indicators. Bull WHO 86, 480–487.
Mogil JS. (2004). Complex trait genetics of pain in the laboratory mouse. Pain Res Manag 28, 123–150.
Mogil JS, and Devor M. (2004). Introduction to pain genetics. Pain Res Manag 28, 1–20.
Mtatiro SN, Singh T, Rooks H., et al. (2014). Genome wide association study of fetal hemoglbin in sickle cell anemia in Tanzania. PLoS One 9, e111464.
Murthy BP, Pollack GM, and Brouwer KL. (2002). Contribution of morphine-6-glucuronide to antinociception following Intravenous Administration of Morphine to Healthy Volunteers. J Clin Pharmacol 42, 569–576.
Nagel RL, Fabry ME, Pagnier J, et al. (1985). Hematologically and genetically distinct forms of sickle cell anemia in Africa: The Senegal type and the Benin type. N Engl J Med 312, 880–884.
Nishank SS, Singh MPSS, and Yadav R. (2013). Clinical impact of factor V Leiden, prothrombin G20210A, and MTHFR C677T mutations among sickle cell disease patients of Central India. Eur J Haematol 91, 462–466.
Olabode JO, and Shokunbi WA. (2006). Types of crises in sickle cell disease patients presenting at the haematology day care unit (HDCU), University College Hospital (UCH), Ibadan. West Afr J Med 25, 284–288.
Orkin SH. (2016). Recent advances in globin research using genome-wide association studies and gene editing. Ann NY Acad Sci 1368, 5–10.
Osborne R, Joel S, Trew D, and Slevin M. (1990). Morphine and metabolite behavior after different routes of morphine administration: Demonstration of the importance of the active metabolite morphine-6-glucuronide. Clin Pharmacol Ther 47, 12–19.
Piel FB, Patil AP, Howes RE, et al. (2013). Global epidemiology of sickle haemoglobin in neonates: A contemporary geostatistical model-based map and population estimates. Lancet 381, 142–151.
Pizzo E, Laverty AA, Phekoo KJ, et al. (2015). A retrospective analysis of the cost of hospitalizations for sickle cell disease with crisis in England, 2010/11. J Public Health (Oxf) 37, 529–539.
Platt OS, Orkin SH, Dover G, Beardsley GP, Miller B, and Nathan DG. (1984). Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. J Clin Invest 74, 652–656.
Platt OS, Thorington BD, Brambilla DJ, et al. (1991). Pain in sickle cell disease: Rates and risk factors. N Engl J Med 325, 11–16.
Pule GD, Mowla S, Novitzky N, Wiysonge CS, and Wonkam A. (2015). A systematic review of known mechanisms of hydroxyurea-induced fetal hemoglobin for treatment of sickle cell disease. Expert Rev Hematol 8, 669–679.
Rahimy MC, Gangbo A, Ahouignan G, et al. (2003). Effect of a comprehensive clinical care program on disease course in severely ill children with sickle cell anemia in a sub-Saharan African setting. Blood 102, 834–838.
Rahimy MC, Gangbo A, Ahouignan G, and Alihonou E. (2009). Newborn screening for sickle cell disease in the Republic of Benin. J Clin Pathol 62, 46–48.
Rakvåg TT, Klepstad P, Baar C, et al. (2005). The Val158Met polymorphism of the human catechol-O-methyltransferase (COMT) gene may influence morphine requirements in cancer pain patients. Pain 116, 73–78.
Rees DC, Williams TN, and Gladwin MT. (2010). Sickle-cell disease. Lancet 376, 2018–2031.
Ross JR, Rutter D, Welsh K, et al. (2005). Clinical response to morphine in cancer patients and genetic variation in candidate genes. Pharmacogenomics J 5, 324–336.
Rumaney MB, Bitoungui VJN, Vorster AA, et al. (2014). The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. PLoS One 9, e100516.
Sheehan VA, Luo Z, Flanagan JM, et al. (2013). Genetic modifiers of sickle cell anaemia in the BABY HUG cohort: Influence on laboratory and clinical phenotypes. Am J Hematol 88, 571–576.
Shenoy S. (2011). Hematopoietic stem cell transplantation for sickle cell disease: Current practice and emerging trends. Hematology ASH Educ 2011, 273–279.
Shiba HF, El-Ghamrawy MK, Shaheen IAE, Ali RAE, and Mousa SM. (2014). Glutathione S-transferase gene polymorphisms (GSTM1, GSTT1, and GSTP1) in Egyptian pediatric patients with sickle cell disease. Pediatr Dev Pathol 17, 265–270.
Slade GD, Diatchenko L, Bhalang K, et al. (2007). Influence of psychological factors on risk of temporomandibular disorders. J Dent Res 86, 1120–1125.
Smith MT, Watt JA, and Cramond T. (1990). Morphine-3-glucuronide-a potent antagonist of morphine analgesia. Life Sci J 47, 579–585.
Solovieff N, Milton JN and Steinberg MH. (2010). Fetal hemaglobin in sickle cell anemia: Genome-wide associated studies suggest a regulatory region in the 5′ olfactory receptor gene cluster. Blood 4, 1815–1825.
Stamer UM, and Stuber F. (2007). Genetic factors in pain and its treatment. Curr Opin Anaesthesiol 20, 478–484.
Steinberg MH, Barton F, Castro O, et al. (2003). Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: Risks and benefits up to 9 years of treatment. JAMA 289, 1645–1651.
Steinberg MH, Lu ZH, Barton FB, Terrin ML, Charache S, and Dover GJ. (1997). Fetal hemoglobin in sickle cell anemia: Determinants of response to hydroxyurea. Multicenter Study of Hydroxyurea. Blood 89,1078–1088.
Steinberg MH, McCarthy WF, Castro OP, et al. (2010). The risks and benefits of long-term use of hydroxyurea in sickle cell anemia: A 17.5 year follow-up. Am J Hematol 85, 403–408.
Tekola-Ayele F, and Rotimi CN. (2015). Translational genomics in low- and middle-income countries: Opportunities and challenges. Public Health Genomics 18, 242–247.
Thein SL, and Menzel S. (2009). Discovering the genetics underlying foetal haemoglobin production in adults. Br J Haematol 145, 455–467.
Thornburg CD, Dixon N, Burgett S, et al. (2009). A pilot study of hydroxyurea to prevent chronic organ damage in young children with sickle cell anemia. Pediatr Blood Cancer 52, 609–615.
Todd KH. (2005). Chronic Pain and aberrant drug-related behavior in the emergency department. J Law Med Ethics 33, 761–769.

Todd KH, Green C, Bonham VL, Haywood C, and Ivy E. (2006). Sickle cell disease related pain: Crisis and conflict. J Pain 7, 453–458.

Tshilolo L, Aissi LM, Lukusa D, et al. (2009). Neonatal screening for sickle cell anaemia in the Democratic Republic of the Congo: Experience from a pioneer project on 31 204 newborns. J Clin Pathol 62, 35–38.

Tubman VN, Marshall R, Jallah W, et al. (2016). Newborn screening for sickle cell disease in Liberia: A Pilot Study. Pediatr Blood Cancer 63, 671–676

Vichinsky EP. (1991). Comprehensive care in sickle cell disease: Its impact on morbidity and mortality. Sem Hematol 28, 220–226.

Voskaridou E, Christoulas D, Bilalis A, et al. (2010). The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: Results of a 17-year, single-center trial (LaSHS). Blood 115, 2354–2363.

Ware RE. (2013). Is sickle cell anemia a neglected tropical disease? PLoS Negl Trop Dis 7, e2120.

Ware RE, Despotovic JM, Mortier NA, et al. (2011). Pharmacokinetics, pharmacodynamics, and pharmacogenetics of hydroxyurea treatment for children with sickle cell anemia. Blood 118, 4985–4991.

Weatherall DJ, and Clegg JB. (2008). *The Thalassaemia Syndromes*, 4th ed. John Wiley & Sons, New Jersey, USA.

Weatherall D, Hofman K, Rodgers G, Ruffin J, and Hrynkow S. (2005). A case for developing North-South partnerships for research in sickle cell disease. Blood 105, 921–923.

Weinshilboum RM, and Raymond FA. (1977). Inheritance of low erythrocyte catechol-o-methyltransferase activity in man. Am J Hum Genet 29, 125–135.

Williams TN, Mwangi TW, Wambua S, et al. (2005). Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases. Int J Infect Dis 192, 178–186.

Wonkam A, Bitoungui VJN, Vorster AA, et al. (2014a). Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. PLoS One 9, e92506.

Wonkam A, Mba CZ, Mbanya D, Ngogang J, Ramesar R, and Angwafo FF. (2014b). Psychosocial stressors of sickle cell disease on adult patients in Cameroon. J Genet Couns 23, 948–956.

Wonkam A, Ponde C, Nicholson N, Fieggen K, Ramesar R, and Davidson A. (2012). The burden of sickle cell disease in Cape Town. SAMJ 102, 753–754.

Wonkam A, Rumaney MB, Ngo Bitoungui VJ, Vorster AA, Ramesar R, and Ngogang J. (2014c). Coinheritance of sickle cell anemia and α-thalassemia delays disease onset and could improve survival in cameroonian's patients (Sub-Saharan Africa). Am J Hematol 89, 664–665.

Xu J, Peng C, Sankaran VG, et al. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science 334, 993–996.

Xu J, Sankaran VG, Ni M, et al. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev 24, 783–798.

Yanni E, Grosse SD, Yang Q, and Olney RS. (2009). Trends in pediatric sickle cell disease-related mortality in the United States, 1983–2002. J Pediatr 154, 541–545.

Yavarian M, Karimi M, Bakker E, Harteveld CL, and Giordano PC. (2004). Response to hydroxyurea treatment in Iranian transfusion-dependent beta-thalassemia patients. Haematologica 89, 1172–1178.

Zago M, Silva W, Dalle B, et al. (2000). Atypical beta(s) haplotypes are generated by diverse genetic mechanisms. Am J Hematol 63, 79–84.

Zhu J, Chin K, Aerbajinai W, Kumkhaek C, Li H, and Rodgers GP. (2014). Hydroxyurea-inducible SAR1 gene acts through the Gialpha/JNK/Jun pathway to regulate gamma-globin expression. Blood 124, 1146–1156.

Zimmerman SA, Schultz WH, Burgett S, Mortier NA, and Ware RE. (2007). Hydroxyurea therapy lowers transcranial Doppler flow velocities in children with sickle cell anemia. Blood 110, 1043–1047.

Zimmerman SA, Schultz WH, Davis JS, et al. (2004). Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. Blood 103, 2039–2045.

Zubieta JK, Heitzeg MM, Smith YR, et al. (2003). COMT val158met genotype affects mu-opioid neurotransmitter responses to a pain stressor. Science 299, 1240–1243.

Zur B. (2016). Increase in genetically determined anemia as a result of migration in Germany. Der Internist 57, 444–445.

Address correspondence to:
Prof. Ambroise Wonkam, MD, DMedSc, PhD
Division of Human Genetics
Department of Medicine
Faculty of Health Sciences
University of Cape Town
Anzio Road Observatory
Cape Town 7925
Republic of South Africa

E-mail: ambroise.wonkam@uct.ac.za

Abbreviations Used

FDA = Food and Drug Administration
HbS = hemoglobin sickle
HU = hydroxyurea
SCD = sickle cell disease
SNP = single-nucleotide polymorphism
SSA = sub-Saharan Africa

Supplementary data

Table S1. Overview of Studies included in the review

Author	*Title*	Year of Study	Study Country	Age Group studied	Sample size	Study type
Adegbola MA	*Can heterogeneity of chronic sickle-cell disease pain be explained by genomics?*	2009	United State of America			Review study
Adekile AD.	*Limitations of Hb F as a phenotypic modifier in sickle cell disease: study of Kuwaiti Arab patients*	2011	Kuwait			Longitudinal clinical studies
Alebouyeh M, et al	*Hydroxyurea in the treatment of major β-thalassemia and importance of genetic screening.*	2004	Germany	6-33 years old	45	Cross-section study
Al-Habboubi HH, et al.	*The relation of vascular endothelial growth factor (VEGF) gene polymorphisms on VEGF levels and the risk of vaso –occlusive crisis in sickle cell disease.*	2012	Bahrain	Children	324	Retrospective case control study
Ballas SK	*Sickle cell pain*	2015	United State of America	Adult		Prospective study
Ballas SK, and Lusardi M.	*Hospital readmission for adult acute sickle cell painful episodes: frequency, etiology, and prognostic significance.*	2005	United State of America	Adult	182	Prospective study longitudinal and observation cohort study
Ballas SK, et al.	*Sickle cell pain: a critical reappraisal.*	2012	United State of America	Adult		Prospective study
Bartolucci P, and Galacteros F	*Clinical management of adult sickle-cell disease.*	2012	France			Review study

Bean CJ, et al.	*Acute chest syndrome is associated with single nucleotide polymorphism-defined beta globin clUnited State of Americater haplotype in Children with sickle cell anaemia.*	2013	North America and Europe	Children	820	Prospective study
Belfer I, et al.	*A GCH1 haplotype confers sex-specifics susceptibility to pain crises and altered endothelial function in adults with sickle cell anemia.*	2014	United State of America	Adult	228	Prospective study
Bergman S	*Psychosocial aspects of chronic widespread pain and fibromyalgia*	2005	Sweden	Adult	3928	cross-sectional study
Beyer A, eat al.	*Effect of the A118G polymorphism on binding affinity, potency and agonist-mediated endocytosis, desensitization, and resensitization of the human mu-opioid receptor*	2004	Germany		293 cells	Prospective study
Bockaert J, and Pin JP	*Molecular tinkering of G protein-coupled receptors: an evolutionary success*	1999	France			Review
Bohn LM, et al.	*Enhanced Morphine Analgesia in Mice Lacking b-Arrestin 2.*	1999	United State of America			Prospective study
Bond C, et al,	*Single-nucleotide polymorphism in the human mu opioid receptor gene alters ß-endorphin binding and activity: Possible implications for opiate addiction.*	1998	New York City	Adults	113 former heroin addicts and 39 no history of drug	Prospective study
Bukar AA, et al.	*Seroprevalence of parvovirus B19 and its clinical effect among anaemic SCA patients in North Eastern Nigeria.*	2013	Nigeria	Adults	45 Case and 45 controls	Prospective study

Canver MC, et al.	*BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis.*	2015	United State of America			Longitudinal study
Charache S, et al.	*Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia.*	1995	United States and Canada	Adults	152 on Hydroxy urea and 147 on Placebo	Longitudinal study
Charache S, et al.	*Hydroxyurea: effects on haemoglobin F production in patients with sickle cell anemia.*	1992	United State of America	Adults	49	Longitudinal study
Chaturvedi S. and DeBaun MR	*Evolution of sickle cell disease from a life-threatening disease of Children to a chronic disease of adults: The last 40 years.*	2016	Tennessee.	Adults		Review
Chou W, et al,	*Human opioid receptor A118G polymorphism affects intravenous patient-controlled analgesia morphine consumption after total abdominal hysterectomy.*	2006	Taiwan	Adults	80	Prospective study
Colombatti R, et al.	*Organizing national responses for rare blood disorders: the Italian experience with sickle cell disease in childhood.*	2013	Italy		54	Longitudinal study
Compton P, et al.	*Association between human μ-opioid receptor gene polymorphism, pain tolerance, and opioid addiction.*	2003	Los Angeles	Adult	50 Cases and 59 controls	Longitudinal study
Cox SE, et al.	*Haptoglobin, alpha-thalassaemia and glucose-6-phosphate dehydrogenase polymorphisms and risk of abnormal transcranial Doppler among patients with sickle cell anaemia in Tanzania.*	2014	Tanzania	Children	601	Longitudinal study

Darbari DS, et al.	*(2008). Pharmacogenetics of morphine: Potential implications in sickle cell disease.*	2008	United State of America			Review study
de Montalembert M, et al.	*Long-term hydroxyurea treatment in Children with sickle cell disease: tolerance and clinical outcomes.*	2006	France.	Children	225	Longitudinal study
DeBaun MR	*Hydroxyurea therapy contributes to infertility in adult men with sickle cell disease.*	2014	Tennessee.	Adult		Review study
Dersh J, et al.	*Chronic pain and psychopathology: research findings and theoretical considerations.*	2002	Texas			Review study
Diatchenko L, et al.	*Genetic basis for individual variations in pain perception and the development of a chronic pain condition.*	2005	North Carolina	Adult	202	Prospective study cohort study
Dixit A, et al.	*Hydroxyurea in thalassemia intermedia—a promising therapy.*	2005	India	Children and Adults	37	Prospective studys
Dzierzynski N, et al	*Enjeux et difficultés de la relation entre soignants et patients drépanocytaires au cours de la crise douloure United State of Americae aiguë.*	2016	France			Review study
Elford HL.	*Effect of hydroxyurea on ribonucleotide reductase.*	1968	Michigan		1	Prospective study
Friedrisch JR, et al.	DNA damage in blood leukocytes of individuals with sickle cell disease treated with hydroxyurea.	2008	Brazil	Children and Adults	18	Prospective study
Galarneau G, et al.	*Gene-centric association study of acute chest syndrome and*	2013	United State of America	Adults	318	Prospective study

	painful crisis in sickle cell disease patients.					
Gibbons C, et al.	*Sickle cell disease: time for a targeted neonatal screening programme.*	2015	United Kingdom	Children	77	Cross-section study
Grosse SD, et al.	*Sickle cell disease in Africa: a neglected cause of early childhood mortality.*	2011	Sweden		62	Cross-section study
Hamideh D,et al.	*Sickle cell disease related mortality in the United States (1999–2009).*	2013	United States	1-85 years	50	Prospective study
Hankins JS, et al.	*2005. Long-term hydroxyurea therapy for infants with sickle cell anemia: the HUSOFT extension study.*	2005	United State of America	Children	21	Longitudinal study
Heeney MM, and Ware RE.	*Hydroxyurea for Children with sickle cell disease.*	2008	United State of America	Children		Cross-section study
Herrick JB.	*Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia.*	2000	United State of America		1	Case study
Höllt V.	*A polymorphism (A118G) in the μ-opioid receptor gene affects the response to morphine-6-glucuronide in humans.*	2002	Germany			Review study
Kanter J, and Kruse-Jarres R.	*Management of sickle cell disease from childhood through adulthood.*	2013	United State of America	Children		Review study
Kinney TR, et al.	*Safety of hydroxyurea in Children with sickle cell anemia: results of the HUG-KIDS study, a phase I/II trial.*	1999	United State of America	Children	84	Prospective study

Klepstad P, et al	*The 118 A> G polymorphism in the human μ-opioid receptor gene may increase morphine requirements in patients with pain caused by malignant disease.*	2004	Norway	Adults	207	Prospective study
Knorring LV.	*The pathogenesis of chronic pain syndromes.*	1989				Review study
Kraus, et al.	*Regulation of mu-opioid receptor gene transcription by interleukin-4 and influence of an allelic variation within a STAT6 transcription factor binding site.*	2001	Germany	Study on cell		Prospective study
Kunz JB,	*Significant prevalence of sickle cell disease in Southwest Germany: results from a birth cohort study indicate the necessity for new-born screening.*	2015	urban and rural areas in Southwest Germany	New-borns	37,838	Cross-section study
Lettre G, et al.	*DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal haemoglobin levels and pain crises in sickle cell disease.*	2008	Brazil		350	Prospective study
Lotsch J, and Geisslinger G.	*Current evidence for a genetic modulation of the response to analgesics*	2006	Germany			Review study
Loh HH, et al.	*μ Opioid receptor knockout in mice: effects on ligand-induced analgesia and morphine lethality.*	2006	Germany			Review study
Makani J, et al.	*Genetics of fetal haemoglobin in Tanzanian and British patients with sickle cell anemia.*	2011	Tanzania	5-45 years	1045	Prospective study

Makani J, et al.	*Sickle cell disease: new opportunities and challenges in Africa.*	2013	Tanzania			Review study
Makubi A, et al.	*Moyamoya Disease, a Rare Cause of Recurrent Strokes in an African Sickle Cell Child: Does hydroxyurea have a Role in this Context?*	2012	Kuwait	14 year old	1	Case study
Mantione KJ, et al.	*Morphine 6beta glucuronide: fortuitous morphine metabolite or preferred peripheral regulatory opiate?*	2005	United State of America			Review study
Matthes HW, et al.	*Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ-opioid-receptor gene.*	1996	France	Study on cell		Cross-section study
McGann PT, et al.	*A Prospective study new-born screening and treatment program for sickle cell anemia in Luanda, Angola.*	2013	Angola	New-borns	36453	Prospective study
Meineke I, et al.	*Pharmacokinetic modelling of morphine, morphine-3-glucuronide and morphine-6-glucuronide in plasma and cerebrospinal fluid of neurosurgical patients after short-term infusion of morphine.*	2002	Germany	Adults	19	Prospective study
Mmbando BP, et al.	*Negative Epistasis between Sickle and Foetal Haemoglobin Suggests a Reduction in Protection against Malaria.*	2015	Tanzania	0-70 years	2049	Prospective study
Modell B, and Darlison M.	*Global epidemiology of haemoglobin disorders and derived service indicators. Bull.*	2008	United Kingdom			Cross-section study

Mogil JS.	*Complex trait genetics of pain in the laboratory moue.*	2004				Prospective study
Mogil JS, and Devor M.	*Introduction to pain genetics.*	2004				Prospective study study/ Cross-section study
Mtatiro SN, et al.	*Genome wide association study of fetal hemoglbin in sickle cell anemia in Tanzania.*	2014	Tanzania	Children	1213	Prospective study
Murthy BP, et al.	*Contribution of Morphine-6-Glucuronide to Antinociception following Intravenous Administration of Morphine to Healthy Volunteers.*	2002	North Carolina	Adult	8	Prospective study
Nagel RL, et al	*Hematologically and genetically distinct forms of sickle cell anemia in Africa: the Senegal type and the Benin type.*	1985	France and Benin			Review study
Nishank SS, et al.	*Clinical impact of factor V Leiden, prothrombin G20210A, and MTHFR C677T mutations among sickle cell disease patients of Central India*	2013	India	Adult	150 cases and 150 Controls	Prospective study
Olabode JO, and Shokunbi WA	*Types of crises in sickle cell disease patients presenting at the haematology day care unit (HDCU), University College Hospital (UCH), Ibadan*	2006	India	Adult	545	Prospective study
Orkin SH.	*Recent advances in globin research using genome-wide association studies and gene editing*	2016	United State of America			Cross-section study

Osborne R, et al	*Morphine and metabolite behaviour after different routes of morphine administration: demonstration of the importance of the active metabolite morphine-6-glucuronide*	1990	England	Adult	10	Prospective study
Piel FB, et al.	*Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates.*	2013	United Kingdom			Cross-section study
Pizzo E, et al	*A retrospective analysis of the cost of hospitalizations for sickle cell disease with crisis in England, 2010/11*	2015	England	Children and Adult	6077	Prospective study
Platt OS, et al.	*Pain in sickle cell disease: rates and risk factors.*	1991	United State of America	0-66 years	3578	Prospective study
Platt OS, et al.	*Hydroxyurea enhances fetal haemoglobin production in sickle cell anemia.*	1984	United State of America	17 years and 23 years	2	Case study
Pule GD, et al	*A systematic review of known mechanisms of hydroxyurea-induced fetal haemoglobin for treatment of sickle cell disease*	2015	South Africa			Revie Study
Rahimy MC, et al.	*Effect of a comprehensive clinical care program on disease course in severely ill Children with sickle cell anemia in a sub-Saharan African setting*	2003	Benin	8 months-12 years	236	Prospective study
Rahimy MC, et al.	*New-born screening for sickle cell disease in the Republic of Benin.*	2009	Benin	New-borns	3000	Prospective study

Rakvåg TT, et al.	*The Val158Met polymorphism of the human catechol-O-methyltransferase (COMT) gene may influence morphine requirements in cancer pain patients.*	2005	Norway	Adults	207	Prospective study
Rees DC, et al.	*Sickle-cell disease*	2010	United Kingdom			Review study
Rumaney MB, et al.	*The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival*	2014	Cameroon	Adults	161 cases and 103 controls	Prospective study
Shenoy S.	*Hematopoietic stem cell transplantation for sickle cell disease: current practice and emerging trends.*	2011	United State of America			Review study
Shiba HF, et al.	*Glutathione S-transferase gene polymorphisms (GSTM1, GSTT1, and GSTP1) in Egyptian paediatric patients with sickle cell disease*	2014	Egypt	3 - 18 years	50	Prospective study
Slade GD, et al.	*Influence of psychological factors on risk of temporomandibular disorders*	2007	Australia	18–34 years	171	Longitudinal study
Smith MT, et al.	*Morphine-3-glucuronide-a potent antagonist of morphine analgesia*	1990	Australia			Longitudinal study
Stamer UM, and Stuber F.	*Genetic factors in pain and its treatment*	2007	Germany			Revie Study
Steinberg MH, et al.	*Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and*	2003	United States and Canada.	Adults	152 on hydroxy urea	Longitudinal study

	benefits up to 9 years of treatment				147 on placebo	
Steinberg MH, et al.	*The risks and benefits of long-term United State of Americae of hydroxyurea in sickle cell anemia: A 17.5 year follow-up.*	2010	United State of America			Longitudinal study
Steinberg MH,et al	*Fetal haemoglobin in sickle cell anemia: determinants of response to hydroxyurea. Multicentre Study of Hydroxyurea*	1997	United State of America.	Adults	150 on hydroxy urea 145 on placebo	Longitudinal study
Tekola-Ayele F, and Rotimi CN.	*Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges.*	2015	United State of America.			Prospective study
Thein SL, and Menzel S.	*Discovering the genetics underlying foetal haemoglobin production in adults*	2009	United Kingdom	Adults		Review study
Thornburg CD,	*A pilot study of hydroxyurea to prevent chronic organ damage in young Children with sickle cell anemia*	2009	United State of America	Children	14	Pilot study
Todd KH	*Chronic Pain and Aberrant Drug-Related Behaviour in the Emergency Department.*	2005	New York City			Review study
Todd KH, et al	*Sickle cell disease related pain: crisis and conflict*	2006	United State of America	17 year old	1	Case study
Tshilolo L, et al.	*Neonatal screening for sickle cell anaemia in the Democratic Republic of the Congo: experience from a pioneer project on 31 204 New-borns*	2009	Congo	New-borns	31204	Longitudinal study

Tubman VN, et al.	*New-born Screening for Sickle Cell Disease in Liberia: A Pilot Study*	2016	United State of America	New-borns	2785	Pilot study
Vichinsky EP.	*Comprehensive care in sickle cell disease: its impact on morbidity and mortality.*	1991	United State of America			Prospective study
Voskaridou E, et al.	*The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: results of a 17-year, single-centre trial*	2010	Greece	20-76 years	330	Prospective study
Wang W, and Thompson B.	*Hydroxyurea treatment of infants with sickle cell anemia: results of the BABY HUG study.*	2010	United State of America	New-borns		Prospective study
Ware RE.	*Is sickle cell anemia a neglected tropical disease?*	2013	United State of America			Review study
Ware RE, et al	*harmacokinetics, pharmacodynamics, and pharmacogenetics of hydroxyurea treatment for Children with sickle cell anemia.*	2011	United State of America	Children	174	Prospective study
Weatherall DJ, and Clegg JB.	*The thalassaemia syndromes*	2008	United State of America			Prospective study
Weatherall D, et al	*A case for developing North-South partnerships for research in sickle cell disease*	2005	United Kingdom			Review study
Weinshilboum RM, and Raymond FA.	*Inheritance of low erythrocyte catechol-o-methyltransferase activity in man.*	1977	United State	16-18 years	373	Prospective study
Williams TN, et al	*Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases*	2005	Kenya	Children	3455	Prospective study cohort study

Wonkam A, et al	*Association of variants at BCL11A and HBS1L-MYB with haemoglobin F and hospitalization rates among sickle cell patients in Cameroon.*	2014	Cameroon	5–54 years	610	Prospective study
Wonkam A, et al	*Psychosocial stressors of sickle cell disease on adult patients in Cameroon*	2014	Cameroon	20–30 years	83	Prospective study
Wonkam A, et al	*The burden of sickle cell disease in Cape Town.*	2012	South Africa	2 - 21 years	58	Prospective study
Wonkam A, et al	*Coinheritance of sickle cell anemia and α-thalassemia delays disease onset and could improve survival in Cameroonian's patients (Sub-Saharan Africa).*	2014	Cameroon	Adults	161 cases and 93 controls	Prospective study
Xu J, et al	*Correction of sickle cell disease in adult mice by interference with fetal haemoglobin silencing.*	2011	United State of America	On Animals		Prospective study
Xu J, et al	*Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6*	2010	United State of America	On cells		Prospective study
Yanni E, et al.	*Trends in paediatric sickle cell disease-related mortality in the United States, 1983-2002.*	2009	United State of America	0-14 year	276 158	Prospective study
Yavarian M, et al.	*Response to hydroxyurea treatment in Iranian transfusion-dependent beta-thalassemia patients.*	2004	Iran		133	Prospective study
Zago M, et al.	*Atypical ss^ s Haplotypes are Generated by Diverse Genetic Mechanisms.*	2000	Brazil			Review study

Zhu J, et al	*Hydroxyurea-inducible SAR1 gene acts through the Gialpha/JNK/Jun pathway to regulate gamma-globin expression.*	2014	United State of America	On cells		Prospective study
Zimmerman SA, et al.	*Hydroxyurea therapy lowers transcranial Doppler flow velocities in Children with sickle cell anemia.*	2007	United State of America	Children	102 for phase 1 and 59 for phase 2	Prospective study
Zimmerman SA, et al.	*Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in Children with sickle cell disease.*	2004	United State of America	Children	122	Prospective study
Zubieta JK, ET AL.	*COMT val158met genotype affects mu-opioid neurotransmitter responses to a pain stressor*	2003	United State of America			Review study
Zur B	*Increase in genetically determined anemia as a result of migration in Germany.*	2016	Germany			Prospective study

Chapter 3: Results: Original Publications

3.1 Wonkam A, **Mnika K**, Ngo Bitoungui VJ,Chemegni BC, Chimusa ER, Dandara C, and Kengne AP. Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. Br J Haematol. 2018 Jan; 180(1): 134–146.

Abstract

We investigated the clinical and genetic predictors of painful vaso-occlusive crises (VOC) in SCD in Cameroon. Socio-demographics, clinical variables/events and haematological indices were acquired. Genotyping was performed for 40 variants in 17 pain-related genes, three fetal haemoglobin (HbF)-promoting loci, two kidney dysfunctions-related genes, and *HBA1*/*HBA2* genes. Statistical models using regression frameworks were performed in R®. A total of 436 HU and opioid-naïve patients were studied; median age was 16 years. Female sex, body mass index, Hb/HbF, blood transfusions, leucocytosis and consultation or hospitalisation rates significantly correlated with VOC. Three pain-related genes variants correlated with VOC (*CACNA2D3*-rs6777055, p = 0 .025; *DRD2*-rs4274224, p = 0 .037; *KCNS1*-rs734784, p = 0.01). Five pain-related genes variants correlated with hospitalisation/consultation rates. (*COMT*-rs6269, p = 0.027; *FAAH*-rs4141964, p = 0.003; *OPRM1*-rs1799971, p = 0.031; *ADRB2-rs*1042713; p < 0.001; *UGT2B7*-rs7438135, p = 0.037). The 3.7 kb *HBA1/HBA2* deletion correlated with increased VOC (p = 0.002). HbF-promoting loci variants correlated with decreased hospitalisation (*BCL11A*-rs4671393, p = 0.026; *HBS1L-MYB*-rs28384513, *p* = 0.01). *APOL1* G1/G2 correlated with increased hospitalisation (p = 0.048). This first study from Africa has provided evidence supporting possible development of a genetic risk model for pain in SCD.

Nature of Publication: Original Full Journal Article

Journal/Publisher: British Journal of Haematology, Peer reviewed

Candidate contribution: clinical and genetic predictors of painful vaso-occlusive crises and analysis of the results. Contributed to writing in part with relation to clinical and genetic predictors analysis.

Co-Author contribution:

AW, KM: Performed the experiment, analysed data, wrote and revised the manuscript.

VJNB: Patients recruitment, sample and clinical data collection and processing

CD, BCC, APK, EC: Revised manuscript

VJNB, BCC, CD, APK, AW: Contributed to reagents/materials/analytic tools

AW, KM: Conceived and designed the experiments, analysed data, wrote and revised and approved manuscript

bjh research paper

Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon

Ambroise Wonkam,[1,*] Khuthala Mnika,[1,*] Valentina J. Ngo Bitoungui,[2] Bernard Chetcha Chemegni,[2] Emile R. Chimusa,[1] Collet Dandara[1] and Andre P. Kengne[3]

[1]*Division of Human Genetics, Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa,* [2]*Faculty of Medicine and Biomedical Sciences, University of Yaoundé, Yaoundé, Cameroon and* [3]*Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa*

Received 25 July 2017; accepted for publication 17 September 2017

Correspondence: Professor Ambroise Wonkam, Division of Human Genetics, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, Cape Town, Republic of South Africa.
E-mail: ambroise.wonkam@uct.ac.za
*These two authors contributed equally to the manuscript.

Summary

We aimed to investigate the clinical and genetic predictors of painful vaso-occlusive crises (VOC) in sickle cell disease (SCD) in Cameroon. Socio-demographics, clinical variables/events and haematological indices were acquired. Genotyping was performed for 40 variants in 17 pain-related genes, three fetal haemoglobin (HbF)-promoting loci, two kidney dysfunctions-related genes, and *HBA1/HBA2* genes. Statistical models using regression frameworks were performed in R®. A total of 436 hydroxycarbamide- and opioid-naïve patients were studied; median age was 16 years. Female sex, body mass index, Hb/HbF, blood transfusions, leucocytosis and consultation or hospitalisation rates significantly correlated with VOC. Three pain-related genes variants correlated with VOC (*CACNA2D3*-rs6777055, $P = 0{\cdot}025$; *DRD2*-rs4274224, $P = 0{\cdot}037$; *KCNS1*-rs734784, $P = 0{\cdot}01$). Five pain-related genes variants correlated with hospitalisation/consultation rates. (*COMT*-rs6269, $P = 0{\cdot}027$; *FAAH*-rs4141964, $P = 0{\cdot}003$; *OPRM1*-rs1799971, $P = 0{\cdot}031$; *ADRB2-rs*1042713; $P < 0{\cdot}001$; *UGT2B7*-rs7438135, $P = 0{\cdot}037$). The 3·7 kb *HBA1/HBA2* deletion correlated with increased VOC ($P = 0{\cdot}002$). HbF-promoting loci variants correlated with decreased hospitalisation (*BCL11A*-rs4671393, $P = 0{\cdot}026$; *HBS1L-MYB*-rs28384513, $P = 0{\cdot}01$). *APOL1* G1/G2 correlated with increased hospitalisation ($P = 0{\cdot}048$). This first study from Africa has provided evidence supporting possible development of genetic risk model for pain in SCD.

Keywords: sickle cell disease, acute vaso-occlusive painful crises, genetics, Cameroon, Africa.

Acute episodes of pain or vaso-occlusive crises (VOC) are hallmarks of sickle cell disease (SCD). Frequent VOC were a marker for disease severity and premature mortality in the Cooperative Study of Sickle Cell Disease (CSSCD) (Platt *et al*, 1991, 1994), and in modern cohorts in the United States of America (USA) (Darbari *et al*, 2013; Elmariah *et al*, 2014). VOC have a major economic impact due to the cost of unscheduled health care, and mostly affect the coping ability of SCD patients (Kanter & Kruse-Jarres, 2013; Wonkam *et al*, 2014a). The pathophysiology of vaso-occlusion involves multiple interrelated processes that have been increasingly linked to inflammation (Owusu-Ansah *et al*, 2016). Erythrocyte sickling and haemolysis trigger acute inflammation, marked by elaboration of inflammatory cytokines which stimulate nociceptors on peripheral nerve endings (Ballas *et al*, 2012) and abnormal expression of endothelial adhesion molecules, such as vascular cell adhesion molecule 1 (VCAM1), E-selectin and P-selectin, that are now targets of new therapies for VOCs in SCD (Ataga *et al*, 2017; Hoppe *et al*, 2017). There are inter-individual variations in frequency and severity of VOC, leading to differential utilization of acute care. In the CSSCD, SCD patients with three to 10 VOC episodes a year represented only 5·2% of the sample, yet accounted for 32·9% of VOC episodes (Platt *et al*, 1991). Similar data were also recently reported despite availability of modern SCD-specific therapies (Darbari *et al*, 2013). Higher haematocrit and lower fetal haemoglobin (HbF) are strong predictors of frequent VOC (Platt *et al*, 1991), and are subject to genetic modifiers.

Genetic variants at three principal loci, including *BCL11A*, *HBS1L–MYB* and *HBB* cluster, account for 10–20% variations in HbF levels among SCD patients in the USA and

First published online 3 December 2017
doi: 10.1111/bjh.15011

Cameroon (Lettre *et al*, 2008; Wonkam *et al*, 2014b), and these variants have been associated with VOC in SCD (Lettre *et al*, 2008; Sheehan *et al*, 2013). Co-inheritance of α-thalassaemia has been inconsistently associated to variable levels of VOC (Platt *et al*, 1991; Tarer *et al*, 2006; Darbari *et al*, 2012). In addition, a few observational studies have explored the associations of VOC with targeted variants in genes coding for enzymes that metabolize analgesics or inflammation-related proteins, with encouraging results (Mendonça *et al*, 2010; Galarneau *et al*, 2013; Belfer *et al*, 2014; Jhun *et al*, 2015; Hu *et al*, 2016). Specifically, one study identified and prioritised a total of 115 single nucleotide polymorphisms (SNPs) in 49 candidate genes that modified pain among African-American SCD patients (Jhun *et al*, 2015); but this has not been followed by genotype to phenotype investigations. We are not aware of a related study conducted in Africa where nearly 80% of new SCD patients are born (Piel *et al*, 2013).

Cameroon is a sub-Saharan African country with approximately 20 million people. The frequency of sickle cell mutation ranges from 8 to 34% in Cameroon (Weatherall & Clegg, 2001). There is currently no provision of universal new-born screening for SCD in the country and the median age of SCD diagnosis is 3·3 years (Wonkam *et al*, 2014b). There are no specialized centres for lifelong medical treatment, resulting in very few patients being exposed to hydroxycarbamide or opioid treatment (Wonkam *et al*, 2014a).

The primary objective of the present study was to investigate targeted genetic variants associated to VOC episodes in a group of patients living with SCD in Cameroon. The secondary objective was to study the association of these variants with health care utilisation (hospitalisations or consultations), considered as direct proxies of VOC.

We have investigated 23 targeted variants in 17 pain-related genes and the correlation of VOC with established genetic modifiers of SCD, namely, the 3.7 *HBA1/HBA2* deletion, variants in HbF-promoting loci, and kidney dysfunction-associated variants *(APOL1* and *HMOX1*), which have been correlated with SCD nephropathy in Cameroon (Geard *et al*, 2017).

Materials and methods

Ethical approval

The study was approved by the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee (HREC REF: 661/2015), Cape Town, South Africa; and the National Ethics Committee of the Ministry of Public Health, Yaoundé, Republic of Cameroon (No. 033/CNE/DNM/07). All patients older than 18 years signed consent forms, while informed consent was given by the parents or guardians for participants younger than 18 years old, in accordance with the declaration of Helsinki.

Patients

Assessment of clinical events. Patients were prospectively recruited at the Yaoundé Central Hospital and Laquintinie Hospital in Douala, between January 2010 and December 2011. Socio-demographic and clinical events were collected by means of a structured questionnaire administered to parents/guardians and adult SCD patients. Patients' medical records were reviewed, to delineate their clinical features over the past 3 years. Specifically, the occurrence of VOC, consultations rates referring to outpatient visits, hospitalisation rates, and blood transfusion history. Painful VOC events were defined as the occurrence of pain in the extremities, back, abdomen, chest or head that lasted at least two hours, and that could not be attributed to causes other than SCD, and required a hospital visit, and treatment with non-opioid antalgics (Platt *et al*, 1991). Body mass index (BMI) and blood pressures (BP) were measured in the outpatient settings.

Only patients older than 5 years of age (to avoid age-related changes in the complete blood count and HbF level), who had not received a blood transfusion or hospitalisation in the past 6 weeks were included. None was currently treated with hydroxycarbamide or opioids.

Control participants. For the purpose of comparative allele frequencies of targeted variants in selected pain-related genes of interest, a total of 105 ethnically matched Cameroonian controls (HbAS and HbAA) were randomly recruited, from apparently healthy blood donors in Yaoundé, for participation in the study.

Measurements of haematological indices and renal functions. Routine blood counts of patients and haemoglobin (Hb) electrophoresis were conducted on arrival at the hospital, at the haematological laboratory of the Centre Pasteur in Yaoundé, as previously described (Wonkam *et al*, 2014b). Routine laboratory tests were performed to measure serum creatinine. The urine albumin level was determined using either the Siemens Clinitek Status test (Erlangen®, Germany) or the Hemocue Albumin 20 system (Angelholm®, Sweden) as describe elsewhere (Geard *et al*, 2017). The glomerular filtration rate (GFR) was estimated (eGFR) using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (Geard *et al*, 2017).

Molecular methods

Sickle cell anaemia mutation, HBB *cluster haplotypes, and 3·7 kb* HBA1/HBA2 *deletion.* DNA was extracted from peripheral blood following the manufacturer's instructions (Puregene Blood Kit; Qiagen, Hilden, Germany). Molecular analysis to determine the presence of the sickle mutation was carried out on 200 ng DNA by polymerase chain reaction (PCR) to amplify a 770 bp segment of the *HBB*, followed by

DdeI restriction analysis of the PCR product (Saiki *et al*, 1985). The present analysis was restricted to sickle cell anaemia (homozygous HbS) due to the well-known differences in laboratory parameters (Platt *et al*, 1991; Darbari *et al*, 2013), and to allow single sickle genotype (HbSS) for genetic associations. Using published primers and methods, five restriction fragment length polymorphism (RFLP) sites in the *HBB* cluster were amplified to analyse the XmnI (5'$^{G}\gamma$), HindIII ($^{G}\gamma$), HindIII ($^{A}\gamma$), HincII (3ψβ') and HinfI (5'β) for the *HBB* haplotype background (Bitoungui *et al*, 2015). The 3·7 kb *HBA1/HBA2* deletion was successfully screened, using the expand-long template PCR (Roche Diagnostics, Basel, Switzerland), as previously published (Rumaney *et al*, 2014).

SNPs in HbF-promoting loci, APOL1 *and* HMOX1. Ten regions containing specific SNPs were amplified: viz, for the *BCL11A* locus, SNPs rs11886868 and rs4671393; for the *HMIP1/2* loci: SNPs rs28384513, rs9376090, rs9399137, rs9389269; rs9402686 and rs9494142; for the *OR51B5/6 loci*: SNP rs5006884, for *HBG2* loci, SNP rs7482144; followed by Sanger sequencing (Wonkam *et al*, 2014b). SNP genotyping of rs60910145 (*APOL1*), rs73885319 (*APOL1*) and rs743811 (*HMOX1*) was performed using predesigned TaqMan genotyping assays (Applied Biosystems, Foster City, CA, USA), and the genotyping of rs3074372 (*HMOX1*) and rs71785313 (*APOL1*) variants using fragment analysis, incorporating fluorescently-labelled forward primers (Geard *et al*, 2017).

Genotyping of targeted SNPs in pain-related genes. Selection of SNPs—We initially performed a thorough review of the literature on pharmacogenomics of SCD therapeutics, and identified a list of variants that are potentially associated with pain in SCD (Mnika *et al*, 2016). Once the SNPs of interests were identified, we investigated their allele frequencies in African populations present in the 1000 Genomes project (http://www.internationalgenome.org/home), and further narrowed the selection to SNPs that showed high frequency among African populations. For the purpose of additional quality control *ADRA2A*-rs3750635, which was monomorphic for all the populations in the 1000 Genomes project, was also genotyped. This resulted in the selection of 23 SNPs from 17 pain-related genes that were investigated in the present study (Table I).

Genotyping—SNPs were genotyped using a TaqMan® SNP Genotyping Assay and TaqMan® Universal Master Mix (Life Technologies, Carlsbad, CA, USA), at the Division of Human Genetics, Faculty of Health Sciences, University of Cape

Table I. Allele frequencies of selected pain-related genes variants among Cameroonian and African American SCD cohorts.

				Cameroon cohort				
Gene	dbSNP ID	Position	Allele change(s)	SCD cases	Controls	African American SCD*	Cameroon SCD *versus* Cameroon control *P* values	Cameroon SCD *versus* African American SCD *P* values
ABCB1	rs1045642	87509329	T>C	0·153	0·205	0·785	0·076	**0·0001**
ADRA1A	rs1048101	26770511	T>C	0·177	0·162	0·776	0·607	0·286
ADRA2A	rs3750635	5750220	T>C	Monomorphic	Monomorphic	Monomorphic	NA	NA
ADRB2	rs1042713	148826877	A>G	0·481	0·5	0·514	0·615	0·409
ARRB2	rs1045280	4719343	C>T	0·342	0·135	0·558	**0·0001**	**0·0001**
AVPR1A	rs10877969	63153459	T>C	0·221	0·371	0·517	**0·0002**	0·305
BDKRB2	rs1799722	96204802	C>T	0·253	0·263	0·715	0·785	0·287
CACNA2D3	rs1851048	54587633	C>T	0·136	0·126	0·162	0·724	0·898
	rs6777055	55039890	A>C	0·195	0·2	0·803	0·863	**0·0001**
COMT	rs4633	19962712	C>T	0·239	0·283	0·620	**0·001**	**0·001**
	rs6269	19962429	A>G	0·44	0·421	0·672	0·634	**0·0008**
	rs4680	19963748	G>A	0·238	0·289	0·319	0·14	0·17
DRD2	rs4274224	113448730	C>T	0·262	0·263	0·290	0·977	0·501
FAAH	rs324419	46406314	T>C	0·128	0·208	0·154	**0·0035**	0·252
	rs2295632	46413890	T>G	0·249	0·31	0·711	0·079	0·485
	rs4141964	46399368	T>C	0·264	0·243	0·716	0·53	0·77
KCNS1	rs734784	45094986	A>G	0·469	0·439	0·548	0·445	0·591
OPRM1	rs1799971	154039662	A>G	0·001	Monomorphic	0·002	NA	0·575
STAT6	rs841718	57099213	C>T	0·317	0·365	0·701	0·191	**0·024**
	rs3024971	57099944	A>C	0·022	0·13	0·952	**0·0001**	0·666
TRPA1	rs920829	72065468	G>A	0·304	0·292	0·708	0·724	**0·0001**
TRPV1	rs222747	3589906	G>C	0·088	0·099	0·874	0·617	**0·0456**
UGT2B7	rs7438135	69095621	G>A	0·3	0·163	0·785	**0·0001**	**0·0001**

Significant *P* values are bolded. dbSNP ID; single nucleotide polymorphism database identification; NA, not applicable; SCD, sickle cell disease.
*Jhun *et al* (2014, 2015).

Town; and by iPLEX GoldSequenom Mass Genotyping Array (Inqaba Biotec, Pretoria, South Africa). Validation was done in a subset of sample (10%), by Sanger sequencing using BigDye terminator mix (Promega, Madison, WI, USA).

Statistical analysis

Descriptive statistics was performed using STATA, version 14.0.370 (StataCorp, College Station, TX, USA). For quality control, a Hardy-Weinberg Equilibrium (HWE) test was performed on all genotype results. Two SNPs were monomorphic in both patients and controls: *HBS1L-MYB*-rs9376090 and *ADRA2A*-rs3750635; *OPRM1*-rs1799971 was monomorphic among controls and very rare (0·001) in patients. Only the 3.7del α-globin gene genotypes ($P = 0·005$) were out of HWE; however, this deviation was expected in view of the strong protective effect of this genetic variant on SCD, as previously reported on Cameroonians (Rumaney *et al*, 2014; Geard *et al*, 2017). The skewness of VOC, and hospitalisation, consultation rates and haematological indices, was corrected by taking their natural logarithm to approximate normal distribution. Prior to log transformation, these variables were all rescaled by systematically adding a constant (one) to allow the inclusion of participants with null values. General linear and multinomial regression frameworks, adjusted for age and sex, were performed to investigate the relationship between genotypes results and clinical data, using the R® statistical software (version 3.3.3, The R Foundation for Statistical Computing, Vienna, Austria). $P < 0·05$ were considered statistically significant. For association analysis with pain-related genes, and modifiers of sub-phenotypes of SCD, the Bonferroni critical *P*-value is also provided to indicated the threshold for significance after accounting for multiple comparisons.

Results

Description of the studied cohort

A total of 436 SCD patients (HbSS) were included; Table II summarizes the participants' characteristics. There was roughly equal numbers of males and females (217 and 219, respectively). The median age was 16 years. The most prevalent β-globin like gene cluster haplotypes was Benin, followed by Cameroon. Up to 41·8% ($n = 151$) of patients had co-inherited a single or double 3·7 kb *HBA1/HBA2* deletion. The median number of VOC per year was 2 (range: 0–40); 46·6% ($n = 185$) of participants had ≥3 VOC per year and 27·2% ($n = 115$) had ≥2 hospitalisations per year.

Clinical and haematological factors associated with acute pain crisis

Several clinical factors significantly correlated with the number of VOC (Figs 1 and S1), including female sex (estimate = 0·073; $P = 0·026$), hospitalisation rates (estimate = 0·41, $P =< 0·0001$), consultation rates (estimate = 0·254, $P =< 0·0001$), BMI (estimate = 0·022; $P = 0·02$) and positive history of blood transfusion (estimate = 0·42; $P = 0·046$; Figure S1).

The number of VOC was also associated with various haematological indices, including Hb level (estimate = −0·074; $P = 0·005$), white blood cell counts (estimate = 0·008; $P = 0·04$) (Fig 1); red blood cell counts (estimate = −0·154; $P = 0·015$; Figure S1) and HbF level (estimate = −0·003; $P = 0·025$; Figure S1). There was no significant association between VOC and age (estimate = −0·003, $P = 0·19$), microalbuminuria (estimate = 0·050; $P = 0·198$), platelet count (estimate = $-2·437e^{-05}$; $P = 0·927$), eGFR (estimate = −0·0001; $P = 0·942$), systolic BP (estimate = 0·0009; $P = 0·794$), and diastolic BP (estimate = 0·0009; $P = 0·851$).

Frequencies of pain-related genes variants across various populations

The differential frequencies across populations of the SNPs investigated are presented in Tables I and SI. When excluding the monomorphic pain-related genes SNPs, a total of 6/21 SNPs (28·6%) were differentially distributed among Cameroonian SCD individuals compared to controls (Table I); all but one (5/6) showed significant or borderline association with VOC or hospital utilisation (Table I). Up to 40·1% of the variants studied (9/22) were differentially frequent when comparing Cameroonian *versus* African American patients living with SCD. Furthermore, comparison with control data extracted from the 1000 Genome Project, showed significant differences in allele frequencies in half of SNPs with Africans (11/22), and for the large majority of SNPs (88·8%; 18/22), with both Europeans and Asians (Table SI).

Correlations of VOC, health care utilisation and pain-related genes variants

Three pain-related gene variants significantly correlated with VOC (*CACNA2D3*-rs6777055, $P = 0·025$; *DRD2*-rs4274224, $P = 0·037$; and *KCNS1*-rs734784, $P = 0·01$); all these three variants, were significantly or borderline associated with hospitalisation rates (Table III; Fig 2). SNPs in four genes were borderline associated with VOC (*ABCB1*-rs1045642, $P = 0·065$; *AVPR1A*-rs10877969; $P = 0·072$; *FAAH*-rs4141964, $P = 0·084$; *TRPA1*-rs920829, $P = 0·078$); two of which were also significantly or borderline associated with hospitalisation rates (Table III).

Five pain-related genes variants correlated with hospitalisation or consultation rates, without any significant association with VOC (Table III; *COMT*-rs6269, $P = 0·027$; *FAAH*-rs4141964, $P = 0·003$; *OPRM1*-rs1799971, $P = 0·031$; *ADRB2* -rs1042713; $P < 0·001$; and *UGT2B7*-rs7438135, $P = 0·037$).

Table II. Description of the studied Cameroonian SCD cohort.

Variable	Median (25th–75th percentiles) or %	Range	Observations (*n*)
Age (years)	16 (9–24)	5–54	436
Gender			
Female/Male (219/216)		–	436
Haematological indices			
RBC ($\times 10^{12}$/l)	2·7 (2·3–3·1)	1·4 –5·5	436
Hb (g/l)	76 (67–85)	35–145	436
MCV (fl)	84 (78–91)	59·0–117·0	436
MCHC (g/l)	338 (316–358)	215–529	436
WBC ($\times 10^9$/l)	12·8 (9·1–16·2)	2·9–49·8	436
Lymphocytes ($\times 10^9$/l)	5·2 (4·0–7·2)	0·2–22·6	436
Monocytes ($\times 10^9$/l)	1·3 (0·9–1·8)	011–7·8	436
Platelet count ($\times 10^9$/l)	374·3 (291·2–448·0)	97–756	436
HbA2 (%)	3·6 (3·0–4·2)	0–18·2	436
HbF (%)	8·8 (2·5–14·1)	0–37·4	436
Clinical events			
VOC (*n*/year)	2 (1–4)	0–40	436
Consultations (*n*/year)	2 (0–4)	0–24	324
Hospitalisation (*n*/year)	1 (0–2)	0–30	422
Blood transfusion (%)	77·8		330/424
Stroke (%)	3·9		17/436
3·7 *HBA1/HBA2* deletion genotypes			
αα/αα	59·8		225/376*
αα/α3·7	30·1		113/376*
α 3·7/α 3·7	10·1		38/376*
HBB haplotype			
Benin/Benin	64·1%		212/331*
Benin/Cameroon	30·8%		102/331*
Cameroon/Cameroon	5·1%		17/331*
Renal functions†			
Crude albuminuria (mg/l)	41 (23–83)	3–1180	407
eGFR (CKD-EPI) (ml/min/1·73 m^2)	135·1 (112·0–154·4)	50·8–250·8	404
Serum creatinine (μmol/l)	7 (5–8·5)	2–13·8	404

CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; eGFR, estimated glomerular filtration rate; Hb, haemoglobin; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; RBC, red blood cell count; SCD, sickle cell disease; VOC, vaso-occlusive crises; WBC, white blood cell count.

*Number of individuals, not alleles.

†Previously reported in Geard *et al* (2017).

Correlations of VOC, hospitalisation rates and variants in established genetic modifiers of SCD

The 3·7 kb *HBA1/HBA2* deletion correlated with increased VOC ($P = 0{\cdot}002$) and related hospitalisation rates ($P = 0{\cdot}02$) (Table IV, Fig 3A). Variants in all the HbF-promoting loci correlated mostly with decreased hospitalisation rates (*BCL11A-rs*4671393, $P = 0{\cdot}026$; *HBS1L-MYB*-rs28384513, $P = 0{\cdot}01$; and *HBS1L-MYB*-rs9494142, $P = 0{\cdot}038$; Fig 3B, C); but *BCL11A-rs*4671393 was also associated with decreased VOC ($P = 0{\cdot}017$). *APOL1* G1/G2 correlated with increased hospitalisation rates ($P = 0{\cdot}048$, Table IV, Fig 3D). Variants in *HMOX1* were not associated with VOC or hospitalisation rates (Table IV).

Discussion

To our knowledge, this is the first study to investigate targeted genomic variants in relation with VOC in SCD in Africa, where the burden of SCD is very high, with a mostly non-advantageous environment for SCD patients. Hydroxycarbamide and opioid medications that are widely used in high income settings, are serious pharmacological modifiers of both pain crisis and health care utilization, and therefore are intrinsic limitations of similar observational studies (Darbari *et al*, 2013). In Cameroon, the non-advantageous physical environment, characterised by high temperature (which could trigger dehydration and VOC) and the endemicity of malaria (which often deteriorates the anaemia in SCD),

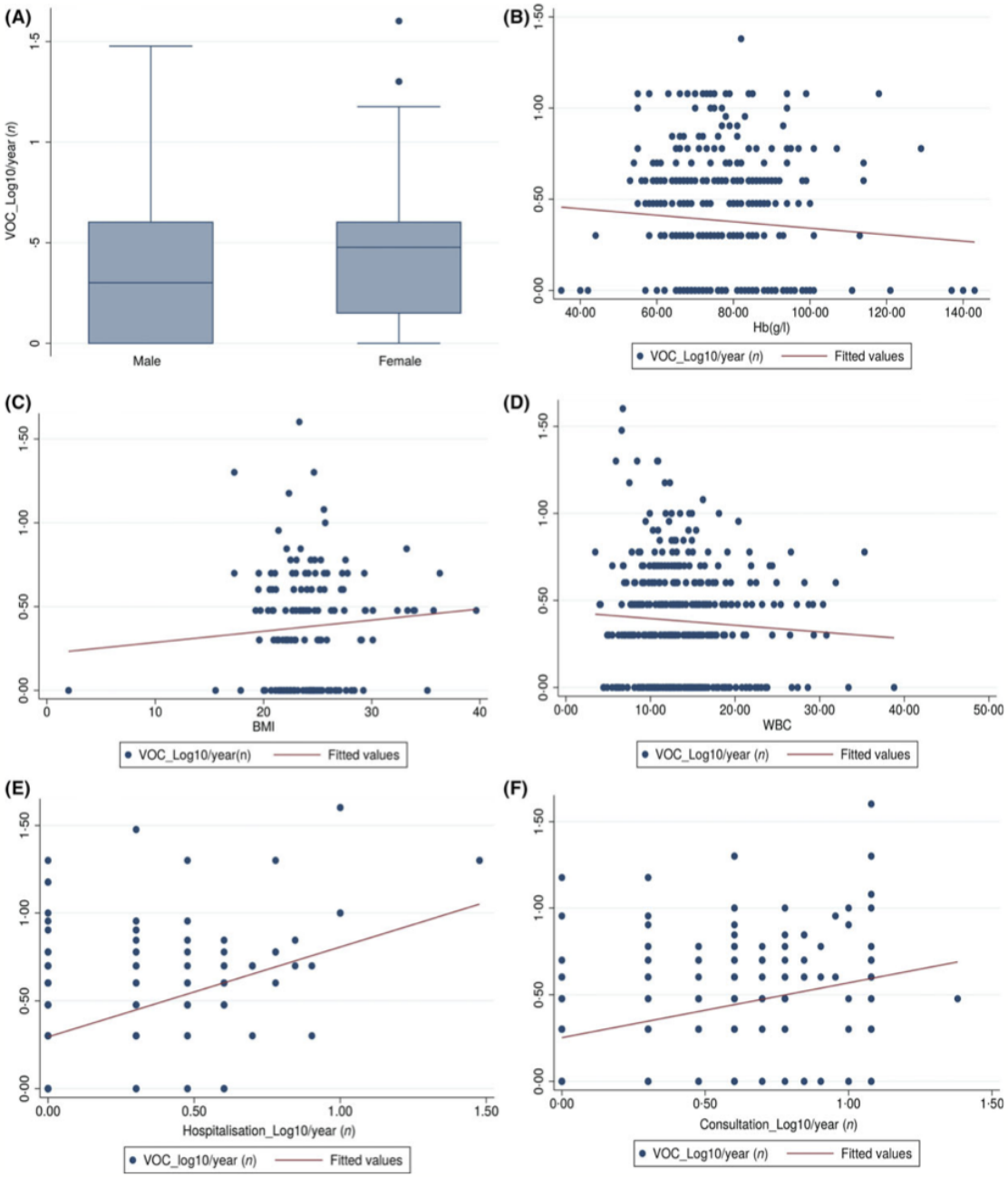


Fig 1. Scatter plot and box and whisker illustrating clinical and haematological factors associated with painful acute VOC episodes. (A) Box and whisker plots showing the correlation of vaso-occlusive crisis (VOC) values with gender (estimate = 0·073; P = 0·026). The horizontal lines that constitute the 'box' correspond to the lower quartile, median and upper quartile parameters. The length of the 'whiskers' that extend from the box in the upwards and downwards direction represent a distance 1·5 times the interquartile range. Values that lie outside this distance are considered outliers, or extreme values. (B) Clinical factors associated with VOC in sickle cell disease with haematological indices. Scatter plot illustrating the negative relationship between log VOC and total haemoglobin (Hb, g/l, estimate = −0·074; P = 0·005); related to this was the association between VOC and red blood cell (RBC) counts (estimate = −0·0154; P = 0·015). (C) There was a positive correlation between body mass index (BMI) and VOC (estimate = 0·022; P = 0·02). (D) white blood cell (WBC) counts ($\times 10^9$/l) was also positively associated with VOC (estimate = 0·008; P = 0·04). Scatter plots illustrating the relationship between VOC values and log hospitalisations (estimated = 0·41; P =< 0·0001). The log hospitalisations variable is displayed on the x-axis, with the VOC values on the y-axis; the red line indicates the line of best fit. (F) Scatter plots illustrating the relationship between VOC values and log consultations (estimate = 0·254; P =< 0·0001); consultations variable is displayed on the x-axis, with the VOC values on the y-axis. The red line indicates a line of best fit. [Colour figure can be viewed at wileyonlinelibrary.com]

Table III. Variants in of selected pain-related genes and VOC, and consultation and hospitalisation rates.

Gene	dbSNP ID	Position	Allele change(s)	MAF	VOC		Consultations		Hospitalisations	
					P values	Effect size (SE)	*P* values	Effect size (SE)	*P* values	Effect size (SE)
ABCB1	rs1045642	87509329	T>C	0·153	**0·065**†	0·152 (0·082)	0·741	0·206 (0·623)	0·417	0·386 (0·474)
ADRA1A	rs1048101	26770511	T>C	0·177	**0·094**	0·122 (0·058)	0·297	0·656 (0·627)	0·793	0·126 (0·477)
ADRA2A	rs3750635	5750220	T>C	Monomorphic	NA	NA	NA	NA	NA	NA
ADRB2	rs1042713	148826877	A>G	0·481	0·277	−0·328 (0·057)	**0·0004**‡	−0·143 (0·045)	0·25	−0·38 (0·329)
ARRB2	rs1045280	4719343	C>T	0·342	0·958	−0·0033 (0·063)	0·875	−0·143 (0·0474)	0·201	−0·464 (0·360)
AVPR1A	rs10877969	63153459	T>C	0·221	0·072†	0·101 (0·056)	0·119	0·672 (0·428)	**0·061**†	−0·168 (0·066)
BDKRB2	rs1799722	96204802	C>T	0·253	0·64	0·034 (0·072)	0·591	−0·289 (0·537)	0·399	0·344 (0·406)
CACNA2D3	rs1851048	54587633	C>T	0·136	0·945	−0·006 (0·086)	0·633	−0·315 (0·658)	0·314	0·505 (0·499)
	rs6777055	55039890	A>C	0·195	**0·025**†	−0·167 (0·074)	0·964	−0·0253 (0·562)	**0·008**	−0·181 (0·068)
COMT	rs4633	19962712	C>T	0·239	0·732	0·0345 (0·100)	0·122	1·185 (0·762)	0·225	−0·707 (0·580)
	rs6269	19962429	A>G	0·44	0·575	−0·0367 (0·065)	0·788	−0·134 (0·494)	**0·027**†	0·138 (0·042)
	rs4680	19963748	G>A	0·238	0·986	0·0018 (0·102)	0·48	−0·553 (0·780)	0·264	0·667 (0·594)
DRD2	rs4274224	1·13E+08	C>T	0·262	**0·037**	0·148 (0·070)	0·078	0·955 (0·537)	**0·091**†	0·695 (0·408)
FAAH	rs324419	46406314	T>C	0·128	0·663	−0·557 (0·227)	**0·076**‡	1·245 (0·908)	0·999	0·0012 (0·692)
	rs2295632	46413890	T>G	0·249	0·916	−0·0098 (0·093)	0·295	−0·733 (0·698)	0·653	0·239 (0·530)
	rs4141964	46399368	T>C	0·264	**0·084**‡	0·221 (0·122)	**0·058**†	0·165 (0·086)	**0·003**†	−185 (0·042)
KCNS1	rs734784	45094986	A>G	0·469	**0·010**§	−0·165 (0·045)	0·581	0·265 (0·479)	**0·002**‡	0·229 (0·074)
OPRM1	rs1799971	1·54E+08	A>G	0·001	0·64	0·299 (0·53)	0·205	−1·047 (0·040)	**0·031**§	−0·135 (0·044)
STAT6	rs841718	57099213	C>T	0·317	0·79	0·0176 (0·066)	0·358	0·465 (0·504)	0·959	−0·02 (0·381)
	rs3024971	57099944	A>C	0·022	0·262	0·213 (0·190)	0·35	−1·347 (1·42)	0·4	0·924 (1·10)
TRPA1	rs920829	72065468	G>A	0·304	**0·078**§	−0·115 (0·050)	0·91	−0·0549 (0·485)	0·312	0·373 (0·368)
TRPV1	rs222747	3589906	G>C	0·088	0·907	0·0132 (0·112)	0·4	0·722 (0·856)	0·695	0·255 (0·650)
UGT2B7	rs7438135	69095621	G>A	0·3	0·805	0·028 (0·062)	**0·037**‡	−0·685 (0·043)	0·209	−0·459 (0·363)

Bonferroni critical $P < 0·003$. Significant/borderline *P* values are bolded. MAF, minor allele frequency; NA, not applicable; SE, standard error; VOC, vaso-occlusive painful crisis rate.

†Dominant model.

‡Recessive model.

§Over dominant model.

Table IV. Variants in known modifiers of sub-phenotypes of SCD and VOC, and hospitalisation rates.

Gene	dbSNP ID	Position	Allele change (s)	MAF	VOC *P* values	Effect Size (SE)	Hospitalisations *P* values	Effect size (SE)
HBA (3·7 Alpha-globin gene deletion)		16	NA	NA	**0·002**‡	0·339 (0·116)	**0·02**†	0·17 (0·073)
APOL1	rs60910145 (G1)	22:36265988	T>G	0·14	0·439	−0·061 (0·071	0·553	−0·075 (0·126)
APOL1	rs73885319 (G1)	22:36265860	T>G	0·13	0·67	−0·034 (0·067)	0·563	−0·070 (0·120)
APOL1	rs71785313 (G2) Indel	22:36266000-36266005	Deletion	0·082	0·784	0·102 (0·400)	0·059†	0·273 (0·1551)
APOL1	G1/G2	NA	NA	NA	0·194	−0·434 (0·114)	**0·048**§	0·339 (0·200)
HMOX1	rs3074372	22:35380894	L>S	0·111	0·756	0·03 (0·084)	0·477	−0·509 (0·346)
	rs743811	22:35396981	T>C	0·111	0·234	−0·010 (0·063)	0·463	0·429 (0·061)
BCL11A	rs11886868	2:60493111	G>A	0·31	**0·081**‡	−0·20 (0·037)	**0·042**†	−0·155 (0·171)
BCL11A	rs4671393	2:60493816	T>C	0·3	**0·017**‡	−0·334 (0·133)	**0·026**†	−0·226 (0·087)
HBS1L-MYB	rs28384513	1·35E+08	A>C	0·217	0·057†	0·136 (0·058)	**0·010**§	0·139 (0·045)
HBS1L-MYB	rs9376090	6:135090090	T>C	0·146	0·403	0·537 (0·033)	0·658	0·447 (0·062)
HBS1L-MYB	rs9399137	6:135097880	T>C	0·043	0·372	0·560 (0·033)	0·744	0·063 (0·100)
HBS1L-MYB	rs9389269	6:135106021	T>C	0·18	0·548	0·043 (0·060)	0·898	0·014 (0·10)
HBS1L-MYB	rs9402686	6:135427817	G>A	0·03	0·355	0·06 (0·11)	0·304	0·0·33 (0·058)
HBS1L-MYB	rs949414 2	6:135431640	T>C	0·11	0·343	−0·08 (0·076)	**0·038**†	−0·163 (0·577)
HBG2	rs7482144	11:5254939	G>A	0·005	0·715	0·126 (0·404)	**0·008**†	0·641 (0·184)
OR51B5/6	rs5006884	11:5352021	C>T	0·08	0·245	0·056 (0·032)	**0·056**‡	1·9 (0·057)

Bonferroni critical $P < 0·029$. Significant/borderline *P* values are bolded. dbSNP ID; Single Nucleotide Polymorphism database identification; MAF, minor allele frequency; NA, not applicable; SCD, sickle cell disease; SE, standard error; VOC, vaso-occlusive painful crisis rate.

†Dominant model.

‡Recessive model.

§Over dominant model.

syndrome, leg ulcers and chronic kidney disease (Higgs *et al*, 1982; Guasch *et al*, 1999; Geard *et al*, 2017), but convey similar or higher rates of VOC (Platt *et al*, 1991; Tarer *et al*, 2006; Darbari *et al*, 2012, 2013; Meier *et al*, 2017). In the present study, we have observed higher rates of VOC with the co-inheritance of α-thalassaemia (Fig 3). In the CSSCD, the slight increase in the pain rate associated with α-thalassaemia was attributable to the higher haematocrit (Platt *et al*, 1991). *APOL1* G1/G2 risk alleles were previously associated with kidney dysfunctions among Cameroonians living with SCD (Geard *et al*, 2017), and were associated with increased hospitalisation rates in the present study (Table IV). This could indicate that some hospitalisations were probably due to other confounding causes. There is evidence that acute kidney injury is common during sickle cell pain crisis (Baddam *et al*, 2017); but, we did not observe any association between VOC and albuminuria, eGFR or variants in *HMOX1*; although variants in *HMOX1* were previously associated with kidney dysfunctions among Cameroonians with SCD (Geard *et al*, 2017), and reduced acute chest syndrome and hospitalisation rates in SCD patients in the USA (Bean *et al*, 2012). Increased health care utilization by SCD patients for VOC is well known (Rees *et al*, 2010; McMillan *et al*, 2015), and these individuals are at particularly high risk for death (Platt *et al*, 1991; Darbari *et al*, 2013; Elmariah *et al*, 2014), and should be vigorously treated; early identification of these individuals could involve a comprehensive risk model including evolving variants specific in pain–related genes.

The majority of SNPs (5/6) in pain-related genes differentially frequent in Cameroonian SCD *versus* control were borderline-to-significantly associated with VOC or healthcare utilizations, indicating possible selection and enrichment of protective variants among SCD patients (Table I), owing to the unfavourable environment. Thus, the allele frequencies reported in the studied patients may not be representative of the entire SCD population in Cameroon, as it is possible that these patients are less severe cases who have survived childhood. Allele frequencies were significantly different among patients from Cameroon *versus* African American living with SCD for 40·1% of the SNPs (9/22 SNPs; Table I), and even more so with African data extracted from the 1000 Genomes Project (11/22, 50%; Table SI). This is in line with the high level of genetic variations in populations of African ancestry (Gurdasani *et al*, 2015).

In total, variants in 8 of the 22 (36·4%) pain-associated genes investigated were significantly associated with VOC or consultations/hospitalisation rates (Table III). These are novel findings. SNPs located in a subunit of the calcium channel gene *CACNA2D3*, were also associated with a higher risk of anaemia, suggesting that calcium channels could potentially be involved in pathways for iron uptake

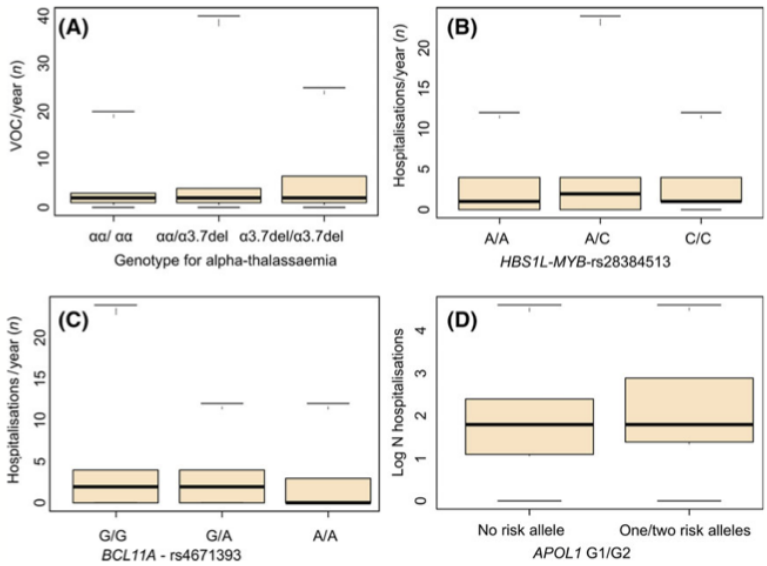


Fig 3. Associations between targeted variants in the alpha-globin gene, HbF-promoting loci and *APOL1*, with painful acute VOC episodes or hospitalisation rates. (A) Box and whisker plots showing the association of 3·7 alpha-globin gene deletions with vaso-occlusive crisis (VOC) ($P = 0{\cdot}002$); a similar association was also reported with the hospitalisation rates ($P = 0{\cdot}02$). (B) and (C) Box and whisker plots showing the associations of HbF- promoting loci variant with the hospitalisation rates: *HBS1L-MYB*-rs28384513 ($P = 0{\cdot}01$); *BCL11A*-rs4671393 ($P = 0{\cdot}026$); *BCL11A*-rs4671393 was also positively associated with VOC ($P = 0{\cdot}017$), while *HBS1L-MYB*-rs28384513 was borderline with VOC ($P = 0{\cdot}057$). (D) Box and whisker plots showing the association of *APOL1* G1/G2 risk alleles with hospitalisation rates ($P = 0{\cdot}048$); *APOL1*-rs71785313 (G2) was borderline with the hospitalisation rates ($P = 0{\cdot}059$). Conventions are as per Fig 1. [Colour figure can be viewed at wileyonlinelibrary.com]

in physiological conditions and in SCD (Baeza-Richer *et al*, 2015). A genome wide meta-analysis showed *DRD2* (Dopamine D2 receptor) genetic variations in the modulation of systolic BP among African Americans with SCD. We also found *DRD2-rs*4274224 to be associated with *VOC* in SCD patients (Table III). In addition, exploratory findings have suggested that *DRD3*-rs6280 (Ser9Gly) may contribute to pain heterogeneity in SCD (Jhun *et al*, 2014). Also related to our findings, variants in *KCNS1* were associated with and multiple chronic pain states in a non-SCD population (Costigan *et al*, 2010). Five pain associated-genes variants correlated with health services utilization only (Table III). These include variants in *COMT* (catechol-O-methyltransferase), *OPRM1* (opioid receptor mu *1* gene), *UGT2B7* (UDP glucuronosyltransferase family 2 member B7) and *ABCB1* (ATP binding cassette subfamily B member 1); all previously suggested as potentially important for SCD VOC (Darbari *et al*, 2008; Joly *et al*, 2012). Other studies have found that *COMT-rs*4680 (158 Met allele or Met/Met genotype) was associated with acute care utilization, an indicator of acute pain (Jhun *et al*, 2014). The presence of the *UGT2B7*-840G allele contributes to the variability in hepatic clearance of morphine in SCD (Eyler *et al*, 2008). Up to 35% of African American SCD patients were previously reported to have variants in *ABCB1*, suggested to potentially influence good morphine exposure (Darbari *et al*, 2008; Joly *et al*, 2012).

Limitations

Possible limitations of the present study are the cross-sectional nature and the hospital-based recruitment. VOC episodes may have been subjected to pain self-tolerance bias, and financial factors could have been also limiting factors for hospital attendance. The issue of chronic pain was difficult to address with the study design, as it seems highly likely that individuals with 40 pain events per year are experiencing chronic pain. However, self-reported VOC in SCD has also been useful as a clinical endpoint in drug trials, patient quality of life measures and as a prognostic marker for mortality (Platt *et al*, 1991; Charache *et al*, 1995; Machado *et al*, 2011; Hoots & Shurin, 2012; Keller *et al*, 2017). The possible poor definition of VOC is also tempered by its strong association with the use of health services, which was more objectively assessed; validating, to some extent, the use of such cost-effective patient reported outcomes for genetic association study. A recent genome-wide association study, which included only VOC episodes requiring hospitalisation, found that *KIAA1109*-rs3115229 approached genome-wide significance in a locus associated with auto-inflammatory disorders (Chaturvedi *et al*, 2017). Therefore, the present study represents an important step forward in understanding clinical and genetic predictors of VOC in sub-Saharan Africa and globally. Lastly, our findings must be interpreted while accounting for the possibility of chance findings in the

context of multiple comparisons. Indeed, based on the Bonferroni corrected threshold *P*-value for significance, half of the pain-related gene variants associations with our outcomes of interest were borderline or non-significant. It is of note however that our study focused on previously characterized genes, and that changes remain: those associations could be significant at a corrected threshold *P*-value in a bigger sample, while correction for multiple comparison always increases the chance of false negative findings.

Conclusion

This study has provided important findings on clinical predictors of acute painful episodes and the use of health care services in a unique group of SCD patients from Cameroon that have not been exposed to hydroxycarbamide and opioid. In addition, the study has identified specific variants of pain-related genes that are associated with acute pain crisis and health care utilization, as well as in established genetic modifiers of SCD, such as *HBA1/HBA2*, HbF- promoting loci and *APOL1*. Altogether, the results may improve our ability to identify SCD patients who are at elevated risk for VOC and other organ complications, and will contribute to refining the elaboration of risk-profiling strategies that integrate both genetic and clinical information. We acknowledge that implementation of any genetic risk model in SCD in Africa is clearly difficult today, due to multiple competing priorities: hopefully, as the costs of genomic tests decreases, it could be possible in future.

Acknowledgements

The molecular experiments of the study were supported by grants from the National Health Laboratory Services (NHLS), South Africa; and the NIH, USA, grant number 1U01HG007459-01, and the Welcome Trust, Developing Excellence in Leadership, Training and Science (DELTAS) Africa, Award 107755Z/15/Z. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. K.M. is a PhD candidate at the University of Cape Town and this work is submitted in partial fulfilment of the requirement for their PhD.

Authorship Contributions

Conceived and designed the experiments: A.W, K.M., C.D. Performed the experiments: K.M., V.J.N.B. Patients recruitment, sample and clinical data collection and processing: V.J.N.B., B.C.C. Data analysis: K.M., A.P.K., E.C., A.W. Contributed reagents/materials/analytic tools: V.J.N.B., B.C.C., C.D., A.P.K., AW. Wrote the paper: A.W K.M., A.W. Revised and approved the manuscript: K.M., V.J.N.B., B.C.C., C.D., E.C., A.P.K.

Conflict of Interest Disclosures

The authors report no conflicts of interest, and take full responsibility for the content and writing of this article.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table SI. Allele frequencies of selected pain-related genes variants among Cameroonian SCD patients and data extracted from the 1000 Genome project.

Figure S1. Scatter plot, and box and whisker illustrating association of haematological factors, and age, with Acute Vaso-occlusive painful (VOC) episodes.

References

Ataga, K.I., Kutlar, A., Kanter, J., Liles, D., Cancado, R., Friedrisch, J., Guthrie, T.H., Knight-Madden, J., Alvarez, O.A., Gordeuk, V.R., Gualandro, S., Colella, M.P., Smith, W.R., Rollins, S.A., Stocker, J.W. & Rother, R.P. (2017) Crizanlizumab for the prevention of pain crises in sickle cell disease. *New England Journal of Medicine*, **376**, 429–439.

Baddam, S., Aban, I., Hilliard, L., Howard, T., Askenazi, D. & Lebensburger, J.D. (2017) Acute kidney injury during a pediatric sickle cell vaso-occlusive pain crisis. *Pediatric Nephrology*, **32**, 1451–1456.

Baeza-Richer, C., Arroyo-Pardo, E., Blanco-Rojo, R., Toxqui, L., Remacha, A., Vaquero, M.P. & López-Parra, A.M. (2015) Genetic contribution to iron status: SNPs related to iron deficiency anaemia and fine mapping of *CACNA2D3* calcium channel subunit. *Blood Cells, Molecules, and Diseases*, **55**, 273–280.

Ballas, S.K., Gupta, K. & Adams-Graves, P. (2012) Sickle cell pain: a critical reappraisal. *Blood*, **120**, 3647–3656.

Bean, C.J., Boulet, S.L., Ellingsen, D., Pyle, M.E., Barron-Casella, E.A., Casella, J.F., Payne, A.B., Driggers, J., Trau, H.A., Yang, G., Jones, K., Ofori-Acquah, S.F., Hooper, W.C. & Debaun, M.R. (2012) Heme oxygenase-1 gene promoter polymorphism is associated with reduced incidence of acute chest syndrome among children with sickle cell disease. *Blood*, **120**, 3822–3828.

Belfer, I., Youngblood, V., Darbari, D.S., Wang, Z., Diaw, L., Freeman, L., Desai, K., Dizon, M., Allen, D. & Cunnington, C. (2014) A GCH1 haplotype confers sex-specific susceptibility to pain crises and altered endothelial function in adults with sickle cell anaemia. *American Journal of Hematology*, **89**, 187–193.

Bitoungui, V.J.N., Pule, G.D., Hanchard, N., Ngogang, J. & Wonkam, A. (2015) Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa?. *Omics: A Journal of Integrative Biology*, **19**, 171–179.

Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S.V., Mcmahon, R.P. & Bonds, D.R. (1995) Effect of hydroxyurea on the frequency of painful crises in sickle cell anaemia. *New England Journal of Medicine*, **332**, 1317–1322.

Chaturvedi, S., Bhatnagar, P., Bean, C.J., Milton, J.N., Casella, J.F., Barron-Casella, E., Arking, D.E. & DeBaun, M.R. (2017) Genome-wide association study to identify variants associated with vaso-occlusive pain in sickle cell anaemia. *Blood*, **130**, 686–688 pii: blood-2017-02-769661.

Costigan, M., Belfer, I., Griffin, R.S., Dai, F., Barrett, L.B., Coppola, G., Wu, T., Kiselycznyk, C.,

Poddar, M. & Lu, Y. (2010) Multiple chronic pain states are associated with a common amino acid–changing allele in *KCNS1*. *Brain*, **133**, 2519–2527.

Darbari, D.S., Van Schaik, R.H., Capparelli, E.V., Rana, S., Mccarter, R. & Van Den Anker, J. (2008) *UGT2B7* promoter variant− 840G> A contributes to the variability in hepatic clearance of morphine in patients with sickle cell disease. *American Journal of Hematology*, **83**, 200–202.

Darbari, D.S., Onyekwere, O., Nouraie, M., Minniti, C.P., Luchtman-Jones, L., Rana, S., Sable, C., Ensing, G., Dham, N. & Campbell, A. (2012) Markers of severe vaso-occlusive painful episode frequency in children and adolescents with sickle cell anaemia. *The Journal of Pediatrics*, **160**, 286–290.

Darbari, D.S., Wang, Z., Kwak, M., Hildesheim, M., Nichols, J., Allen, D., Seamon, C., Peters-Lawrence, M., Conrey, A. & Hall, M.K. (2013) Severe painful vaso-occlusive crises and mortality in a contemporary adult sickle cell anaemia cohort study. *PLoS One*, **8**, e79923.

Elmariah, H., Garrett, M.E., Castro, L.M., Jonassaint, J.C., Ataga, K.I., Eckman, J.R., Ashley-Koch, A.E. & Telen, M.J. (2014) Factors associated with survival in a contemporary adult sickle cell disease cohort. *American Journal of Hematology*, **89**, 530–535.

Eyler, C.E., Jackson, T., Elliott, L.E., De Castro, L.M., Jonassaint, J., Ashley-Koch, A. & Telen, M.J. (2008) beta(2)-Adrenergic receptor and adenylate cyclase gene polymorphisms affect sickle red cell adhesion. *British Journal of Haematology*, **14**, 105–108.

Galarneau, G., Coady, S., Garrett, M.E., Jeffries, N., Puggal, M., Paltoo, D., Soldano, K., Guasch, A., Ashley-Koch, A.E., Telen, M.J., Kutlar, A., Lettre, G. & Papanicolaou, G.J. (2013) Gene-centric association study of acute chest syndrome and painful crisis in sickle cell disease patients. *Blood*, **122**, 434–442.

Geard, A., Pule, G.D., Chetcha Chemegni, B., Ngo Bitoungui, V.J., Kengne, A.P., Chimusa, E.R. & Wonkam, A. (2017) Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *British Journal of Haematology*, https://doi.org/10.1111/bjh.14724

Guasch, A., Zayas, C.F., Eckman, J.R., Muralidharan, K., Zhang, W. & Elsas, L.J. (1999) Evidence that microdeletions in the alpha globin gene protect against the development of sickle cell glomerulopathy in humans. *Journal of the American Society of Nephrology*, **10**, 1014–1019.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O. & Choudhury, A. (2015) The African genome variation project shapes medical genetics in Africa. *Nature*, **517**, 327–332.

Higgs, D.R., Aldridge, B.E., Lamb, J., Clegg, J.B., Weatherall, D.J., Hayes, R.J., Grandison, Y., Lowrie, Y., Mason, K.P. & Serjeant, B.E. (1982) The interaction of alpha-thalassemia and homozygous sickle-cell disease. *New England Journal of Medicine*, **306**, 1441–1446.

Hoots, W.K. & Shurin, S.B. (2012) Future directions of sickle cell disease research: the NIH perspective. *Pediatric Blood & Cancer*, **59**, 353–357.

Hoppe, C., Jacob, E., Styles, L., Kuypers, F., Larkin, S. & Vichinsky, E. (2017) Simvastatin reduces vaso-occlusive pain in sickle cell anaemia: a pilot efficacy trial. *British Journal of Haematology*, **177**, 620–629.

Hu, X., Jhun, E.H., Yao, Y., He, Y., Molokie, R.E., Wilkie, D.J. & Wang, Z.J. (2016) *IL1A rs1800587* associates with chronic noncrisis pain in sickle cell disease. *Pharmacogenomics*, **17**, 1999–2006.

Jhun, E., He, Y., Yao, Y., Molokie, R.E., Wilkie, D.J. & Wang, Z.J. (2014) Dopamine D3 receptor Ser9Gly and catechol-o-methyltransferase Val158Met polymorphisms and acute pain in sickle cell disease. *Anesthesia and Analgesia*, **119**, 1201–1207.

Jhun, E.H., Yao, Y., He, Y., Kyle, M.A., Wilkie, D.J., Molokie, R.E. & Wang, Z.J. (2015) Prevalence of pain-related single nucleotide polymorphisms in patients of African origin with sickle cell disease. *Pharmacogenomics*, **16**, 1795–1806.

Joly, P., Gagnieu, M., Bardel, C., Francina, A., Pondarre, C. & Martin, C. (2012) Genotypic screening of the main opiate-related polymorphisms in a cohort of 139 sickle cell disease patients. *American Journal of Hematology*, **87**, 534–536.

Kanter, J. & Kruse-Jarres, R. (2013) Management of sickle cell disease from childhood through adulthood. *Blood Reviews*, **27**, 279–287.

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.

Keller, S., Yang, M., Treadwell, M.J. & Hassell, K.L. (2017) Sensitivity of alternative measures of functioning and wellbeing for adults with sickle cell disease: comparison of PROMIS® to ASCQ-Me℠. *Health and Quality of Life Outcomes*, **15**, 117.

Leonardo, F.C., Brugnerotto, A.F., Domingos, I.F., Fertrin, K.Y., Albuquerque, D.M., Bezerra, M.A., Araújo, A.S., Saad, S.T., CostA, F.F. & MenzeL, S. (2016) Reduced rate of sickle-related complications in Brazilian patients carrying HbF-promoting alleles at the BCL11A and HMIP-2 loci. *British Journal of Haematology*, **173**, 456–460.

Lettre, G., Sankaran, V.G., Bezerra, M.A., Araujo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N. & Orkin, S.H. (2008) DNA polymorphisms at the *BCL11A, HBS1L-MYB*, and beta-globin loci associate with foetal haemoglobin levels and pain crises in sickle cell disease. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11869–11874.

Machado, R.F., Barst, R.J., Yovetich, N.A., Hassell, K.L., Kato, G.J., Gordeuk, V.R., Gibbs, J.S., Little, J.A., Schraufnagel, D.E., Krishnamurti, L., Girgis, R.E., MorriS, C.R., Rosenzweig, E.B., Badesch, D.B., Lanzkron, S., Onyekwere, O., Castro, O.L., Sachdev, V., Waclawiw, M.A., WoolsoN, R., Goldsmith, J.C. & Gladwin, M.T.; Walk-PHaSST Investigators and Patients. (2011) Hospitalisation for pain in patients with sickle cell disease treated with sildenafil for elevated TRV and low exercise capacity. *Blood*, **118**, 855–864.

McMillan, J.E., Meier, E.R., Winer, J.C., Coco, M., Daymont, M., Long, S. & Jacobs, B.R. (2015) Clinical and geographic characterization of 30-day readmissions in pediatric sickle cell crisis patients. *Hospital Pediatrics*, **5**, 423–431.

Meier, E.R., Fasano, R.M. & Levett, P.R. (2017) A systematic review of the literature for severity predictors in children with sickle cell anemia. *Blood Cells, Molecules, and Diseases*, **65**, 86–94.

Mendonça, T., Oliveira, M., Vasconcelos, L., Pereira, L., Moura, P., Bezerra, M., Santos, M., Araújo, A. & Cavalcanti, M. (2010) Association of variant alleles of *MBL2* gene with vasoocclusive crisis in children with sickle cell anaemia. *Blood Cells, Molecules, and Diseases*, **44**, 224–228.

Mnika, K., Pule, G.D., Dandara, C. & Wonkam, A. (2016) An expert review of pharmacogenomics of sickle cell disease therapeutics: not yet ready for global precision medicine. *OMICS: A Journal of Integrative Biology*, **20**, 565–574.

Mtatiro, S.N., Singh, T., Rooks, H., Mgaya, J., MarikI, H., Soka, D., Mmbando, B., MsakI, E., Kolder, I. & Thein, S.L. (2014) Genome wide association study of foetal haemoglobin in sickle cell anaemia in Tanzania. *PLoS One*, **9**, e111464.

Njamnshi, A., Mbong, E., Wonkam, A., Ongolo-Zogo, P., Djientcheu, V., Sunjoh, F., Wiysonge, C., Sztajzel, R., Mbanya, D. & Blackett, K.N. (2006) The epidemiology of stroke in sickle cell patients in Yaounde, Cameroon. *Journal of the Neurological Sciences*, **250**, 79–84.

Owusu-Ansah, A., Ihunnah, C.A., Walker, A.L. & Ofori-Acquah, S.F. (2016) Inflammatory targets of therapy in sickle cell disease. *Translational Research*, **167**, 281–297.

Piel, F.B., PatiL, A.P., Howes, R.E., NyangirI, O.A., Gething, P.W., Dewi, M., Temperley, W.H., Williams, T.N., Weatherall, D.J. & Hay, S.I. (2013) Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet*, **381**, 142–151.

Platt, O.S., Thorington, B.D., Brambilla, D.J., Milner, P.F., Rosse, W.F., Vichinsky, E. & Kinney, T.R. (1991) Pain in sickle cell disease: rates and risk factors. *New England Journal of Medicine*, **325**, 11–16.

Platt, O.S., Brambilla, D.J., Rosse, W.F., Milner, P.F., Castro, O., Steinberg, M.H. & Klug, P.P. (1994) Mortality in sickle cell disease–life expectancy and risk factors for early death. *New England Journal of Medicine*, **330**, 1639–1644.

Rees, D.C., Williams, T.N. & Gladwin, M.T. (2010) Sickle-cell disease. *The Lancet*, **376**, 2018–2031.

Rumaney, M.B., Bitoungui, V.J.N., Vorster, A.A., Ramesar, R., Kengne, A.P., Ngogang, J. & Wonkam, A. (2014) The co-inheritance of alpha-thalassemia and sickle cell anaemia is associated with better haematological indices and lower

consultations rate in Cameroonian patients and could improve their survival. *PLoS One*, **9**, e100516.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. (1985) Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science*, **230**, 1350–1354.

Sheehan, V.A., Luo, Z., Flanagan, J.M., Howard, T.A., Thompson, B.W., Wang, W.C., Kutlar, A. & Ware, R.E. (2013) Genetic modifiers of sickle cell anaemia in the BABY HUG cohort: influence on laboratory and clinical phenotypes. *American Journal of Hematology*, **88**, 571–576.

Tarer, V., Etienne-Julan, M., Diara, J., Belloy, M.S., Mukizi-Mukaza, M., Elion, J. & Romana, M. (2006) Sickle cell anaemia in Guadeloupean children: pattern and prevalence of acute clinical events. *European Journal of Haematology*, **76**, 193–199.

Weatherall, D. & Clegg, J. (2001) Inherited haemoglobin disorders: an increasing global health problem. *Bulletin of the World Health Organization*, **79**, 704–712.

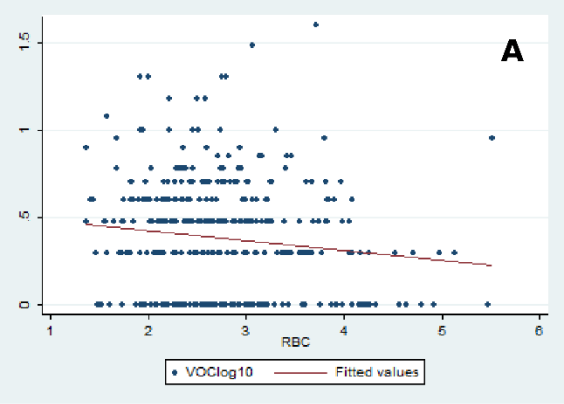
Wonkam, A., Mba, C.Z., Mbanya, D., Ngogang, J., Ramesar, R. & Angwafo, F.F. III (2014a) Psychosocial stressors of sickle cell disease on adult patients in Cameroon. *Journal of Genetic Counseling*, **23**, 948–956.

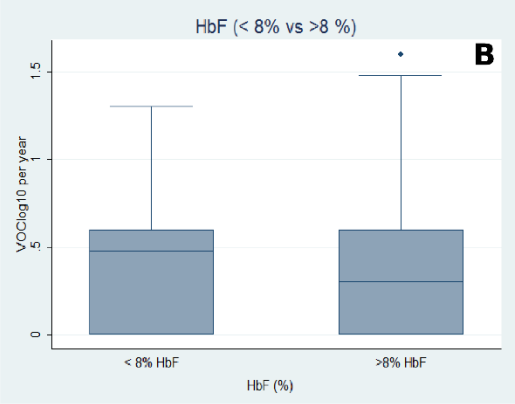
Wonkam, A., BitounguI, V.J.N., Vorster, A.A., Ramesar, R., Cooper, R.S., Tayo, B., Lettre, G. & Ngogang, J. (2014b) Association of variants at *BCL11A* and *HBS1L-MYB* with haemoglobin F and hospitalisation rates among sickle cell patients in Cameroon. *PLoS One*, **9**, e92506.

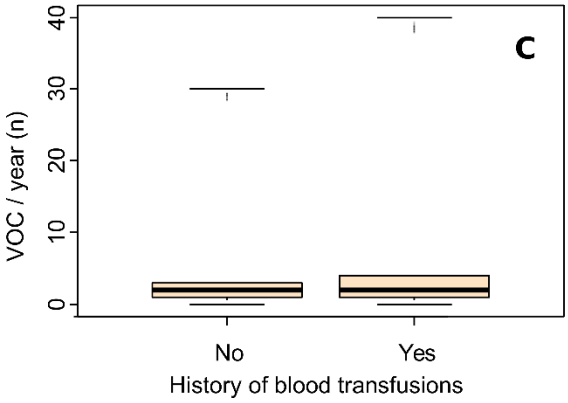
World Health Organization. (2010) Trends in Maternal Mortality: 1990 to 2008. Estimates Developed by WHO, UNICEF, UNFPA and the World Bank. World Health Organization, Geneva, Switzerland.

Zivot, A., Apollonsky, N., Gracely, E. & Raybagkar, D. (2017) Body mass index and the association with vaso-occlusive crises in pediatric sickle cell disease. *Journal of Pediatric Hematology/Oncology*, **39**, 314–317.

Supplementary data







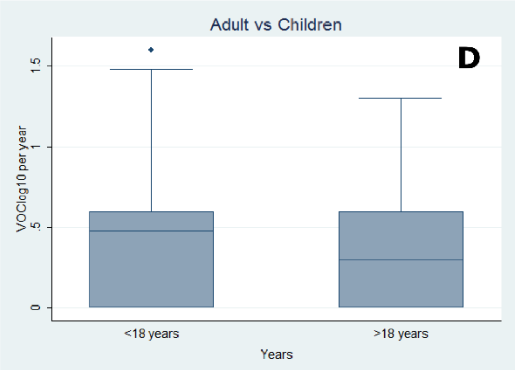


Figure S1: Scatter plot, and box and whisker illustrating association of haematological factors, and age, with Acute Vaso-occlusive painful (VOC) episodes. (A) Scatter plot illustrating the negative relationship between log VOC and red blood cell count (RBC) (estimate = -0.0154; p = 0.015), in line with Hb level of figure 1 (Hb, estimate = -0.074; p = 0. 005); (B) Box and whisker plots showing the correlation of VOC values with HbF level (estimate = -0.003; p = 0.025). (C) There was a positive correlation between history of blood transfusion and VOC (estimate = 0.42; p = 0.046). (D) Younger age also tended non-significantly associated with VOC (estimate = -0.003, p = 0.19). Conventions are as per figure 1.

Table S1. Allele frequencies of selected pain-related genes variants among Cameroonian SCD patients and data extracted from the 1000 Genome project

GENE	dbSNP ID	Position	Allele Changes	Cameroon SCD (Current study)	Africans	East Asians	Europeans	Cameroon SCD vs Africans P values	Cameroon SCD vs East Asians P values	Cameroon SCD vs Europeans P values
ABCB1	rs1045642	87509329	T>C	0.847	0.150	0.398	0.518	0.825	**0.00001**	**0.00001**
ADRA1A	rs1048101	26770511	T>C	0.823	0.216	0.107	0.567	0.056	**0.00001**	**0.00001**
ADRA2A	rs3750635	57502020	T>C	Monomorphic	Monomorphic	Monomorphic	Monomorphic	*NA*	*NA*	*NA*
ADRB2	rs1042713	148826877	A>G	0.481	0.48	0.451	0.614	0.964	0.206	**0.00001**
ARRB2	rs1045280	47193043	C>T	0.342	0.595	0.175	0.312	**0.0032**	**0.00001**	0.162
AVPR1A	rs10877969	63153459	T>C	0.779	0.613	0.154	0.131	**0.0001**	**0.00001**	**0.00001**
BDKRB2	rs1799722	96204802	C>T	0.253	0.284	0.473	0.418	0.119	**0.00001**	**0.00001**
CACNA2D3	rs1851048	54587633	C>T	0.136	0.123	0.059	0.338	0.387	**0.00001**	**0.00001**
	rs6777055	55039890	A>C	0.195	0.208	0.095	0.175	0.441	**0.00001**	0.275
COMT	rs4633	19962712	C>T	0.761	0.293	0.270	0.499	**0.0001**	**0.00001**	**0.00001**
	rs6269	19962429	A>G	0.44	0.371	0.342	0.413	**0.0014**	**0.00001**	0.231
	rs4680	19963748	G>A	0.238	0.281	0.28	0.5	**0.032**	**0.0454**	**0.00001**

DRD2	rs4274224	113448730	C>T	0.738	0.722	0.188	0.455	0.404	**0.00001**	**0.00001**
FAAH	rs324419	46406314	T>C	0.128	0.141	0.005	0.148	0.374	**0.00001**	0.212
	rs2295632	46413890	T>G	0.751	0.682	0.464	0.256	**0.0006**	**0.00001**	**0.00001**
	rs4141964	46399368	T>C	0.736	0.691	0.465	0.37	**0.026**	**0.00001**	**0.00001**
KCNS1	rs734784	45094986	A>G	0.531	0.620	0.214	0.464	**0.0001**	**0.00001**	**0.0039**
OPRM1	rs1799971	154039662	A>G	0.001	0.009	0.393	0.162	**0.017**	**0.00001**	**0.00001**
STAT6	rs841718	57099213	C>T	0.683	0.294	0.331	0.556	0.247	0.508	**0.00001**
	rs3024971	57099944	A>C	0.022	0.035	0.087	0.107	0.082	**0.00001**	**0.00001**
TRPA1	rs920829	72065468	G>A	0.304	0.295	0.285	0.127	**0.0001**	**0.00001**	**0.0002**
TRPV1	rs222747	3589906	G>C	0.088	0.11	0.558	0.248	0.083	**0.00001**	**0.00001**
UGT2B7	rs7438135	69095621	G>A	0.7	0.227	0.275	0.486	**0.0001**	0.217	**0.00001**

Significant P values are bolded; *NA* = not applicable

3.2 Mnika K, Bope CD, Mazandu GK, Chimusa ER, Nembaware V, Dandara C, Wonkam A. Analyses of 299 Pharmacogenes in Sickle Cell Disease Patients in Africa. (under review)

Nature of Publication: Original Full Journal Article
Journal/Publisher: Frontiers in Pharmacology, section Pharmacogenetics and Pharmacogenomics, Peer reviewed

Candidate contribution: Performed experiments pertaining the present analysis of 299 pharmacogenes in Sickle Cell Disease patients in Africa.

Co-Author contribution:

AW: Conceived and designed the experiments
KM, AW, CDB, GKM, ERC: Performed the experiments
AW, GP, VBN, KM: Patients' recruitment, samples and clinical data collection and processing
CDB, GKM, ERC, KM, AW: Analysed the data
AW, ERC, GKM, CDB, VN, CD: Contributed reagents/materials/analysis tools
KM, GKM, AW: Wrote the paper
KM, GKM, CDB, ERC, VN, GP, VBN, CD, AW: Revised and approved the manuscript

Analyses of 299 Pharmacogenes in Sickle Cell Disease Patients in Africa

Khuthala Mnika[1,2], Christian D. Bope[1,3], Gaston K. Mazandu[1, 4], Gift Pule[1] , Valentina Josiane Ngo Bitoungui [1] ,Victoria Nembaware[1], Collet Dandara[1,2], Emile R. Chimusa[1,2], and Ambroise Wonkam[1,2]

[1]Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [2]Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [3]University of Kinshasa, Faculty of Sciences, Departments of Mathematics and Computer Sciences, Kinshasa, Democratic Republic of Congo,[4] African Institute for Mathematical Sciences (AIMS), Cape Town, South Africa.

Corresponding author:

Ambroise Wonkam, MD, PhD
Division of Human Genetics, Department of Medicine, and
Institute of Infectious Disease and Molecular Medicine
Faculty of Health Sciences, University of Cape Town
Anzio Road, Observatory, 7925, Cape Town, South Africa
E-mail: ambroise.wonkam@uct.ac.za

Paper type: original paper
Number of Tables: 2
Number of Figure: 4
Supplementary tables: 1
Word count: 4297

Abstract

Sickle cell disease (SCD) is caused by a mutation that leads to polymerization of hemoglobin S in erythrocytes. Inflammation, haemolysis, microvascular obstruction, and multiple organ damage characterize SCD. Globally, an estimated 305 800 babies with SCD are born annually, with 75% from Africa. Most SCD patients are subjected to lifelong medications, with variable individual responses to therapies. However, polymorphisms in genes involved in drug absorption, distribution, metabolism and excretion (ADME) have seldom been explored in SCD.

In this study, we used the Affymetrix PharmacoScan® Array to investigate variations in 299 pharmacogenes (32 ADME core and 267 extended pharmacogenes) in a group of SCD patients from Cameroon (*n* = 123), with comparisons with ethnically matched controls (*n* = 40), and to an independent group of SCD patients (*n* = 25) from the Democratic Republic of the Congo (DRC). We also compared the pharmacogenes variants from these African groups, to data extracted from the 1000 genomes Project. Moreover, association studies were carried out on pharmacogenes variants with SCD clinical variability. Additionally, knowledge-based unified protein-protein interaction (PPI) network and enriched biological processes and pathways were investigated.

Principal component analysis showed no difference in the distribution of pharmacogenes variants among Cameroonian SCD and matched controls. Cameroonians data clustered with other Africans, but are significantly distinct from American, European and Asian populations data. Variants allele frequencies in 21/32 core pharmacogenes were significantly different between the two SCD groups (Cameroon vs. Congo). Single nucleotide polymorphisms (SNPs) in 50 genes have significant associations with vaso-occlusive painful (VOC) episodes in SCD patients. Among them, four genes were identified to be essential: *NTSR1, LRMDA, SMAD SMAD4 and CDH2,* and all interacted

with three core pharmacogenes equally associated with VOC: *UGT1A8, UGT1A10* and *SLCO4A1.* Three biological processes and four pathways were enriched with VOC-associated genes variants, including xenobiotic glucuronidation (GO:0052697, p = 2.3e-03), and drug metabolism - other enzymes (p = 2.1e-02), that are relevant to SCD pathophysiology.

Our data reveals that the commercially available pharmacogenes arrays investigated might not be suitable for Africans. We found novel associations among pharmacogenes variants with pain in SCD, that deserve further investigations.

Keywords: ADME; Pharmacogenomics; Sickle Cell Disease; Africa.

Introduction

Sickle cell disease (SCD) is caused by a single point mutation (c.20A>T, p. Glu6Val) in beta-globin gene, that leads to polymerisation of hemoglobin S (HbS) and the sickling of erythrocytes. HbS is associated with inflammation, haemolysis, microvascular obstruction, and multiple organ damage, specifically in the cardiovascular system, leading to acute and chronic complications in SCD such as stroke, acute chest syndrome, and kidney dysfunctions (Piel et al., 2017). The pathophysiology of SCD is influenced by both environmental and genetic factors. The mutation causing sickle cell is documented to have appeared in Africa about 7000 years ago (Shriner and Rotimi, 2018) and its high prevalence in Africa is driven by the protection against malaria conferred on mutation carriers. It is estimated that more than 300 000 babies are born each year with the disease worldwide, with nearly 75% of these births in sub-Saharan Africa (Piel et al., 2013). Most SCD patients are subjected to long-term treatment, mainly using hydroxyurea (HU), antibiotics, and opioids (Mnika et al., 2016). High variability in individual responses to these therapies has been reported (Mnika et al., 2016; Wonkam et al., 2018). While environmental factors affecting responses to treatment have been reported, little is known about their pharmacogenomic drivers of treatment response in SCD (Yahouédéhou et al., 2018).

Drug responses are highly variable both within and among populations groups (Ahmed et al., 2016; Ma and Lu, 2011), even after taking into account demographic factors such as age, gender and geographical location. An individual's response to drug treatment is a crucial aspect of therapeutic outcomes (Li et al., 2011). Thus, it is important to decode genomic drivers of such variability. Polymorphisms in genes involved in the absorption, distribution, metabolism, and excretion (ADME) processes, play an important role in determining the pharmacokinetic profiles of drugs, and contribute to the heterogeneity observed on drug responses in human populations (Clark et al., 2005; Ahn and Park, 2017; Li et al., 2014; Ramos et al., 2014; Birmingham et al., 2015; Kurose et al., 2012). Population specific data is needed, because of the qualitative and quantitative differences

in the distribution of variants of pharmacogenomic relevance. For example, *CYP3A5*3* is present in less than 10% in African populations but has been reported at frequencies of up to 80% among populations of European extraction and *CYP2D6*17* is an African specific allele (Chen et al., 2009, Wooding et al., 2002). *CYP3A5* and *CYP2D6* between them, metabolise nearly 50% of commonly used medication, that's, the effects of the variants will differ in different populations. The core and extended ADME genes, as defined by the PharmaADME Consortium, represent the most important genes directly involved in the majority of drug metabolism (PharmaADME.org, 2019). It is particularly critical that these genes be investigated in common conditions for which patients are subjected to lifelong treatment, such as SCD in the context of Africa.

Comprehensive clinical care programmes have reduced premature childhood deaths related to sickle cell disease by 70% in the USA and increased adult life expectancy to the 6th decade (Chaturvedi and Debaun, 2016). Unfortunately, sickle cell disease mortality in adults has not decreased in the last 30 years because of acute and chronic end-organ damage, for example, cardiovascular complications. Over the past decade, there have been numerous promising clinical trials that address all aspects of potential therapy based on known sickle cell pathophysiology e.g. hemoglobin sickling (Vichinsky et al., 2019), endothelial dysfunction (Ataga et al., 2017), and oxidative stress (Niihara et al., 2018). Moreover, major progress has been made in the area of bone marrow transplantation (Bernaudin et al., 2019; Bolaños-Meade et al., 2019), and gene therapy (Ribeil et al., 2017), all of which show potential to benefit sickle cell patients in the near future. Therefore, as part of complementing research on the development of effective life-long, accessible and efficient therapies for patients with SCD, our group is investigating the pharmacogenetics of SCD in order to map the profile of patients likely to benefit from new therapeutic drugs. To date no study has focused on comprehensive pharmacogenomics genes in SCD patients.

In this study we investigated variation in 32 ADME core and 267 extended pharmacogenes in a selected group of SCD patients from Cameroon used a commercially available Affymetrix array named PharmacoScan®. Specifically, 1) we assessed variant minor allele frequency (MAF) differences in SCD patients versus ethnically matched controls, and 2) compared these to an independent SCD group from the Democratic Republic of Congo (DRC); moreover, 3) we compared these Cameroonian data to that of other African and non-African populations extracted from the 1000 genomes Project; and lastly, 4) we evaluated pharmacogene variant associations with varied sub-groups of SCD patients i.e. a "long survivor group" (age over 40 years), a "stroke group" (at least one episode of overt stroke), and a "random group" (patients younger than 40 years with no cerebrovascular disease), and analysed associations between pharmacogene variants and specific SCD clinical events and endophenotypes. e.g. VOC, stroke, and fetal haemoglobin (HbF) levels.

Materials and Methods

Ethical approvals

The study was performed in accordance with the Declaration of Helsinki. Ethical approval was granted by the National Ethical Committee Ministry of Public Health, Republic of Cameroon (No 033/CNE/DNM/07); and the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee (HREC REF: 132/2010 and HREC REF: 475/2019). Written and signed informed consent was obtained from all participants who are 18 years of age or older, and from parents or guardians in cases of minors, with verbal assent from participants aged 7 years or older.

Study participants and assessment of clinical events

A total of 148 SCD patients was investigated, comprising of 123 Cameroonian SCD patients recruited from Yaounde Central Hospital and Laquintinie Hospital in Douala (Wonkam et al., 2014) , 3 patients were on HU and 25 recent migrants from the DRC

SCD patients, recruited at the Haematology Clinic, Groote Schuur Hospital in Cape Town, South Africa (Pule et al., 2017a), 23 patients were on HU. Socio-demographic and clinical data were collected by means of a structured questionnaire. Patients' medical records were reviewed to describe their clinical features over the past 3 years. Specifically, the occurrences of VOC, hospital outpatient visits, hospitalisations, overt strokes, blood transfusions, and administration of hydroxyurea were recorded. Height, weight, body mass index (BMI), systolic and diastolic blood pressures (SBP and DBP) were measured. Detailed descriptions of patients and sampling methods used in the Cameroonian patients have been reported previously (Wonkam et al., 2014; Pule et al., 2017). For the purpose of comparing frequencies of pharmacogenes variants, a total of 40 ethnically matched Cameroonian controls were randomly recruited from apparently healthy blood donors in Yaounde for participation in the study. All blood samples were collected for genomic characterisation and analysis.

Molecular methods

Characterisation of the sickle cell anaemia mutation, hemoglobin gene (HBB) cluster haplotypes, and the 3.7kb HBA1/HBA2 deletion

Molecular analysis to determine the presence of the sickle mutation was carried out on 100 ng DNA by polymerase chain reaction (PCR) amplification of a 770 bp segment of the *HBB* gene, followed by *Dde*I restriction analysis of the PCR product (Saiki et al., 1985). Using published primers and methods, five restriction fragment length polymorphism (RFLP) sites in the *HBB* cluster were amplified to perform analysis using the restriction enzymes *Xmn*I (5'Gc), *Hind*III (Gc), *Hind*III (Ac), *Hinc*II (3wb') and *Hinf*I (5'b) for the *HBB* haplotype background (Bitoungui et al., 2015). The 3.7kb *HBA1/HBA2* deletion was successfully screened using expand-long template PCR (Roche Diagnostics, Basel, Switzerland), as previously published (Rumaney et al., 2014).

Determination of pharmacogenomics variants using PharmacoScan® SNP

Genotyping

A commercially available Affymetrix SNP array, PharmacoScan® was used in in the genotyping of participants DNA for variants on 267 pharmacogenomics-related genes. The pharmacogenes characterisation was carried out at the Centre for proteomics and genomic research (CPGR) in Cape Town. The genotyping method uses molecular inversion probe (MIP) technology on the Affymetrix PharmacoScan® (Affymetrix Inc., Santa Clara, CA, USA), enabling the genotyping of 4,627 markers in 1,191 genes of known pharmacogenomic value. This includes core functional pharmacogenomic content, containing markers recommended by the clinical pharmacogenomics implementation consortium (CPIC) guidelines (Caudle et al., 2014), pharmacogenomics knowledge base (PharmGKB) markers in very important pharmacogenes (VIP), PharmGKB markers with clinical annotations, and pharmaADME core markers (Whirl-Carrillo et al., 2012). PharmacoScan also contain all markers from the Applied Biosystems™ DMET™ Plus Solution (1,936 genetic variants across 231 relevant genes), and ADME markers in genes targeted for European populations drawn from the Applied Biosystems™ UK Biobank Axiom™ Array (Deeken, 2009).

PharmacoScan® SNP calling and quality control

A total of 188 samples were genotyped. The experimental quality control of the chip was assessed using the apt-geno-qc in Affymetrix Power Tools (APT). Only raw intensity data (.cel files) with quality rate greater than 90% were retained for further analysis (*n*=188). To generate genotyping calls, APT based on apt-probeset-genotype 1.20 was implemented using the Bayesian Robust Linear Model with Mahalanobis (BRLMM) distance classifier by assessing multiple raw intensity data (.cel files) at once. The probe IDs were converted to rs IDs using the PharmacoScan array annotation files in a custom Python script. The output from this was then used to created plink-format files using another custom Python script.

PharmacoScan-specific population structure

Since 1000 genomes data is from whole genome sequencing, we extracted pharmaco-SNPs that overlapped with our dataset in 299 genes (32 ADME core and 267 extended pharmacogenes). We used Plink® 1,9 to merge resulting specific pharmaco-SNPs from 2504 samples which are part of the 1000 genomes, and 188 samples from our current cohort. Principal Component Analysis (PCA) was performed on the merged datasets (*n*=2692) using smartpca, which is part of the EIGENSOFT 3.0 package (Patterson et al., 2006).

Minor allele frequency differentiation and distribution

To examine the extent to which our population samples differ from other populations, we investigated the distribution of minor allele frequencies for variants genotyped with the PharmacoScan® array. To this end, the proportion of minor alleles were categorised into 6 bins (0-0.05, >0.05-0.1, >0.1-0.2, >0.2-0.3, >0.3-0.4, >0.4-0.5) with respect to the patient and other control group frequencies. The MAF for each category was computed using Plink®. Thus, the fraction of SNPs corresponding to each bin was computed. Genotypes for SNPs within core (32) and extended (272) pharmacogenes were extracted from the annotated pharmacogene data using ANNOVAR (Wang et al., 2010), cataloguing closest pharmacogenes to each variant (40kb downstream/upstream) and mapping each SNP to corresponding closest pharmacogenes. SNP-specific unusual allele frequency summary statistics were also aggregated as described in (Chimusa et al., 2015) to obtain gene-specific differences in SNP frequencies. The obtained p-values were adjusted using the Benjamini-Hochberg multiple comparison model and a Manhattan plot was generated using R (R Development Core Team, 2011). The significance level cut-off was set to 0.001.

Association analysis

Given the incidence of SCD in Cameroon is 0.8% (WHO, 2010) and the incidence of heterozygous HbAS patients in Cameroon is 8–34 % (Weatherall and Clegg, 2001), the

required sample size was calculated using the normal approximation to the binomial model (Fosgate, 2009). For an error estimation of 10%, approximately 162 samples are required with a confidence level of 95% to achieve sufficient statistical power. For an approximate sample size of n = 188, a desired level of statistical power was guaranteed and the association was then performed using genome-wide complex trait analysis (GCTA) (Yang et al., 2011) with the significance cut-off set to 3.9e-7 (using Bonferroni multiple corrections with cut-off = 0.05/number of SNPs).

Systems level Analysis for Identifying Potential Biological Targets

We analysed variant genes associated to VOC as an integrated system using topological properties of knowledge-based unified protein-protein interaction (PPI) network retrieved from different sources, including STRING (Szklarczyk et al., 2019), database of interacting protein (DIP) (Salwinski et al., 2004), IntAct (Orchard et al., 2014), molecular interaction database (MINT) (Licata et al., 2012), MIPS (Mewes et al., 2006), human protein reference database (HPRD) (Keshava Prasad et al., 2009) and biological general repository for interaction database (BioGRID) (Oughtred et al., 2019) databases, as well as functional interactions derived from protein sequence (Mazandu and Mulder, 2011) similarity and conserved protein domains from the InterPro database (Mitchell et al., 2019). This enabled the identification of essential variant genes by computing different network centrality measures (betweenness, closeness and degree) and mapping variant genes to the PPI network. We also retrieved enriched biological processes and pathways using a hypergeometric test adjusted with the Bonferroni multiple corrections as implemented in the A-DaGO-Fun tool (Mazandu et al., 2016). For this enrichment analysis, biological processes, human protein and process mapping, and biological pathways were derived from the Gene Ontology (The Gene Ontology Consortium, 2019), UniProt Gene Ontology Annotation (Huntley et al., 2015) and the Kyoto encyclopeadia of genes and genomes (KEGG) pathway (Kanehisa et al., 2019) online databases, respectively.

Results

Cameroonian Sickle cell disease patients' description

A total of 122 SCD patients (HbSS) and 40 controls from Cameroon were included. Among the Cameroonian sample, there was roughly an equal number of males (*n*=63) and females (*n*=59) and the median age was 26 years. Only 31.1% (*n=38/122*) of SCD patients had co-inherited a single or double 3·7 kb *HBA1/HBA2* deletion, and 2.5 % of patients were on HU treatment (see **Table 1**).

Pharmacogenes' variants in Cameroonian and global populations variability

Comparing the Cameroonian (combined SCD and controls) with other world populations, it is observed that the Cameroonians clustered together regardless of SCD status as presented on the output using two principal components (PCs), PC2 vs PC1 (see **Figure 1**). Data indicate that the two groups are homogenous populations, except for the four control individuals which are far away from the main cluster. Those four individuals were considered outliers and excluded from downstream analyses.

We also merged data from the SCD patients with available 1000 genome datasets. Population abbreviations as per the 1000 genome project are provided in **Figure 1A**. We pruned the merged dataset using the Plink® tool to ensure that SNPs in linkage disequilibrium (LD) are excluded. The pruned set of SNPs contained 115,481 SNPs and PCA was performed using the first two PCs. Results suggest the studied SCD participants are clustered with other African ancestry populations, such as Yoruba in Ibadan, Nigeria (YRI), Esan in Nigeria (ESN), African Caribbeans in Barbados (ACB), Gambian in Western Divisions in the Gambia (GWD), Americans of African Ancestry in South West USA (ASW), Luhya in Webuye, Kenya (LWK) and Mende in Sierra Leone (MSL), with distinct differences from other American, European and Asian populations. This observation is further confirmed by the heatmap shown in **Figure 1B**.

Pharmacogenes variants in two SCD patients from Africa: Cameroon vs Congo

To evaluate the correlation between pharmacogene variants and variable SCD clinical severity, we first intersected common significant SNPs in all populations, namely SCD patients (Cameroonian and DRC). We then sub-categorised Cameroonian SCD patients according to a variable range of clinical severity: a "long survivor group" (age over 40 years) (*n*=26), a "stroke group" (at least one episode of overt stroke) (*n*=22), and a "random group" (patients younger than 40 years with no cerebrovascular disease) (*n*=74). Allele frequencies within each sub-group were categorised into bins and the distributions (frequencies) of SNPs falling into each bin plotted. **Figure 2** reveals the differences between these SCD sub-groups. The results show that there are differences between these populations based on allele frequencies for common variants with MAF > 0.1, in the 21/32 core genes described in Table S1. These differences are more remarkable between SCD patients from the "stroke" DRC and the Cameroonian sub-group. **Figure 2B** illustrates the difference between patients from DRC vs patients from Cameroon samples, supporting high genetic diversity between these two SCD African populations, and suggesting that ancestry, geographical origin and/or location of patients could significantly influence SCD therapeutic outcomes.

In addition, the core pharmacogene integrated MAFs (see **Figure 2C** for all populations under consideration and **Figure 2D** displaying only SCD clinical phenotype in patients from Cameroon vs patients from the DRC) confirm the differences, indicating that there are intra-African differences in MAF in these core genes (p-value = 1.178e-05). This will be important to consider in future studies on drug response.

Pharmacogenes variants in various clinical sub-group of Cameroonian SCD

Investigation of associations between variants in pharmacogenes [core (n=32) and extended (n=267)], within Cameroonian various SCD sub-groups: "stroke", "long survival" and "random" shows no significant difference. However, in two-ways comparisons, the MAF of *HBB*-rs33930165 (p = 8.99e-05) was significantly different between controls and

members of the “long survivor” group. The MAFs of *HBB*-rs33930165 (p = 9.29e-06), *OR51L1*-rs2499984 (p = 3.5e-05), *OR52A1*-rs10768634 (p = 8.6e-04), *L3MBTL*-rs12373231 (p = 8.7e-04), *TRIM22*-rs16934565 (p = 9.29e-04) were also significantly different between the “random” SCD and control groups. The MAF of *EYA2*-rs6018337 (p = 6.6e-04) was also different between “long survivor” and “random” SCD groups when analysing all gene variants included in the PharmacoScan® panel.

Pharmacogenes variants and SCD clinical events/endophenotypes

Association analysis between pharmacogenes’ variants and specific clinical events/endophenotypes among Cameroonian patients i.e. vaso-occlusive painful crisis (VOC), hospitalisation rates, use of HU, and HbF level, indicated that up to 50 genetic variants on the PharmacoScan panel were significantly associated with VOC, two of which are core genes. **Table 2** describes all significant pharmacogenes’ variants that are highlighted in the Manhattan plot in **Figure 3**. No significant differences were observed with hospitalisation rates, use of HU, and HbF level.

Vaso-occlusive painful crisis (VOC) associated variants genes, processes and pathways

Among 50 genes predicted to be associated with VOC in SCD patients, four genes have been identified to be essential: neurotensin receptor 1 (*NTSR1*), leucine-rich repeat protein *(LRMDA),* SMAD family member 4 *(SMAD4)* and cadherin 2 (*CDH2*). Among these four essential genes, one gene, namely *NTSR1*, appeared be particularly important based on the network centrality measures (betweenness, closeness and degree). In addition, these essential genes interact with core pharmacogenes *UGT1A8*, *UGT1A10* and *SLCO4A1*. A summary of a biological sub-network of the identified essential and the network centrality measures of essential genes are shown in **Figure 4**.

Furthermore, these VOC-associated genes variants were used to elucidate enriched biological processes and pathways in which they are involved. Three biological processes

and four pathways have been identified. These enriched processes are: (1) flavonoid glucuronidation (GO:0052696, p = 2.3e-03), (2) xenobiotic glucuronidation (GO:0052697, p = 2.3e-03) and (3) flavone metabolic process (GO:0051552, p = 8.20e-05). The four biological pathways are: (1) pentose and glucuronate interconversions (p = 8.4e-03), (2) porphyrin and chlorophyll metabolism (p =1.4e-02), (3) ascorbate and aldarate metabolism (p = 4.4e-03) and (4) drug metabolism - other enzymes (p = 2.1e-02).

Discussion

This is the first study to use the PharmacoScan® array on two African populations, and will serve as a benchmark for future studies aiming at investigating pharmacogenomic variants in Africa. Principal component analysis (PCA) showed that data from Cameroonian SCD patients and control groups were homogeneous, and clustered with data from other populations of African ancestry, but were clearly distinct from American, European and Asian populations (see **Figure 1**). This is an important finding that points to the possible limits of using this commercially available array among Africans, which may elicit different drug responses, as compared to populations of European and Asian ancestries, without considering African population stratifications. This also indicates the opportunity to design and perform appropriate pharmacogenetic and dynamic studies in Africa in relation to pharmacogenetic variants that still need to be properly studied in Africans, likely through Whole Exome Sequencing (WES) of multiple African participants from diverse ethnolinguistic and geographical locations. This task is severely complicated by the scarcity of WES data representative of genetic diversity of populations across the continent in public data bases (Abboud et al., 2017). Indeed, populations of African ancestry have an evolutionary history spanning nearly 300 000 years. This has resulted in many more variations being present in African genomes than in any other population in the world. A recent deep-sequenced dataset of 910 individuals of African descent demonstrated that the African pan-genome contains ~10% more DNA than the current human reference genome (Sherman et al., 2019). Moreover, we found that core ADME gene variants have different patterns in two different SCD groups of patients from

Cameroon and the DRC (see **Figure 2B, Table S1**). This is likely to due to the fact that African populations are genetically more diverse than non-African populations (Campbell and Tishkoff, 2008), which may in part explain the current genomic focus on Africa. These differences could have been shaped over the years by specific local adaptation, possibly resulting in differentiation in the ADME genes across populations in multiple African populations from different geographical locations. Differences observed between stroke patients and other groups (see **Figure 2A**) could potentially explain the enrichment of specific variants from an evolutionary viewpoint, as they could be protective or increase susceptibility to a specific SCD phenotypes. This can be explored in future studies appropriately designed for genetic markers to perform selection analysis.

In the present study, we reported for the first time 50 genes variants associated with VOC (see **Table 2**) that are the most critically challenging clinical expression for SCD patients and parents. There findings will add the rare available studies on pharmacogenetics of SCD in Africa (Mnika et al., 2016; Wonkam et al., 2018). Three of these genes are among core-pharmacogenes: *UGT1A8, UGT1A10 and SLCO4A1.* The *UGT1A8* gene variant was previously associated with bilirubin levels in SCD patients (Milton et al., 2012). Unconjugated or indirect bilirubin is responsible for all of the toxic effects of bilirubin, clinically observed in SCD patients. To permit excretion, bilirubin complexes with albumin are transported via the bloodstream to the liver, where it undergoes glucuronidation by the UGT family of enzymes (Perera et al., 2008). Genetic alterations in the glucuronidation pathway have been linked to abnormalities in the hepatic metabolism of certain medications, predisposition to cardiovascular disease (Hunt et al., 2001; Lammert and Matern, 2005; Lin et al., 2006; Portincasa et al., 2006), and kidney dysfunctions in SCD (Tsai and Tarng, 2019; Geard et al., 2017; Yawn et al., 2014). These data suggest a clinical importance of bilirubin and related genes, beyond cholelithiasis. The second SNP associated with VOC was *SLCO4A1*-rs118042746. Organic-anion-transporting polypeptides (OATPs), are known as transmembrane proteins with 643–722 amino acids and twelve transmembrane domains, *SLCO4A1* protein being one of them, and has been

reported to play a role in expression of drug transporters and drug metabolizing enzymes in the bladder urothelium in humans (Bexten et al., 2015). The full extent of the roles of the 50 genes associated with VOC in the present study will need more investigation with bigger sample size studies, ideally in a prospective cohort with some pharmacodynamic measures.

Four essential genes variants: *NTSR1*, *LRMDA*, *SMAD4* and *CDH2*, were revealed to be associated with VOC, and all interacting with core pharmacogenes named above, and equally associated with VOC (**Figure 4).** These further highlights the importance of these genes in pain, particularly in SCD patients (Feldman et al., 2008). Specifically, *NTSR1* gene was shown to be associated with trigeminal nerve neoplasm and neurilemmoma on the fifth cranial nerve, mediating the multiple functions of neurotensin (Stelzer et al., 2016), and predicted to be involved in the regulation of sensory perception of pain (Bateman et al., 2017). In addition, variants in *SMAD4* and *CDH2* are associate with aneurysms-osteoarthritis phenotype, a chronic joint disease mainly characterized by pain (Zhang et al., 2015; Shang et al., 2019). Finally, mutations in the *LRMDA* gene have been associated with photophobia in autosomal recessive oculocutaneous albinism 7 (Zhong et al., 2019), which is a pain in the eyes due to exposure to bright light (Kohler et al., 2019).

The essential processes found in the present study, in which VOC-associated genes variants are involved, included flavonoid and xenobiotic glucuronidations. These processes are implicated in flaviods undergoing phase II metabolism or extensive glucuronidation process (Kosaka et al., 2011), and constitutes opioid analgesics substrates commonly used in the management of SCA pain (Nagar et al., 2004). Xenobiotic metabolism and glucuronidation are involved in response to iron chelators (Stephanou et al., 2019), used as therapy for iron overload, an inevitable clinical limit point of RBC transfusion-dependent hemoglobinopathies (Vichinsky et al., 2008). Moreover, pathway-based enrichment analysis revealed that these VOC-associated

pharmacogenes participate to signalling cascades: pentose and glucuronate interconversions, porphyrin and chlorophyll metabolism, ascorbate and aldarate metabolism, and Drug metabolism-other enzymes, that are known to be relevant in the SCD pathophysiology (Desai et al., 2017; Darghouth et al., 2011). Taken together these findings may potentially expand the current knowledge of pathophysiology of SCD, and specifically the genetic basis of variable responses to therapy among patients.

To our knowledge, this is the first comprehensive investigation of variants in the 299 core and extended-pharmacogenes in any African population using a widely available commercial array. Strengths of the study include the two independent SCD groups, inclusion ethnically matched controls, comparative analysis with wide range of data from the 1000 genome project, attempt to link the variations to SCD clinical phenotypes, complementary analytical approaches, and linking the identified significant genes variants to biological pathways. However, the study has some limitations. The sample sizes are relatively modest, and there is no pharmacokinetics data to link the variants to specific drug metabolisms, that will require future studies. Nonetheless, the total number of genes/variants provided in the study represents, to date, the only available of such datasets for SCD, and from Africa.

Conclusion

The present study reveals significant differences between African and non-African populations in allele frequencies of core and extended pharmacogenes contained in a commercially available array. In addition, the data has described differences between pharmacogene variants among two different groups of SCD patients according to their geographical origin (Cameroon vs Congo), as well as specific associations of variants in 50 genes, including two that are involved in bilirubin metabolism and with vaso-occlusive painful crisis. The data emphasises the urgent need to perform comprehensive investigation of population-specific pharmacogene variants among Africans, in order to adequately inform clinical management. The present data indicate that the

appropriateness of using commercially available pharmacogene arrays in African populations needs supporting evidence, as the important differential allele frequencies suggest differential anticipated responses among Africans (SCD patients or not) as compared to populations of European and Asian ancestries.

Authors' Contributions

Conceived and designed the experiments: AW. Performed the experiments: KM, AW, CDB, GKM, ERC. Patients' recruitment, samples and clinical data collection and processing: AW, GP, VBN, KM. Analysed the data: CDB, GKM, ERC, KM, AW. Contributed reagents/materials/analysis tools: AW, ERC, GKM, CDB, VN, CD. Wrote the paper: KM, GKM, AW. Revised and approved the manuscript: KM, GKM, CDB, ERC, VN, GP, VBN, CD, AW.

Acknowledgements

Our acknowledgment to all patients and participants.

Dr Gloudi Agenbag and Miss Caitlin McIntosh for helping with grammar edits.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding

The molecular experiments of the study were funded by the National Health Laboratory Services (NHLS), South Africa; and the NIH, USA, grant number 1U01HG007459-01, and NIH/NHLBI U24HL135600 to AW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Abboud, H. E., Abecasis, G., Aguilar-Salinas, C. A., Campos, O. A.-, Atzmon, G., Aukrust, I., et al. (2017). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi:10.1038/nature19057.Analysis.

Ahmed, S., Zhou, Z., Zhou, J., and Chen, S. Q. (2016). Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine. *Genomics, Proteomics Bioinforma.* doi:10.1016/j.gpb.2016.03.008.

Ahn, E., and Park, T. (2017). Analysis of population-specific pharmacogenomic variants using next-generation sequencing data/6. *Sci. Rep.* doi:10.1038/s41598-017-08468-y.

Ataga, K. I., Kutlar, A., Kanter, J., Liles, D., Cancado, R., Friedrisch, J., et al. (2017). Crizanlizumab for the prevention of pain crises in sickle cell disease. *N. Engl. J. Med.* doi:10.1056/NEJMoa1611770.

Benet, L. Z. (2009). The drug transporter-metabolism alliance: Uncovering and defining the interplay. *Mol. Pharm.* doi:10.1021/mp900253n.

Bernaudin, F., Verlhac, S., De Latour, R. P., Dalle, J. H., Brousse, V., Petras, E., et al. (2019). Association of Matched Sibling Donor Hematopoietic Stem Cell Transplantation with Transcranial Doppler Velocities in Children with Sickle Cell Anemia. *JAMA - J. Am. Med. Assoc.* doi:10.1001/jama.2018.20059.

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., et al. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* doi:10.1093/nar/gkw1099.

Bexten, M., Oswald, S., Grube, M., Jia, J., Graf, T., Zimmermann, U., et al. (2015). Expression of drug transporters and drug metabolizing enzymes in the bladder urothelium in man and affinity of the bladder spasmolytic trospium chloride to transporters likely involved in its pharmacokinetics. *Mol. Pharm.* doi:10.1021/mp500532x.

Birmingham, B. K., Bujac, S. R., Elsby, R., Azumaya, C. T., Wei, C., Chen, Y., et al. (2015). Impact of ABCG2 and SLCO1B1 polymorphisms on pharmacokinetics of

rosuvastatin, atorvastatin and simvastatin acid in Caucasian and Asian subjects: A class effect? *Eur. J. Clin. Pharmacol.* doi:10.1007/s00228-014-1801-z.

Bitoungui, V. J. N., Pule, G. D., Hanchard, N., Ngogang, J., and Wonkam, A. (2015). Beta-Globin Gene Haplotypes Among Cameroonians and Review of the Global Distribution: Is There a Case for a Single Sickle Mutation Origin in Africa? *Omi. A J. Integr. Biol.* doi:10.1089/omi.2014.0134.

Bolaños-Meade, J., Reshef, R., Fraser, R., Fei, M., Abhyankar, S., Al-Kadhimi, Z., et al. (2019). Three prophylaxis regimens (tacrolimus, mycophenolate mofetil, and cyclophosphamide; tacrolimus, methotrexate, and bortezomib; or tacrolimus, methotrexate, and maraviroc) versus tacrolimus and methotrexate for prevention of graft-versus-host disease with haemopoietic cell transplantation with reduced-intensity conditioning: a randomised phase 2 trial with a non-randomised contemporaneous control group (BMT CTN 1203). *Lancet Haematol.* doi:10.1016/S2352-3026(18)30221-7.

Campbell, M. C., and Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genomics Hum. Genet.* doi:10.1146/annurev.genom.9.081307.164258.

Caudle, K., Klein, T., Hoffman, J., Muller, D., Whirl-Carrillo, M., Gong, L., et al. (2014). Incorporation of Pharmacogenomics into Routine Clinical Practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Curr. Drug Metab.* doi:10.2174/1389200215666140130124910.

Chaturvedi, S., and Debaun, M. R. (2016). Evolution of sickle cell disease from a life-threatening disease of children to a chronic disease of adults: The last 40 years. *Am. J. Hematol.* doi:10.1002/ajh.24235.

Chen, X., Wang, H., Zhou, G., Zhang, X., Dong, X., Zhi, L., et al. (2009). Molecular population genetics of human CYP3A locus: Signatures of positive selection and implications for evolutionary environmental medicine. *Environ. Health Perspect.* doi:10.1289/ehp.0800528.

Chimusa, E. R., Meintjies, A., Tchanga, M., Mulder, N., Seioghe, C., Soodyall, H., et al. (2015). A Genomic Portrait of Haplotype Diversity and Signatures of Selection in Indigenous Southern African Populations. *PLoS Genet.* doi:10.1371/journal.pgen.1005052.

Darghouth, D., Koehl, B., Madalinski, G., Heilier, J., Bovee, P., Xu, Y., et al. (2011). Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. Blood 2011, 117(6), e57-e66. doi:10.1182/blood-2010-07-29963

Deeken, J. (2009). The affymetrix DMET platform and pharmacogenetics in drug development. *Curr. Opin. Mol. Ther.*

Desai, A. A., Lei, Z., Bahroos, N., Maienschein-Cline, M., Saraf, S. L., Zhang, X., et al. (2017). Association of circulating transcriptomic profiles with mortality in sickle cell disease. Blood, 129(22), 3009-3016. https://doi.org/10.1182/blood-2016-11-752279

Feldman, I., Rzhetsky, A., and Vitkup, D. (2008). Network properties of genes harboring inherited disease mutations. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0701722105.

Fosgate, G. T. (2009). Practical sample size calculations for surveillance and diagnostic investigations. *J. Vet. Diagnostic Investig.* doi:10.1177/104063870902100102.

Geard, A., Pule, G. D., Chetcha Chemegni, B., Ngo Bitoungui, V. J., Kengne, A. P., Chimusa, E. R., et al. (2017). Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *Br. J. Haematol.* doi:10.1111/bjh.14724.

Hunt, S. C., Kronenberg, F., Eckfeldt, J. H., Hopkins, P. N., Myers, R. H., and Heiss, G. (2001). Association of plasma bilirubin with coronary heart disease and segregation of bilirubin as a major gene trait: The NHLBI family heart study. *Atherosclerosis*. doi:10.1016/S0021-9150(00)00420-2.

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Res., 43, D1057-D1063.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M. (2019). New approach

for understanding genome variations in KEGG. Nucleic Acids Res., 47, D590-D595. doi: 10.1093/nar/gky962.

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., et al. (2009). Human Protein Reference Database—2009 update. Nucleic Acids Res., 37, D767-772.

Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J. P., et al. (2019). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* doi:10.1093/nar/gky1105.

Kosaka, K., Sakai, N., Endo, Y., Fukuhara, Y., Tsuda-Tsukimoto, M., Ohtsuka, T., et al. (2011). Impact of intestinal glucuronidation on the pharmacokinetics of raloxifene. Drug Metabolism and Disposition, 39(9), 1495-1502.

Kurose, K., Sugiyama, E., and Saito, Y. (2012). Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: Implications in the clinical trials for novel drug development. *Drug Metab. Pharmacokinet.* doi:10.2133/dmpk. DMPK-11-RV-111.

Lammert, F., and Matern, S. (2005). The genetic background of cholesterol gallstone formation: An inventory of human lithogenic genes. *Curr. Drug Targets Immune, Endocr. Metab. Disord.* doi:10.2174/1568008054064841.

Li, J., Lou, H., Yang, X., Lu, D., Li, S., Jin, L., et al. (2014). Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *J. Med. Genet.* doi:10.1136/jmedgenet-2014-102530.

Li, J., Zhang, L., Zhou, H., Stoneking, M., and Tang, K. (2011). Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* doi:10.1093/hmg/ddq498.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Lannuccelli, M., Galeota, E., et al. (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res., 40, D857-D861.

Lin, J. P., O'Donnell, C. J., Schwaiger, J. P., Cupples, L. A., Lingenhel, A., Hunt, S. C., et

al. (2006). Association between the UGT1A1*28 allele, bilirubin levels, and coronary heart disease in the Framingham Heart Study. *Circulation*. doi:10.1161/CIRCULATIONAHA.106.633206.

Ma, Q., and Lu, A. Y. H. (2011). Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacol. Rev.* doi:10.1124/pr.110.003533.

Mazandu, G. K., Chimusa, E. R., Mbiyavanga, M., Mulder, N. J. (2016). A-DaGO-Fun: an adaptable Gene Ontology semantic similarity-based functional analysis tool. Bioinformatics, 32(3), 477-479.

Mazandu, G. K., Mulder, N. J. (2011) Scoring protein relationships in functional interaction networks predicted from sequence data. PLOS ONE, 6, e18607.

Mewes, H. W., Ruepp, A., Theis, F., Rattei, T., Walter, M., Frishman, D., et al. (2006). MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic Acids Res., 39, D220-D224. https://doi.org/10.1093/nar/gkq1157

Milton, J. N., Sebastiani, P., Solovieff, N., Hartley, S. W., Bhatnagar, P., Arking, D. E., et al. (2012). A genome-wide association study of total Bilirubin and Cholelithiasis risk in sickle cell anemia. *PLoS One*. doi:10.1371/journal.pone.0034741.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res., 47, D351-D360.

Mnika, K., Pule, G. D., Dandara, C., and Wonkam, A. (2016). An expert review of pharmacogenomics of sickle cell disease therapeutics: Not yet ready for global precision medicine. *Omi. A J. Integr. Biol.* doi:10.1089/omi.2016.0105.

Nagar, S., Remmel, R. P., Hebbel, R. P., Zimmerman, C. L. (2004). Metabolism of opioids is altered in liver microsomes of sickle cell transgenic mice. Drug Metabolism and Disposition, 32(1), 98-104

Niihara, Y., Miller, S. T., Kanter, J., Lanzkron, S., Smith, W. R., Hsu, L. L., et al. (2018). A phase 3 trial of l-glutamine in sickle cell disease. *N. Engl. J. Med.* doi:10.1056/NEJMoa1715971.

Obaidat, A., Roth, M., and Hagenbuch, B. (2012). The Expression and Function of

Organic Anion Transporting Polypeptides in Normal Tissues and in Cancer. *Annu. Rev. Pharmacol. Toxicol.* doi:10.1146/annurev-pharmtox-010510-100556.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*., 42(D1), D358-363.

Oughtred, R., Stark, C., Breitkreutz, B., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res*., 47(D1), D529–D541. https://doi.org/10.1093/nar/gky1079

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* doi:10.1371/journal.pgen.0020190.

Perera, M. A., Innocenti, F., and Ratain, M. J. (2008). Pharmacogenetic testing for uridine diphosphate glucuronosyltransferase 1A1 polymorphisms: Are we there yet? *Pharmacotherapy*. doi:10.1592/phco.28.6.755.

Piel, F. B., Steinberg, M. H., and David C. Rees, D. C. (2017). Sickle Cell Disease. *N Engl J Med*. doi: 10.1056/nejmra1510865.

Piel, F. B., Hay, S. I., Gupta, S., Weatherall, D. J., and Williams, T. N. (2013). Global Burden of Sickle Cell Anaemia in Children under Five, 2010-2050: Modelling Based on Demographics, Excess Mortality, and Interventions. *PLoS Med.* doi:10.1371/journal.pmed.1001484.

PharmaADME.org (2019). Core ADME Gene List. *ADME Gene List*. Available at: http://pharmaadme.org/joomla/index.php?option=com_content&task=view&id=12&Itemid=27 [Accessed October 30, 2019].

Portincasa, P., Moschetta, A., and Palasciano, G. (2006). Cholesterol gallstone disease. *Lancet*. doi:10.1016/S0140-6736(06)69044-2.

Pule, G. D., Mnica, K., Joubert, M., Mowla, S., Novitsky, N., and Wonkam, A. (2017a). Burden, genotype and phenotype profiles of adult patients with sickle cell disease in Cape Town, South Africa. *South African Med. J.* doi:10.7196/samj.2017.v107i2.10849.

Ramos, E., Doumatey, A., Elkahloun, A. G., Shriner, D., Huang, H., Chen, G., et al.

(2014). Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J.* doi:10.1038/tpj.2013.24.

R Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing*. doi:10.1007/978-3-540-74686-7.

Ribeil, J.-A., Hacein-Bey-Abina, S., Payen, E., Magnani, A., Semeraro, M., Magrin, E., et al. (2017). Gene Therapy in a Patient with Sickle Cell Disease. *N. Engl. J. Med.* doi:10.1056/NEJMoa1609677.

Rumaney, M. B., Ngo Bitoungui, V. J., Vorster, A. A., Ramesar, R., Kengne, A. P., Ngogang, J., et al. (2014). The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLoS One* 9, 1–10. doi:10.1371/journal.pone.0100516.

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science (80-.).* 230, 1350–1354. doi:10.1126/science.2999980.

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*. 2004, 32, D449-D451.

Shang, H., Hao, Y., Hu, W., Hu, X., and Jin, Q. (2019). CDH2 gene rs11564299 polymorphism is a risk factor for knee osteoarthritis in a Chinese population: A case-control study. *J. Orthop. Surg. Res.* doi:10.1186/s13018-019-1256-0.

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., et al. (2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* doi:10.1038/s41588-018-0273-y.

Shriner, D., and Rotimi, C. N. (2018). Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *Am. J. Hum. Genet.* doi:10.1016/j.ajhg.2018.02.003.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al.

(2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinforma.* doi:10.1002/cpbi.5.

Stephanou, C., Tamana, S., Minaidou, A., Papasavva, P., Kleanthous, M., and Kountouris, P. (2019). Genetic Modifiers at the Crossroads of Personalised Medicine for Haemoglobinopathies. *J. Clin. Med, 8, 1927.* doi:10.3390/jcm8111927

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res., 47(D1), D607-D613. https://doi.org/10.1093/nar/gky1131

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*., 47(D1), D330-D338. doi: 10.1093/nar/gky1055.

Tsai, M. T., and Tarng, D. C. (2019). Beyond a measure of liver function—bilirubin acts as a potential cardiovascular protector in chronic kidney disease patients. *Int. J. Mol. Sci.* doi:10.3390/ijms20010117.

Vichinsky, E., Hoppe, C. C., Ataga, K. I., Ware, R. E., Nduba, V., El-Beshlawy, A., et al. (2019). A Phase 3 Randomized Trial of Voxelotor in Sickle Cell Disease. *N. Engl. J. Med.* doi:10.1056/nejmoa1903212.

Vichinsky, E. (2008). Oral Iron Chelators and the Treatment of Iron Overload in Pediatric Patients With Chronic Anemia. *Pediatrics*, 121(6), 1253. DOI: 10.1542/peds.2007-1824

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* doi:10.1093/nar/gkq603.

Weatherall, D. J., and Clegg, J. B. (2001). Inherited haemoglobin disorders: An increasing global health problem. *Bull. World Health Organ.* 79, 704–712. doi:10.1590/S0042-96862001000800005.

Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* doi:10.1038/clpt.2012.96.

World Health Organization Regional Office for Africa. WHO; Geneva, Switzerland: 22 June 2010. Sickle-cell disease: a strategy for the WHO African Region: Report of the Regional Director. AFR/RC60/8.

Wonkam, A., Mnika, K., Ngo Bitoungui, V. J., Chetcha Chemegni, B., Chimusa, E. R., Dandara, C., et al. (2018). Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. *Br. J. Haematol.* 180. doi:10.1111/bjh.15011.

Wonkam, A., Rumaney, M. B., Ngo Bitoungui, V. J., Vorster, A. A., Ramesar, R., and Ngogang, J. (2014). Coinheritance of sickle cell anemia and α-thalassemia delays disease onset and could improve survival in cameroonian's patients (Sub-Saharan Africa). *Am. J. Hematol.* doi:10.1002/ajh.23711.

Wooding, S. P., Watkins, W. S., Bamshad, M. J., Dunn, D. M., Weiss, R. B., and Jorde, L. B. (2002). DNA sequence variation in a 3.7-kb noncoding sequence 5′ of the CYP1A2 gene: Implications for human population history and natural selection. *Am. J. Hum. Genet.* doi:10.1086/342260.

Yahouédéhou, S. C. M. A., Adorno, E. V., da Guarda, C. C., Ndidi, U. S., Carvalho, S. P., Santiago, R. P., et al. (2018). Hydroxyurea in the management of sickle cell disease: pharmacogenomics and enzymatic metabolism. *Pharmacogenomics J.* doi:10.1038/s41397-018-0045-1.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* doi:10.1016/j.ajhg.2010.11.011.

Yawn, B. P., Buchanan, G. R., Afenyi-Annan, A. N., Ballas, S. K., Hassell, K. L., James, A. H., et al. (2014). Management of sickle cell disease: Summary of the 2014 evidence-based report by expert panel members. *JAMA - J. Am. Med. Assoc.* doi:10.1001/jama.2014.10517.

Zhang, W., Zhou, M., Liu, C., Leu, C., Qiao, T., Huang, D., et al. (2015). A Novel Mutation of SMAD3 Identified in a Chinese Family with Aneurysms-Osteoarthritis Syndrome. *Biomed Res. Int.* doi:10.1155/2015/968135.

Zhong, Z., Gu, L., Zheng, X., Ma, N., Wu, Z., Duan, J., et al. (2019). Comprehensive analysis of spectral distribution of a large cohort of Chinese patients with non-syndromic oculocutaneous albinism facilitates genetic diagnosis. *Pigment Cell Melanoma Res.* doi:10.1111/pcmr.12790.

Figures Legend

Figure 1: Principal component (PC) based clustering and scaled MAF heatmap results. (A) PC plot (PC2 vs PC1) projection of pharmacogenes, data from Cameroonian sickle cell samples (patient_SC) and Controls onto the 1000 genomes. Similarly, to (A), (B) shows the scaled MAFs (proportions) in different populations used in (A). The data indicate a major difference between African population and the other population from European and Asian ancestry, suggesting that using PharmacoScan in clinical practices in Africa, may lead to different anticipated outcomes, and that this could not be specifically suitable for African populations.

Figure 2: Distributions of SNPs per MAF bin groups (A and B), and specific core pharmacogenes integrated MAF (C and D) showing population variabilities within SCD groups in Africa (Cameroon and DRC). Up to 21 Core Pharmacogenes in (C) that are significantly different in MAF between these two SCD African groups are described in detailed in **Table S1**.

Figure 3: Manhattan plot showing significant variants in selected genes variants associated with Vaso-occlusive painful crisis (VOC). The 50 genes with significant SNPs are beyond the red line and all described in **Table S2**, and the two which are within Core Pharmacogenes are described in **Table 1**.

Figure 4: Summary sub-network of the essential VOC-associated genes with interacting genes along a path at less than three steps. A) A sub-network of the essential genes including *NTSR1*, *LRMDA*, *SMAD4* and *CDH2* in green colour, interacting with core

pharmacogenes: *UGT1A8*, *UGT1A10* and *SLCO4A1*, in gold colour. B) Network centrality measures of essential genes.

Figure 1

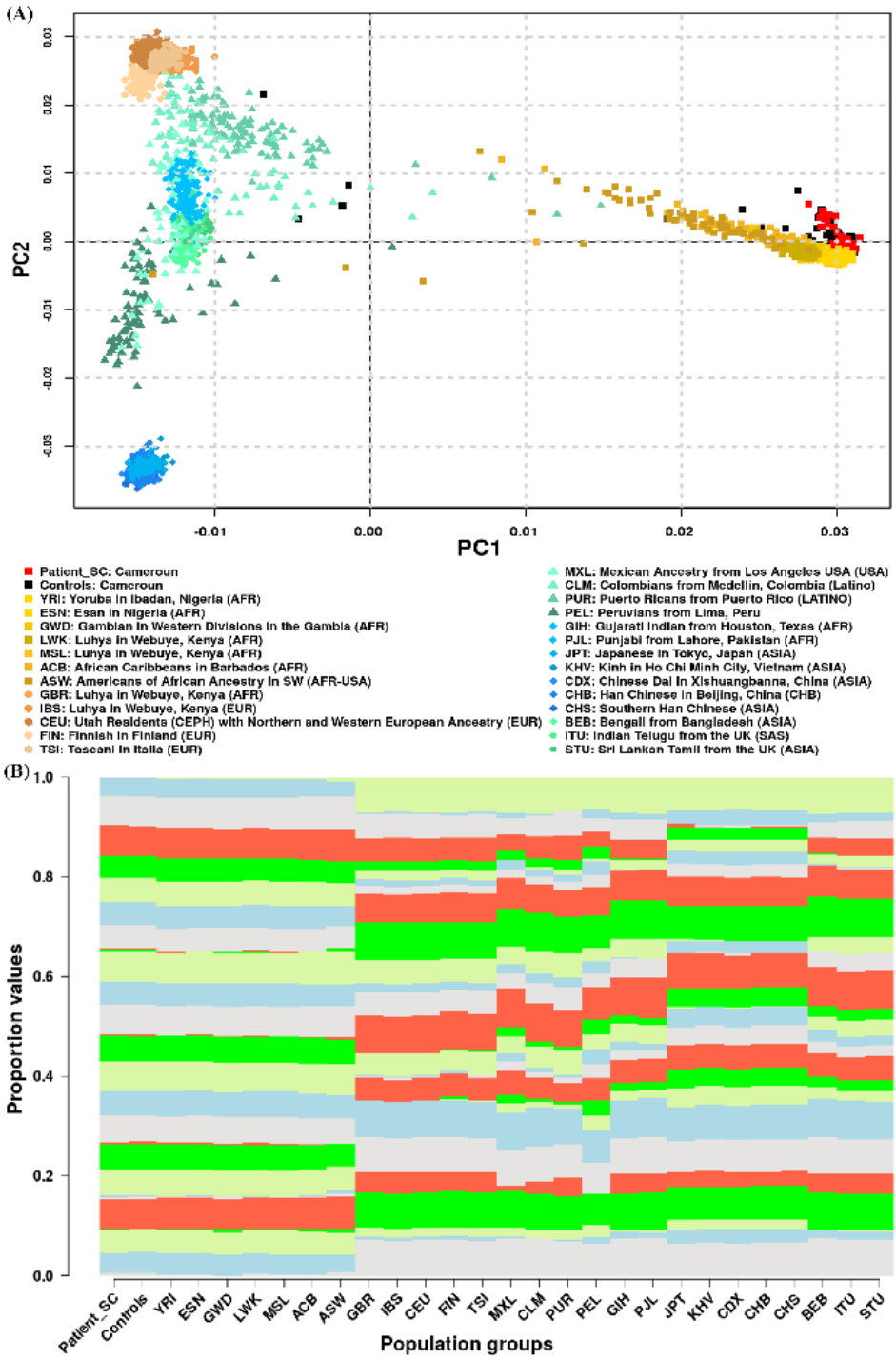

(A)
PC2
PC1
-0.01
0.00
0.01
0.02
0.03
Patient_SC: Cameroun
Controls: Cameroun
YRI: Yoruba in Ibadan, Nigeria (AFR)
ESN: Esan in Nigeria (AFR)
GWD: Gambian in Western Divisions in the Gambia (AFR)
LWK: Luhya in Webuye, Kenya (AFR)
MSL: Luhya in Webuye, Kenya (AFR)
ACB: African Caribbeans in Barbados (AFR)
ASW: Americans of African Ancestry in SW (AFR-USA)
GBR: Luhya in Webuye, Kenya (AFR)
IBS: Luhya in Webuye, Kenya (EUR)
CEU: Utah Residents (CEPH) with Northern and Western European Ancestry (EUR)
FIN: Finnish in Finland (EUR)
TSI: Toscani in Italia (EUR)
MXL: Mexican Ancestry from Los Angeles USA (USA)
CLM: Colombians from Medellin, Colombia (Latino)
PUR: Puerto Ricans from Puerto Rico (LATINO)
PEL: Peruvians from Lima, Peru
GIH: Gujarati Indian from Houston, Texas (AFR)
PJL: Punjabi from Lahore, Pakistan (AFR)
JPT: Japanese in Tokyo, Japan (ASIA)
KHV: Kinh in Ho Chi Minh City, Vietnam (ASIA)
CDX: Chinese Dai in Xishuangbanna, China (ASIA)
CHB: Han Chinese in Beijing, China (CHB)
CHS: Southern Han Chinese (ASIA)
BEB: Bengali from Bangladesh (ASIA)
ITU: Indian Telugu from the UK (SAS)
STU: Sri Lankan Tamil from the UK (ASIA)
(B)
Proportion values
0.0
0.2
0.4
0.6
0.8
1.0
Patient_SC
Controls
YRI
ESN
GWD
LWK
MSL
ACB
ASW
GBR
IBS
CEU
FIN
TSI
MXL
CLM
PUR
PEL
GIH
PJL
JPT
KHV
CDX
CHB
CHS
BEB
ITU
STU
Population groups

Figure 2

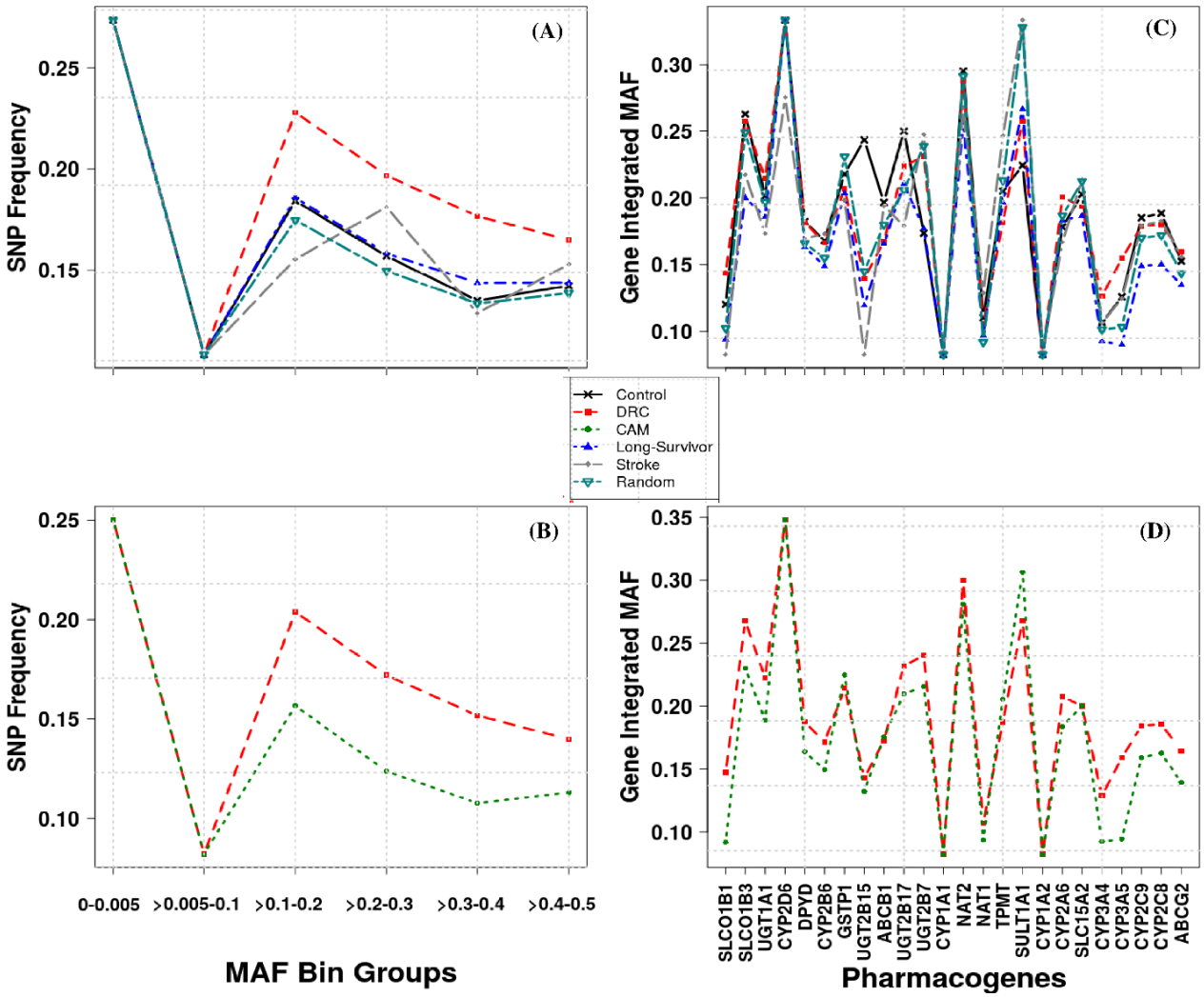

(A)
SNP Frequency
0.25
0.20
0.15
(C)
Gene Integrated MAF
0.30
0.25
0.20
0.15
0.10
Control
DRC
CAM
Long-Survivor
Stroke
Random
(B)
SNP Frequency
0.25
0.20
0.15
0.10
0-0.005
>0.005-0.1
>0.1-0.2
>0.2-0.3
>0.3-0.4
>0.4-0.5
MAF Bin Groups
(D)
Gene Integrated MAF
0.35
0.30
0.25
0.20
0.15
0.10
SLCO1B1
SLCO1B3
UGT1A1
CYP2D6
DPYD
CYP2B6
GSTP1
UGT2B15
ABCB1
UGT2B17
UGT2B7
CYP1A1
NAT2
NAT1
TPMT
SULT1A1
CYP1A2
CYP2A6
SLC15A2
CYP3A4
CYP3A5
CYP2C9
CYP2C8
ABCG2
Pharmacogenes

Figure 3

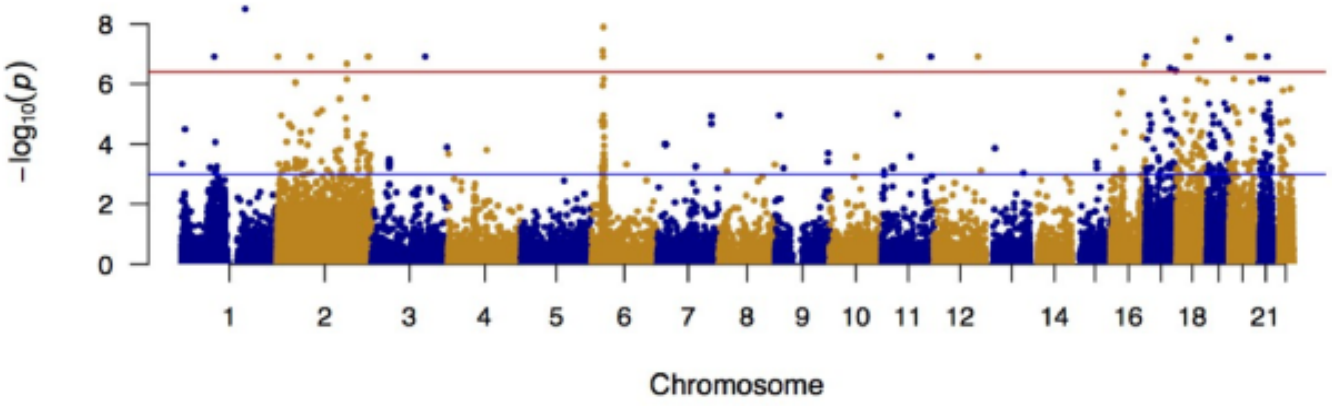


Figure 4

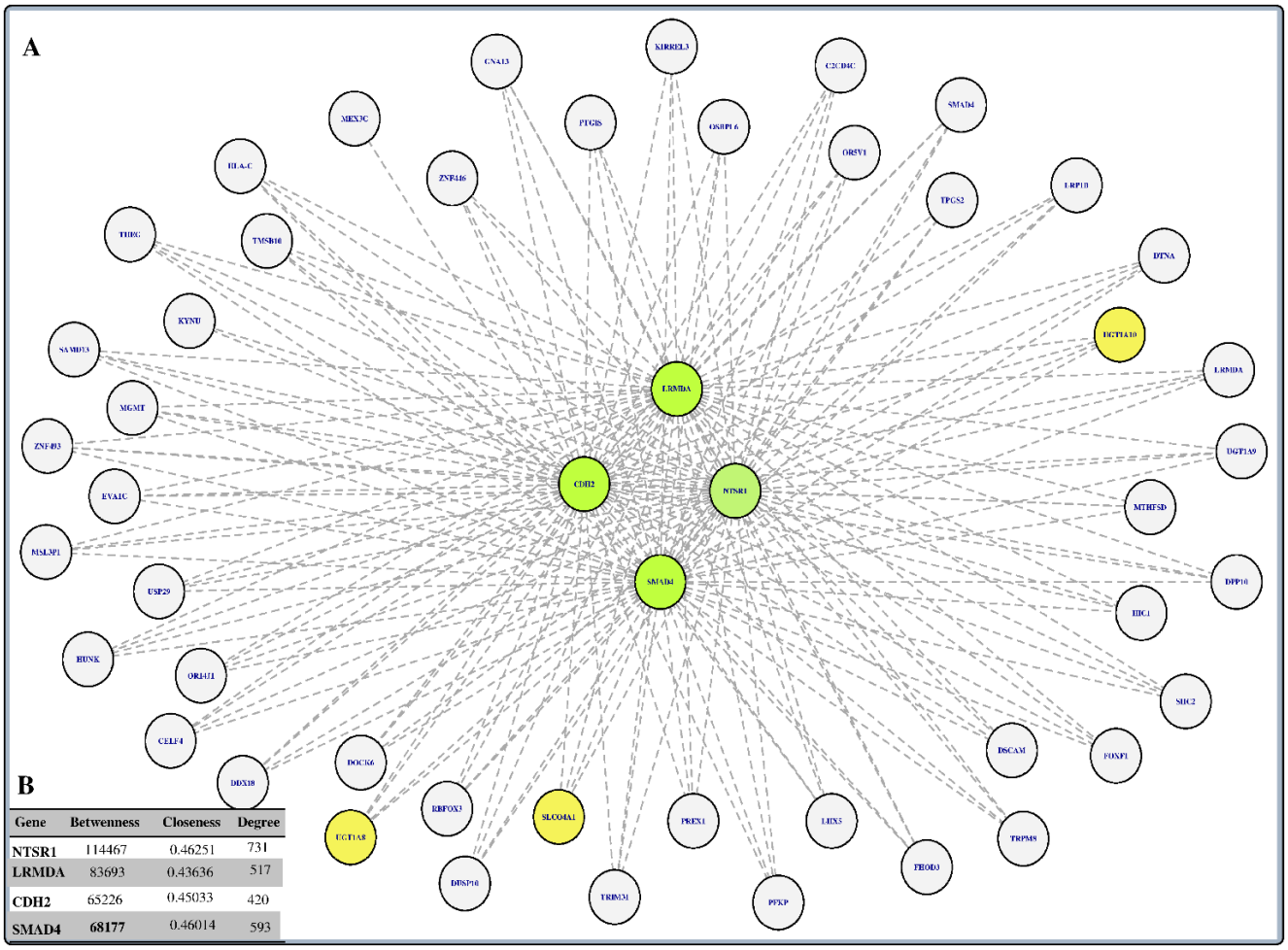

A
KIRREL3
GNA13
C2CD4C
MEX3C
PTGIS
SMAD4
OSBPL6
OR5V1
HLA-C
ZNF446
TPCS2
LRP1B
THOC
TMSB10
DTNA
KYNU
DGTTA10
SAMD13
LRMDA
LRMDA
MGMT
ZNF493
DGTLA9
CDH2
NTSR1
EVA1C
MTHFSD
MSL3P1
SMAD4
DPP10
USP29
IDC1
BUNK
ORL4J1
SHC2
CELF4
DSCAM
FOXP1
DOCK6
DDX10
B
RBFOX3
SLCO4A1
PREX1
LHX3
TRPM8
UGT1A8
FBOD3
DUSP10
TRIM31
PFKP
Gene
Betwenness
Closeness
Degree
NTSR1
114467
0.46251
731
LRMDA
83693
0.43636
517
CDH2
65226
0.45033
420
SMAD4
68177
0.46014
593

Table 1: Description of the Cameroonians SCD participants

Variables		Mean ± SD	Range	Number of observations
Age (years)		27.0±16.6	5-56	119
Gender	F/M (63/59)	0.50±0.5	-	122
Haematological indices	RBC ($\times 10^{12}$/l)	3.0±0.8	1.4-5.5	117
	Hb (g/l)	8.2±1.6	5-14.3	118
	MCV (fl)	84.9±11.7	59.0-112	119
	MCHC (g/l)	33.2±3.5	27.0-50.3	118
	WBC(x 10^9/l)	11.7±5.0	4.0-13.0	119
	Lymphocytes (x10^9/l)	4.79±2.61	0.6-16.0	118
	Monocytes (x10^9/l)	1.2±0.7	0.0-3.6	119
	Platelets count (x10^9/l)	367.4±141.9	58-934	119
	HbA2 (%)	4.0±1.9	0,6-18	117
	HbF (%)	9.8±7.7	0-31.6	117
Clinical events	VOC (n/year)	3.56±5.2	0-36	116
	Consultations (n/year)	2.4±3.28	0-12	89
	Hospitalization (n/year)	1.4±1.9	0-14	115
	Blood transfusion (%)	73.7		90/122
	Stroke (%)	11.4		14/122
	Hydroxyurea (%)	2.5		3/122
3.7 del alpha-globin gene	αα / αα (%)	29.5		36/122
	αα/ α3.7 (%)	24.6		30/122
	α 3.7/α 3.7 (%)	6.6		8/122
	- (%)	39.3		48/122

Hb: haemoglobin; MCHC: mean corpuscular haemoglobin concentration; MCV: mean corpuscular volume; RBC: red blood cell count; VOC: vaso-occlusive crises; WBC: white blood cell count; F: female; M: Male.

Table 2: Pharmacogenes variants associated with the number of Vaso-occlusive painful crisis in SCD patients.

Chr	SNP	Position	Gene	Minor Allele	Major Allele	MAF	SE	P-values
21	rs9984312	33428049	*HUNK,LINC00159*	C	A	0.013	2.95	4.40e-12
19	rs11670764	58992896	*ZNF446*	A	G	0.013	2.95	1.014e-11
18	rs28375234	69314871	*LINC01541,LINC01899*	C	T	0.004	5.06	8.93e-11
18	rs610846	34920165	*CELF4)*	A	C	0.004	5.06	8.93e-11
18	rs72893891	34154236	*FHOD3*	T	A	0.004	5.06	8.93e-11
18	rs74565398	34402135	*TPGS2*	A	G	0.004	5.06	8.93e-11
21	rs9982383	42305604	*DSCAM,LINC00323*	A	C	0.004	5.06	8.95e-11
19	rs34301174	11348098	*DOCK6*	A	G	0.004	5.06	8.95e-11
19	rs3951363	21619651	*ZNF493,LINC00664*	A	G	0.009	2.53	8.95e-11
19	rs72988423	391125	*THEG,C2CD4C*	A	G	0.004	5.06	8.95e-11
19	rs740871	454430	*SHC2*	A	G	0.004	5.06	8.95e-11
10	rs17194091	2824453	*LOC105376351,PFKP*	T	C	0.009	2.53	8.98e-11
10	rs72813546	77795251	*LRMDA*	C	T	0.004	5.06	8.98e-11
22	rs5770370	49785676	*LINC01310*	G	A	0.004	5.06	9.004e-11
1	rs12091463	221898783	*DUSP10*	C	T	0.004	5.06	9.01e-11
17	rs12453207	77383748	*RBFOX3*	A	G	0.004	5.06	9.05e-11
17	rs76679230	43253551	*LOC105371795,LOC339192*	A	G	0.004	5.06	9.05e-11
17	rs76965506	69323448	*CASC17,LINC02095*	G	A	0.004	5.06	9.05e-11
2	rs116777676	66988064	*LINC01797,LINC01799*	T	A	0.004	5.06	9.12e-11
2	rs17630500	67003923	*LINC01797,LINC01799*	G	A	0.004	5.06	9.12e-11
2	rs17679928	66919316	*LINC01798*	C	T	0.004	5.06	9.12e-11
2	rs74867338	117333802	*DPP10,DDX18*	C	T	0.004	5.06	9.12e-11
2	rs79422048	143082947	*LRP1B,KYNU*	A	G	0.004	5.06	9.12e-11
1	rs4657449	165465281	*LOC400794*	A	G	0.066	1.40	3.15e-09
6	rs9368666	31229644	*HCG27,HLA-C*	G	A	0.013	2.27	1.26e-08
19	rs4801391	57392265	*MIMT1,USP29*	A	G	0.018	2.59	2.96e-08
18	rs17663994	48666798	*SMAD4,MEX3C*	G	C	0.013	2.27	3.59e-08
6	rs3117436	29315419	*OR14J1,OR5V1*	A	G	0.036	1.86	7.77e-08
20	rs112154394	47169718	*LINC00494,PREX1*	G	A	0.004	5.06	1.21e-07
20	rs118042746	61314972	*SLCO4A1*[#]*,NTSR1*	G	A	0.004	5.06	1.21e-07
20	rs354751	58905494	*MIR646HG,LOC101928048*	A	G	0.004	5.06	1.21e-07
20	rs5621	48164497	*PTGIS*	A	C	0.004	5.06	1.21e-07
10	rs7101109	30422319	*LINC01163,MGMT*	G	A	0.004	5.06	1.21e-07
12	rs76767467	13937518	*LHX5-AS1,LINC01234*	T	G	0.004	5.06	1.21e-07
17	rs8065820	957893	*HIC1*	A	G	0.004	5.06	1.22e-07

3	rs192689	39290811	*LOC100507291*	G	A	0.004	5.06	1.22e-07
6	rs9261418	76660	*TRIM31-AS1*	A	G	0.004	5.06	1.22e-07
1	rs72714713	4764993	*SAMD13*	T	C	0.004	5.06	1.22e-07
21	rs2211789	3782887	*URB1-AS1,EVA1C*	C	T	0.004	5.06	1.21e-07
18	rs117009663	2198857	*DTNA*	A	G	0.004	5.06	1.22e-07
18	rs79423255	5303850	*LOC105372038,CDH2*	G	A	0.004	5.06	1.22e-07
2	rs10176426	34579915	*UGT1A10*[#]*, UGT1A8*[#]*,,UGT1A9*	T	C	0.004	5.06	1.22e-07
2	rs12465676	10975	*LINC01874,LINC01875*	G	T	0.004	5.06	1.22e-07
2	rs12473004	34808569	*MSL3P1,TRPM8*	G	A	0.004	5.06	1.22e-07
2	rs13409738	5133861	*TMSB10*	T	C	0.004	5.06	1.22e-07
11	rs111418068	26343227	*KIRREL3*	C	T	0.004	5.06	1.22e-07
16	rs72818139	6554138	*FOXF1,MTHFSD*	G	A	0.018	2.08	2.10e-07
2	rs13017264	79156556	*OSBPL6*	C	T	0.128	1.04	2.12e-07
17	rs62062488	3071911	*GNA13,LOC100507002*	A	C	0.022	2.32	2.95e-07
17	rs1077692	77270544	*RBFOX3*	T	C	0.018	2.08	3.49e-07

SE = Standard Error; # Corephamacogene

Supplementary Data

Table S1: Different Core genes shown in Figure 2, which were mapped to SNPs found in PharmacoScan array annotation files. MIMAF is the minimum value of the integrated MAF and Sample is the population achieving this minimum value. p-values are obtained by testing whether IMAP are homogeneous across different samples (CAM, DRC, Stroke, Long Survivor, Random) using Chi-square test.

Gene	Description	MIMAF	Sample	p-value
SLCO1B1	Solute carrier organic anion transporter family, member 1B1	0.08	Random	3.07e-07
SLCO1B3	Solute carrier organic anion transporter family, member 1B3	0.22	DRC	0.008
UGT1A1	UDP glucuronosyltransferase 1 family, polypeptide A1	0.18	Random	0.15
CYP2D6	Cytochrome P450, family 2, subfamily D, polypeptide 6	0.29	Random	0.005
DPYD	Dihydropyrimidine dehydrogenase	0.17	Survivor	0.88
CYP2B6	Cytochrome P450, family 2, subfamily B, polypeptide 6	0.16	Survivor	0.62
GSTP1	Glutathione S-transferase P1	0.21	Random	0.36
UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17	0.19	Random	0.002
CYP1A1	Cytochrome P450, family 1, subfamily A, polypeptide 1	0.08	CAM	0.55
NAT2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	0.28	DRC	0.8
NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	0.10	Survivor	0.03
TPMT	Thiopurine S-methyltransferase	0.19	Control	9.2e-05
SULT1A1	Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1	0.22	CAM	4.24e-14
CYP1A2	Cytochrome P450, family 1, subfamily A, polypeptide 2	0.08	CAM	0.55
CYP2A6	Cytochrome P450, family 2, subfamily A, polypeptide 6	0.17	Random	0.14
SLC15A2	Solute carrier family 15 (H+/peptide transporter), member 2	0.20	Control	0.49
CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4	0.10	DRC	0.19
CYP3A5	Cytochrome P450, family 3, subfamily A, polypeptide 5	0.10	DRC	1.6e-06

CYP2C9	Cytochrome P450, family 2, subfamily C, polypeptide 9	0.16	DRC	0.44
CYP2C8	Cytochrome P450, family 2, subfamily C, polypeptide 8	0.16	DRC	0.37
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	0.15	DRC	0.71
ABCB1	ATP Binding Cassette Subfamily B Member 1	0.17	Control	0.26
UGT2B15	UDP Glucuronosyltransferase Family 2 Member B15	0.08	Random	5.6e-36
UGT2B7	UDP Glucuronosyltransferase Family 2 Member B7)	0.17	CAM	3.2e-09

Chapter 4: Results: Original Publication

4.1 Mechanisms of hydroxyurea-induced γ-globin expression in Sickle Cell Disease

Synopsis: Fetal hemoglobin (HbF) is the central modifier of Sickle Cell Disease and is inherited as a quantitative trait under the regulation of variants at principal HbF-promoting loci. Understanding the genetic predispositions to hereditary persistence of HbF in adulthood will explain the vast heterogeneity of the disease and improve management. The present chapter includes one original publications (reporting on miRNA expression level) that; demonstrate differential HU-induced global miRNA expression *in vivo*.

4.1.1 **Mnika K**, Mazandu GK, Pule G, Jonas M, Chimusa ER, Hanchard N, and Wonkam A. Using miRNA expression level to analyse the effect of Hydroxyurea in SCD patient at Groote Schuur Hospital.

Abstract

Hydroxyurea (HU) is clinically beneficial in Sickle Cell Disease (SCD) through foetal haemoglobin (HbF) induction; however, the mechanism of HU is not yet fully elucidated. Selected miRNAs have been associated with HU-induced HbF production. We have investigated differential HU-induced global miRNA expression in peripheral blood of adult SCD patients in patients from Congo, living in South Africa. We found 22 of 798 miRNAs evaluated that were differentially expressed under HU treatment, with the majority (13/22) being functionally associated with HbF-regulatory genes, including BCL11A (miR-148b-3p, miR-32-5p, miR-340-5p, miR-29c-3p), MYB (miR-105-5p), and KLF-3 (miR-106b-5), and SP1 (miR-29b-3p, miR-625-5p, miR-324-5p, miR-125a-5p, miR-99b-5p,miR-374b-5p, miR-145-5p). The study provides additional miRNA candidates for therapeutic exploration.

Nature of Publication: Original Full Journal Article

Journal/Publisher: Frontiers in Genetics

Candidate contribution: clinical and genetic predictors of painful vaso-occlusive crises and analysis of the results. Contributed to writing in part with relation to clinical and genetic predictors analysis.

Co-Author contribution:

KM, GP, AW, NH: Performed the experiment, analysed data, wrote and revised the manuscript.

KM, GP, AW: Patients recruitment, sample and clinical data collection and processing

AW, EC, GM, MJ: Contributed to reagents/materials/analytic tools

AW, KM, MJ, GM, EC, NH, AW: Conceived and designed the experiments, analysed data, wrote and revised and approved manuscript.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 May 2019
doi: 10.3389/fgene.2019.00509

Hydroxyurea-Induced miRNA Expression in Sickle Cell Disease Patients in Africa

*Khuthala Mnika[1], Gaston K. Mazandu[1,2], Mario Jonas[1], Gift D. Pule[1], Emile R. Chimusa[1], Neil A. Hanchard[3] and Ambroise Wonkam[1]**

[1] *Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa,* [2] *African Institute for Mathematical Sciences, Cape Town, South Africa,* [3] *Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States*

Hydroxyurea (HU) is clinically beneficial in sickle cell disease (SCD) through fetal hemoglobin (HbF) induction; however, the mechanism of HU is not yet fully elucidated. Selected miRNAs have been associated with HU-induced HbF production. We have investigated differential HU-induced global miRNA expression in peripheral blood of adult SCD patients in patients from Congo, living in South Africa. We found 22 of 798 miRNAs evaluated that were differentially expressed under HU treatment, with the majority (13/22) being functionally associated with HbF-regulatory genes, including *BCL11A* (miR-148b-3p, miR-32-5p, miR-340-5p, and miR-29c-3p), *MYB* (miR-105-5p), and *KLF-3* (miR-106b-5), and *SP1 (*miR-29b-3p, miR-625-5p, miR-324-5p, miR-125a-5p, miR-99b-5p, miR-374b-5p, and miR-145-5p). The preliminary study provides potential additional miRNA candidates for therapeutic exploration.

Keywords: sickle cell disease, fetal hemoglobin, hydroxyurea, miRNA, Africa

OPEN ACCESS

Edited by:
Zané Lombard,
University of the Witwatersrand, Johannesburg, South Africa

Reviewed by:
Neneh Sallah,
London School of Hygiene and Tropical Medicine (LSHTM), United Kingdom
Fan Jin,
Zhejiang University, China

****Correspondence:***
Ambroise Wonkam
ambroise.wonkam@uct.ac.za

Specialty section:
This article was submitted to Genetic Disorders, a section of the journal Frontiers in Genetics

Received: *16 November 2018*
Accepted: *10 May 2019*
Published: *28 May 2019*

Citation:
Mnika K, Mazandu GK, Jonas M, Pule GD, Chimusa ER, Hanchard NA and Wonkam A (2019) Hydroxyurea-Induced miRNA Expression in Sickle Cell Disease Patients in Africa. Front. Genet. 10:509. doi: 10.3389/fgene.2019.00509

INTRODUCTION

Hydroxyurea (HU), the only food and drug administration (FDA) – approved treatment for sickle cell disease (SCD), is beneficial primarily through its ability to induce fetal hemoglobin (HbF) (Platt et al., 1984; Charache et al., 1992; Zimmerman et al., 2004). Clinical trials have shown hydroxyurea to be efficacious for increasing HbF in children, adolescents, and adults with SCA (Charache et al., 1992; Lee and Ambros, 2001; Thornburg et al., 2009). However, the precise mechanism by which HU can induces HbF in patients with SCA is not fully defined. Three main molecular pathways have been reported in HU-mediated response in increase HbF: (i) Epigenetic modifications, and transcriptional events, (ii) Signaling pathways, and (iii) Post-transcriptional pathways with regulation by Small non-coding RNA oligonucleotides (miRNA) (Pule et al., 2015).

miRNA have emerged as ubiquitous and potent molecular regulators that modulate the expression of many protein-coding genes by inhibiting mRNA translation (Lee and Ambros, 2001; Friedman et al., 2009). Multiple miRNAs have been implicated in the regulation of cell differentiation and maturation during hematopoiesis and erythropoiesis (Havelange and Garzon, 2010; Lawrie, 2010; Zhao et al., 2010). A few studies have demonstrated post-transcriptional regulation of HU-mediated γ-globin expression through miRNA in SCD patients; for example, miR-15a and miR-16-1 have been linked via the transcription factor *MYB*3 to elevated HbF

(Zhu et al., 2014; Pule et al., 2016), and expression of miR-26b and miR-151-3p have both been associated with HbF levels at the maximum tolerated dose (MTD) (Walker et al., 2011).

Studies have shown that miRNA expression of erythrocytes contributes to the majority of the miRNA expressions in whole blood (Juzenas et al., 2017). Because recent studies have identified miRNAs in mature erythrocytes that may reflect miRNA regulated processes during early erythropoiesis (Walker et al., 2011), we investigated differential HU-induced miRNA expressions using peripheral blood isolated from SCA patients before starting HU and after reaching the MTD. We identified novel miRNA expression changes after HU treatment, and their associated pathways, which mainly implicate HbF-regulatory genes. Our findings thus, provide novel insights into post-transcriptional mechanisms of actions of HU.

MATERIALS AND METHODS

Ethics Statement

The study was performed in accordance with the Declaration of Helsinki and with the approval of the Faculty of Health Sciences Human Research Ethics Committee, University of Cape Town (HREC Ref. No. 132/2010). Informed and written consent was obtained from all patients that were all adult participants (> 18 years).

Patients and HU Exposure

Ten patients were enrolled in this study, all attending adult hematological clinic of Groote Schuur Hospital in Cape Town (South Africa), denoted as GS01 to GS10. All consenting patients were selected, socio-demographic and clinical data were collected by means of a structured questionnaire. Adult SCA patients were interviewed; patients' medical records were reviewed, to delineate their clinical features over the past 3 years. Anthropomorphic variables (body mass Index (BMI), and blood pressures (BP) were measured in the outpatient setting. No incentive was provided for participation in the study. Only patients who, who was at steady clinical state, without current acute such as vaso-occlusive painful crisis and had not received a blood transfusion or hospitalization in the past 6 weeks where included. The hematological measures were those reported at the first visit to the hospital (**Supplementary Table S1**). Two patients GS01 and GS04, were investigated at two stage: before administration and after HU at MTD (indexed as H); Six patients were already on HU at MTD at the time of the study (GS02, GS03, GS07 GS08, and GS09, and GS10); and lastly, two patients (GS05 and GS06) had never been on HU.

Molecular Method

Genotyping: Sickle Cell Disease Mutation, β-Globin Gene Cluster Haplotypes, and 3.7 kb α-Globin Gene Deletion

DNA was extracted from peripheral blood, following instructions on the available commercial kit [QIAamp DNA Blood Maxi Kit. ® (Qiagen, United States)]. Molecular analysis to determine the presence of the sickle mutation was carried out by polymerase chain reaction (PCR), followed by DdeI restriction analysis (Saiki et al., 1985). Using published primers and methods, five restriction fragment length polymorphism (RFLP) sites in the β-globin gene cluster were amplified to analyze the *HBB* haplotype background (Bitoungui et al., 2015). The 3.7 kb α-globin gene deletion was screened using expand-long template PCR, as previously reported (Rumaney et al., 2014).

RNA Extraction and miRNA Expression Profile

Total RNA was isolated using the miRNeasy kit according to protocol of the Manufacturer (QIAGEN, Hilden, Germany); and sequenced by the Genomic and RNA Profiling Core at Baylor College of Medicine, United States, using the NanoString Platform (NanoString Technologies, Inc., Seattle, WA, United States), according to manufacturer's instructions. miRNA expression profile analyses were performed using the significance analysis of microarrays (SAM) tool (Thusher et al., 2001). A cross-sectional analysis was performed for differential expression for all the patients without HU and those under HU at MTD (**Table 1** and **Figure 1A**). In addition, a pair-wise analysis was performed for patients GS01 and GS04, before and after treatment of HU at MTD for each patient alone (**Figures 1B1–B4**), and for both patients together (**Figure 1C**), looking mainly for miRNAs that were over or under-expressed, using the paired Wilcoxon rank test. Differences in expression counts of differentially expressed miRNAs were tested using one-factor analysis of variance (ANOVA), after normalizing different samples based on their Fisher-Pearson skewness coefficient scores (Doane and Seward, 2011), adjusted for multiple comparisons with the significance level set to 0.05. We refer the interested readers to the **Supplementary File (Section S2, sub-section 2)** for more information. Specially, for pair-wise analysis, we extracted sets of over- and under-expressed miRNAs in different sample pairs e.g., GS01-GS01_H, GS01-GS04_H, and GS04-GS01_H using Pearson-Chi square scores and these sets were assessed using sample randomization to check whether the identified sets of over- and under-expressed miRNAs were more than expected by chance (Yocgo et al., 2017).

TABLE 1 | Differentially expressed microRNAs between SCD patients on HU and off HU in cross-sectional analysis.

microRNA-ID Fold-change *q*-value (%) *p*-values			
microRNA	**microRNA-ID Fold-change**	***q*-values**	***p*-values**
hsa-miR-105-5p	1.566	35.526	0.01644
hsa-miR-188-3p	1.525	35.526	0.01737
hsa-miR-561-3p	1.451	35.526	0.01757
hsa-miR-3074-3p	1.509	35.526	0.01871
hsa-miR-892a	1.327	35.526	0.02010
hsa-miR-490-5p	1.532	35.526	0.02534
hsa-miR-1258	1.339	35.526	0.02519
hsa-miR-629-5p	1.183	35.526	0.02802

In order to avoid possible residual effect of previous HU exposure Patient GS01 was remove from this analysis for non-compliance at initial administration of HU that he stopped for 3 months, before resuming treatment to achieve MTD.

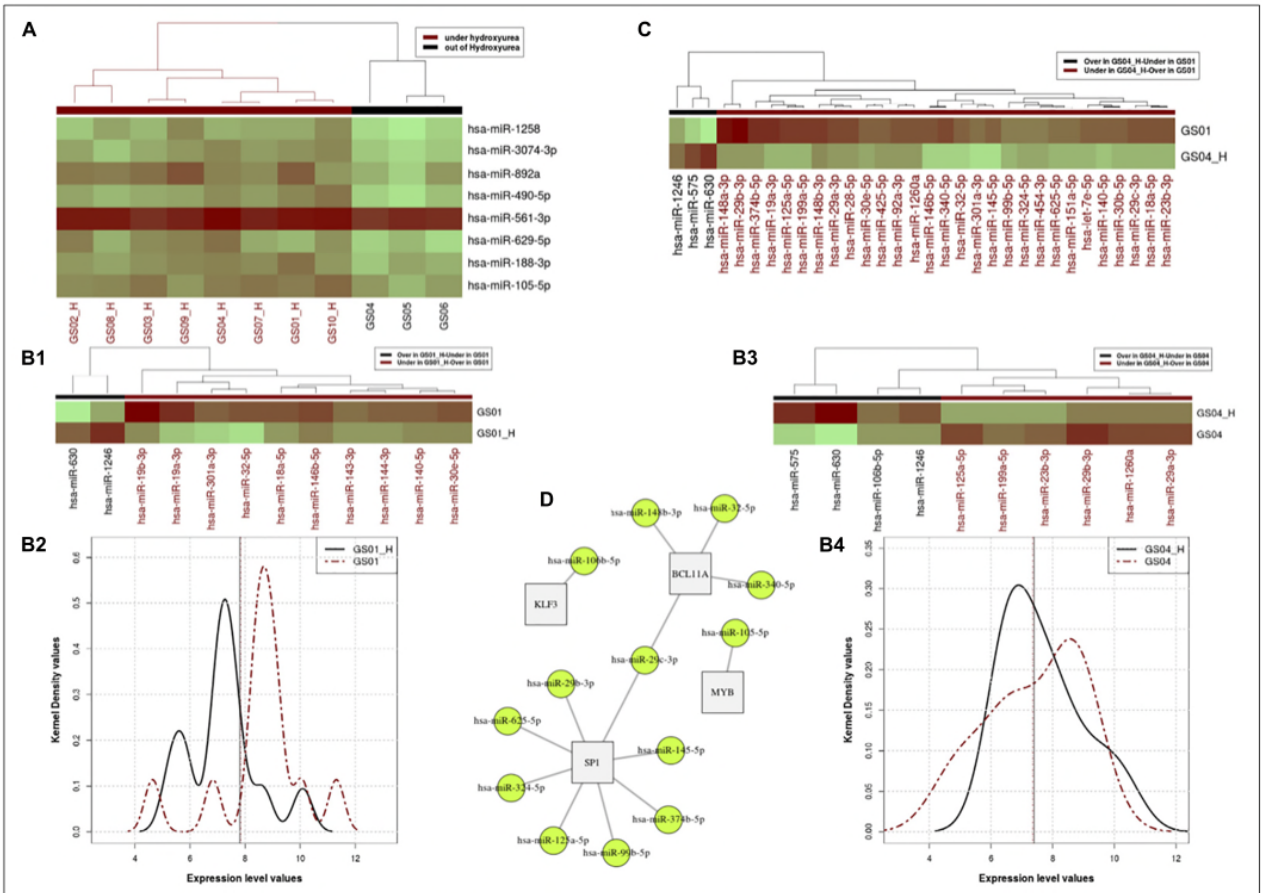


FIGURE 1 | Profiles of Hydroxyurea-induced miRNA expression in selected SCD patients. Panel **(A)**. Heatmap of the miRNA expressing profile for patients that are under HU and off HU in the cross-sectional analysis. Panels **(B,C).** miRNA expressing profile for patients that are under HU and out of HU, excluding the relapsed patient in pair-wire analysis of patients GS01 and GS04. Using the paired Wilcoxon rank test, we found a significant difference between expression profiles of the two states (p = 0.032e-2), with distinctive micro-RNAs differentially over or expressed in patients GS01 (Panel **B1**) and GS04 (Panel **B3**), as shown in the heat map (Panel **B2**) and the kernel density distributions (Panel **B4**) plots. Panel **(C).** miRNA expression levels of 29 miRNA that are over or under-expressed in GS04-HU vs. GS01. Panel **(D).** Networks linked to over and under expressed miRNA under HU treatment. In the figure 13 miRNAs target mainly 4 genes *BCL11A*, *MYB*, *KLF3*, and *SP1*, that belong to the same network and influence erythropoiesis and HbF expression.

Bioinformatics Pathway Analysis: HU Effects and Identifying Potential Biological Targets

All the differentially expressed miRNAs in cross-sectional analysis (**Figure 1A**), and miRNAs over- or under-expressed miRNAs in pair-wise analysis (**Figures 1B1,B2,C**) were used to retrieve potential post-transcriptionally regulated gene targets, from the miRTarBase database (Chou et al., 2015), which stores experimentally validated miRNA-target interactions. For specific miRNAs that were over- and/or under-expressed in different sample pairs, we performed enrichment analyses, using Gene Ontology (GO) process, the protein GO Annotation (GOA) mapping and the Kyoto encyclopeadia of genes and genomes (KEGG) pathway datasets, in order to identify enriched biological processes and pathways in which gene targets are involved (Mazandu and Mulder, 2013).

RESULTS

Patients' Description

Ten patients were investigated, all migrant from Democratic Republic of Congo, with a median age of 25 (95% CI: 23–26). All patients are homozygous for the Sickle cell mutation (HbSS). Patients receiving HU had higher HbF levels than those without HU treatment (13 vs. 4.4%). Most patients had at least a Bantu haplotype, in the beta-globin genes' cluster; four patients were heterozygous for the 3.7 kb alpha-globin gene deletion; detailed clinical characteristics are shown in **Supplementary Table S1**.

Cross-Sectional Analysis of the miRNA Expression Profiling

A total of 829 miRNAs were sequenced, and 798 that passed quality control were analyzed (**Supplementary File,Section S2**

for more details). The cross-sectional analysis identified 8 miRNAs differentially (over-) expressed with statistical characteristics and expression levels shown in **Table 1**, and the heat map in **Figure 1A**, respectively.

Pair-Wise Analysis of Differential Over-Under-Expressed miRs in Two SCD Patients

With or without HU exposure, we found a significant difference between expression profiles of the two states (*p*-value = 0.03266e-2), with 12 and 10 distinctive micro-RNAs differentially (over or under) in patients GS01 and GS04, respectively (**Figures 1B1–B4**). In order to elucidate miRNAs that influence the difference between the two patients' expression level profiles, we investigated miRNA expression levels which are over or under-expressed in GS04_H vs. GS01. A total of 29 miRNAs met these criteria, most were under-expressed in GS04_H (**Figure 1C**).

Genes Targets and Biological Pathways of miRNAs That Are Differentially Expressed Under HU Treatment

Next, we used miRNAs that were differentially expressed in cross-sectional analysis of all patients, alongside over- and under-expressed miRNAs identified in pair-wise analysis of GS01 and GS04, to retrieve potential post-transcriptionally regulated genes using datasets extracted from the miRTarBase database (Chou et al., 2015). We found 13 miRNAs that mainly targeted mainly 4 genes *BCL11A*, *MYB*, *KLF3*, and *SP1*, belonging to the same network and predicted to influence erythropoiesis and HbF expression (**Figure 1D**); most of these miRNAs were under-expressed with the exposure to HU at MTD (**Supplementary Table S4**).

Additionally, we used genes targeted by miRNAs that were differentially expressed, to identify enriched biological processes and pathways in which targeted genes are involved. We mostly found association with cancer pathways. Other enriched biological pathways identified were *pyrimidine metabolism* (*p*-value = 0.00986), *pathogenic Escherichia coli infection* (*p*-value = 0.00072) and *Oxidative phosphorylation* (*p*-value = 0.00032). Enriched biological process identified with p-adjusted using Bonferroni multiple corrections was miRNA mediated inhibition of translation, the main post-transcriptional mode of action of miRNA (GO: 0035278 with *p*-value adjusted = 0.0274).

DISCUSSION

The present study is the first to investigate *in vivo* miRNA expression in SCD patients in Africa, exposed to HU. MiRNA expression of erythrocytes is different from that of reticulocytes and leukocytes, but contribute to the majority of the microRNA expression in whole blood (Chen et al., 2008; Juzenas et al., 2017). This supports the most practical approach of using peripheral blood, in this study. Most of the miRNAs found to be differentially expressed under HU treatment in the current study, were also previously shown to be preferentially expressed in erythrocyte in SCD patients (Chen et al., 2008).

A major finding of the present study is the identification of specific and novel miRNA that are targeting HF- regulating genes (**Figure 1D**), i.e., miR-125b (*SP1*), mi199a, miR-7e, miR-106a, and miR-106b (*KLF3*), miR-140 miR-146; miR-188, miR-143, miR-125a, miR-19b, and miR-105 (*MYB*), miR-23b and miR-29a (*BCL11A* and *SP1*). We replicated previous findings that miR-148a, miR-29a, and mi151-3p, are differentially expressed in CD71+ erythroid cells, both before HU and after HU treatment at MTD in SCD-HbSS patients (Walker et al., 2011). Several other miRNAs are able to increase γ-globin gene expression, such as Lin28B, miR-486-3p, with let-seven family participating in the regulation of fetal to adult erythroid development process by increasing γ-*globin* gene expression through inhibitory effects on *BCL11A* (Lee et al., 2013; Ginder, 2015). miR-15a/16-1 restrain the MYB factor which then cause loss of the inhibitory effect on γ-gene and induce HbF in early erythroid progenitors (Sankaran et al., 2011).

Multiple miRNAs that target *SP1* and *KLF3* were differentially expressed under the HU treatment; several of these are novel (**Figure 1D**) and will require further functional investigation. *KLF3* and *SP1* are transcription factors, that belong to the family of β-*like globin* gene transcription regulation that act by binding to the LCR regions of the ε, γ, and β-*globin* promoters (Hu et al., 2007). *SP1* has been shown to be the main target for miR-23a which increases γ and ε globin expression by SP1 inhibition and repression. KLF3 factor, a negative regulator of erythropoiesis process, is also specifically inhabited by miR-27a (Ma et al., 2013). Therefore, this translational study provides additional candidates miRNAs that may contribute to globin gene expression and subsequent HbF production, and thus stand as prospects for future post-transcriptional therapeutic approaches that could minimize the alterations of the whole cellular transcriptome and related HU sides effects.

Association of differentially expressed miRNA with cancer pathways might be because cancer pathways are over – represented in the supporting literature. Other enriched biological pathways included "biological process' associations with Pathogenic *Escherichia coli* infection, Pyrimidine metabolism and Oxidative phosphorylation. These pathways could be related to the known increased susceptibility to bacterial infection in patients with SCD, folate acid metabolism that is important erythropoiesis, or the oxidative stress associated with recurrent vaso-occlusive crisis (VOC), and deserve additional investigations in much larger samples.

There are a few limitations to the present study, the first of which is the modest sample size that might result false positive associations and over-claiming significance. With a larger sample size, it is also possible that additional microRNA and biological pathways would be identified. We have provided a simulation of the needed statistical power in future studies in the section S3 of the **Supplementary File** provided. Even though the current pilot study did not achieve the expected statistical power for its modest sample size, the results obtained are consistent with the

literature, biologically relevant, and provide strong hypothesis for future studies. The second possible limitation is that by analyzing miRNAs that are likely from late-stage erythroblasts instead of erythroid progenitors from the bone marrow, epigenetic or molecular changes resulting from hydroxyurea treatment may have been missed. Lastly, the observed associations with targeted HbF genes regulators in pathway analysis, do not provide direct evidence for miRNA expression with HbF production. Despite these limitations, the significant associations of a limited number of differentially expressed miRNAs that potentially target HbF gene regulators provide preliminary hypothesis-generating results that can be used to design future functional experiments. These results also emphasize the need for future studies to investigate epigenetic processes in mechanisms of HbF expression and induction.

CONCLUSION

The study has shown that the global analysis of microRNA expression in peripheral blood of SCD patients, in the African context, can provide valuable insights into the mechanism of action of HU treatment. The study has identified novel HU-induced miRNA that specifically target HbF regulatory genes (*BCL11A*, *MYB*, *KLF-3,* and *SP1*), and are therefore strong candidates for post-transcriptional therapeutic exploration in SCD.

ETHICS STATEMENT

The study was performed in accordance with the Declaration of Helsinki and with the approval of the Faculty of Health Sciences Human Research Ethics Committee, University of Cape Town (HREC Ref. No. 132/2010). Informed and written consent was obtained from all patients that were all adult participants (> 18 years).

AUTHOR CONTRIBUTIONS

AW conceived and designed the experiments. AW, KM, GP, and NH performed the experiments. AW, GP, and KM patient recruitment, samples and clinical data collection and processing. MJ, GM, EC, KM, NH, and AW analyzed the data. AW, EC, MJ, and GM contributed reagents, materials, and analysis tools. KM, GP, GM, and AW wrote the manuscript. All authors revised and approved the manuscript.

FUNDING

The molecular experiments of the study were funded by the National Health Laboratory Services (NHLS), South Africa, and the NIH, United States, grant number 1U01HG007459-01, and NIH/NHLBI U24HL135600 to AW.

ACKNOWLEDGMENTS

We acknowledge to all patients' participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00509/full#supplementary-material

REFERENCES

Bitoungui, V. J. N., Pule, G. D., Hanchard, N., Ngogang, J., and Wonkam, A. (2015). Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? *OMICS* 19, 171–179. doi: 10.1089/omi.2014.0134

Charache, S., Dover, G. J., Moore, R. D., Eckert, S., Ballas, S. K., Koshy, M., et al. (1992). Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia. *Blood* 79, 2555–2565.

Chen, S., Wang, Y., Telen, M. J., and Chi, J. (2008). The genomic analysis of erythrocyte microRNA expression in sickle cell diseases. *PLoS One* 3:e2360. doi: 10.1371/journal.pone.0002360

Chou, C., Chang, N., Shrestha, S., Hsu, S., Lin, Y., Lee, W., et al. (2015). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* 44, D239–D247. doi: 10.1093/nar/gkv1258

Doane, D. P., and Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *J. Stat. Educ.* 19, 1–18. doi: 10.1080/10691898.2011.11889611

Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105. doi: 10.1101/gr.082701.108

Ginder, G. D. (2015). Epigenetic regulation of fetal globin gene expression in adult erythroid cells. *Transl. Res.* 165, 115–125. doi: 10.1016/j.trsl.2014.05.002

Havelange, V., and Garzon, R. (2010). MicroRNAs: emerging key regulators of hematopoiesis. *Am. J. Hematol.* 85, 935–942. doi: 10.1002/ajh.21863

Hu, X., Eszterhas, S., Pallazzi, N., Bouhassira, E. E., Fields, J., Tanabe, O., et al. (2007). Transcriptional interference among the murine beta-like globin genes. *Blood* 109, 2210–2216. doi: 10.1182/blood-2006-06-029868

Juzenas, S., Venkatesh, G., Hübenthal, M., Hoeppner, M. P., Du, Z. G., Paulsen, M., et al. (2017). A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res.* 45, 9290–9301. doi: 10.1093/nar/gkx706

Lawrie, C. H. (2010). MicroRNA expression in erythropoiesis and erythroid disorders. *Br. J. Haematol.* 150, 144–151. doi: 10.1111/j.1365-2141.2009.07978.x

Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science* 294, 862–864. doi: 10.1126/science.1065329

Lee, Y. T., de Vasconcellos, J. F., Yuan, J., Byrnes, C., Noh, S. J., Meier, E. R., et al. (2013). LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo. *Blood* 122, 1034–1041. doi: 10.1182/blood-2012-12-472308

Ma, Y., Wang, B., Jiang, F., Wang, D., Liu, H., Yan, Y., et al. (2013). A feedback loop consisting of microRNA 23a/27a and the beta-like globin suppressors KLF3 and SP1 regulates globin gene expression. *Mol. Cell. Biol.* 33, 3994–4007. doi: 10.1128/MCB.00623-613

Mazandu, G. K., and Mulder, N. J. (2013). DaGO-Fun: tool for gene ontology-based functional analysis using term information content measures. *BMC Bioinformatics* 14:284. doi: 10.1186/1471-2105-14-284

Platt, O. S., Orkin, S. H., Dover, G., Beardsley, G. P., Miller, B., and Nathan, D. G. (1984). Hydroxyurea enhances fetal hemoglobin production

in sickle cell anemia. *Eur. J. Clin. Invest.* 74, 652–656. doi: 10.1172/JCI11 1464

Pule, G. D., Mowla, S., Novitzky, N., Wiysonge, C. S., and Wonkam, A. (2015). A systematic review of known mechanisms of hydroxyurea-induced fetal hemoglobin for treatment of sickle cell disease. *Expert Rev. Hematol.* 8, 669–679. doi: 10.1586/17474086.2015.1078235

Pule, G. D., Mowla, S., Novitzky, N., and Wonkam, A. (2016). Hydroxyurea down-regulates BCL11A, KLF-1 and MYB through miRNA-mediated actions to induce γ-globin expression: implications for new therapeutic approaches of sickle cell disease. *ı Clin. Transl. Sci.* 5:15. doi: 10.1186/s40169-016-0092-97

Rumaney, M. B., Bitoungui, V. J. N., Vorster, A. A., Ramesar, R., Kengne, A. P., Ngogang, J., et al. (2014). The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLoS One* 9:e100516. doi: 10.1371/journal.pone.0100516

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354. doi: 10.1126/science.2999980

Sankaran, V. G., Menne, T. F., Scepanovic, D., Vergilio, J. A., Ji, P., Kim, J., et al. (2011). MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1519–1524. doi: 10.1073/pnas.1018384108

Thornburg, C. D., Dixon, N., Burgett, S., Mortier, N. A., Schultz, W. H., Zimmerman, S. A., et al. (2009). A pilot study of hydroxyurea to prevent chronic organ damage in young children with sickle cell anemia. *Pediatr. Blood Cancer* 52, 609–615. doi: 10.1002/pbc.21738

Thusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to ionising radiation response. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5116–5521. doi: 10.1073/pnas.091062498

Walker, A. L., Steward, S., Howard, T. A., Mortier, N., Smeltzer, M., Wang, Y. D., et al. (2011). Epigenetic and molecular profiles of erythroid cells after hydroxyurea treatment in sickle cell anemia. *Blood* 118, 5664–5670. doi: 10. 1182/blood-2011-07-368746

Yocgo, R. E., Geza, E., Chimusa, E. R., and Mazandu, G. K. (2017). A post-gene silencing bioinformatics protocol for plant-defence gene validation and underlying process identification: case study of the *Arabidopsis thaliana* NPR1. *BMC Plant Biol.* 17:218. doi: 10.1186/s12870-017-1151-y

Zhao, G., Yu, D., and Weiss, M. J. (2010). MicroRNAs in erythropoiesis. *Curr. Opin. Hematol.* 17, 155–162. doi: 10.1097/MOH.0b013e328337ba6c

Zhu, J., Chin, K., Aerbajinai, W., Kumkhaek, C., Li, H., and Rodgers, G. P. (2014). Hydroxyurea-inducible SAR1 gene acts through the gialpha/JNK/Jun pathway to regulate gamma-globin expression. *Blood* 124, 1146–1156. doi: 10.1182/blood-2013-10-534842

Zimmerman, S. A., Schultz, W. H., Davis, J. S., Pickens, C. V., Mortier, N. A., Howard, T. A., et al. (2004). Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. *Blood* 103, 2039–2045. doi: 10.1182/blood-2003-07-2475

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary data

Section S1: General description of patients' information

The clinical characteristics and haematological indices of the miRNA array cohort collected from the Groote Schuur Hospital in Cape Town (South Africa) are shown in **Table S1**.

All consenting patients were selected, socio-demographic and clinical data were collected by means of a structured questionnaire. Adult SCA patients were interviewed; patients' medical records were reviewed, to delineate their clinical features over the past three years. Anthropomorphic variables (body mass Index (BMI), and blood pressures (BP) were measured in the outpatient setting. No incentive was provided for participation in the study. Only patients who, who was at steady clinical state, without current acute such as vaso-occlusive painful crisis and had not received a blood transfusion or hospitalization in the past 6 weeks where included.

Cohort consisted of 10 SCD patients identified GS01 to GS10: with two patients GS01 and GS04 investigated at two stages: before HU and after HU administration at the maximum tolerated dose (MTD) indexed by H; Six patients were already on HU at MTD at the time of the study (GS02, GS03, GS07 GS08 and GS09, and GS10); and lastly, two patients (GS05 and GS06) had never been on HU. Note that for GS01, before HU administration stage corresponds to the non-compliance with the HU treatment for the patient GS01 for about three months

The median age of the patients on HU was 33 years whereas that of patients off HU was 24 years. There was 50 % distribution of females and males on the patients off HU and males were over-represented in the group that was on HU 66.67% were on HU at the time of study enrolment.

The Bootstrap technique [1] was used to compute statistical parameters and p-values between the two independent cohorts of SCD patients on HU and off HU under the null

hypothesis that on HU clinical parameter values are greater than those of off HU. Results suggest that some of on HU clinical parameter values, including age, RBC, HB and HbF, were significantly greater than those of off HU (with p-values > 0.5). This is not the case for the following clinical parameters: MCV, MCH and PLT, for which data did not show evidence that on HU parameter values are greater than off HU parameter values. Note that there was not enough data to perform similar approach for HbA and HbA2. Finally, a χ^2 independence test for homogeneity of proportions was performed for Alpha-thalassemia and Haplotype variables, revealing that the two groups are not homogeneous.

Table S1: Descriptive data for patients who are on HU and off HU.

		No HU (N=4)	Under HU at MDT (N=8)	
Variables		Median (25th- 75th percentiles) or %	Median (25th- 75th percentiles) or %	P-value
Age (Years)		24 (23.2-25.5)	33 (23.2-33.0)	0.861
Gender (N)	M/F	2/2	1/7	
Haematological Index	RBC (1X1012/ul)	3.0 (2.7-3.9)	3.2 (2.3-3.2)	0.636
	HB (g/dL)	8.4 (7.9-9.1)	9.0 (7.4-9.05)	0.720
	MCV (fL)	83.8 (65.9-85.9)	80.9 (74.2-80.9)	0.374
	MCH (pg)	29.8 (22.5-30.0)	28.5 (25.2-28.5)	0.364
	PLT (1X109/ul)	368.5 (308.5-622)	315.5 (188.5-315.5)	0.073
	HbA (%)	3.2	3	-
	HbA2 (%)	-	6.6	-
	HbF (%)	4.4 (4.1-4.7)	13 (6.9–14·1)	> 0.99
Alpha-thalassemia (%)	αα/αα	25 (n = 1)	80 (n = 5)	2.07×10^{-14}
	αα/α3.7	75 (n = 3)	20 (n = 1)	5.86×10^{-05}
Haplotype (%)	Bantu/Benin	25	20	
	Bantu	50	40	
	Bantu/Atypical	25	20	
	Atypical		20	

RBC: red blood cell counts; Hb: hemoglobin; MCV: mean corpuscular volume; MCHC: mean corpuscular hemoglobin concentration; WBC: white blood cell counts; PLT: platelet; HbA: adult hemoglobin; HbA2: hemoglobin A2; HbF: fetal hemoglobin;Bantu/Benin; Bantu; Bantu/Atypical.

Section S2: miRNA expression based on knowledge inference

Different analyses were performed to account for the non-compliance with the HU treatment for the patient GS01 for about three months. Three assumptions were made: (1) the effect of the therapy is not known, in this case, this patient is removed from the dataset to avoid potential uncontrollable biases, (2) the therapy still have an effect and we have assumed that the patient still under HU and (3) the patient is at the state where the effect of therapy has completely waned, in which case, the patient is considered to be out of treatment. Out of 828 miRNAs initially sequenced, only 798 that passed quality control based on the variability of expression levels tested using Muller statistic [2], testing coefficient of variation (C_V) with $C_V < 1$ for low variability and $C_V > 1$ for high variability. These 798 were analysed to identify differentially expressed miRNAs and used to predict post-transcriptionally regulated genes.

1. Identifying differentially expressed miRNA

In each of the three assumptions stated above, we performed the expression profile analyses using the SAM tool [3] in order to predict differentially expressed profiles or microRNAs.

For the first assumption, we obtained 8 differentially (over-) expressed microRNAs with statistical characteristics and expression levels are shown in Table 1 and in the heat map in Figure 1A (Main manuscript), respectively.

For (2), results obtained show evidence of 9 differentially (over-) expressed microRNAs, with statistical features in **Table S2** and heat map expression levels in (**Figure S2**)

Table S2: Differentially expressed microRNAs between SCD patients on HU and off HU including GS01.

microRNA-ID Fold-change q-value (%) p-values

microRNA	microRNA-ID Fold-change	q-values	p-values
hsa-miR-561-3p	1.457	23.183	0.01395
hsa-miR-105-5p	1.544	23.183	0.011616
hsa-miR-892a	1.328	23.183	0.01641
hsa-miR-188-3p	1.504	23.183	0.01902
hsa-miR-1258	1.351	23.183	0.01976
hsa-miR-1299	1.303	23.183	0.02015
hsa-miR-3074-3p	1.577	23.183	0.02189
hsa-miR-490-5p	1.603	23.183	0.02442
hsa-miR-1275	1.510	30.911	0.03089

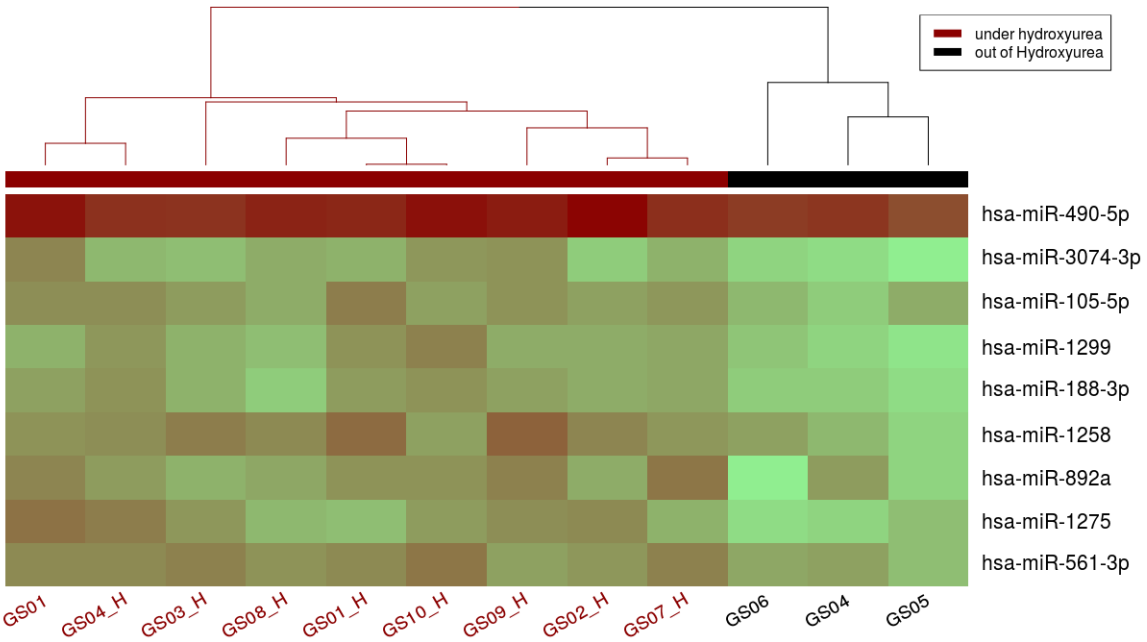


Figure S1: miRNA expressing profile for patients that are under HU and out of HU, including the relapsed patient (GS01).

For third assumption, 10 differentially (under-) expressed microRNAs were identified with statistical features in (**Table S3)** and heat map expression levels in (**Figure S3**).

Table S3: Differentially expressed microRNAs between SCD patients on HU and off HU assuming that GS01 is off HU.

microRNA-ID Fold-change q-value (%) p-values

microRNA	microRNA-ID Fold-change	q-values	p-values
hsa-miR-1827	0.757	54.85	0.00364
hsa-miR-330-5p	0.635	54.85	0.00565
hsa-miR-1204	0.768	54.85	0.00698
hsa-miR-422a	0.633	54.85	0.00960
hsa-miR-579-3p	0.719	54.85	0.01300
hsa-miR-95-3p	0.756	54.85	0.01524
hsa-miR-146b-5p	0.469	54.85	0.01762
hsa-miR-150-5p	0.59	54.85	0.02043
hsa-miR-613	0.795	54.85	0.02413
hsa-miR-433-5p	0.761	54.85	0.024403

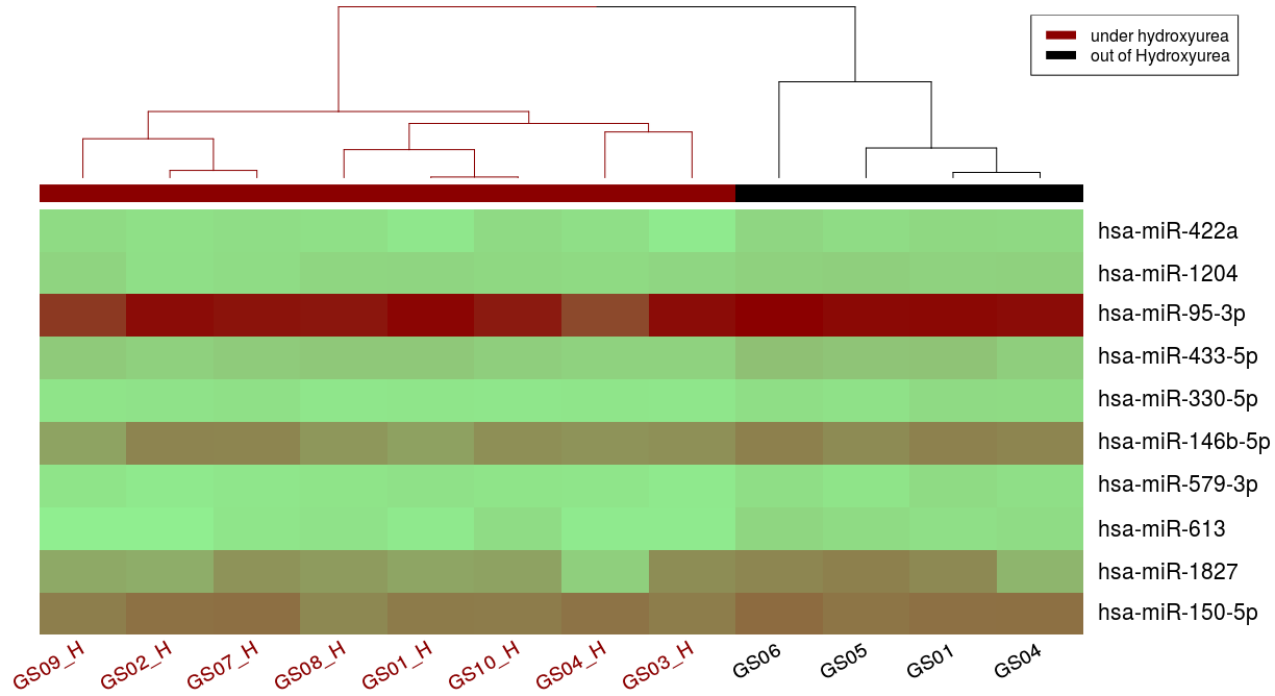


Figure S3: miRNA expressing profile for patients that are under HU and out of HU, assuming GS01 is under treatment.

In addition, for the second and third assumptions above, we also focused on the two different states of patient GS01 and GS04, checking whether is difference between the significant differences in expression levels of identified over-expressed or under-expressed miRNAs between the two states using the paired Wilcoxon signed rank test under the null hypothesis that there is no difference between the two state profiles for these two patients. For the context of the second assumption, this mainly aims to check whether results obtained show significant difference in expression levels of identified differentially expressed miRNAs between the two states for GS01. For this patient, the p-value score of 0.7263, which is greater than the significant level set to 0.05, and for GS04, however, the p-value was 0.01427. This suggests that there is no significant difference for GS01, indicating that the previous therapy might still have an effect on the patient.

2. miRNA expression based on analysis of HU concentration effect

In order to get more insights on whether GS01 previous therapy still had an influence on the patient and to provide evidence that predicted factors might have contributed to the difference in drug concentration between the two different states of the patient, we explore different datasets based on the differentially expressed miRNAs and results are shown in Figures 3(a) and 3(b), respectively, for GS01 patient after non-complying to the HU therapy and under treatment (GS01_H), the patient GS04 under the treatment (GS04_H) and before the treatment, and two other patients who were still out of treatment (GS05 and GS06).

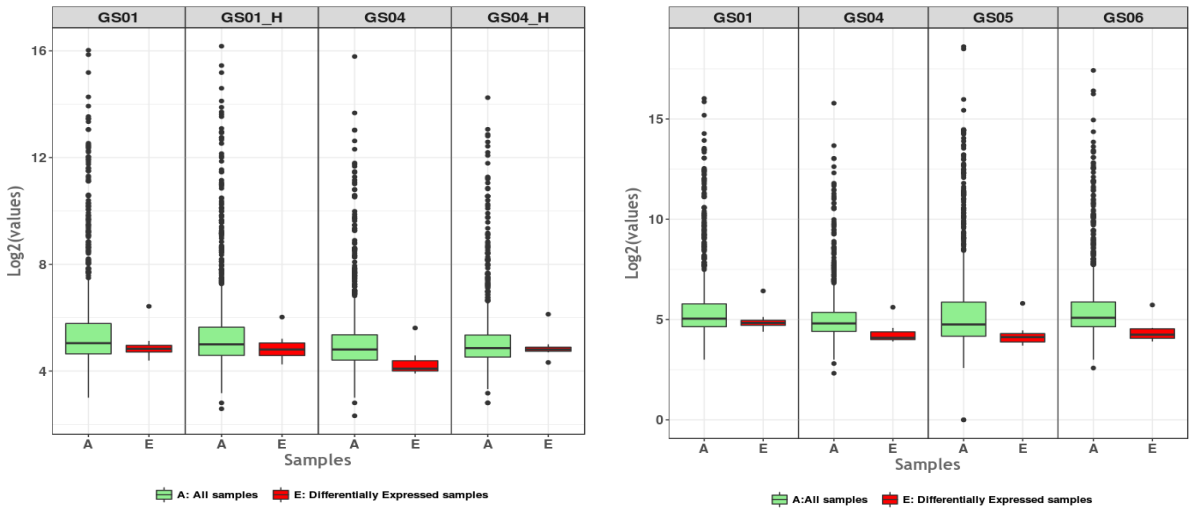


Figure S4: Box and whisker plots showing differences in expression level profiles of differentially expressed miRNAs targeting different states of patient GS01 and GS04.

To further confirm this observation, we used graphical representations (boxplots) to explore different datasets based on the differentially expressed miRNAs and results are shown in Figures S3, respectively, for GS01 patient after no-complying to the HU treatment, getting back to treatment (GS01_H) and the patient 4 under treatment (GS04_H), as well as with the patient 4 before the treatment GS04) and two other patients who were still out of treatment (GS05 and GS06). Furthermore, we performed one-way analysis of variance (ANOVA) or one-factor ANOVA to double confirm these results by transforming dataset values in log10 values driven by the Fisher-Pearson skewness coefficient scores [4, 5], which are 1.49025 and 1.29908 in the influence of prior GS01 HU administration or not, respectively, to bring different data subsets into agreement with the normality assumption. These figures and results from the one-way ANOVA still confirmed that there was no significance difference within data subsets (p–value = 0.7723), but there exists a significant difference in miRNA expression levels between patient GS01, who did not comply to the HU treatment for two months, and those who never got into the HU treatment (p–value = 0.01178). This partly provides evidence that

that the effect of HU may still have an influence on a patient relapsing from the treatment for some time.

3. Mapping differentially expressed miRNAs to gene targets

Here we used over-under expressed miRNAs identified to retrieve potential post-transcriptionally regulated genes using datasets extracted from the miRTarBase database [6] storing experimentally validated miRNA-target interactions. Different genes targeted and associated miRNAs in miRNA-gene associations shown in Figure 1D in the main manuscript. Table S4 maps differentially expressed miRNAs to gene targets and also provides their expression levels (over or under) based on the exposure to HU at MTD.

Table S4: Expression profile of miRNAs and their HbF-related target genes

microRNA	Target Gene	HU effect on miRs expression level
miR-106b-5p	*KFL3*	Over-expressed
miR-148b-3p	*BCL11A*	Under-expressed
miR-32-5p	*BCL11A*	Under-expressed
miR-340-5p	*BCL11A*	Under-expressed
miR-29c-3p	*BCL11A*	Under-expressed
miR-29b-3p	*SP1*	Under-expressed
miR-625-5p	*SP1*	Under-expressed
miR-324-5p	*SP1*	Under-expressed
miR-125a-5p	*SP1*	Under-expressed
miR-99b-5p	*SP1*	Under-expressed
miR-374b-5p	*SP1*	Under-expressed
miR-145-5p	*SP1*	Under-expressed
miR-105-5p	*MYB*	Over-expressed

Section S3: Discussing statistical power of the inferred information

The modest sample size constitutes the main limitation to the present pilot study, which should be addressed in near future as more data are currently being collected. With this larger sample size, it is possible that additional and relevant differentially expressed microRNAs would be identified with a reduced likelihood of selecting false positives. For now, we compute the statistical power score achieved with the current cohort and estimate the sample size needed to achieve the power that may significantly reduce false positive with an optimal effect size.

We computed the statistical power score of two samples with unequal size at the level of significance of 0.05 used throughout our analyses and varying effect size: small (d = 0.2), medium (d = 0.5) and large (d = 0.8) [7]. At these different effect sizes, the power scores achieved were very small with values of 0.05854, 0.10448 and 0.19258 for small. medium and large effect sizes, respectively. The general overview of power score versus sample size is shown in Figure S5, which suggests that, in order to achieve an effective power score of approximately 0.95 for a small effect size (d = 0.2), the size of 650 is needed for each sample type (under HU and off HU administration).

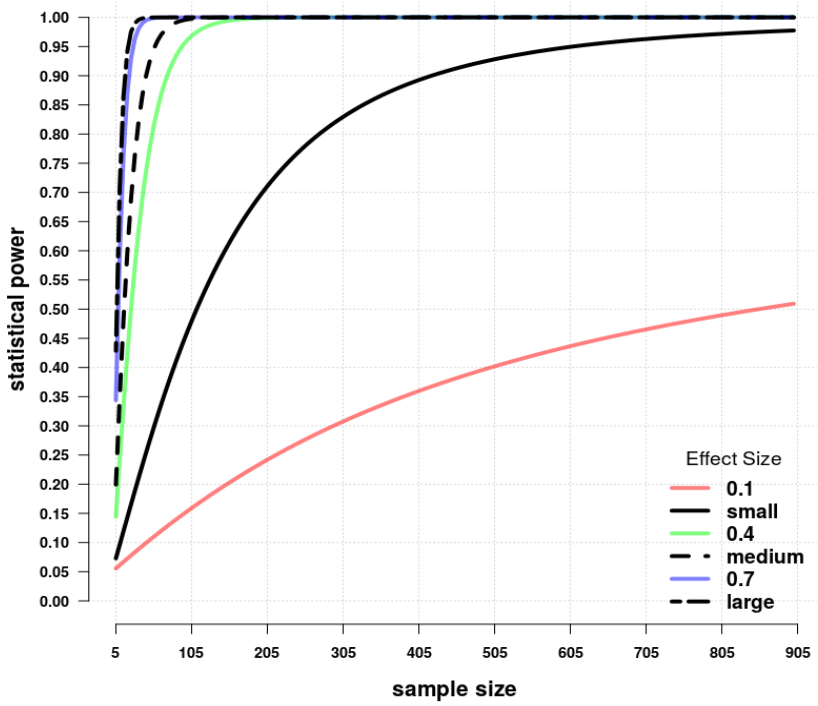


Figure S5: Relationships between statistical power and sample size while varying the effect size

It is worth noting that, even though the current pilot study did not achieve the expected statistical power for its modest sample size, the results obtained are consistent with the literature, biologically relevant and promising. These preliminary results are likely to be consistently confirmed on the large data set. As shown in the present study, the differential miRNA expression is largely different for two sample sizes; therefore, we will suggest the use of the median supplement approach in future studies [8].

References

[1] Li J., Tai B. C., Nott D. J. (2009). Confidence interval for the bootstrap P-value and sample size calculation of the bootstrap test, Journal of Nonparametric Statistics, 21(5):649– 661

[2] Miller, G. E. (1991). Asymptotic test statistics for coefficients of variation. Communications in Statistics-Theory and Methods, 20:3351–3363.

[3] Thusher, V. G., and Tibshirani, R. (2001) Significance analysis of microarrays applied to ionising radiation response. Proc Natl Acad Sci U S A. 98, 5116–5221. doi:10.1073/pnas.091062498.
[4] Howell D. C. Statistical Methods for Psychology, 6th edition. Belmont:Thomson Wadsworth; 2007, pp. 318–324.
[5] Tabachnick B. G, Fidell L. S. (2007) Using Multivariate Statistics, 5th edition. Boston: Allyn and Bacon; 2007, pp. 86–89.
[6] Chou, C., Chang, N., Shrestha, S., Hsu, S., Lin, Y., Lee, W., Yang, C., Hong, H. et al. (2015). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 44, D239–D247. doi: 10.1093/nar/gkv1258.
[7] Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic.
[8] Adabor E. S., Acquaah-Mensah, G. K. (2017). Machine learning approaches to decipher hormone and HER2 receptor status phenotypes in breast cancer, Briefings in Bioinformatics, bbx138, https://doi.org/10.1093/bib/bbx138

Chapter 5. Summary Discussion and Conclusions

Sickle cell disease (SCD) is a common monogenetic disorder and was first reported in 1910 (Herrick, 1910) and is a worldwide health problem with a disease burden comparable to that of communicable diseases, and major non-communicable diseases such as diabetes and hypertension. The disease burden in Africa well exceeds that of any other region in the world, with an estimated 75% of the 305 800 effected children that are born each year being in sub-Saharan Africa (Modell & Darlison, 2008) The high incidence of SCA in sub-Saharan Africa is likely due to carrier resistance to *Plasmodium falciparum* malaria, inferred from the geographical co-occurrence of these diseases (Flint et al., 1998; Williams et al., 2005). Other genetic variants associated with SCD such as hemoglobin C (HbC), β-thalassemia, and haplotypes have undergone similar selection (Pule et al., 2016).

Up to recently, SCD mutation was said to have evolved independently in 5 region of the world, and exists on a variable genetic haplotype background consisting in the β-globin genes-like cluster: Benin, Bantu/Central African Republic (CAR), Senegal, Cameroon and the Indian-Arab haplotypes (Pagnier et al., 1984; Elion et al., 1992). These haplotypes likely have multi-centric origins of the HbS gene and have also been associated with disease severity and clinical course of patients (Steinberg, 2009; Alsultan et al., 2012), bantu, CAR haplotype being reported to be associate with high severity of SCD. However, a recent study of sickle *HBB* haplotypes based on sequence data rather than restriction site data led to a reclassification of the classical haplotypes and found sub-structuring of haplotypes which may have confounded previous associations of the haplotypes with clinical severity (Shriner & Rotimi, 2018). In the present study, the most prevalent β-globin like gene cluster haplotypes was Benin, followed by Cameroon, that were not found to be associate with VOC (Wonkam et al., 2018). Early diagnosis has been shown to decrease disease-related mortalities (Gaston et al., 1986; Rahimy et al., 2003) and also significantly improve survival rates in the Caribbean (Lee et al., 1995), United States (Quinn et al.,

2004), the United Kingdom (Telfer et al., 2007) and Africa (Rahimy et al., 2009). In the absence of SCD new-born screening Cameroon, such could not be evaluated. In our data, the oldest patient with SCD was 54 years old, indicating that despite the unfavourable environment with poor and inadequate clinical care for SCD patients, a few patients still live tot eh 5th decade. The key to improving clinical management and therapeutic interventions is to systematically understand the genetic complexities of the SCD, genetic promoting loci of quantitative trait such as HbF (Wonkam et al., 2018), the molecular processes involved in drug-induced disease improvement and the effects of the vast genetic variation in African patients on symptom predisposition to, drug-metabolism (Mnika et al., 2016b).

We set out to investigate the pharmacogenomics of SCD patients of African origin and to identify novel miRNA expression changes after HU treatment, and the associated pathways, which mainly implicate HbF-regulatory genes. This study provides novel insights into post-transcriptional HU mechanisms of action by identifying novel miRNA that could mediate HU HBF related actions (Pule et al., 2016; Mnika et al., 2019). The associate pathways found included “biological process’ associated with Pathogenic *Escherichia coli* infection, Pyrimidine metabolism and Oxidative phosphorylation. These pathways could be related to the known increased susceptibility to bacterial infection in patients with SCD, folate acid metabolism that is important erythropoiesis, or the oxidative stress associated with recurrent vaso-occlusive crisis (VOC). The significance of this work will be discussed within the context of the currently available data on pharmacogenomics genes in the following parts: (1) Clinical and genetic factors (2) Genes involved in ADME, (3) Differential HU-induced miRNA expression using peripheral blood isolated from SCA patients before starting HU and after reaching the MTD.

5.1 Sickle Cell Disease and possible genetic prediction models for pain in Africa

This disease is common throughout most of SSA, affecting up to 2% of births in some parts of the continent. However, it remains a low priority for many health ministries (Grosse et al., 2011). The most common form of SCD is caused by homozygosity for the β-globin S gene mutation. It is widely believed that this condition is associated with very high child mortality, but unfailing contemporary data is lacking (Rees, Williams & Gladwin, 2010). We have reviewed available African data on the pharmacogenomics of SCD therapeutics (Mnika et al., 2016c) that shaped the four specific aims in this thesis, in addressing the need of data from Africa.

Therefore, we investigated genes that are significantly associated with SCD therapeutics, focusing on targeted genetic variants associated to VOC episodes in a group of patients living with SCD, and a broad scale of genes and variants related to pharmacogenomics to understand the drug metabolism and SCD-related polymorphisms among African populations (Wonkam et al., 2018 & Mnika et al, 2019). Furthermore, we studied pharmacogenes variants association with SCD clinical phenotypes (Mnika et al, 2019). We also investigated the profile of miRNA expression under-exposure of HU with the primary aim of understanding the mechanism of HU SCD therapeutics in a group of SCD patients from Africa (Mnika et al., 2019).

This study has revealed important findings on clinical predictors of acute painful episodes and the use of health care services in a unique group of SCD patients from Cameroon that have not been exposed to HU and opioids. Including female sex ($p = 0.026$), hospitalisation rates ($p = < 0.0001$), consultation rates ($p = <0.0001$), BMI ($p = 0.02.$) and positive history of blood transfusion ($p = 0.046$). Also various haematological indices, including Hb level ($p = 0.005$), white blood cell counts ($p = 0.04$), red blood cell counts ($p = 0.015$) and HbF level (estimate = $p = 0.025$. In addition, the study has identified specific variants in pain-related genes that are associated with acute pain crises and health care utilization, as well as in established genetic modifiers of SCD, such as *HBA1/HBA2*, HbF-

promoting loci and *APOL1*. This data confirmed previous data reported on *APOL1* G1/G2 risk alleles associated with kidney dysfunctions among Cameroonians living with SCD (Geard et al., 2017). In other cohorts, variants in HbF-promoting loci have been associated with higher total haemoglobin concentrations and lower leucocyte counts (Sheehan et al., 2013; Mtatiro et al., 2014), as well as lower VOC and composite endpoints, such as hospitalisations (Lettre et al., 2008; Sheehan et al., 2013; Leonardo et al., 2016). Co-inheritance of α-thalassaemia is protective against some SCD-related complications, such as acute chest syndrome, leg ulcers and chronic kidney disease (Guasch et al., 1999; Geard et al., 2017), but convey similar or higher rates of VOC (Meier et al, 2017; Platt et al*,* 1991; Darbari et al, 2013; Darbari et al*,* 2012; Tarer et al*,* 2006). In the present study, we have observed higher rates of VOC with the co-inheritance of α-thalassaemia.

Other major finding of our study is the identification of specific and novel miRNA that are targeting HF- regulating genes, i.e., miR-125b (*SP1*), mi199a, miR-7e, miR-106a, and miR-106b (*KLF3*), miR-140 miR-146; miR-188, miR-143, miR-125a, miR-19b, and miR-105 (*MYB*), miR-23b and miR-29a (*BCL11A* and *SP1*). We replicated previous findings that miR-148a, miR-29a, and mi151-3p, are differentially expressed in CD71+ erythroid cells, both before HU and after HU treatment at MTD in SCD-HbSS patients (Walker et al., 2011). Several other miRNAs are able to increase *γ*-globin gene expression, such as Lin28B, miR-486-3p, with let-seven family participating in the regulation of fetal to adult erythroid development process by increasing *γ-globin* gene expression through inhibitory effects on *BCL11A* (Lee et al., 2013; Ginder, 2015). miR-15a/16-1 restrain the MYB factor which then cause loss of the inhibitory effect on γ-gene and induce HbF in early erythroid progenitors (Sankaran et al., 2011).

Multiple miRNAs that target *SP1* and *KLF3* were differentially expressed under the HU treatment; several of these are novel and will require further functional investigation. *KLF3* and *SP1* are transcription factors, that belong to the family of *β-like globin* gene transcription regulation that act by binding to the LCR regions of the ε, γ,

and *β-globin* promoters (Hu et al., 2007). *SP1* has been shown to be the main target for miR-23a which increases γ and ε globin expression by SP1 inhibition and repression. KLF3 factor, a negative regulator of erythropoiesis process, is also specifically inhabited by miR-27a (Ma et al., 2013). Therefore, this translational study provides additional candidates miRNAs that may contribute to globin gene expression and subsequent HbF production, and thus stand as prospects for future post-transcriptional therapeutic approaches that could minimize the alterations of the whole cellular transcriptome and related HU sides effects. This supports the most practical approach of using peripheral blood, in this study. Most of the miRNAs found to be differentially expressed under HU treatment in the current study, were also previously shown to be preferentially expressed in erythrocyte in SCD patients (Chen et al., 2008). Pathways identified are enriched biological pathways included "biological process' associations with Pathogenic *Escherichia coli* infection, Pyrimidine metabolism and Oxidative phosphorylation. These pathways could be related to the known increased susceptibility to bacterial infection in patients with SCD, folate acid metabolism that is important erythropoiesis, or the oxidative stress associated with recurrent VOC. Altogether, the results may improve our ability to identify SCD patients who are at elevated risk for VOC and other organ complications and will contribute to refining the risk-profiling strategies that integrate both genetic and clinical information. We acknowledge that currently, the implementation of any genetic risk model in SCD in Africa may be challenging due to multiple competing priorities. However, we are hopeful that as the cost of genomic tests decreases, it could be possible in future. Furthermore, we recommend that this study be replicated in other African populations.

5.2 Sickle Cell Disease, Pharmacogenes, and Precision Medicine in Africa

Pharmacogenomics aims to identify the effects of genetic variations on drug response, with the goal of optimizing drug therapy and development. Some of the pharmacogenes such as drug metabolizing enzymes also have endogenous substrates contributing to

human pathophysiology. In this sense, pharmacogene variants may potentially display associations with both drug treatment outcomes and susceptibility to SCD and other human diseases. Taking advantage of the technological advances in genomics, a growing list of clinical biomarkers of drug response and adverse drug reactions have been identified. Although pharmacogenomic relevant markers have been identified and improved our understanding of the underlying mechanisms behind drug treatments, these are often studied in patients of European ancestry and do not always replicate in African populations (Simón-Sánchez & Singleton, 2008; Turner et al., 2008; Perera et al., 2013) This is a result of the differing allelic frequencies, linkage disequilibrium (LD), and confounding environmental factors across populations (Haga, 2010).

To date, none of the studies that have reported on variations in ADME pharmacogenes using a commercially available Affymetrix array named PharmacoScan®, genes with any disease pathophysiology. In our study we have investigated variation in 32 ADME core and 267 extended pharmacogenes in a selected group of SCD patients from Cameroon using PharmacoScan®. Our study showed no difference in the distribution of pharmacogenes variants among Cameroonian SCD and matched controls using principal component analysis (PCA). Cameroonians clustered with data from other populations of African ancestry but are significantly distinct from data from American, European and Asian populations. Variants allele frequencies in 21/32 core pharmacogenes were significantly different between the two SCD groups (Cameroon vs. Congo), highlighting high degree of variations among Africans. Single nucleotide polymorphisms (SNPs) in 50 genes have significant associations with vaso-occlusive painful episodes in SCD patients, including two that are core pharmacogenes (*SLCO4A1*- rs118042746, p = 1.21e-07; *UGT1A10*, *UGT1A8* - rs10176426, p = 1.22e-07); there are novel findings the deserve further pathophysiologic explorations. This data will add the rare available studies in Africa (Mnika et al., 2016; Wonkam et al., 2018). Other genes found to be associated were *FLT1, MAP3K5*, *PDE7B*, *ASS*, *TOX*, *ARG1*, *ARG2*, *NOS2A*, and *NOS1* (Husain et al., 2017*)*. Also, this study reveals that the commercially available pharmacogenes arrays

that was investigated might not be suitable for Africans and emphasize the need for urgent investigations of African-specific variants and their relation to pharmacokinetics/dynamics for various drugs. In addition, we found novel associations with pharmacogenes variants with pain in SCD, that deserve further research in other SCD patients' populations. A 2009 analysis on disparities in human genomics revealed that 96% of participants in GWAS were of European descent (Need & Goldstein, 2009). Since then, the proportion of non-European individuals included in GWAS has increased to approximately 20%. Much of this increase is the result of studies conducted in populations of Asian ancestry, with relatively small increase in representation of Africans or African Americans (Popejoy & Fullerton, 2016). Although the present study a targeted analysis, it will contribute to reduce the gap, as it the first such data from Africa. The inclusion of African populations in genomic studies is essential for evaluating the accuracy and wider relevance of findings that would help us understand the genetic heterogeneity in Africa and the creation of a reasonable distribution of personalized medicine.

5.3 Hydroxyurea pharmacogenomics and mechanism of action in SCD

Hydroxyurea (HU) is one of the US FDA approved treatments of SCD in adults and children. HU is a ribonucleotide reductase inhibitor that increases the HbF level. Patients respond differently to HU due to key genomic variants, mainly in HbF promoting loci (*BCL11A, HBG2* and *HBS1L–MYB*; (Friedrisch et al., 2008; Adekile, 2011; Ware et al., 2011). However, the complete picture of pharmacogenomics determinants of HU remains incomplete (Pule, et al., 2015). Another novelty of this these lies in the global analysis of microRNA expression in peripheral blood of SCD patients in an African context (Mnika et al 2019), providing valuable insights into the mechanism of action of HU treatment. This is seen through the identification of novel HU-induced miRNAs that specifically target HbF regulatory genes [*BCL11A* (miR-148b-3p, miR-32-5p, miR-340-5p, and miR-29c-3p), *MYB* (miR-105-5p), and *KLF-3* (miR-106b-5), and *SP1 (*miR-29b-3p, miR-625-5p, miR-324-5p, miR-125a-5p, miR-99b-5p, miR-374b-5p, and miR-145-5p)] (Mnika et al 2019). Our study confirmed previous findings that miR-148a, miR-29a, and mi151-3p, are

differentially expressed in CD71+ erythroid cells, before HU and after HU treatment at MTD in SCD-HbSS patients (Walker et al., 2011). These miRNAs are strong candidates for post-transcriptional therapeutic explorations in SCD.

5.4 Perspectives: investing in pharmacogenomics in SCD in Africa

Regardless of the increasing global burden of SCD and major progress in understanding the genetic modifiers of SCD clinical phenotype and response to HU, there are limited studies to support pharmacogenetics of SCD therapeutics for precision medicine (Mnika et al, 2016). It is expected that most patients living with SCD will have access to painkillers, antibiotics and HU treatment worldwide; which could increase the global importance in the field of pharmacogenetics of SCD. As the understanding of the pathophysiology of SCD advances and more is known about regulatory mechanisms, associated pathways, genetic modifiers and associated gene expressions, it is likely that more medications that are outside the realm of HbF induction will be developed. For example, the induction of stress haematopoiesis or endothelial nitric oxide release, to reduce leukocytes counts or red blood cells adhesion to the endothelium or inflammation processes (Pule et al., 2015; Mnika et al., 2016b). Investing in this field collaboratively in Africa will advocate for the lack of data in pharmacogenomics and improve population-specific drug design, targeting and efficacy, as well as the clinical management of patients.

Understanding the genetic basis of severity of SCD and its pharmacogenetics is a major challenge, given the pathophysiological complexity and overlapping nature of the biological processes culminate into the SCD clinical phenotype which has been reported in this project. However, longitudinal cohort studies are the most scientifically robust methods and are essential in understanding both environmental and genetics risk factors, and health and disease outcomes. While high regional disease prevalence in SSA would be expected to provoke epidemiological, translational and clinical research studies on SCD; there remains a lack of integration and coordination of emerging efforts from a few African countries, including the implementation of new-born screening and

comprehensive care, and genetic research (Wonkam & Makani, 2019). This highlights the need for global interdisciplinary research projects using sufficiently phenotyped and large cohorts of patients with SCD from different populations and environments. This will aid in the development of effective therapies SCD patients, which clearly could be best established in Africa, the core of SCD. This will assist in careful exploration and validation of multiple genetic variants that modulate some of the common clinical phenotype of SCD, using high-throughput genotyping methods, whole exome and whole genome sequencing, ADME genotype, together with innovative bioinformatics, biostatics and geocoding analytical techniques. This has the potential to generate genetic and environmental markers that could be used in an integrated model, to anticipate guidance in order to improve management, quality of life and ultimately survival of patients living with SCD (Wonkam & Makani, 2019).

Study limitations

Possible limitations of this study include its cross-sectional nature and the hospital-based recruitment. VOC episodes may have been subjected to pain self-tolerance bias, while financial factors could also have been limiting factors for hospital attendance. Chronic pain was difficult to address within the study design as it seems highly likely that individuals with 40 pain events per year are experiencing chronic pain. However, self-reported VOC in SCD has also been useful as a clinical endpoint in drug trials, patient quality of life measures and as a prognostic marker for mortality (Platt *et al.*, 1991; S Charache *et al.*, 1995; Machado *et al.*, 2011; Hoots and Shurin, 2012; Keller *et al.*, 2017). The strong association with the use of health services also tempers the possibly inadequate definition of VOC. However, the present study represents an important step in understanding clinical and genetic predictors of VOC in SSA and globally. For our pharmacogenes study: possible limitation will be small samples size and no data for pharmacokinetics to link variants to specific drug metabolism.

In the functional experiments, limitations included, the first being the sample size. With a larger sample size, it is possible that additional microRNAs and biological pathways would be identified. By analysing miRNAs that are likely from late-stage erythroblasts instead of erythroid progenitors from the bone marrow, epigenetic or molecular changes resulting from HU treatment may have been missed. Moreover, the observed associations with targeted HbF gene regulators in pathway analysis, do not provide direct evidence for miRNA expression changes with HbF production. The sample sizes for pharmacogenetic studies are relatively modest, and there is no pharmacokinetics data to link the variants to specific drug metabolisms, that will require future studies. Nonetheless, the total number of genes/variants provided in the study represents, to date, the only available of such datasets for SCD, and from Africa. Furthermore strengths of the study include the investigations in two independent SCD groups, inclusion ethnically matched controls for genetics studies, comparative analysis with wide range of data from the 1000 genome project, attempt to link the variations to SCD clinical phenotypes, complementary analytical approaches, and linking the identified significant genes variants to biological pathways. Despite these limitations, the significant associations with targeted genes and VOC, and the findings of a limited number of differentially expressed miRNAs that potentially target HbF gene regulators, and pharmacogenes study findings provide preliminary hypothesis-generating results that can be used to design future functional experiments.

Implications for clinical practice and future research

There are several implications of the research presented here: (i) The interrogation of genetic polymorphisms that are responsible for drug metabolism in a cohort of African patients with severe clinical events such as stroke and vaso-occlusive crises will improve our understanding of the genomics and clinical management of African patients. Furthermore, the hypothesis of using miRNA as agents for SCD therapeutics in African

populations will help us to understand the mechanism of HU action. (ii) The work presented on evaluating differences in variants of drugs metabolizing enzymes and transporter genes in SCD patients versus controls and the general population, prediction of the likelihood of the identified drug responses genes to known SCD medications, especially HU. Also identifying the contribution of pharmacogenomics variants to level of pain using pain thresholds as clinical phenotype. (iii) The expert review and comprehensive synthesis of the present analysis offers the latest reference on known drug metabolism genes related to HU and will serve as a reference for future studies on the progress made to date. Contribute to an understanding of the effectiveness of pharmacogenomics/genetics of pain management in SCD, with specific focus on HU and opioids. This may be a promising future area of research towards developing novel therapeutic approaches. (iv) The global analysis of microRNA expression in peripheral blood of SCD patients, in the African context, provides valuable preliminary insights about the mechanism of action of HU treatment. Furthermore, we report on novel HU-induced miRNAs that specifically target HbF regulatory genes (*BCL11A, MYB, KLF-3* and *SP1*) which are strong candidates for post-transcriptional therapeutic exploration in SCD.

Conclusion and perspectives

To the best of our knowledge, the present study is the first study in SSA to assess the pharmacogenomics of sickle cell therapeutics, as well as to map the landscape of the important pharmacogenomic genes associated with the hallmark phenotype of SCD, VOC. The data presented in this thesis is unique because of the inclusion of a large number of clinical variables and selected SNPs in both specific pain-related genes and established modifiers. It is therefore reasonable to envisage expending future explorations with genome-wide association studies, and targeted deep sequencing of genes in the inflammatory pathways, in SCD patients living in Africa. The study has identified specific variants of pain-related genes that are associated with acute pain crisis and health care utilization (hospitalisation and consultation), as well as in established genetic modifiers of SCD, such as *HBA1/HBA2*, HbF- promoting loci and *APOL1*.

Altogether, the results may improve our ability to identify SCD patients who are at elevated risk for VOC and other organ complications and will contribute to refining the elaboration of risk-profiling strategies that integrate both genetic and clinical information.

In the perspective of HU mechanisms, the study has shown that the global analysis of microRNA expression in peripheral blood of SCD patients, in the African context, can provide valuable insights into the mechanism faction of HU treatment. The study has identified novel HU-induced miRNA that specifically target HbF regulatory genes (*BCL11A, MYB, KLF-3* and *SP1*), and are therefore strong candidates for post-transcriptional therapeutic exploration in SCD. The molecular and cell biology component of this study have a capacity building dimension, as they were performed, designed, experimented, analysed and reported on, on the African continent.

References (Introduction and Discussion's Chapters)

Adekile, A.D. 2011. Limitations of Hb F as a phenotypic modifier in sickle cell disease: Study of Kuwaiti Arab patients. *Hemoglobin*. DOI: 10.3109/03630269.2011.617230.

Alsultan, A., Aleem, A., Ghabbour, H., Algahtani, F.H., Al-Shehri, A., Osman, M.E., Kurban, K., Alsultan, M.S., et al. 2012. Sickle cell disease subphenotypes in patients from Southwestern Province of Saudi Arabia. *Journal of Pediatric Hematology/Oncology*. DOI: 10.1097/MPH.0b013e3182422844.

Ansari, J., Moufarrej, Y. E., Pawlinski, R., & Gavins, F. N. (2018). Sickle cell disease: a malady beyond a hemoglobin defect in cerebrovascular disease. *Expert review of hematology*. doi: 10.1080/17474086.2018.1407240.

Ballas, S.K., Kesen, M.R., Goldberg, M.F., Lutty, G.A., Dampier, C., Osunkwo, I., Wang, W.C., Hoppe, C., et al. 2012. DOI: 10.1100/2012/949535.

Bockaert, J. 1999. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *The EMBO Journal*. DOI: 10.1093/emboj/18.7.1723.

Bukar, A. & Abjah, U. 2013. Seroprevalence of parvovirus B19 and its clinical effect among anaemic SCA patients in Northeastern Nigeria. *American Journal of Scientific and Industrial Research*. DOI: 10.5251/ajsir.2013.4.2.195.200.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., et al. 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. DOI: 10.1038/nature15521.

Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S. V.,

McMahon, R.P. & Bonds, D.R. 1995a. Effect of Hydroxyurea on the Frequency of Painful Crises in Sickle Cell Anemia. *New England Journal of Medicine*. 332(20):1317–1322. DOI: 10.1056/NEJM199505183322001.

Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S. V, McMahon, R.P. & Bonds, D.R. 1995b. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. *The New England journal of medicine*. DOI: 10.1056/NEJM199505183322001.

Chaturvedi, S. & Debaun, M.R. 2016. Evolution of sickle cell disease from a life-threatening disease of children to a chronic disease of adults: The last 40 years. *American Journal of Hematology*. DOI: 10.1002/ajh.24235.

Chou, W.Y., Wang, C.H., Liu, P.H., Liu, C.C., Tseng, C.C. & Jawan, B. 2006. Human opioid receptor A118G polymorphism affects intravenous patient-controlled analgesia morphine consumption after total abdominal hysterectomy. *Anesthesiology*. DOI: 10.1097/00000542-200608000-00016.

Colombatti, R., Montanaro, M., Guasti, F., Rampazzo, P., Meneghetti, G., Giordan, M., Basso, G. & Sainati, L. 2012. Comprehensive care for sickle cell disease immigrant patients: A reproducible model achieving high adherence to minimum standards of care. *Pediatric Blood and Cancer*. DOI: 10.1002/pbc.24110.

Cox, S.E., Makani, J., Soka, D., L'Esperence, V.S., Kija, E., Dominguez-Salas, P., Newton, C.R.J., Birch, A.A., et al. 2014. Haptoglobin, alpha-thalassaemia and glucose-6-phosphate dehydrogenase polymorphisms and risk of abnormal transcranial Doppler among patients with sickle cell anaemia in Tanzania. *British Journal of Haematology*. DOI: 10.1111/bjh.12791.

Elion, J., Berg, P.E., Lapoumeroulie, C., Trabuchet, G., Mittelman, M., Krishnamoorthy, R., Schechter, A.N. & Labie, D. 1992. DNA sequence variation in a negative control region

5’ to the (beta)-globin gene correlates with the phenotypic expression of the (beta)(s) mutation. *Blood*.

Flint, J., Harding, R.M., Boyce, A.J. & Clegg, J.B. 1998. The population genetics of the haemoglobinopathies. *Bailliere’s Clinical Haematology*. DOI: 10.1016/S0950-3536(98)80069-3.

Friedrisch, J.R., Prá, D., Maluf, S.W., Bittar, C.M., Mergener, M., Pollo, T., Kayser, M., da Silva, M.A.L., et al. 2008. DNA damage in blood leukocytes of individuals with sickle cell disease treated with hydroxyurea. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. DOI: 10.1016/j.mrgentox.2007.09.005.

Fleming A.F. 1989. The presentation, management and prevention of crisis in sickle cell disease in Africa. *Blood Rev*. (3):18–28.

Gaston, M.H., Verter, J.I., Woods, G., Pegelow, C., Kelleher, J., Presbury, G., Zarkowsky, H., Vichinsky, E., et al. 1986. Prophylaxis with Oral Penicillin in Children with Sickle Cell Anemia. *New England Journal of Medicine*. DOI: 10.1056/NEJM198606193142501.

Geard, A., Pule, G.D., Chetcha Chemegni, B., Ngo Bitoungui, V.J., Kengne, A.P., Chimusa, E.R. & Wonkam, A. 2017. Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *British Journal of Haematology*. 1–11. DOI: 10.1111/bjh.14724.

Grosse, S.D., Odame, I., Atrash, H.K., Amendah, D.D., Piel, F.B. & Williams, T.N. 2011. Sickle cell disease in Africa: a neglected cause of early childhood mortality. *American journal of preventive medicine*. DOI: 10.1016/j.amepre.2011.09.013.

Guasch, a, Zayas, C.F., Eckman, J.R., Muralidharan, K., Zhang, W. & Elsas, L.J. 1999. Evidence that microdeletions in the alpha globin gene protect against the development of sickle cell glomerulopathy in humans. *Journal of the American Society of Nephrology : JASN*. 10(5):1014–1019.

Haga, S.B. Impact of limited population diversity of genome-wide association studies. *Genetics in Medicine*. 2010. DOI: 10.1097/GIM.0b013e3181ca2bbf.

Hamideh, D. & Alvarez, O. 2013. Sickle cell disease related mortality in the United States (1999-2009). *Pediatric Blood and Cancer*. DOI: 10.1002/pbc.24557.

Herrick, J.B. 1910. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. Archives of Internal Medicine. DOI: 10.1001/archinte.1910.00050330050003.

Hoots WK, Shurin SB. Future directions of sickle cell disease research: the NIH perspective. 2012. *Pediatric Blood & Cancer*. 59(2):353-357. DOI: 10.1002/pbc.24180.

Husain, M., Hartman, A. D., & Desai, P. 2017. Pharmacogenomics of sickle cell disease: steps toward personalized medicine. Pharmacogenomics and Per Med, 10, 261–265. https://doi.org/10.2147/PGPM.S123427

Keller, S., Yang, M., Treadwell, M.J. & Hassell, K.L. 2017. Sensitivity of alternative measures of functioning and wellbeing for adults with sickle cell disease: Comparison of PROMIS® to ASCQ-MeSM. *Health and Quality of Life Outcomes*. DOI: 10.1186/s12955-017-0661-5.

Lee, A., Thomas, P., Cupidore, L., Serjeant, B. & Serjeant, G. 1995. Improved survival in homozygous sickle cell disease: Lessons from a cohort study. *BMJ*. DOI: 10.1136/bmj.311.7020.1600.

Leonardo, F.C., Brugnerotto, A.F., Domingos, I.F., Fertrin, K.Y., de Albuquerque, D.M., Bezerra, M.A.C., Araújo, A.S., Saad, S.T.O., et al. 2016. Reduced rate of sickle-related complications in Brazilian patients carrying HbF-promoting alleles at the BCL11A and HMIP-2 loci. *British Journal of Haematology*. 173(3):456–460. DOI: 10.1111/bjh.13961.

Lettre, G., Sankaran, V.G., Bezerra, M.A.C., Araujo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., et al. 2008. DNA polymorphisms at the BCL11A, HBS1L-MYB, and -

globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.0804799105.

Machado, R.F., Barst, R.J., Yovetich, N.A., Hassell, K.L., Kato, G.J., Gordeuk, V.R., Gibbs, J.S.R., Little, J.A., et al. 2011. Hospitalization for pain in patients with sickle cell disease treated with sildenafil for elevated TRV and low exercise capacity. *Blood*. DOI: 10.1182/blood-2010-09-306167.

Makani, J., Ofori-Acquah, S.F., Nnodu, O., Wonkam, A. & Ohene-Frempong, K. 2013. Sickle Cell Disease: New Opportunities and Challenges in Africa. *The Scientific World Journal*. DOI: 10.1155/2013/193252.

Makubi, A., Soka, D. & Makani, J. 2012. Moyamoya disease, a rare cause of recurrent strokes in an African sickle cell child: Does hydroxyurea have a role in this context?. *International Journal of Child Health and Nutrition*.

McGann, P.T., Santos, B., Oliveira, V. de, Bernardino, L., Ware, R.E. & Grosse, S.D. 2013. Cost-Effectiveness Of Neonatal Screening For Sickle Cell Disease In The Republic Of Angola (Abstract). *Blood*.

Mmbando, B.P., Mgaya, J., Cox, S.E., Mtatiro, S.N., Soka, D., Rwezaula, S., Meda, E., Msaki, E., et al. 2015. Negative epistasis between Sickle and foetal haemoglobin suggests a reduction in protection against malaria. *PLoS ONE*. DOI: 10.1371/journal.pone.0125929.

Mnika, K., Pule, G.D., Dandara, C. & Wonkam, A. 2016b. An Expert Review of Pharmacogenomics of Sickle Cell Disease Therapeutics: Not Yet Ready for Global Precision Medicine. *OMICS: A Journal of Integrative Biology*. DOI: 10.1089/omi.2016.0105.

Mnika, K., Mazandu, G.K., Jonas, M., Pule, G.D., Chimusa, E.R., Hanchard, N.A. & Wonkam, A. 2019. Hydroxyurea-induced miRNA expression in sickle cell disease patients in Africa. *Frontiers in Genetics*. DOI: 10.3389/fgene.2019.00509.

Modell, B. & Darlison, M. 2008. Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization*.

Mtatiro, S.N., Singh, T., Rooks, H., Mgaya, J., Mariki, H., Soka, D., Mmbando, B., Msaki, E., et al. 2014. Genome wide association study of fetal hemoglobin in sickle cell Anemia in Tanzania. *PLoS ONE*. 9(11):7–14. DOI: 10.1371/journal.pone.0111464.

Need, A.C. & Goldstein, D.B. 2009. Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics*. DOI: 10.1016/j.tig.2009.09.012.

Pagnier, J., Mears, J.G., Dunda-Belkhodja, O., Schaefer-Rego, K.E., Beldjord, C., Nagel, R.L. & Labie, D. 1984. Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proceedings of the National Academy of Sciences of the United States of America*.

Perera, M.A., Cavallari, L.H., Limdi, N.A., Gamazon, E.R., Konkashbaev, A., Daneshjou, R., Pluzhnikov, A., Crawford, D.C., et al. 2013. Genetic variants associated with warfarin dose in African-American individuals: A genome-wide association study. *The Lancet*. DOI: 10.1016/S0140-6736(13)60681-9.

Platt OS, Thorington BD, Brambilla DJ, Milner PF, Rosse WF, Vichinsky E, K.T. 1991. Pain in sickle cell disease. Rates and risk factors. *N Engl J Med*.

Popejoy, A.B. & Fullerton, S.M. 2016. Genomics is failing on diversity. *Nature*. DOI: 10.1038/538161a.

Pule, G.D., Ngo Bitoungui, V.J., Chetcha Chemegni, B., Kengne, A.P., Antonarakis, S. & Wonkam, A. 2015. Association between Variants at BCL11A Erythroid-Specific Enhancer and Fetal Hemoglobin Levels among Sickle Cell Disease Patients in Cameroon: Implications for Future Therapeutic Interventions. *OMICS: A Journal of Integrative Biology*. DOI: 10.1089/omi.2015.0124.

Pule, G. D., Mowla, S., Novitzky, N., Wiysonge, C. S., & Wonkam, A. 2015. A systematic review of known mechanisms of hydroxyurea-induced fetal hemoglobin for treatment of sickle cell disease. *Expert review of hematology*, *8*(5), 669-679. DOI: 10.1586/17474086.2015.1078235.

Pule, G.D., Mowla, S., Novitzky, N. & Wonkam, A. 2016. Hydroxyurea down-regulates BCL11A, KLF-1 and MYB through miRNA-mediated actions to induce γ-globin expression: implications for new therapeutic approaches of sickle cell disease. *Clinical and Translational Medicine*. DOI: 10.1186/s40169-016-0092-7.

Pule, G.D., Mnika, K., Joubert, M., Mowla, S., Novitzky, N. & Wonkam, A. 2017. Burden, genotype and phenotype profiles of adult patients with sickle cell disease in Cape Town, South Africa. *South African Medical Journal*. 107(2). DOI: 10.7196/SAMJ.2017.v107i2.10849.

Rahimy, M.C., Gangbo, A., Ahouignan, G., Adjou, R., Deguenon, C., Goussanou, S. & Alihonou, E. 2003. Effect of a comprehensive clinical care program on disease course in severely ill children with sickle cell anemia in a sub-Saharan African setting. *Blood*. DOI: 10.1182/blood-2002-05-1453.

Rahimy, M.C., Gangbo, A., Ahouignan, G. & Alihonou, E. 2009. Newborn screening for sickle cell disease in the Republic of Benin. *Journal of Clinical Pathology*. DOI: 10.1136/jcp.2008.059113.

Rakvåg, T.T., Klepstad, P., Baar, C., Kvam, T.M., Dale, O., Kaasa, S., Krokan, H.E. & Skorpen, F. 2005. The Val158Met polymorphism of the human catechol-O-methyltransferase (COMT) gene may influence morphine requirements in cancer pain patients. *Pain*. 116(1–2):73–78. DOI: 10.1016/j.pain.2005.03.032.

Rees, D.C., Williams, T.N. & Gladwin, M.T. 2010. Sickle-cell disease. *The Lancet*. 376(9757):2018–2031. DOI: 10.1016/S0140-6736(10)61029-X.

Sheehan, V.A., Luo, Z., Flanagan, J.M., Howard, T.A., Thompson, B.W., Wang, W.C., Kutlar, A. & Ware, R.E. 2013. Genetic modifiers of sickle cell anemia in the baby hug cohort: Influence on laboratory and clinical phenotypes. *American Journal of Hematology*. DOI: 10.1002/ajh.23457.

Shenoy, S. 2011. Hematopoietic stem cell transplantation for sickle cell disease: current practice and emerging trends. *ASH Education Program Book*. DOI: 10.1182/asheducation-2011.1.273.

Shriner, D. & Rotimi, C.N. 2018. Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *American Journal of Human Genetics*. DOI: 10.1016/j.ajhg.2018.02.003.

Simón-Sánchez, J., & Singleton, A. 2008. Genome-wide association studies in neurological disorders. *The Lancet Neurology*. DOI: 10.1016/S1474-4422(08)70241-2.

Slade, G.D., Diatchenko, L., Bhalang, K., Sigurdsson, A., Fillingim, R.B., Belfer, I., Max, M.B., Goldman, D., et al. 2007. Influence of psychological factors on risk of temporomandibular disorders. *Journal of Dental Research*. DOI: 10.1177/154405910708601119.

Steinberg, M.H. 1997. Determinants of fetal hemoglobin response to hydroxyurea. *Seminars in hematology*.

Steinberg, M. H. 2009. Genetic etiologies for phenotypic diversity in sickle cell anemia. *The Scientific World Journal*, *9*, 46-67. DOI: 10.1100/tsw.2009.10.

Steinberg, M.H., McCarthy, W.F., Castro, O., Ballas, S.K., Armstrong, F.D., Smith, W., Ataga, K., Swerdlow, P., et al. 2010. The risks and benefits of long-term use of hydroxyurea in sickle cell anemia: A 17.5 year follow-up. *American Journal of Hematology*. DOI: 10.1002/ajh.21699.

Thein, S.L. & Menzel, S. 2009. DOI: 10.1111/j.1365-2141.2009.07650.x.

Tshilolo, L., Aissi, L.M., Lukusa, D., Kinsiama, C., Wembonyama, S., Gulbis, B. &

Vertongen, F. 2009. Neonatal screening for sickle cell anaemia in the Democratic Republic of the Congo: Experience from a pioneer project on 31 204 newborns. *Journal of Clinical Pathology*. DOI: 10.1136/jcp.2008.058958.

Turner, S.T., Bailey, K.R., Fridley, B.L., Chapman, A.B., Schwartz, G.L., Chai, H.S., Sicotte, H., Kocher, J.P., et al. 2008. Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension*. DOI: 10.1161/HYPERTENSIONAHA.107.104273.

Ware, R.E., Despotovic, J.M., Mortier, N.A., Flanagan, J.M., He, J., Smeltzer, M.P., Kimble, A.C., Aygun, B., et al. 2011. Pharmacokinetics, pharmacodynamics, and pharmacogenetics of hydroxyurea treatment for children with sickle cell anemia. *Blood*. DOI: 10.1182/blood-2011-07-364190.

Williams, T.N., Mwangi, T.W., Wambua, S., Alexander, N.D., Kortok, M., Snow, R.W. & Marsh, K. 2005. Sickle Cell Trait and the Risk of *Plasmodium falciparum* Malaria and Other Childhood Diseases. *The Journal of Infectious Diseases*. DOI: 10.1086/430744.

Wonkam, A., & Makani, J. 2019. Sickle cell disease in Africa: an urgent need for longitudinal cohort studies. *The Lancet Global Health*. DOI: 10.1016/S2214-109X(19)30364-X.

Wonkam, A., Mba, C.Z., Mbanya, D., Ngogang, J., Ramesar, R. & Angwafo, F.F. 2014. Psychosocial burden of sickle cell disease on parents with an affected child in cameroon. *Journal of Genetic Counseling*. 23(2):192–201. DOI: 10.1007/s10897-013-9630-2.

Wonkam, A., Rumaney, M.B., Ngo Bitoungui, V.J., Vorster, A.A., Ramesar, R. & Ngogang, J. 2014. Coinheritance of sickle cell anemia and α-thalassemia delays disease onset and could improve survival in cameroonian's patients (Sub-Saharan Africa). *American Journal of Hematology*. DOI: 10.1002/ajh.23711.

Wonkam, A., Ngo Bitoungui, V.J., Vorster, A.A., Ramesar, R., Cooper, R.S., Tayo, B., Lettre, G. & Ngogang, J. 2014a. Association of variants at BCL11A and HBS1L-MYB with

hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLoS ONE*. 9(3). DOI: 10.1371/journal.pone.0092506.

Wonkam, A., Mnika, K., Ngo Bitoungui, V.J., Chetcha Chemegni, B., Chimusa, E.R., Dandara, C. & Kengne, A.P. 2018. Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. *British Journal of Haematology*. 180(1). DOI: 10.1111/bjh.15011.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., et al. 2011. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science*. DOI: 10.1126/science.1211053.

Yanni, E., Grosse, S.D., Yang, Q.H. & Olney, R.S. 2009. Trends in Pediatric Sickle Cell Disease-Related Mortality in the United States, 1983-2002. *Journal of Pediatrics*. DOI: 10.1016/j.jpeds.2008.09.052.

Zimmerman, S.A., Schultz, W.H., Davis, J.S., Pickens, C. V., Mortier, N.A., Howard, T.A. & Ware, R.E. 2004. Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. *Blood*. DOI: 10.1182/blood-2003-07-2475.

Annexes: Published paper excluded, with significant contribution of the candidate

1. Pule GD, **Mnika K**, Jourbert M, Mowla S, Novitzky N, Wonkam A. Increasing burden of adult patients living with Sickle Cell Disease in Cape Town. S. Afr. med. j. 2017 Feb; 107 (2). (Status: *published*)

Contribution to authorship

GP and AW conceived and designed the experiments.
GP, **KM (Khuthala Mnika)** and MJ recruited and sampled the patients.
GP and **KM(Khuthala Mnika)** performed the experiments.
GP, **KM(Khuthala Mnika)** and AW analysed the data.
AW, MJ, SM and NN contributed reagents/materials/analysis tools.
GP and AW wrote the article.
GP, **KM (Khuthala Mnika)**, MJ, SM, NN and AW revised and approved the manuscript.

Burden, genotype and phenotype profiles of adult patients with sickle cell disease in Cape Town, South Africa

G D Pule,[1] PhD; **K Mnika**,[1] BSc Hons; **M Joubert**,[2,4] MB ChB; **S Mowla**,[3] PhD; **N Novitzky**,[2,3,4] MD, PhD; **A Wonkam**,[1] MD, DMedSc, PhD

[1] *Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa*
[2] *Haematology Clinic, Groote Schuur Hospital, Cape Town, South Africa*
[3] *Division of Haematology, Departments of Internal Medicine and Pathology, Faculty of Health Sciences, University of Cape Town, South Africa*
[4] *National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa*

Corresponding author: *A Wonkam (ambroise.wonkam@uct.ac.za)*

Background. An exponential increase in the number of sickle cell disease (SCD) patients in paediatric services in Cape Town, South Africa, has been reported. The trend in adult/adolescent services has not been investigated.
Objectives. To evaluate epidemiological trends of SCD and the profile of patients affected by SCD attending the Haematology Clinic at Groote Schuur Hospital (GSH), Cape Town.
Methods. (*i*) A retrospective review of the number of SCD patients over the past 20 years; (*ii*) a cross-sectional analysis of clinical and haematological characteristics of SCD patients; and (*iii*) molecular analysis of the haemoglobin S mutation, the haplotype in the β-globin-like genes cluster, the 3.7 kb α-thalassaemia gene deletion and 19 selected single-nucleotide polymorphisms (SNPs) associated with fetal haemoglobin (HbF) levels.
Results. From 1995 to 2016, 81 adolescent/adult patients with SCD were registered, mostly originating from other African countries (*n*=61, 75.3%). There was an increase of over 200% in new cases *(n*=47) during the last quarter of the two decades investigated. Data from 34 of 58 regular attendees (58.6%) were analysed. The mean age of the patients was 26.1 years (standard deviation (SD) 9.8), and 70.6% were male. With the exception of four patients with sickle/β-thalassaemia, all the patients had SCD (haemoglobin SS). The co-inheritance of a single 3.7 kb α-globin deletion was found in 42.3% of cases (*n*=11). The Bantu haplotype was the most observed (65.4% of chromosomes). Most HbF-promoting SNPs were not associated with variable levels of haematological indices.
Conclusions. There is an increasing burden of adult SCD patients at GSH. National health and academic institutions need to adapt policies and healthcare professional training accordingly.

S Afr Med J 2017;107(2):149-155. DOI:10.7196/SAMJ.2017.v107i2.10849

Sickle cell disease (SCD) was the first well-documented molecular disease[1] and is the most prevalent monogenic disease in the world. SCD is an accepted worldwide health problem that is comparable to other major global diseases such as diabetes, hypertension and communicable diseases.[2] Sub-Saharan Africa (SSA) has the highest burden of SCD disease, with in excess of 300 000 new affected births annually, accounting for 80% of all annual affected child births globally.[3] In spite of the high burden of disease in SSA, SCD is often associated with limited to poor medical resources, infrastructure and quality of care, so estimates of neonatal and childhood mortality remain high, with up to 90% of affected children dying by 5 years of age.[4] SCD is caused by the polymerisation and precipitation of the β-globin chains (HbS) during deoxygenation and dehydration of erythrocytes.[5] The altered structure of erythrocytes (normal biconcave shape to a crescent shape) is the basis of the vascular pathology of the disease, which includes abnormality of platelet and leucocyte adhesion and hypercoagulation leading to microvascular occlusion, haemolysis, hypoxia, failed nitric oxide production and multiorgan damage.[5] The hallmark phenotypes of the disease include vaso-occlusive crises, stroke and acute chest syndrome.[5-7] The phenotype of SCD is influenced by both environmental and genetic factors. Variants at three principal loci, *BCL11A*, *HBS1L-MYB* intergenic polymorphism and the β-globin haplotype, have been shown to account for 10 - 20% of the variance of fetal haemoglobin (HbF) levels and to be associated with the amelioration of SCD symptoms.[8-10] Other variants in the *BCL11A* erythroid-specific enhancer (rs1427407 and rs7606173) have been shown to account for 8% and 6.2% of HbF variance, respectively, among SCD patient cohorts in the USA,[11,12] Tanzania[13] and Cameroon.[14] The co-inheritance of α-thalassaemia has also been associated with improved clinical manifestations of SCD.[15-18] Although the multiple independent origins of the HbS mutation have been questioned recently,[19] the SCD mutation is classically associated with five region-defined β-globin gene haplotypes, Benin, Bantu or Central African (CAR), Cameroon, Senegal and Indian-Arab,[20-23] four of which are from Africa and associated with malaria incidence.[24]

Because of the low incidence of malaria, the incidence of SCD in South Africa (SA) is equally extremely low; the HbS allele can be found in some indigenous SA ethnic groups (Venda and Shangaan) at an approximated frequency of 0.2%.[25-26] However, this is changing with the socioeconomically motivated influx of immigrants from other African countries, especially those within the equatorial malaria-endemic belt, resulting in a 300 - 400% increase in new cases of SCD over the past 10 years at Red Cross War Memorial Children's Hospital (RCWMCH) in Cape Town, SA.[27] The existence of similar trends in adult SCD patient services has not been investigated. Following our previous report at RCWMCH, we report in the present study the trend of new cases of adolescent and adult SCD over the past 20 years, having studied the clinical, haematological and genetic profiles of a cohort of 34 adolescent and adult SCD patients at the Haematology Unit at Groote Schuur Hospital (GSH), Cape Town.

Methods

Ethical approval

The study was performed in accordance with the Declaration of Helsinki and with the approval of the Faculty of Health Sciences Human Research Ethics Committee, University of Cape Town (HREC ref. no. 132/2010). Informed and written consent was obtained from adult participants (≥18 years), and for one patient aged 15 years informed consent was obtained from the guardian with assent from the participant.

Patients

The Haematology Clinic runs weekly every Wednesday. Most patients are seen at least once a month, and clinically stable patients every 3 months, with the exception of crisis-related hospitalisation. A retrospective review of the number of SCD patients attending the clinic over the past 20 years and a cross-sectional analysis of patients who regularly attend the clinic were performed. Clinical events and haematological indices were retrospectively collected from hospital records. The haematological measures were those reported at the first visit to the hospital.

Molecular methods

DNA extraction

DNA was isolated from the peripheral blood using the AllPrep DNA/RNA/miRNA universal kit (Qiagen, USA) according to the manufacturer's instructions.

Genotyping

HbS mutation and β-globin haplotypes

Polymerase chain reaction (PCR) and DdeI restriction analysis were used to confirm the presence of the HbS mutation using 100 ng DNA.[28] Published primers and methods[29] genotyping five restriction fragment length polymorphic regions in the β-globin gene cluster were used to analyse XmnI (5′Gγ), HindIII (Gγ), HindIII (Aγ), HincII (3′Ψβ) and HinfI (5′β) to determine the β-globin haplotype background.[19]

Single-nucleotide polymorphisms (SNPs)

Using a reported method,[9] SNaPshot genotyping, capillary electrophoresis and direct cycle sequencing were used to assay five selected HbF-associated variants: rs8176703, rs372091, rs2334880, rs1427407 and rs7606173. In addition, 18 other variants were analysed using the iPLEX Gold Sequenom Mass Genotyping Array (Inqaba Biotec, SA): X12_123681790, X16_391593, rs10468869, rs10756993, rs113267280, rs11754265, rs141494605, rs148706947, rs183437571, rs192197462, rs570013781, rs59329875, rs62573842, rs6466533, rs6590706, rs67104793, rs7163278 and rs76901220.

Statistical analysis

Descriptive statistics were obtained for all quantitative data using SPSS version 21.0 (IBM, USA). A χ^2 test with one degree of freedom was used to perform the Hardy-Weinberg Equilibrium (HWE) test on the SNP genotypes with all variants in HWE ($p>0.05$).

Results

Patients' origin and trends

A total of 128 patients' files from 1995 to March 2016 were reviewed. Among them, 47 patients were diagnosed with some form of α- or β-thalassaemia (Fig. 1). Of the remaining 81 patients affected by SCD, 61 (75.3%) were from other SSA countries. Over the last quarter (2011 - 2016) of the past two decades, there was an approximately 200% increase in new cases of SCD registered at the GSH Haematology Clinic (*n*=47) (Fig. 2). Fig. 3 shows the number of patients seen at GSH (A) and countries of origin (B), with 16.4% (*n*=21) SA patients, most of whom are of mixed/Indian ancestry (*n*=15), and 21.9% (*n*=28) from the Democratic Republic of Congo (DRC).

Of the 58 patients who regularly attend the Haematology Clinic, 34 (58.6%) consented to inclusion in the study (Fig. 1). Over 75% of the patients at GSH were referrals from RCWMCH, with others coming from neighbouring secondary-level hospitals in Cape Town and a minority of internal referrals of relatives of attending patients.

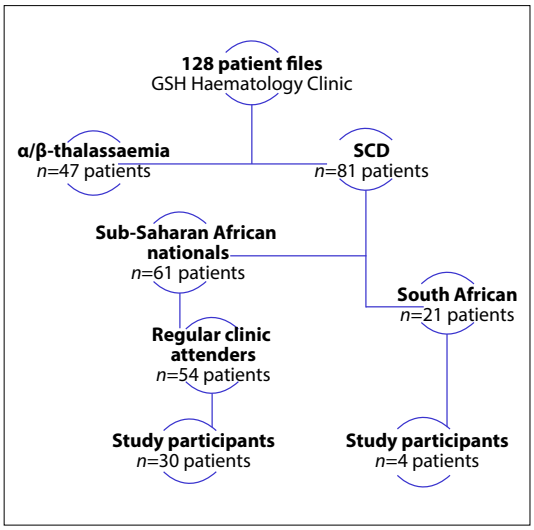


Fig. 1. Patient recruitment flow chart.

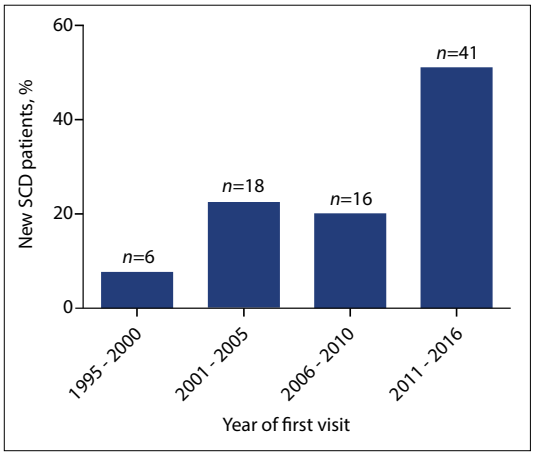


Fig. 2. Incidence trend of SCD at GSH over 20 years.

Clinical and haematological profile

The mean age was 26.1 years (standard deviation (SD) 9.8, range 15 - 51), and 70.6% were male. The rate of co-inheritance of a single 3.7 kb α-globin gene deletion was 42.3% (*n*=11).

Table 1 summarises the haematological and clinical events recorded for all patients. Information obtained from the anamnesis indicated that the majority of the patients were diagnosed relatively late (mean

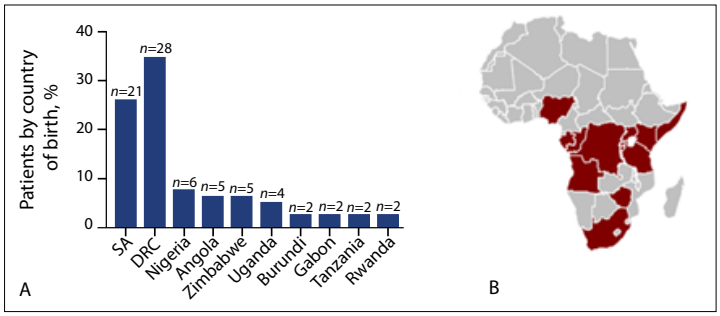


Fig. 3. Numbers and percentages of patients by nationality (A) and the distribution of countries of birth for patients at GSH (B).

Table 1. Description of haematological indices, clinical events and globin genes of the study cohort

Variables	% or mean (SD)	Value range	*n*
Gender			
Female	29.4	-	10
Male	70.6	-	24
Age (years)	26.1 (9.8)	15 - 51	34
Haematological indices			
Haemoglobin (g/dL)	9.1 (1.8)	5.6 - 14.0	27
MCV (fL)	91.5 (31.9)	53.6 - 146.4	27
WBCs (× 10^9/L)	8.1 (3.0)	3.0 - 13.6	27
Platelets (× 10^9/L)	334 (166.5)	133 - 737	27
Neutrophils (× 10^9/L)	4.6. (1.5)	2.0 - 10.3	27
Clinical events			
Age at diagnosis (years)	6.8 (7.1)	1 - 39	30
Vaso-occlusive crisis (number/year)	1.5 (1.2)	0 - 5	30
Stroke	13.3	-	4/30
Leg ulcers	16.7	-	5/30
Hospitalisations (number/year)	2.1 (1.4)	0 - 5	30
Treatment			
Blood transfusions	33.3	-	10/30
HU (mg/d)	544.1 (144)	500 - 1 000	30
β-globin genotype			
HbSS	85.2	-	23/27
HbAS	14.8	-	4/27
β-globin haplotype			
Bantu/Bantu	50.0	-	13/26
Bantu/Senegal	7.7	-	2/26
Bantu/Benin	7.7	-	2/26
Bantu/Atypical	15.4	-	4/26
Atypical	19.2	-	5/26
α-globin gene deletion			
αα/αα	57.7	-	15/26
αα/α3.7	42.3	-	11/26

MCV = mean corpuscular volume; WBCs = white blood cells; HbSS = haemoglobin SS; HbAS = haemoglobin AS.

age at diagnosis 6.8 years (SD 7.1), range 1 - 39), as a result of the presentation of the initial clinical manifestations of SCD, mainly pain episodes.

Clinical management

With regard to treatment, 33.3% (*n*=10) of the patients in the cohort had received at least one blood transfusion. About 16.7% (*n*=5) of the patients had been enrolled in a hypertransfusion programme, ranging from a fortnightly to monthly transfusion regimen, to manage complications such as stroke, chronic pain crises and non-healing chronic leg ulcers. The frequency of transfusions is largely dependent on symptom severity and availability of blood units from the Western Cape Blood Transfusion Service, Cape Town, SA.

Of the patient cohort, 86.6% (*n*=26) were at maximum tolerated dose of hydroxyurea (HU) with dosages ranging between 500 and 1 000 mg/d. From our patient survey, 30 - 50% of the participants were fully compliant with HU treatment, 20 - 30% reported partial compliance (tending to forget to take the treatment two to three times a week), and some patients refused treatment. Reasons for refusal included potential cancer development in the future, family planning, particularly for men afraid of treatment-related infertility, and self-perceived improvement of symptoms without the treatment.

Genetic characteristics

Sickle cell genotypes: β-globin haplotypes and co-inheritance of α-thalassaemia

The description of the HbS allele frequency, β-globin haplotype background and α-globin gene deletion for the patients is given in Table 1. Genotyping for the HbS mutation revealed that 85.2% (*n*=23) of the patients were homozygous for the mutation (haemoglobin SS), with the rest being heterozygous (haemoglobin AS) with a possibility of β^0-thalassaemia (HbS/β^0) (*n*=4), all of whom were South African with mixed and Indian ancestry. Fig. 4 shows the distribution of the SCD β-globin gene haplotypes: the Bantu and Atypical haplotypes accounted for 65.4% and 26.9%, respectively, whereas the Senegal and Benin haplotypes accounted for 3.8% each, with no observation of the Cameroon and Indian-Arab haplotypes. In combination, the Bantu/Bantu haplotype represented 50.0% of the patients (Table 1). The heterozygous 3.7 kb α-globin gene deletion was observed in 42.3% of the patients.

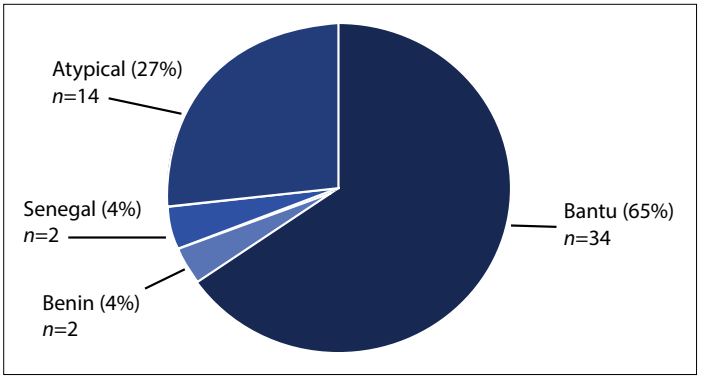


Fig. 4. Distribution of SCD haplotypes by number and percentage of chromosomes.

Frequency of genetic variants associated with HbF levels

Table 2 shows the observed alleles and minor allele frequency (MAF) of genetic variants previously associated with HbF, some recently in a Sardinian population.[30] All variants were in HWE (p>0.05) with the exception of four loci (X12_123681790, rs141494605, rs183437571 and rs192197462) that presented monomorphic alleles in all patients. Tests of association between the variants and all haematological indices including Hb levels were conducted; however, no significant associations were observed except for rs6466533 and the SCD haplotype combinations. The CC genotype (rs6466533) was associated with higher platelet counts than the heterozygous TC genotype (p<0.05). Similarly, the Bantu/Bantu haplotype combination was associated with higher platelet counts than the Atypical/Atypical genotype (p<0.05). Table 3 shows the MAF of the above variants in African populations: Esan (Nigeria), Luhya (Kenya), Mandinka (Gambia) and Mende (Sierra Leone); American (including African-American), European and both East and South Asian populations.

Table 2. Genetic variants previously associated with HbF

SNP	Chromosome loci	Alleles	MAF	*n*
rs8176703	136135863*	G/A	0.130	27
rs372091	11:5496926	G/A	0.148	27
rs2334880	16:71619734	C/T	0.400	25
rs1427407	2:60490908	C/A	0.204	27
rs7606173	2:60498316	C/G	0.426	27
X12_123681790	X12_123681790	A†	-	31
X16_391593	X16_391593	T/C	0.015	33
rs10468869	18:51795403	A/G	0.400	30
rs10756993	9:18839726	C/A	0.016	31
rs113267280	6:41952511	T/G	0.016	31
rs11754265	6:135356216	C/G	0.172	29
rs141494605	16:216593	T†	-	27
rs148706947	16:342218	C/T	0.047	32
rs183437571	19:13121899	C†	-	30
rs192197462	14:57116065	A†	-	30
rs570013781	16:149539	G/A	0.016	31
rs59329875	20:44547672	C/T	0.296	27
rs62573842	9:32264314	A/G	0.250	30
rs6466533	7:79163576	T/C	0.283	30
rs6590706	11:133465011	A/G	0.217	30
rs67104793	3:142444839	A/DEL.A	0.190	29
rs7163278	15:93345162	T/C	0.121	29
rs76901220	ss131769967†	G/A	0.016	31

*No chromosome location provided in dbSNP, reference ID provided.
†Monomorphic.

Discussion

To the best of our knowledge, this is the first study describing the clinical and genetic backgrounds of SCD patients at GSH and reporting on adult patients with SCD in SA. The results of this study indicate a similar trend of a rapid increase in the number of cases of SCD that was previously reported at RCWMCH in Cape Town.[27] This was also the result of migration from SSA countries where SCD is most prevalent. Related to this was a specific administrative difficulty in taking care of some patients who lack the up-to-date and correct paperwork for immigrants and asylum seekers. This was an indirect indication that most patients arrived as adults in SA, contrary to the observation in the second part of the last decade at RCWMCH, where most patients were SA born. It is therefore expected that the adult SCD population at GSH will continue to grow from the compounded effects of future referrals from neighbouring paediatric hospitals and the arrival of new adult patients from migrant populations, as migration, particularly into SSA, continues to be the reality for many people seeking political asylum, economic opportunities and better healthcare. Concomitant with this migration, the improved clinical management and healthcare of paediatric SCD patients is expected to increase the pool of adult patients living with SCD. This will increase the number of patients who will survive well beyond reproductive age, which is likely to increase the frequency of the HbS allele in the population.

Newborn screening and comprehensive clinical care programmes, which are also possible in SA, have reduced SCD-related premature childhood deaths by 70% in high-income nations such as the USA,[31,32] and most patients can survive into adulthood.[33] A similar increasing trend of SCD in countries previously not affected by the disease has been observed in Ireland,[34] Italy,[35] Germany,[36] England[37] and France.[38] Therefore, the evidence that the SCD burden is comparable to that of communicable diseases and other major global diseases such as hypertension and diabetes[2] will have increasing resonance. The marked increase

Table 3. Minor allele frequencies of select HbF-promoting SNPs in various populations in the 1000G project

SNPs	Our data	African	Esan (Nigeria)	Luhya (Webuye, Kenya)	Mandinka (The Gambia)	Mende (Sierra Leone)	America	African Caribbean (Barbados)	African American (southwestern USA)	Europe	East Asia	South Asia
rs8176703	A = 0.13	T = 0.041	T = 0.0354	T = 0.0202	T = 0.053	T = 0.0353	T = 0.006	T = 0.0573	T = 0.0164	T = 0.002	T = 0.0000	T = 0.0000
rs372091	A = 0.148	A = 0.068	A = 0.1414	A = 0.0707	A = 0.013	A = 0.0235	A = 0.001	A = 0.0469	A = 0.0410	A = 0.000	A = 0.000	A = 0.000
rs2334880	T = 0.4	A = 0.398	A = 0.3990	A = 0.4697	A = 0.403	A = 0.4000	A = 0.102	A = 0.3594	A = 0.3115	A = 0.160	A = 0.009	A = 0.019
rs1427407	A = 0.204	T = 0.238	T = 0.1970	T = 0.2020	T = 0.301	T = 0.3353	T = 0.226	T = 0.2135	T = 0.2295	T = 0.151	T = 0.256	T = 0.120
rs7606173	G = 0.426	C = 0.467	G = 0.4697	G = 0.4545	C = 0.403	C = 0.4176	C = 0.290	C = 0.4115	G = 0.4754	C = 0.435	C = 0.013	C = 0.173
rs10468869	G = 0.4	G = 0.341	G = 0.3283	G = 0.3737	G = 0.274	G = 0.3529	A = 0.324	G = 0.3542	G = 0.4180	A = 0.251	G = 0.377	A = 0.369
rs10756993	A = 0.016	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.000	A = 0.0000	A = 0.0000	A = 0.000	A = 0.003	A = 0.000
rs113267280	G = 0.016	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.0000	G = 0.008	G = 0.0000	G = 0.004
rs11754265	G = 0.172	G = 0.234	G = 0.3081	G = 0.1717	G = 0.195	G = 0.1882	G = 0.448	G = 0.2552	G = 0.2951	C = 0.493	G = 0.466	G = 0.468
rs141494605	*	C = 0.0000	C = 0.0000	C = 0.0000	C = 0.0000	C = 0.0000	C = 0.012	C = 0.0000	C = 0.0000	C = 0.012	C = 0.0000	C = 0.002
rs148706947	T = 0.047	T = 0.0000	T = 0.0000	T = 0.0000	T = 0.0000	T = 0.0000	T = 0.001	T = 0.0000	T = 0.0000	T = 0.013	T = 0.0000	T = 0.003
rs183437571	*	T = 0.008	T = 0.0000	T = 0.0051	T = 0.027	T = 0.0118	T = 0.006	T = 0.0000	T = 0.0082	T = 0.001	T = 0.0000	T = 0.0000
rs570013781	A = 0.016	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.0000	A = 0.009	A = 0.0000	A = 0.0000	A = 0.003	A = 0.0000	A = 0.001
rs59329875	T = 0.296	C = 0.191	C = 0.2222	C = 0.2374	C = 0.133	C = 0.1412	C = 0.114	C = 0.2292	C = 0.1639	C = 0.198	C = 0.026	C = 0.252
rs62573842	G = 0.25	G = 0.160	G = 0.1313	G = 0.1970	G = 0.150	G = 0.0941	G = 0.329	G = 0.1927	G = 0.2295	G = 0.437	A = 0.459	A = 0.473
rs6466533	C = 0.283	C = 0.188	C = 0.2071	C = 0.1717	C = 0.146	C = 0.1235	T = 0.419	C = 0.2552	C = 0.3279	T = 0.171	C = 0.486	T = 0.497
rs6590706	G = 0.217	G = 0.222	G = 0.1667	G = 0.1717	G = 0.221	G = 0.1588	A = 0.091	G = 0.2865	G = 0.3934	A = 0.026	A = 0.154	A = 0.143
rs67104793	DEL.A = 0.19	T = 0.227	T = 0.1364	T = 0.2828	T = 0.279	T = 0.2529	T = 0.218	T = 0.2031	T = 0.2295	T = 0.197	T = 0.319	T = 0.231
rs7163278	C = 0.121	C = 0.137	C = 0.1111	C = 0.0909	C = 0.168	C = 0.1353	C = 0.311	C = 0.1719	C = 0.2459	T = 0.383	C = 0.300	C = 0.221

*Monomorphic.

in patients between 2001 and 2005 could also be associated with the ending of a civil war, a transitional government and political instability in DRC, the effects of which had spread into neighbouring states. The more recent increase (2011 - 2016) is more likely to be due to economic and health-motivated migration. The increase of SCD in both paediatric and adult settings will impose a new burden on the healthcare system in SA concomitant with a new need for training at all levels of medical education, as well as the need for policies from health authorities for the prevention, management and care of haemoglobinopathies.

The number of annual vaso-occlusive crises was similar to that reported in Cameroon, as was the mean number of vaso-occlusive crises per year.[9] Major challenges faced by healthcare professionals at GSH were patient compliance with HU treatment, compliance with supportive medication such as folic acid and patient clinic attendance. There were several barriers to HU treatment, including the financial implications of taking time off work to attend the clinic and receive it (maximum one month's supply), and misconceptions about the treatment and its possible carcinogenic effects, as well as potential impotence in male patients.[39] Most patients who fail to comply with clinic appointments do so simply because they feel better and therefore see no need to go to the hospital.

The novel aspect of this article is the report on the genetic background of four key modifiers of HbF: variants at the *BCL11A* erythroid-specific enhancer, β-globin haplotypes, α-thalassaemia 3.7 kb gene deletion and several other known HbF-promoting polymorphisms. It is imperative to gain a better understanding of genetic variants affecting the predisposition to specific complications such as stroke and acute chest syndrome, and polymorphisms affecting susceptibility to pain, as well as the pharmacogenomics of commonly prescribed treatments such as HU, malaria prophylaxis and pain medication. The dominance of the Bantu haplotype in this cohort is in accordance with the Congolese origin of most patients. Indeed, the Bantu

haplotype is most prevalent below the equatorial malaria belt across southern African countries.[19] Most variants were not associated with haematological indices except the CC genotype at rs6466533 and Bantu/Bantu haplotype combination with platelet counts, probably because of the modest sample size. The differences in MAF of the recently identified HbF-promoting loci in a Sardinian population[30] among African populations from the Human 1000 Genome Project (1000G) (Table 3) emphasises the necessity of large-scale genomic analyses on various populations across the continent, as there are vast variations between any two African populations.

It should be made clear that the intention of this article is not to stigmatise SCD nor immigrant patients, but to inform and prepare medical care providers and healthcare officials of the increasing need for management of haemoglobinopathies in SA. This trend is not restricted to SA, with countries such as Italy[40] the Republic of Ireland,[34] England[41] and Germany[36] affected by the reality of population movement and new burden of disease, developing neonatal screening programmes and establishing SCD centres in response to similar increases in SCD prevalence.

Study limitations

The final sample size of patients included in the present study was limited because of poor patient compliance with clinic attendance, self-transfers to other hospitals and the shortened period of recruitment. The sample size did not allow for robust statistical analyses to reveal potent markers of specific phenotypes or clinical measures. HbF levels were measured for a handful of patients performed using high performance liquid chromatography before initiation of HU treatment. This haematological measure would have been ideal to check for association with genetic variants as baseline HbF.

Conclusions

Over the past 10 years, the number of adult patients living with SCD has increased considerably, imposing the creation of a weekly outpatient service at GSH. The genetic profile is similar to that of many other SCD patients from the other SSA countries from where most patients originate. The trend has a number of implications, particularly for medical education at academic and training institutions, policy action on prevention and care at the National Department of Health, and research in haemoglobinopathies in SA.

Acknowledgments. We would like to thank the nursing staff of the Haematology Service at GSH for their support and the patients for their participation.

Contribution to authorship. GP and AW conceived and designed the experiments. GP, KM and MJ recruited and sampled the patients. GP and KM performed the experiments. GP, KM and AW analysed the data. AW, MJ, SM and NN contributed reagents/materials/analysis tools. GP and AW wrote the article. GP, KM, MJ, SM, NN and AW revised and approved the manuscript.

Funding statement. The molecular experiments of the study were funded by the National Health Laboratory Services, SA and the National Institutes of Health, USA, grant no. 1U01HG007459-01. The students' bursaries were supported by the Oppenheimer Memorial Trust, the National Research Foundation, and FirstRand Laurie Dippenaar Scholarship, SA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1. Herrick JB. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. 1910. Yale J Biol Med 2001;74(3):179-184. http://dx.doi.org/10.1001/archinte.1910.00050330050003
2. Weatherall DJ. Hemoglobinopathies worldwide: Present and future. Curr Mol Med 2008;8(7):592-599. http://dx.doi.org/10.2174/156652408786241375
3. Piel FB, Patil, AP, Howes RE, et al. Global epidemiology of sickle haemoglobin in neonates: A contemporary geostatistical model-based map and population estimates. Lancet 2013;381(9861):142-151. http://dx.doi.org/10.1016/S0140-6736(12)61229-X
4. Makani J, Menzel S, Nkya S, et al. Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. Blood 2011;117(4):1390-1392. http://dx.doi.org/10.1182/blood-2010-08-302703
5. Bartolucci P, Galacteros F. Clinical management of adult sickle-cell disease. Curr Opin Hematol 2012;19(3):149-155. http://dx.doi.org/10.1097/MOH.0b013e328351c35f
6. Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. Blood 2010;115(22):4331-4336. http://dx.doi.org/10.1182/blood-2010-01-251348
7. Malowany JI, Butany J. Pathology of sickle cell disease. Semin Diagn Pathol 2012;29(1):49-55. http://dx.doi.org/10.1053/j.semdp.2011.07.005
8. Thein SL, Menzel S, Peng X, et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci USA 2007;104(27):11346-11351. http://dx.doi.org/10.1073/pnas.0611393104
9. Wonkam A, Bitoungui VJN, Vorster AA, et al. Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. PLoS One 2014;9(3):e92506. http://dx.doi.org/10.1371/journal.pone.0092506
10. Lettre G, Sankaran VG, Bezerra MA, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci USA 2008;105(33):11869-11874. http://dx.doi.org/10.1073/pnas.0804799105
11. 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. http://dx.doi.org/10.1126/science.1242088
12. Galarneau G, Palmer CD, Sankaran VG, et al. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet 2010;42(12):1049-1051. http://dx.doi.org/10.1038/ng.707
13. Mtatiro SN, Mgaya J, Singh T, et al. Genetic association of fetal-hemoglobin levels in individuals with sickle cell disease in Tanzania maps to conserved regulatory elements within the MYB core enhancer. BMC Med Genet 2015;16:1. http://dx.doi.org/10.1186/s12881-015-0144-7
14. Pule GD, Ngo Bitoungui VJ, Chetcha Chemegni B, et al. Association between variants at BCL11A erythroid-specific enhancer and fetal hemoglobin levels among sickle cell disease patients in Cameroon: Implications for future therapeutic interventions. OMICS 2015;19(10):627-631. http://dx.doi.org/10.1089/omi.2015.0124
15. Hsu LL, Miller ST, Wright E, et al. Alpha thalassemia is associated with decreased risk of abnormal transcranial Doppler ultrasonography in children with sickle cell anemia. J Pediatr Hematol Oncol 2003;25(8):622-628. http://dx.doi.org/10.1097/00043426-200308000-00007
16. Rumaney MB, Bitoungui VJN, Vorster AA, et al. The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. PLoS One 2014;9(6):e100516. http://dx.doi.org/10.1371/journal.pone.0100516
17. Saleh-Gohari N, Mohammadi-Anaie M. Co-inheritance of sickle cell trait and thalassemia mutations in south central Iran. Iran J Public Health 2012;41(10):81-86.
18. Wambua S, Mwacharo J, Uyoga S, et al. Co-inheritance of α+-thalassaemia and sickle trait results in specific effects on haematological parameters. Br J Haematol 2006;133(2):206-209. http://dx.doi.org/10.1111/j.1365-2141.2006.06006.x
19. Bitoungui VJN, Pule GD, Hanchard N, et al. Beta-globin gene haplotypes among Cameroonians and review of the global distribution: Is there a case for a single sickle mutation origin in Africa? OMICS 2015;19(3):171-179. http://dx.doi.org/10.1089/omi.2014.0134
20. Flint J, Harding, RM, Boyce, AJ, et al. The population genetics of the haemoglobinopathies. Bailliere Clin Hematol 1998;11(1):1-51. http://dx.doi.org/10.1016/s0950-3536(98)80069-3
21. Elion J, Berg PE, Lapoumeroulie C, et al. DNA sequence variation in a negative control region 5′ to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. Blood 1992;79(3):787-792.
22. Pagnier J, Mears JG, Dunda-Belkhodja O, et al. Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. Proc Natl Acad Sci USA 1984;81(6):1771-1773. http://dx.doi.org/10.1073/pnas.81.6.1771
23. Labie D, Pagnier J, Lapoumeroulie C, et al. Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. Proc Natl Acad Sci USA 1985;82(7):2111-2114. http://dx.doi.org/10.1073/pnas.82.7.2111_
24. Williams TN, Mwangi TW, Wambua S, et al. Sickle cell trait and the risk of *Plasmodium falciparum* malaria and other childhood diseases. J Infect Dis 2005;192(1):178-186. http://dx.doi.org/10.1086/430744
25. Bonafede RP, Botha MC, Beighton P. Inherited anaemias in the Greek community of Cape Town. J Med Genet 1979;16(3):197-200. https://dx.doi.org/10.1136/jmg.16.3.197
26. Beighton P, Botha M. Inherited disorders in the black population of southern Africa. S Afr Med J 1986;69(4):247-249.
27. Wonkam A, Ponde C, Nicholson N, et al. The burden of sickle cell disease in Cape Town. S Afr Med J 2012;102(9):752-754. http://dx.doi.org/10.7196SAMJ.5886
28. Saiki RK, Scharf S, Faloona F, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985;230(4732):1350-1354. http://dx.doi.org/10.1126/science.2999980
29. Steinberg M, Lu Z, Nagel R, et al. Hematological effects of atypical and Cameroon β-globin gene haplotypes in adult sickle cell anemia. Am J Hematol 1998;59(2):121-126. http://dx.doi.org/10.1002/(sici)1096-8652(199810)59:2<121::aid-ajh4>3.0.co;2-#
30. Danjou F, Zoledziewska M, Sidore C, et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. Nat Genet 2015;47(11):1264-1271. http://dx.doi.org/10.1038/ng.3307
31. Yanni E, Grosse SD, Yang Q, et al. Trends in pediatric sickle cell disease-related mortality in the United States, 1983-2002. J Pediatr 2009;154(4):541-545. http://dx.doi.org/10.1016/j.jpeds.2008.09.052
32. Vichinsky EP. Comprehensive care in sickle cell disease: Its impact on morbidity and mortality. Semin Hematol 1991;28(3):220-226.
33. Chaturvedi S, DeBaun MR. Evolution of sickle cell disease from a life-threatening disease of children to a chronic disease of adults: The last 40 years. Am J Hematol 2016;91(1):5-14. http://dx.doi.org/10.1002/ajh.24235
34. Gibbons C, Geoghegan R, Conroy H, et al. Sickle cell disease: Time for a targeted neonatal screening programme. Ir Med J 2015;108(2):43-45.
35. Colombatti R, Perrotta S, Samperi,P, et al. Organizing national responses for rare blood disorders: The Italian experience with sickle cell disease in childhood. Orphanet J Rare Diseases 2013;8:169. http://dx.doi.org/10.1186/1750-1172-8-169
36. Zur B. Increase in genetically determined anemia as a result of migration in Germany. Internist (Berl) 2016;57(5):444-451. http://dx.doi.org/10.1007/s00108-016-0030-1
37. Pizzo E, Laverty AA, Phekoo KJ, et al. A retrospective analysis of the cost of hospitalizations for sickle cell disease with crisis in England, 2010/11. J Public Health (Oxf) 2015;37(3):529-539. http://dx.doi.org/10.1093/pubmed/fdu026

38. Habibi A, Arlet JB, Stankovic K, et al. French guidelines for the management of adult sickle cell disease: 2015 update. Rev Med Interne 2015;36(5 Suppl 1):5S3-S84. http://dx.doi.org/10.1016/S0248-8663(15)60002-9
39. Pule GD, Mowla S, Novitzky N, et al. A systematic review of known mechanisms of hydroxyurea-induced fetal hemoglobin for treatment of sickle cell disease. OMICS 2015;8(5):669-679. http://dx.doi.org/10.1586/17474086.2015.1078235
40. Colombatti R, Montanaro M, Guasti F, et al. Comprehensive care for sickle cell disease immigrant patients: A reproducible model achieving high adherence to minimum standards of care. Paed Blood Cancer 2012;59(7):1275-1279. http://dx.doi.org/10.1002/pbc.24110
41. Daniel YA, Henthorn J. Newborn screening for sickling and other haemoglobin disorders using tandem mass spectrometry: A pilot study of methodology in laboratories in England. J Med Screen 2016;23(4):175-178. http://dx.doi.org/10.1177/0969141316631008

Accepted 10 June 2016.

2. Pule GD, **Mnika K**, Chimusa E, Mhandire K, Kampira E, Dandara C, Wonkam A. Beta globin gene haplotypes and tag variants for malaria resistance among black southern African populations. Global Health, Epidemiology and Genomics. 2017; 2, (17); 1-11. (Status: *published*)

Contribution to authorship

GP, EC, AW conceived and designed the experiments.
EK, **KM (Khuthala Mnika)**, GP performed the experiments.
GP, EC, AW analysed the data.
AW, CD, **KM (Khuthala Mnika)**, EK contributed reagents/materials/ analysis tools.
GP, EC, AW wrote the paper.
GP, **KM (Khuthala Mnika)**, EC, CD, EK, AW revised and approved the manuscript.

global health, epidemiology and genomics

GENETICS
ORIGINAL RESEARCH ARTICLE

Beta-globin gene haplotypes and selected Malaria-associated variants among black Southern African populations

G. D. Pule[1], E. R. Chimusa[1], K. Mnika[1], K. Mhandire[2], E. Kampira[3], C. Dandara[1] and A. Wonkam[1]*

[1] Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[2] Departments of Chemical Pathology, University of Zimbabwe, Harare, Zimbabwe
[3] Malawi College of Health Sciences, University of Malawi, Blantyre, Malawi

Global Health, Epidemiology and Genomics (2017), **2**, e17, page 1 of 11. doi:10.1017/gheg.2017.14

Partial carrier-resistance to *Plasmodium falciparum* malaria conferred by the sickle cell (*HbS*) mutation has resulted in the local amplification and positive selection of sickle cell disease (SCD) in malaria-endemic regions and particularly in sub-Saharan Africa (SSA). The present study investigated the β-globin gene haplotypes, and selected malaria-associated variants among three cohorts of Bantu-speaking individuals from Malawi, Zimbabwe and South Africa compared with reports with data from others SSA populations. The data suggest a south-ward frequency decrease of malaria-associated variants in SSA linked to the evolutionary dynamics of various African populations' genomes through selective pressure of malaria. These selected genomics differences, positive selection of SCD in malaria-endemic regions among 'Bantus' from various part of Africa emphasise the evidence of the dissociation between genetics, anthropology and culture. The present study also showed a relatively prevalent Benin haplotype, which is mostly found in West Africa, among Southern African Blacks and very low Bantu haplotype, which could suggest a major migration route, of Southern Africa Bantu, along the African west coast, post-occurrence of the Sickle cell mutation, which date remain to be fully elucidated.

Received 6 October 2016; Revised 27 September 2017; Accepted 16 October 2017

Key words: β-globin haplotypes, malaria resistance, sickle cell disease, Southern Africa.

Introduction

Sickle cell disease (SCD) is a monogenic, hematological and multi-organ disorder affecting the structure of erythrocytes by altering the normal biconcave shape to a crescent [1]. The sickling results from the polymerization and precipitation of the β-globin chains (*HbS*) during deoxygenation and dehydration of erythrocytes [2]. The vascular pathology of the disease includes platelet and leukocyte adhesion abnormality and hypercoagulation leading to microvascular occlusion, hemolysis and hypoxia and ultimately, multi-organ damage.

There is a strong correlation between the frequency of the HbS gene and the historical distribution and incidences of malaria [3] because of the partial carrier-resistance to *Plasmodium falciparum* malaria. The geographical co-occurrence of SCD and malaria and the partial carrier-resistance is believed to have resulted in the local amplification and positive selection of SCD in malaria-endemic regions [4, 5]. A GWAS for severe malaria in Ghana and the Gambia reported four loci with genome-wide significant single-nucleotide polymorphisms (SNPs) associated with the disease. Two of them were tag SNPs of previously known causal variants (*rs8176703* in *ABO*, causal variant *rs8176719*; *rs372091* in *HBB*, causal variant

* Address for correspondence: Professor A. Wonkam, Division of Human Genetics, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, Cape Town, Republic of South Africa.
(Email: ambroise.wonkam@uct.ac.za)

rs334), whereas the other two were novel loci with unknown causal SNPs. The ABO locus has the previous indication of a protective effect conferred by the blood group O against severe malaria [6–8]. The variant *rs2334880*, was one of the novel resistance loci identified and was mapped to 6.4 kb upstream of the MARVEL domain-containing protein 3 gene (*MARVELD3*; MIM ID*614094), which forms part of multiple tight-junction of epithelial and vascular endothelial cells [9–11] and is strongly associated with severe malaria [12]. It is, however, noteworthy that no function mutation at *MARVELD3* is known and, in the current literature, evidence of association is conflicting [5, 6, 10]. The endemicity of malaria in sub-Saharan Africa (SSA) and the associated *HbS* mutations, has resulted in the highest SCD burden with nearly 80% of the approximately 300 000 new affected births that occur in SSA annually [13].

The *HbS* mutation is believed to have evolved independently in five regions of the world, classically associated with five region-defined haplotypes, four of which are African, based on conserved patterns of polymorphisms across the β-globin gene cluster, namely Benin, Central African (CAR) or Bantu; Cameroon; Senegal and Indian-Arab [4, 14, 15]. A recent review of the global distribution and frequencies of these haplotypes has provided a glimpse into population dynamics and migration within and out of Africa that has prompted the hypothesis of a single origin of *HbS* mutation [16]. In this context, the study of malaria associated variants among Southern African populations, specifically among South African Blacks that have been living outside the malaria-endemic equatorial belt for 3–5000 years [17, 18], could provide new insight into the within-Africa migration patterns, and some perspectives into the dissociation between genetics and anthropology, with regard to differential allele frequencies related to various conditions such as malaria, susceptibility and resistance among Bantu-speaking groups from various parts of Africa.

In this present study, we investigated the β-globin gene haplotypes and selected malaria-associated variants among three cohorts of healthy Southern African populations from Malawi, Zimbabwe and South Africa and compared the frequencies of these variants to that of other SSA populations, and data extracted from the 1000 Genome Project.

Methods

Ethics approval

The study was performed with the approval of the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee (HREC REF: 132/2010 and HREC REF: 1094/2009).

Populations

A total of 158 DNA samples (50 Zimbabweans; 58 Malawians and 50 South Africans) all of Bantu origin were randomly selected for the Division of Human Genetics bio-repositories, Faculty of Health Sciences, University of Cape Town. These participants were randomly sampled from a cohort of the unrelated and apparently healthy individual, initially recruited for a population genetics study.

Genotyping

HbS mutation and β-globin haplotypes

Using the participants from the three southern African populations, PCR and Dde I restriction analysis were used to confirm the absence of the HbS mutation [19] and published primers and methods [20] genotyping five restriction fragment length polymorphic (RFLP) regions in the β-globin gene cluster were used to analyse the XmnI (5'Gγ), HindIII (Gγ), HindIII (Aγ), HincII (3''$\Psi\beta$) and HinfI (5'β) loci for the HbS haplotype background (online Supplementary Table S1) [16]. Restriction endonuclease cutting patterns that represent each of the five most common atypical β-globin gene haplotypes are represented in online Supplementary Table S2.

Selection of Malaria associated SNPs

To compare the Minor Allele Frequencies (MAF) of malaria-associated SNPs between African living outside (mainly our three cohorts) and the malaria-endemic equatorial populations. We selected among recently identified malaria SNPs in [21], SNPs under linkage equilibrium, mostly with a great number of LD proxy variants in both Western (YRI, Yoruba) and eastern (LWK, Luhya in Webuye, Kenya) African Bantu. In doing so, three SNPs include *rs8176703*, *rs372091* and *rs2334880* that meet the above criteria (online Supplementary Figs S1–S3). These three SNPs are in fairly low LD ($r^2 < 0.2$) with the primary functional mutations [1], but being the most-associated markers in the GWAS conducted in [21]. To genotype these targeted SNPs, SNaPshot multiplex genotyping (based on the incorporation of a single ddNTP to an extension primer designed to anneal 1 bp upstream of the target SNP), and followed by capillary electrophoresis were used, according to a previously reported method [22]. Up to 10% of the genotypes' results were confirmed, by direct Sanger sequencing.

Data analysis and bioinformatics analysis using data extracted from the 1000G

Genotyping at the characterised loci conformed to Hardy–Weinberg Equilibrium (HWE) (p values > 0.05). Leveraging the moderated sample size and the accurate publicly phased data from 1000 Genomes Project, we compared the MAF of the selected SNPs to those of other African and non-African populations, and analysed the diversity of the beta-globin haplotype in five other African populations. We have used a custom python script to extract the data of five African

populations from 1000 Genome project phase3 on chromosome 11 in a 100 kb region around *HBB*. The data included 108 samples from Yoruba (YRI) in Nigeria, 99 from Esan (ESN) in Nigeria, 113 from Gambia (GWD) in Western Divisions in the Gambia, 99 Luhya (LWK) in Webuye, Kenya and 85 from Mende (MSL) in Sierra Leone. Plink software [23] was used to compute the haplotype blocks in each of those populations. Each inferred haplotype blocks was utilised in plink to estimate the haplotype frequency within the specific population. Similarly, the LD blocks were computed using Plink based on LD r2, and the LD pattern was visualised using Haploview [24]. From a custom R script, we have made use of 20 haplotypes from each population to plot the haplotype bifurcation at the variant *rs334*.

Results

Sickle cell genotype frequencies

The description of the HbS allele frequency, β-globin haplotype background and selected malaria-related SNPs for the study cohorts are given in Table 1. All participants from South Africa (100%, $n = 50$); and the majority from Zimbabwe (88%, $n = 50$) and Malawi (93.5%, $n = 58$) were determined to be homozygous unaffected (HbAA), with the rest being heterozygous for the sickle mutation (HbAS).

Haplotypes in the β-globin gene cluster

SCD exists in Africa on disparate haplotype backgrounds [25] and is described by a specific pattern of five SNPs across the β-globin gene cluster [16]. This pattern confers four haplotypes associated with the HbS mutation in Africa; Benin, Bantu/Central African Republic (CAR), Senegal and Cameroon, with the fifth haplotype arising in the Indian/Arabian peninsula (Arab/Hindu) [15, 26]. Any recombination of the defining SNPs results in recombinant haplotypes referred to as 'atypical'. The SCD haplotypes were described using a previously published method and the global distribution of the haplotypes reviewed [16]. The haplotypes were described based on the analysis of chromosomes from the South Africa, Zimbabwe and Malawi cohorts (78, 64 and 70 chromosomes respectively), the most prevalent of the β-globin gene haplotypes was the atypical form; 67.9, 65.6 and 51.4%, respectively. Specifically, atypical I was common across all three populations at similar frequencies, (32.1% South Africa; 38.1% Zimbabwe and 38.9% Malawi) (online Supplementary Table S3). The two second most prevalent haplotypes were the Benin and Cameroon forms. In combination, the atypical/atypical haplotype was most frequent in the South Africa and Zimbabwe cohorts (41.0 and 37.5%, respectively) whereas the Benin/atypical was the most frequent combination in Malawi (41.2%). Figure 1 shows the distribution of the

Table 1. *Frequencies of the HbAA; β-globin haplotypes and malaria-related SNPs*

		South Africa *N* (%)	Zimbabwe *N* (%)	Malawi *N* (%)
β-globin mutation	HbAA	50 (100)	44 (88.0)	58 (93.5)
	HbAS	0 (0.0)	6 (12.0)	4 (6.5)
β-globin haplotypes[a]	Atypical	53 (68.0)	42 (65.7)	36 (51.4)
	Benin	13 (16.6)	8 (12.5)	19 (27.1)
	Bantu	4 (5.1)	2 (3.1)	4 (5.7)
	Cameroon	5 (6.4)	10 (15.6)	5 (7.1)
	Senegal	3 (3.9)	2 (3.1)	6 (8.7)
β-globin haplotype recombinants[b]	Atypical/Atypical	16 (41.0)	12 (38.0)	8 (23.5)
	Benin/Atypical	13 (33.3)	8 (25.0)	14 (41.2)
	Bantu/Atypical	4 (10.3)	2 (6.3)	2 (5.9)
	Senegal/Atypical	2 (5.1)	1 (3.1)	4 (11.8)
rs8176703	GG	35 (0.97)	48 (0.96)	48 (0.98)
	AG	1 (0.03)	2 (0.04)	1 (0.02)
	AA	0 (0.0)	0 (0.0)	0 (0.0)
rs372091	GG	34 (0.94)	42 (0.93)	48 (0.98)
	AG	1 (0.03)	3 (0.07)	1 (0.02)
	AA	1 (0.03)	0 (0.0)	0 (0.0)
rs2334880	CC	7 (0.23)	10 (0.21)	14 (0.29)
	CT	19 (0.63)	23 (0.48)	24 (0.50)
	TT	4 (0.13)	15 (0.31)	10 (0.21)

[a] β-globin haplotype frequencies are given as the number of chromosomes presenting with a specific haplotype.
[b] β-globin haplotype recombinants: the pair of haplotypes inherited in two separate chromosomes in an individual.

SCD β-globin gene haplotypes amongst the study cohorts compared with the haplotypes reported in SCD patients in other African countries [16].

Targeted Malaria-related variants

Malaria has slight low incidence in Southern compare with Western-central (equatorial region) Africa and given that sickle cell anemia patients are known for potential resistance to the parasite that causes malaria [1]; it is therefore worth to investigate the population allele frequency at these resistance loci between populations in Southern and Western-central Africa. After discarding associated variants under LD and prioritizing associated variants with high proxy LD variants, three SNPs were selected (online Supplementary Figs S1–S3); *rs8176703* (9q34.2; ABO), *rs2334880* (16q22.2; *MARVELD3*) and *rs372091* (11p15.5; *HBB*) from the GWAS results in [21], to probe the allele frequencies and relative geographic distribution around and below the equatorial malaria belt. Despite the conflict in literature regarding the role of *MARVELD3*, this approach was driven by the hypothesis that, such resistance loci even at the level of single nucleotide polymorphisms, which confer clinically significant resistance to severe malaria would undergo strong positive selection in malaria-endemic regions and to a gradual lesser extent, regions around the equatorial belt. Therefore, the allele frequencies of three variants at resistance loci [21] were investigated among three sub-Saharan African populations (Malawi, Zimbabwe and South Africa) at varying proximity to the equatorial malaria endemicity belt. SCD unaffected populations were selected in order to eliminate the possible effect of co-inheritance of malaria resistance loci and the *HbS* allele, as a result of the HbS allele-conferred partial resistance to *P. falciparum*. The genotype frequencies for the *rs8176703* (GG), *rs372091* (GG) and *rs2334880* (CT) among South African, Malawian and Zimbabwean populations were largely similar. However, when comparing MAFs at these loci with other populations from the Human 1000 Genome Project, 1000 Genomes Phase III, there was an apparent gradient of the MAF for *rs8176703* and *rs372091*, highest in countries within the equatorial malaria belt (Gambia, Nigeria and Kenya) and lowest in the sub-equatorial populations investigated in this study (Table 2, Fig. 2). However, this pattern was not observed in the MAF at *rs2334880*. When comparing the measure of frequency differentiation among the genotyped SNPs and the corresponding frequencies of these SNPs in the 1000 Genomes data, the frequency of the genotyped SNPs were highest among the Southern African populations and the African populations extracted from the 1000 Genomes Project (Esan, Luhya, Yoruba, Mende and Mandinka) (Fig. 3). As expected, the frequencies were lowest among American, East-Asian and European populations, consistent with the fact that these geographic regions do not have a problem with malaria, and that the incidence of sickle cell anemia is decreasing [27]. Table 3 and online Supplementary Fig. S4 show the frequency of the *HbS* allele across African populations [13, 27–36]. When investigating the LD between these variants in the African 1000 Genomes phase3 data, these variants were found to be in linkage equilibrium with their respective

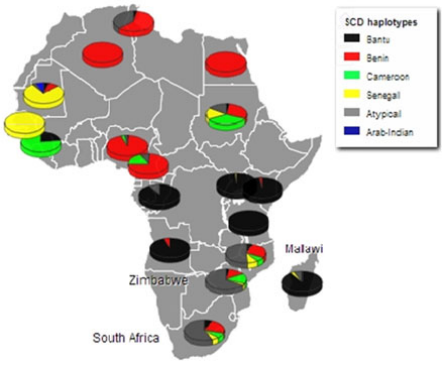


Fig. 1. Population frequencies of all haplotypes for the study cohorts (South Africa, Zimbabwe and Malawi), versus frequencies conditional on being homozygous for HbS in the other population's groups across the continent. (with adaptation of previously reported from [16]).

Table 2. *Minor allele frequencies of study cohorts and several populations from the 1000Genomes Project*

Region	Variants		
	rs8176703	rs372091	rs2334880
South Africa[a]	0.013	0.023	0.472
Malawi[a]	0.01	0.01	0.457
Zimbabwe[a]	0.02	0.033	0.552
African	0.398	0.068	0.398
African Caribbean (Barbados)	0.359	0.047	0.359
Southwest US (African American)	0.311	0.041	0.311
Nigeria (Esan)	0.399	0.141	0.399
Kenya (Luhya)	0.470	0.071	0.470
Kenya (Yoruba)	0.407	0.125	0.407
Mende (Sierra Leone)	0.400	0.024	0.400
Gambia (Mandinka)	0.403	0.013	0.403
America	0.102	0.001	0.102
Europe	0.160	1.000	0.160
East Asia	0.009	1.000	0.009
South Asia	0.019	1.000	0.019

[a] Study populations from the current study. Other data was sourced from the 1000G project.

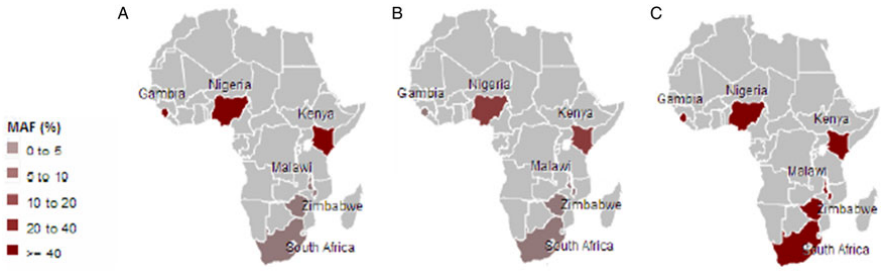


Fig. 2. Minor frequencies of malaria-restriction SNPs amongst southern African populations and three populations from the 1000Genomes Project within the malaria-endemic central Africa. A: *rs8176703*; B: *rs372091*; C: *rs2334880*.

functional mutations, suggesting deep sequencing to potentially prioritise novel mutation variants.

Pattern of linkage disequilibrium and haplotype blocks at rs334 in African populations

We have computed the haplotype blocks, block of linkage disequilibrium and the haplotype frequency in a 100 kb region around *HBB*, targeting the variant *rs334* in that region, which is well known of alleles A/T, encoding the *Hb A* form of (adult) hemoglobin and the sickling form of hemoglobin, *Hb S*, respectively. The results in Figs 4 and 5 show differing pattern of LD between Western and Eastern African Bantu.

Discussion

The present data confirm the evolutionary dynamics of various African Bantu genomes through selective pressure of malaria, and prompt the persecution of the dissociation between genetics and anthropology and culture, and lastly illustrated the importance of understanding the migration path of southern African populations, as a result of the past 1200 years southern African Bantu migration and various contact with sea-borne immigrants from Europe, Asia and Indonesia [37, 38].

South-ward frequency decrease of malaria-associated SNPs in SSA

As a result of the known partial resistance conferred by the *HbS* allele to malaria *Plasmodium falciparum* infection, the HbS allele is highly prevalent in malaria-endemic regions particularly around the tropical equatorial belt in SSA [3, 4]. The study confirms the accepted notion of low *HbS* allele frequency in populations outside malaria-endemic regions (online Supplementary Fig. S4; Table 3). In addition to the *HbS* mutation, whose association with malaria is extensively

	South Africa*	Malawi*	Zimbabwe*	Nigeria (Esan)	Kenya (Luhya)	Kenya (Yoruba)	Sierra Leone (Mende)	Gambia (Mandinka)	America	Europe	East Asia	South Asia
South Africa*												
Malawi*	1.07E-05											
Zimbabwe*	2.66E-04	3.83E-04										
Esan (Nigeria)	4.45E-03	4.43E-03	5.09E-03									
Luhya (Kenya)	6.39E-03	6.46E-03	6.54E-03	4.90E-04								
Yoruba (Kenya)	4.58E-03	4.57E-03	5.17E-03	1.01E-05	3.69E-04							
Mende (Sierra Leone)	4.30E-03	4.26E-03	4.95E-03	1.73E-04	4.18E-04	1.21E-04						
Mandinka (The Gambia)	4.37E-03	4.33E-03	5.00E-03	1.95E-04	3.91E-04	1.36E-04	1.09E-06					
America	4.75E-03	4.29E-03	7.50E-03	5.74E-03	9.25E-03	6.03E-03	5.57E-03	4.38E-04				
Europe	4.36E-02	4.42E-02	4.50E-02	3.40E-02	4.27E-02	3.55E-02	4.35E-02	3.43E-03	3.22E-03			
East Asia	4.63E-02	4.68E-02	4.88E-02	3.84E-02	4.87E-02	4.01E-02	4.80E-02	3.77E-03	3.22E-03	4.52E-05		
South Asia	4.61E-02	4.66E-02	4.86E-02	3.81E-02	4.83E-02	3.98E-02	4.77E-02	3.75E-03	3.22E-03	4.15E-05	3.54E-08	
	South Africa*	**Malawi***	**Zimbabwe***	Nigeria (Esan)	Kenya (Luhya)	Yoruba (Kenya)	Sierra Leone (Mende)	Gambia (Mandinka)	America	Europe	East Asia	South Asia

Fig. 3. Distribution of frequency differentiation of targeted SNPs *rs8176703*; *rs372091* and *rs2334880* across various African populations. When comparing the measure of frequency differentiation among the genotyped SNPs and the corresponding frequencies of these SNPs in the 1000Genomes data, the frequency of the genotyped SNPs were highest among the Southern African populations and the African populations (Esan, Luhya, Yoruba, Mende and Mandinka) (Fig. 3). The frequencies were lowest between American, Asian and European populations. * Populations studied in from the current paper (South Africa, Zimbabwe and Malawi); other data were extracted from the 1000G project. The values provided are F-statistics calculated between each MAF for the three SNPs (*rs8176703*; *rs372091* and *rs2334880*) and colored coded grey (genetically proximal) to red (genetically distal). Populations with less genetic distance have lower F-st and shown in grey whereas populations with greater genetic distance have higher F-st and are shown in red.

Table 3. *HbS allele frequencies by country in Africa*

Country	Study years	Population	Age group	HbS allele frequency	Reference
Angola	1950–2010	18 994	All ages	0.137	[13]
Benin	1950–2010	9219	All ages	0.159	[13]
Botswana	1950–2010	1977	All ages	0.003	[13]
Burkina Faso	1950–2010	16 250	All ages	0.056	[13, 30, 31]
	1997–1999	9201	New-borns	0.0025	
			Median 9 years	0.0013	
Burundi	1950–2010	8519	All ages	0.040	[13]
Cameroon	1950–2010	19 957	All ages	0.120	[13]
Cape Verde	1950–2010	513	All ages	0.026	[13]
Central African Republic	1950–2010	4506	All ages	0.077	[13]
Chad	1950–2010	11 509	All ages	0.051	[13]
Comoros	1950–2010	691	All ages	0.018	[13]
Congo	1950–2010	3760	All ages	0.145	[13]
Democratic Republic of the Congo	1950–2010	67 829	All ages	0.165	[13]
Djibouti	1950–2010	879	All ages	0.001	[13]
Equatorial Guinea	1950–2010	693	All ages	0.192	[13]
Eritrea	1950–2010	5204	All ages	0.003	[13]
Ethiopia	1950–2010	84 996	All ages	0.003	[13]
Gabon	1950–2010	1501	All ages	0.280	[13]
Gambia	1950–2010	1751	All ages	0.075	[13, 29]
	2003	536	New-borns	0.012	
			10–72 months	0.003	
Ghana	1950–2010	24 339	All ages	0.087	[13]
	2002	1266	0–4 years	0.0039	
		842	5–10 years	0.0012	
Guinea	1950–2010	10 324	All ages	0.168	[13]
Guinea-Bissau	1950–2010	1647	All ages	0.041	[13]
Kenya	1950–2010	40 835	All ages	0.038	[13, 15, 28]
	1998–1999	2774	New-borns	0.016	
		782	0–3 years	0.006	
	1998–2008	282	0–11 months	0.01	
		415	12–23 months	0.0035	
		3677	3–5 years	0.0024	
			6–13 years	0.0009	
Lesotho	1950–2010	2064	All ages	0.001	[13]
Liberia	1950–2010	4102	All ages	0.046	[13]
Madagascar	1950–2010	20 146	All ages	0.061	[13]
Malawi	1950–2010	15 690	All ages	0.033	[13]
Mali	1950–2010	13 362	All ages	0.057	[13]
Mauritania	1950–2010	3359	All ages	0.050	[13]
Mozambique	1950–2010	23 418	All ages	0.027	[13]
Namibia	1950–2010	2212	All ages	0.010	[13]
Niger	1950–2010	15 885	All ages	0.080	[13]
Nigeria	1950–2010	158 255	All ages	0.171	[13, 26]
	1970–1972	534	New-borns	0.021	
		259	1–4 years	0.004	
		637	5–14 years	0.002	
Rwanda	1950–2010	10 277	All ages	0.023	[13]
Sao Tome and Principe	1950–2010	165	All ages	0.094	[13]
Senegal	1950–2010	12 866	All ages	0.067	[13, 24, 25]
Senegal (rural kegoudou)	2002–2003	432	New-borns	0.005	
			2–10 years	None	
			Newborn	0.01	
Sierra Leone	1950–2010	5837	All ages	0.164	[13]

(*Continued*)

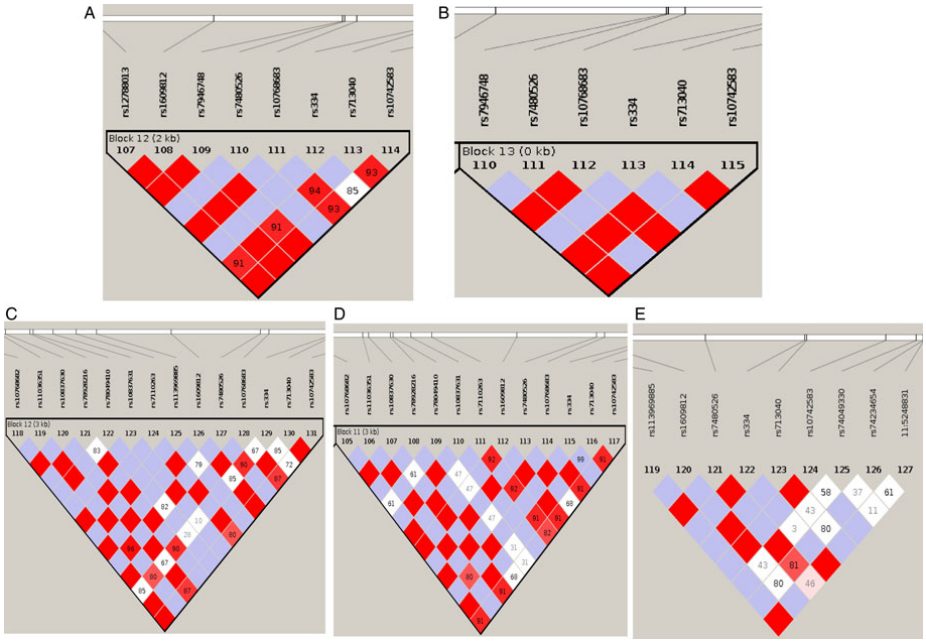


Fig. 5. Pattern of linkage disequilibrium. A linkage disequilibrium (LD) block of polymorphisms in a tight region around *rs334*. (A) ESN (B) YRI (C) GWD (D) MSL and (E) LWK.

and Kenya (1000 Genomes data) and showed a decreasing gradient of MAF for two of the loci (*rs8176703* and *rs372091*), which were highest within the equatorial malaria belt and lowest in all three study cohorts. However, there was no such gradient for the *rs2334880* variant, with similar MAF across all six populations, therefore not specifically linked with malaria endemicity, which is concomitant with results from an independent study where this variant failed to replicate its association [2, 3]. This trend suggests that although all three variants were highlighted with low LD to known functional mutations [1] and have been associated with resistance to severe malaria [21], only two (*rs8176703* and *rs372091*) are largely restricted to the equatorial malaria belt and possibly confer greater resistance to severe malaria as compared with the less equator-bound variant (*rs2334880*). The observed gradient for *rs8176703* and *rs372091* could indeed reflect that these variants are also associated with causal variants outside of West Africa, however, the absence of a gradient for *rs2334880* can just indicate that this variant is a poor proxy for the putative causal variant there. Furthermore, it is noteworthy that the *ATP2B4* [1] and *FREM3* [2] loci were not genotyped in this study due to a limitation of resources and challenges with assay optimization. Furthermore, several other loci have been associated with malaria-resistance and were not included in the present study. Future work should consider determining the frequency and effect of not only LD SNPs but include functional mutations at several loci to improve our understanding of the malaria-resistance genotype profile of Southern African populations, largely unexposed to malaria.

Haplotype blocks at rs334 in African populations

The differing pattern of LD between Western and Eastern African Bantu can be explained by the fact that Eastern African Bantu undergone admixture due to the various contacts with sea bone migrants in their past history [37]. However, the result in online Supplementary Table S4 suggests a similar pattern of beta globin haplotype diversity at the variant rs334 across all the five African populations (Western and Eastern Bantu). This illustrates the importance of investigating the origin and age of the *HbS* mutation following the past southern Bantu migration, admixture and population sub-structure.

Implications for genetics, anthropology and culture: Bantu is not equal to Bantu

Taken in sum; the frequency of *HbS* mutation and the decreasing MAF gradient of the targeted SNPs from the

tropical malaria-endemic regions towards the South suggest that the specific combination and pattern of multiple malaria resistant variants could allow the broad determination of the regional origins of an individual as Western, Central or Southern African. Although the vast majority of differentiated loci among Bantu populations are no more differentiated than would be expected from population drift, the modest data presented here support the proposed notion of dissociation between genetic background and ethno-linguistic attributes and classifications. Indeed, there are several indicators of a linguistic and cultural similarity among Bantus; for instance (i) 'muntu' for 'human' is the same in Xhosa (South African Bantu language) and Ewondo in Cameroon, (ii) the Ewondo and Xhosa tribes also share similar cultural and rite of passage practices such as the ritual of male circumcision and the burial of the umbilical cord or placenta of new-born as part of welcoming the new-born and introduction to the ancestors; (iii) and religious beliefs such as the 'cult of ancestry' and reincarnation are common amongst Bantu-speakers. Despite these and many other shared cultural, linguistic and anthropological attributes, the present data further support the notion that Bantu-speakers from Central and West Africa are no more genetically similar to those in Southern Africa, as previously illustrated with differential prevalence of HIV resistant genes amongst SSA populations [39, 40] and in this paper with malaria-associated variants. Given the vast genetic diversity within the continent and amongst any two SSA populations, the present research further emphasises the need to redefine the classifications of various groups in Africa by region-defined genomic attributes, as this approach could better serve Genomic medicine practice, as opposed to the classical ethno-linguistic population classification approach; the modest data presented here illustrate that at the genetic levels, Bantu is not equal to Bantu.

SCD β-globin haplotype: insights into the migration of Southern African blacks

The third question of this study was to investigate the degree of conservation of the five SCD haplotype-conferring loci in populations both largely unaffected by the disease and void of the environmental pressure of malaria. The most apparent, although not surprising result, was the high frequency of the atypical haplotypes in all the study cohorts leading to the hypothesis that in such populations, the five loci of the β-globin gene cluster may be under less evolutionary pressure to remain conserved. This could be due to several reasons; there is no apparent clinical benefit to retaining an otherwise unfavorable haplotype in the absence of malaria and potentially its strongest environmental positive selector, malaria. Furthermore, this could be as a result of genetic drift and recombination at the β-globin gene cluster. The next frequent haplotype in all study cohorts was the Benin form, suggesting that the Southern African Bantu-speakers migrated southwards, post-occurrence of the *HbS* mutation and is consistent with their West African origins [16]. The data showing the classification of the *HbS* haplotypes in these Southern African populations and the degree of similarity among the haplotype distributions in South Africa, Zimbabwe and Malawi is novel. The data suggest some insight into the evolutionary dynamics at the β-globin loci with regard to recombination of the classical *HbS* haplotypes and expansion of the atypical form in malaria-devoid regions in Africa.

Indeed, the result confirms to anthropological data detailing the most significant events of the geographic expansion of the Bantu Niger-Kordofanian-speakers out of Cameroon and Nigeria [17, 18]. It was previously hypothesised that the migration path was first through rainforest equatorial Africa and later into Eastern and Southern Africa. This is supported by the widespread distribution of Bantu-related linguistic groups and the presence of Niger-Kordofanian genetic ancestry in many African populations. However, the present result with a prevalent Benin haplotype and very low Bantu haplotype that is characteristic of SCD patients from Central and West Africa [16] could be due to a myriads of possible reasons that remain to be investigated: (i) the migration through East Africa of modern Southern African Bantu-speaking populations was transient with limited admixture with populations found locally in East Africa; (ii) some Bantu haplotypes may have been lost during recombination events at the β-globin gene locus potentially leading to the expansion of the highly prevalent atypical form; (iii) during the early migration events through the equatorial rainforests, the migrating populations from Central, East and West Africa encountered largely unoccupied regions, therefore expanding the Benin, Cameroon and Senegal haplotypes; (iv) the Bantu haplotype could be a recent haplotype of SCD, only recently expanding in Central Africa and subsequently in some parts of North and South America through slave trade; and lastly (v) the continuous socio-economically motivated migration from Central, East and West Africa into Southern African countries could have led to the relatively higher frequencies of the Benin, Cameroon and Senegal haplotypes although unlikely as this has become a significant migration phenomenon only in the past two to three decades. Beyond the concept of the dissociation between genetic background and ethno-linguistic attributes and classifications, the present data also complements previous studies on migrations of Southern African populations from West and/or Central Africa [38].

A limitation of the present study includes the number and the selection of malaria-associated variants selected. Given that the two SNPs in *HBB* and ABO are only weakly linked to the causative SNPs in the Ghana [21], it is likely that they are poor tags for the known causative SNP in the South African population. In addition, future studies should investigate the full distribution of 'atypical' haplotypes among

HbAA and *HbAS* individuals from both malaria endemic and non-endemic areas, to have a full profile of *HBB* haplotypes in Africa.

Conclusion

These selected malaria-associated variants in SSA suggest differences among 'Bantus' from various part of Africa, and emphasise the evidence of the dissociation between genetics, anthropology and culture. The present study also showed a relatively prevalent Benin haplotype, which is mostly found in West Africa, among Southern African Blacks and very low Bantu haplotype, which could suggest a major migration route, of Southern Africa Bantu, along the African west coast, post-occurrence of the sickle cell mutation. The data are indicative of the importance of the inclusion of Southern African populations when studying the age and origin of the HbS mutation, that remains to be fully elucidated [16]; Future studies should include Khoi and San populations, some of which may not have been exposed to malaria, and sequence around regions where our present results indicate LD with functional mutation to unravel novel candidates. Furthermore, these data also provide additional genetic evidence indicating the independent and continuous waves of migration of West and East African Bantu-speaking groups into Southern Africa. Beyond the data presented here, the high proportion of atypical haplotypes in Southern African populations, together with the data from diverse populations on the African continents could suggest various level of genetic diversification of African populations, whether attributable to recent and/or more ancient admixture, that did not probably result from a single North to South migration path nor a specific era, but rather through several independent and associated, multi-directional migration events [41–43]. It can be anticipated that modern-day continuous immigration, will further reinforce the African genomic diversity, by allowing the redistribution of gene pools previously restricted to specific geographical location, such as malaria-related mutations, across the continent.

Contribution to authorship

GP, EC, AW conceived and designed the experiments. EK, KM, GP performed the experiments. GP, EC, AW analysed the data. AW, CD, KM, EK contributed reagents/materials/analysis tools. GP, EC, AW wrote the paper. GP, KM, EC, CD, KM, EK, AW revised and approved the manuscript.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/gheg.2017.14

Acknowledgements

The molecular experiments of the study were funded by the National Health Laboratory Services (NHLS), South Africa; and the NIH, USA, grant number 1U01HG007459–01. The students' bursaries were supported by the Oppenheimer Memorial Trust, the National Research Foundation, and FirstRand Laurie Dippenaar Scholarship, South Africa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. **Weatherall D**, ***et al.*** A case for developing North-South partnerships for research in sickle cell disease. *Blood* 2005; **105**: 921–923.
2. **Bartolucci P**, ***et al.*** Clinical management of adult sickle-cell disease. *Current Opinion in Hematology* 2012; **19**: 149–155.
3. **Williams TN**, ***et al.*** Sickle cell trait and the risk of *Plasmodium falciparum* malaria and other childhood diseases. *The Journal of Infectious Diseases* 2005; **192**: 178–186.
4. **Flint J**, ***et al.*** 1 The population genetics of the haemoglobinopathies. *Baillière's Clinical Haematology* 1998; **11**: 1–51.
5. **Flint J**, ***et al.*** High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature* 1986; **321**: 744–750.
6. **Jallow M**, ***et al.*** Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genetics* 2009; **41**: 657–665.
7. **Loscertales M**, ***et al.*** ABO blood group phenotypes and *Plasmodium falciparum* malaria: unlocking a pivotal mechanism. *Advances in Parasitology* 2007; **65**: 1–50.
8. **Fry AE**, ***et al.*** Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. *Human Molecular Genetics* 2008; **17**: 567–576.
9. **Steed E**, ***et al.*** Identification of MarvelD3 as a tight junction-associated transmembrane protein of the occludin family. *BMC Cell Biology* 2009; **10**: 1.
10. **Malaria Genomic Epidemiology Network**, ***et al.*** A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 2015; **526**: 253–257.
11. **Idro R**, ***et al.*** Cerebral malaria: mechanisms of brain injury and strategies for improved neurocognitive outcome. *Pediatric Research* 2010; **68**: 267–274.
12. **Cinel I**, ***et al.*** Advances in pathogenesis and management of sepsis. *Current Opinion in Infectious Diseases* 2007; **20**: 345–352.
13. **Piel FB**, ***et al.*** Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet* 2013; **381**: 142–151.
14. **Elion J**, ***et al.*** DNA sequence variation in a negative control region 5' to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. *Blood* 1992; **79**: 787–792.
15. **Pagnier J**, ***et al.*** Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proceedings of the*

National Academy of Sciences of the United States of America 1984; **81**: 1771–1773.

16. **Bitoungui VJN**, **et al.** Beta-globin gene haplotypes among Cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? *Omics: a Journal of Integrative Biology* 2015; **19**: 171–179.
17. **Campbell MC**, **et al.** African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics* 2008; **9**: 403–433.
18. **Blench R**, **et al.** (eds) *The origins and development of African livestock: archaeology, genetics, linguistics and ethnography*, pp. 127–162. Routledge 2006, Abingdon-on-Thames, United Kingdom.
19. **Saiki RK**, **et al.** Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science *(*New York, N.Y.*)* 1985; **230**: 1350–1354.
20. **Steinberg M**, **et al.** Hematological effects of atypical and Cameroon *β*-globin gene haplotypes in adult sickle cell anemia. *American Journal of Hematology* 1998; **59**: 121–126.
21. **Timmann C**, **et al.** Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* 2012; **489**: 443–446.
22. **Wonkam A**, **et al.** Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLoS ONE* 2014; **9**: e92506.
23. **Purcell S**, **et al.** PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 2007; **81**: 559–575.
24. **Barrett JC**, **et al.** Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
25. **Labie D**, **et al.** Common haplotype dependency of high G gamma-globin gene expression and high HbF levels in beta-thalassemia and sickle cell anemia patients. *Proceedings of the National Academy of Sciences* 1985; **82**: 2111–2114.
26. **Elion J**, **et al.** DNA sequence variation in a negative control region 5'to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. *Blood* 1992; **79**: 787–792.
27. **Mulumba LL**, **et al.** Sickle cell disease among children in Africa: an integrative literature review and global recommendations. *International Journal of Africa Nursing Sciences* 2015; **3**: 56–64.
28. **Barclay GP.** A mortality study of sickle cell anaemia in Central Africa. *Journal of Clinical Pathology* 1971; **24**: 768.
29. **Sarr JB**, **et al.** Impact of red blood cell polymorphisms on the antibody response to *Plasmodium falciparum* in Senegal. *Microbes and Infection* 2006; **8**: 1260–1268.
30. **Saurin W.** Natural-selection study in small size populations - the drepanocytosis case in Madinka and Bedik from the Kedougou region (Eastern Senegal). *Bulletins et Memoires de la societe d Anthropologie de Paris* 1984; **11**: 37–56.
31. **Fleming AF**, **et al.** Abnormal haemoglobins in the Sudan savanna of Nigeria. I. Prevalence of haemoglobins and relationships between sickle cell trait, malaria and survival. *Annals of Tropical Medicine and Parasitology* 1979; **73**: 161–172.
32. **McAuley CF**, **et al.** High mortality from *Plasmodium falciparum* malaria in children living with sickle cell anemia on the coast of Kenya. *Blood* 2010; **116**: 1663–1668.
33. **Desai PC**, **et al.** Alloimmunization is associated with older age of transfused red blood cells in sickle cell disease. *American Journal of Hematology* 2015; **90**: 691–695.
34. **Cox SE**, **et al.** Haptoglobin genotype, anaemia and malaria in Gambian children. *Tropical Medicine & International Health* 2008; **13**: 76–82.
35. **Simpore J**, **et al.** Biological and clinical presentation of patients with hemoglobinopathies attending an urban hospital in Ouagadougou: confirmation of the modification of the balance between Hb S and Hb C in Burkina Faso. *Hemoglobin* 2002; **26**: 121–127.
36. **Simpore J**, **et al.** Modification in the frequency of Hb C and Hb S in Burkina Faso: an influence of migratory fluxes and improvement of patient health care. *Hemoglobin* 2002; **26**: 113–120.
37. **Chimusa ER**, **et al.** A genomic portrait of haplotype diversity and signatures of selection in indigenous Southern African populations. *PLoS Genetics* 2015; **11**: p.e1005052.
38. **Campbell MC, Sarah A.** Tishkoff. The evolution of human genetic and phenotypic variation in Africa. *Current Biology* 2010; **20**: R166-R173.
39. **Kinomoto M**, **et al.** HIV-1 proteases from drug-naive West African patients are differentially less susceptible to protease inhibitors. *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America* 2005; **41**: 243–251.
40. **Jenabian M**, **et al.** Pitfalls of antiretroviral drug resistance genotyping of HIV-1 group M and group N from Cameroon by sequenced-based assays. *Nigerian Medical Journal* 2015; **56**: 420.
41. **Li S, Schlebusch C, Jakobsson M.** Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proceedings of the Royal Society B: Biological Sciences* 2014; **281**: pii: 20141448.
42. **Ansari Pour N, Plaster CA, Bradman N.** Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *European Journal of Human Genetics* 2013; **21**: 423–429.
43. **Grollemund R**, **et al.** Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 2015; **112**: 13296–13301.

3. Sickle Cell Disease Ontology Working GroupThe Sickle Cell Disease Ontology: enabling universal sickle cell-based knowledge representation Database, Volume 2019, 2019, baz118, https://doi.org/10.1093/database/baz118 (*Status: published*)

Authors' contributions

1. Project leads: N.M., A.W. 2.
2. Project coordinator: V.N.
3. SCDO key curators: J.H., A.G., **K.M (Khuthala Mnika)**
4. SCDO Ontologists and Developers: G.K.M., S.J., K.G., C.B.H., M.H.
5. SCDO Content Providers: K.O.-F., all working group chairs and members
6. SCDO Working Group Chairs: K.A., A.C., F.C., C.C.- L., N.H., J.K.-M., D.M., O.N., B.T., M.T., C.R., S.O.-A., K.O.-F.

DATABASE
The Journal of Biological Databases and Curation

Database, 2019, 1–12
doi: 10.1093/database/baz118
Original article

Original article

The Sickle Cell Disease Ontology: enabling universal sickle cell-based knowledge representation

Sickle Cell Disease Ontology Working Group[1,2,†]

[1]Computational Biology Division, Institute of Infectious Disease and Molecular Medicine, N1.05, Werner Beit North, Faculty of Health Sciences, Anzio Road, Observatory, 7925 Cape Town, South Africa and [2]Division of Human Genetics, Department of Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925 Cape Town, South Africa

*Corresponding author: Tel: 0027 21 4066058; Email: nicola.mulder@uct.ac.za.
†Working group members who contributed to this work are listed in alphabetical order at the end of the manuscript.

Citation details: Sickle Cell Disease Ontology Working Group. The Sickle Cell Disease Ontology: enabling universal sickle cell-based knowledge representation. *Database* (2019) Vol. 2019: article ID baz118; doi:10.1093/database/baz118

Received 1 May 2019; Revised 4 September 2019; Accepted 5 September 2019

Abstract

Sickle cell disease (SCD) is one of the most common monogenic diseases in humans with multiple phenotypic expressions that can manifest as both acute and chronic complications. Although described more than a century ago, challenges in comprehensive disease management and collaborative research on this disease are compounded by the complex molecular and clinical phenotypes of SCD, environmental and psychosocial factors, limited therapeutic options and ambiguous terminology. This ambiguous terminology has hampered the integration and interoperability of existing SCD knowledge, and SCD research translation. The SCD Ontology (SCDO), which is a community-driven integrative and universal knowledge representation system for SCD, overcomes this issue by providing a controlled vocabulary developed by a group of experts in both SCD and ontology design. SCDO is the first and most comprehensive standardized human- and machine-readable resource that unambiguously represents terminology and concepts about SCD for researchers, patients and clinicians. It is built around the central concept 'hemoglobinopathy', allowing inclusion of non-SCD haemoglobinopathies, such as thalassaemias, which may interfere with or influence SCD phenotypic manifestations. This collaboratively developed ontology constitutes a comprehensive knowledge management system and standardized terminology of various SCD-related factors. The SCDO will promote interoperability of different research datasets, facilitate seamless data sharing and collaborations, including meta-analyses within the SCD community,

and support the development and curation of data-basing and clinical informatics in SCD. **Availability:** Ontology URL https://bioportal.bioontology.org/ontologies/SCDO.

Contact: nicola.mulder@uct.ac.za, ambroise.wonkam@uct.ac.za

Introduction

Haemoglobinopathies in general are the most common monogenic diseases of human, with sickle cell disease (SCD) the most common recessive condition (1), having variable incidence and prevalence across countries (2). SCD is mainly caused by a single-point mutation yielding a single amino acid substitution in the beta-subunit of haemoglobin, the principal oxygen transporter in red blood cells (3). Because of the protective effect of the sickle cell mutation against malaria, SCD has the highest incidence and prevalence in tropical regions, particularly in Sub-Saharan African (SSA) countries, where more than 70% of SCD patients live (4), affecting ~300 000 newborn babies every year (5). SCD affects more than 20 million people globally (6), with associated healthcare cost exceeding a billion US dollars annually (7). While it was previously thought that the SCD mutation arose independently in different regions, recent evidence has suggested a single origin of the sickle cell allele over 7000 years ago (8). The widespread increase in population migration patterns has changed the distribution of SCD frequencies in different countries, making it a global health concern (9). SCD is a chronic disease of variable phenotypes, associated with increased morbidity and mortality, and with limited effective drugs to address the clinical manifestations (10); this is particularly true in the developing world, posing a substantial burden to the healthcare system of affected countries (11). Fortunately, over the past decade, there have been numerous promising clinical trials that address all aspects of potential therapy based on sickle cell pathophysiology: from haemoglobin sickling (12), to endothelial dysfunction (13), to oxidative stress (14), to transplant (15, 16) and to gene therapy (17).

In the last decades, SCD researchers have utilized large datasets and biomedical knowledge discovery to develop a better understanding of the molecular mechanisms of the disease (18). However, a model that can incorporate results from basic biological research into clinical decision-making processes is still needed (19). Most scientific knowledge is still held in natural language text that is generally unstructured, ambiguous and subjective. With the constant evolution and growing complexity of biomedical knowledge (20), a system is needed for standardized and well-defined knowledge representation. This standard knowledge representation system should enable efficient and reliable exchange of information and integration of new knowledge by (i) more concisely defining SCD concepts, (ii) ensuring common understanding amongst scientists and clinicians and (iii) fostering more efficient data integration and interoperability with other existing systems.

Ontologies are commonly used in biomedical research to represent knowledge in a given domain, in a structured and computable format (21). An ontology defines a formal/standardized common vocabulary and relations that exist in a domain and provides an unambiguous means of communication amongst humans and between humans and computers (22). Several biomedical ontologies exist that have captured knowledge that cover some aspects of SCD, such as the Human Phenotype Ontology (HPO), which describes abnormal phenotypes encountered in human diseases (23), and the Human Disease Ontology (DO), which describes various disease domains (24). However, these ontologies do not capture all the elements that are specific to SCD, including more granular phenotypes and therapeutics that are specific to this disease state. This indicates that no ontology presently exists that can serve as a definitive and comprehensive source of SCD knowledge.

We have constructed an SCD ontology (SCDO), an integrative and universal knowledge representation system for SCD that was collaboratively developed by domain experts, including clinicians, basic scientists, bioinformaticians and data scientists. The SCDO provides a consistent vocabulary that describes key concepts and properties that establish hierarchical relationships between concepts and axioms (evidence or truths). These data are collated into a human- and machine-readable format in order to help process, reuse and reapply knowledge in biomedical research and in healthcare systems. SCDO is built around the key component, haemoglobinopathy, by linking it to phenotypes, modes of inheritance, therapeutics, diagnostics, diseases and other environmental and behavioural information for patients (personal attributes, quality of life and care). The comprehensiveness in content and structure of the SCDO makes it a model for other disease-specific ontologies.

Methods

Ontology development process

The development of the SCDO was a collaborative and interactive process that included three workshops with

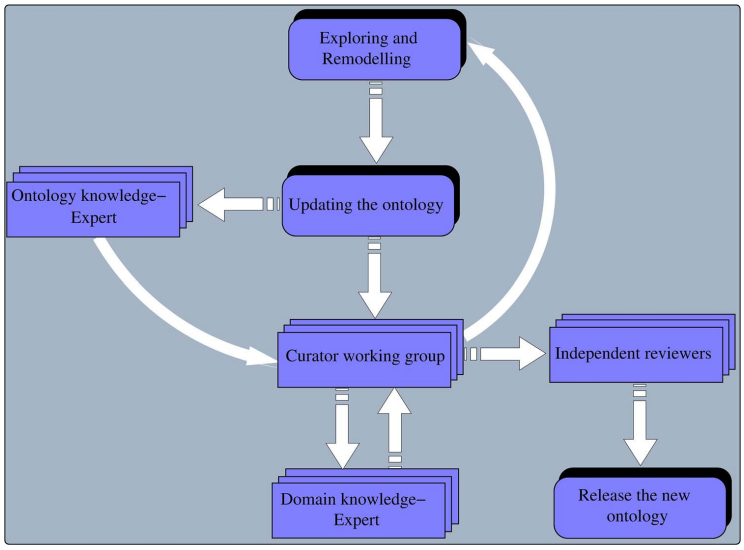


Figure 1. Overview of different steps in the remodelling, reviewing and release of the SCDO by curators, domain and ontology experts and independent reviewers.

subject matter experts from diverse backgrounds, including SCD and ontology experts, geneticists, adult and paediatric clinicians, specialists in organ systems involved in SCD, bioinformaticians and social and data scientists. At each of the annual workshops organized over 3 years, participants were split into five groups according to their expertise (phenotype, diagnostics, therapeutics, quality of life and care, disease modifiers) and each group contained 8 to 12 contributors. In between each workshop, follow-up sessions were organized in subgroups, either online or face-to-face, to continue reviewing the SCDO following steps shown in Figure 1. The work produced at each workshop was submitted to specific curation and ontology teams for further curation.

First workshop activity and outcome

The aim of the first workshop was to (i) develop competency questions that could be addressed by the SCDO and (ii) extract the initial concepts to be included in the SCDO. The initial set of SCDO concepts was retrieved from existing ontologies and knowledge resources, including the HPO, the DO and the Online Mendelian Inheritance in Man (OMIM) database (25), as well as Genotype Ontology (GENO) at https://www.ebi.ac.uk/ols/ontologies/geno. The second set of concepts was manually extracted by SCDO curators from existing SCD guidelines and standards of care (Ghana, Nigeria, Tanzania, Jamaica, UK, Europe, Canada and USA Sickle Cell Disease Management Guidelines), as well as from literature. Different SCDO classes were merged in a spreadsheet and reviewed in two successive steps by a review team consisting of curators and SCD domain experts, including researchers and clinicians. In addition, an ontology team, which consisted of members from the European Bioinformatics Institute (EBI), Oregon State University and the SCDO team, ensured that best practices, such as OBO Foundry principles, were followed and the OWL file was produced and uploaded to BioPortal. The dynamic and iterative ontology development process is shown in Figure 1. As an outcome of this workshop, an initial list of concepts was created and further reviewed and refined by the curation team to further refine the data (26).

Second workshop activity and outcome

After the curation team reviewed, cleaned up and fleshed out the original list of concepts to be included in the SCDO, a second workshop was convened, during which attendees reviewed the current version of the SCDO and further refined the classes and definitions. Annotation and object properties were also reviewed, and additional relation-

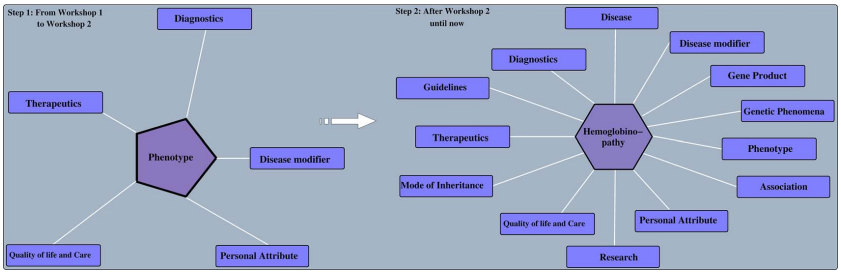


Figure 2. SCDO evolution before and after the second workshop. Before the ontology was built around the 'Phenotype' class and after, the 'hemoglobinopathy' class became the central class.

ships between classes were made using the list of approved object properties. This process led to the identification of new classes that needed to be included in the ontology, probably the most notable being the 'SCD Genotype' and 'SCD Causal Mutation' classes. Discussions around the structuring of the haemoglobinopathy class and naming of haemoglobinopathies and SCD genotypes highlighted subtle differences between concepts and how ambiguous and even misleading some of the names are, thus underscoring the difficulty and importance of the SCDO's endeavour to define these concepts.

Third workshop activity and outcome

The third workshop was convened to conduct final edits on the ontology terms, with different working groups tackling the various classes. Each group had a curator who edited the terms as they were discussed. The groups were tasked with accepting the term label and text definition. Appropriate modifications were made based on the review. Some terms and relationships were added to the existing ontology; others were reclassified upon review. These classes included 'Research' (which includes 'Ethnolinguistics' as a subclass), 'Mode of Inheritance', 'Association', 'Disease' and 'Gene'. Discussions after the workshop led to the 'Gene' class being replaced by 'Genetic Phenomena' and 'Gene Product' and to the inclusion of the upper level 'Guidelines' class.

Refinement and evolution of the ontology

An ontology is always dynamic and evolving as new domain knowledge discovered is added. This results in changes within an ontology, depending on the nature of transformations applied to ontology objects. These transformations may consist of updating or removing an old object or adding a novel object, leading to ontology extension, refinement and enrichment (27). The SCDO has undergone significant enrichment, in which the central concept 'hemoglobinopathy', which includes SCD, has been linked to phenotypes, diagnostics, therapeutics, disease modifiers, modes of inheritance, SCD-related diseases, genetic phenomena and gene product, as well as other environmental data (personal attribute, quality of life and care for patients and research), as described in Figure 2. Properties and axioms were built mostly around this central concept in connection to other SCDO concepts, as illustrated in Figure 3.

Results

The SCDO describes the 'hemoglobinopathy' class as a key aspect linking the various classes through SCDO axioms and properties as illustrated in Figure 3. This has enabled the incorporation of other haemoglobinopathies, which are related to SCD, to be included into the SCDO. These include, but are not limited to, thalassaemias and other haemoglobinopathies, which may interact or even interfere with the phenotypic manifestation of SCD.

Current states of specific SCDO classes

The current SCDO has 1477 well-described terms of which 300 are terms specific to SCD and not defined previously in existing ontologies, and 1177 were retrieved terms from other ontologies (see online supplementary material for Table S1). These terms are categorized based on the upper-level classes shown in Figure 2 above, with the number of subclasses in each upper-level class shown in Figure 4. They are topologically linked by 1676 associations.

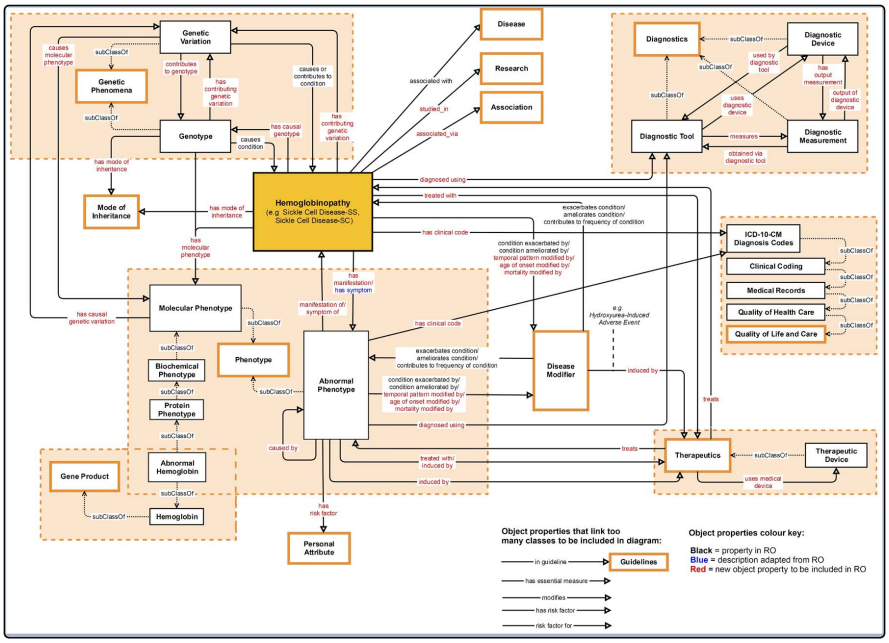


Figure 3. Association between the central class 'hemoglobinopathy' and other upper-level classes (close to the root of the ontology) in the SCDO. Properties with an asterisk also link other classes within the ontology, but these could not be shown in detail here.

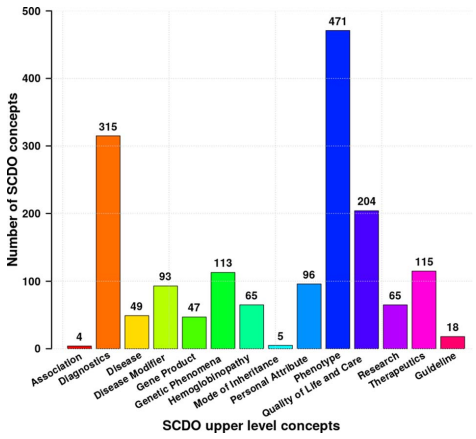


Figure 4. Distribution of different SCDO concepts per upper-level class with the number of associated terms in the ontology (see Supplementary Section 1 for more details about these classes).

It is worth mentioning a few notable changes to some terms suggested by SCD experts. For example, the term 'beta plus thalassemia' has been replaced with 'beta minus thalassemia' based on the description of the concept 'beta-zero' (See Supplementary File Section 1 for more details). Table 1 provides some new terms unique to the SCDO in the phenotype class for illustration, highlighting the SCDO contribution to advancing existing knowledge and expanding scientific content in this field.

Evaluation of the ontology

Generally, data-driven quality evaluation of a given knowledge-based system, such as an ontology, is based on its performance in associated applications, for which there is a need for an independent specification against which the ontology should be assessed. However, since SCDO is only in its infancy, instead, we used rules and questions sketched by SCD experts, referred to as competency questions, to check whether the ontology addresses its scope by ensuring that it contains appropriate information to satisfy these rules or answer these questions.

Table 1. Example terms specific to the SCDO phenotype upper-level class that may also be important for other ontologies, such as HPO and DO

SCDO ID	SCDO term	Definitions or description
SCDO:0000162	Breakthrough pain	Originally used to describe patients with cancer pain who were maintained on a stable dose of analgesics, breakthrough pain was defined as a flair-up of sudden pain unresponsive to usual therapy. Such a flare-up is usually sudden and incidental and can last from a few seconds to a few hours. There are currently no data that clearly describe or can be used to define breakthrough pain in SCD.
SCDO:0006461	Acutely severe anemia	Aplasia or haemolysis may be precipitated by another illness/infection in patients with sickle cell disease. Acutely severe anaemia is defined as Hb < 5 g/dl or a recent acute drop in HB > 2 g/dl below the individual's steady state value. If no steady state value is available, it can be detected by the presence of acutely symptomatic anaemia (i.e. tachycardia, cardiac failure, shock).
SCDO:0007900	Embryonic hemoglobin	The type of haemoglobin present within an embryo in the first 8 weeks of gestation. Two haemoglobins, Gower 1 and Gower 2, are found in embryos of up to 8 weeks of gestation, and Haemoglobin Portland is a third normal embryonic haemoglobin found at lower levels in an embryo.
SCDO:0007161	Acute splenic sequestration crisis	Significant change in blood picture characterized by a precipitous fall in the haemoglobin level of at least 2 g/dl and accompanied by a rapidly enlarging spleen or liver (greater than 2 cm from the steady state level) and reticulocytosis above the steady state level for each individual patient. Signs of acute circulatory insufficiency, such as tachypnoea, tachycardia and hypotension, may or may not be present. It is the earliest life-threatening complication seen in patients with SCD besides pneumococcal infections.
SCDO:0006847	Aplastic crisis	An acute form of acquired red cell aplasia. A significant change in blood picture is observed, characterized by a precipitous fall in the haemoglobin level (>2 g/dl beyond steady state level) and reduced (<1%) or absent reticulocytes in the peripheral blood. The total white blood cell or platelet counts may or may not be affected. In addition, there is no significant increase in the unconjugated fraction of serum bilirubin.
SCDO:0009664	Abdominal vaso-occlusive crisis	Abdominal distension with generalized abdominal tenderness (no rebound tenderness) and reduced bowel sounds. Abdomen moves with respiration. Vomiting/diarrhoea is not common. It is thought to occur secondary to the occlusion of mesenteric vessels.
SCDO:0000812	Non-specific acute lower respiratory tract episode	Includes acute respiratory episodes with lower respiratory tract signs that do not meet the criteria for other diagnoses. May include episodes that would have been diagnosed as ACS were radiographic facilities available.
SCDO:0008625	Right upper quadrant syndrome	Characterized by pain and discomfort in the right upper quadrant (RUQ) of the abdomen caused by a number of possible aetiologies in sickle cell disease. Causes of RUQ pain may be divided into pain originating from the liver or gall bladder versus other origins of abdominal pain in that region.
SCDO:0006909	Hyperhemolytic Crisis	Significant change in blood picture characterized by a precipitous fall in the haemoglobin level associated with jaundice, marked reticulocytosis and polychromasia on the blood smear, increased unconjugated hyperbilirubinemia and increased urobilinogen content in urine above the steady state level for each individual patient.
SCDO:0002039	Zinc deficiency	A deficiency of the essential metal zinc; an essential cofactor for many enzymes. Zinc deficiency is caused by a lack of zinc in the diet, loss of zinc after absorption, for example through loss through burns, inability to absorb zinc or increased loss through exercise.
SCDO:0008623	Vaso-occlusion	The obstruction of blood vessels by altered erythrocytes that can result in pain, anaemia and tissue ischemia.
SCDO:0004888	Functional hyposplenism	An acquired disorder caused by several haematological and immunological diseases and characterized by impairment of splenic function.
SCDO:0006835	Normal hemoglobin	Haemoglobins that present no inherited health condition phenotype susceptible to undergo alterations in the red blood cells.

(Continued)

Table 1. Continued

SCDO ID	SCDO term	Definitions or description
SCDO:0004187	Hemolytic crisis	May be caused by an acute VOC, malarial infection or oxidant drug exposure in individuals with concomitant glucose-6-phosphate dehydrogenase (G6PD) deficiency. Haemolytic crisis may be distinguished from aplastic crisis by the finding of a reticulocytosis as opposed to a reticulocytopenia.
SCDO:0001229	Vaso-occlusive crisis	Pain resulting from tissue ischemia as a result of blockage of blood vessels, occurring in a variety of vascular beds, but most commonly in the bone or bone marrow and requiring analgesic medication.
SCDO:0004576	Chronic hypersplenism	Chronic splenic sequestration associated with enlarged spleen and cytopenia with anaemia and reduction in white blood cells and platelets. The anaemia is usually chronic in nature and patients rarely present with signs of heart failure.
SCDO:0000225	Chronic complications of sickle cell disease	A condition that co-exists or follows from sickle cell disease and that has a slow, creeping onset, slow progress and long continuance of disease manifestations.
SCDO:0000895	Phenotype of sickle cell disease	A (combination of) quality(ies) of some or all sickle cell disease individuals, determined by the interaction of the genetic make-up of these individuals (with regard to sickle cell disease) and their environment.
SCDO:0001135	Acute sickle cell crisis	Refers to a worsening, over a short period of time, of the symptoms and signs of SCD; usually associated with pain and/or shortage of blood (anaemia). Can be suspected in a person with sickle cell disease who presents with a sudden onset of pain, infection, anaemia or other symptoms such as stroke or priapism. Acute pain frequently occurs spontaneously, but may be precipitated by infections, skin cooling, dehydration or stress.
SCDO:0001023	SCD related pain	Pain resulting from the presence of sickle cell disease (SCD). Such pain can be acute, chronic or a mixture of the two.
SCDO:0000233	Chronic sickle cell pain	Pain that does not resolve and lasts for more than 3 months.
SCDO:0001604	Steady state	This is a period when the patient with sickle cell anaemia is free of infection, pain or other disease processes.
SCDO:1000061	Altered level of normal hemoglobin present in SCD	An altered level of normal haemoglobin (Haemoglobin A (Hb A), Haemoglobin A2 (Hb A2) or Haemoglobin F (Hb F, fatal haemoglobin) in the blood, which may be seen in those suffering with sickle cell disease (SCD).
SCDO:0000007	Abnormal hemoglobin structure present in SCD	A structurally abnormal haemoglobin that occurs in one or more forms of sickle cell disease (SCD).

Competency questions (see Supplementary File Section 4) were related to SCD signs, symptoms and complications, focusing specifically on pain, skin, chest infection, stroke and kidney disease to explore potential connections for knowledge discovery. To enable SCDO to answer these questions and its structure to contain different rules and constraints, we have modelled SCDO in such a way that it contains concepts specific to SCD (see Table 1), appropriate relations and properties or axioms as shown in Figure 5 and illustrated in the SCDO schema (Figure 3). We expect that SCDO can serve as a template for knowledge acquisition and reuse and is applicable to future SCD research applications, specifying a standardized common vocabulary verified by SCD experts from diverse specialized backgrounds.

SCDO release and licence

SCDO is released every 2 months with possible special releases when there are significant incidental changes. It is freely available under the Creative Commons Attribution 4.0 Unported License (CC:https://creativecommons.org/licenses/by/4.0/legalcode) and further copyrighted to maintain the quality and integrity of the vocabularies, meaning that any modification to the SCDO can only be done by SCDO developers and curators.

Accessing the ontology

A website was built to serve the SCDO at http://scdontology.h3abionet.org, hosted on servers at the Pasteur Institute in Tunis (http://tesla.pasteur.tn). To facilitate the viewing and searching of terms, the Ontology Lookup Service (OLS) auto-complete widget was installed, and its associated search bar was placed on the main page. A global comment system using Disqus was integrated under each term's web page, to enable discussion and to connect conversations across the website. Search and viewing mechanisms were also integrated in the EBI OLS, which is directly accessible via https://www.ebi.ac.uk/ols/

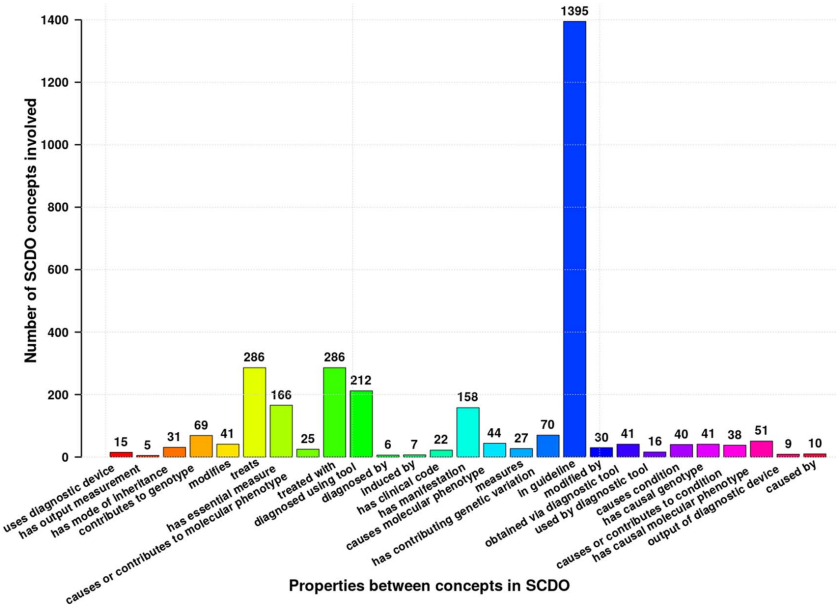


Figure 5. Different properties and axioms defined between different SCDO concepts to satisfy rules set by SCD experts or to answer competency questions. Numbers at the top of bars represent the occurrence frequency of the association in the ontology. *Authors: Sickle Cell Disease Ontology Working Group.

ontologies/scdo. The current OWL and OBO files are accessible via the GitHub repository (https://github.com/scdodev/scdo-ontology).

Discussion

SCDO is the first and most comprehensive standardized human- and machine-readable resource that unambiguously represents terminology and concepts about SCD and other haemoglobinopathies for researchers, patients and clinicians. The SCDO was derived through an exhaustive, iterative and collaborative process drawing on expertise of individuals across multiple disciplines worldwide. Prior to its development, existing ontologies included items relevant to SCD, but none of these ontologies fully encompassed all elements of this uniquely complex and frequent genetic disease. The utilization of the SCDO can have far-reaching outcomes in areas where the burden of disease is greatest, narrowing the gap of evidence-based SCD care management.

SCDO and other disease ontologies

Disease ontologies provide coverage of disease and disorder domains and help in disease annotations in different biomedical applications, supporting clinical and research applications, including clinical data aggregation of electronic health records, clinical decision processes and literature-based mining. In most cases, existing disease-related ontologies (http://www.obofoundry.org/) attempt to describe global disease domains and abnormal phenotypes encountered in disease conditions, including the DO (24), the HPO (23), the Mouse Pathology Ontology (MPATH—http://www.pathbase.net/), the Mammalian Phenotype Ontology (MPO—http://obofoundry.org/ontology/mp.html) (26) and Phenotype And Trait Ontology (PATO—http://obofoundry.org/ontology/pato). As pointed out previously, these ontologies do not capture concepts specific or unique to SCD. The SCDO team is submitting terms and properties for inclusion in existing ontologies including the HPO, the OBO Relations Ontology (RO) (27) and the Data Usage Ontology (DUO) (http://www.ontobee.org/ontology/DUO).

There are specialized and disease-specific ontology projects, such as the Infectious Disease Ontology (IDO—http://www.infectiousdiseaseontology.org/), which aims at providing a set of interoperable ontologies that should cover entities related to specific pathogens and diseases, including human immunodeficiency virus (HIV), dengue fever, influenza, malaria and tuberculosis. Another

hematology-related ontology project is the Blood ontology (BLO—http://mba.eci.ufmg.br/BLO/, http://bioportal.bioontology.org/projects/Blood_Ontology), integrating different existing blood terminologies. More recently, an ontology of Type 2 diabetes (http://purl.bioontology.org/ontology/DIAB), DIAB, has been developed around Phenotype (28) as have other existing disease-specific ontologies. However, there is neither deterministic nor systematic mapping between phenotype (manifestation) and disease (condition), except for some specific 'pathognomonic' manifestations. Thus, for consistency, SCDO is built around the 'hemoglobinopathy' concept, which is then linked to other disease-related aspects, such as diagnostics, therapeutics and phenotypes, as well as several other aspects pertinent to SCD research, patient outcomes and care, including Personal Attribute with Ethnolinguistic groups, Quality of Life and Care, Guidelines and Disease Modifier, which may orient phenotypic manifestations and inform clinical management. This suggests that the SCDO development approach may constitute a model that can drive implementation of other disease-specific ontologies.

Challenges encountered in the development

The main aim is to provide an ontology that will conceptualize SCD disease management and research domains and be effectively applicable across a range of biomedical applications (29). Designing such an ontology is a tedious and daunting process, requiring expertise from varying specialized backgrounds, including geneticists, adult and paediatric clinicians, specialists in organ systems involved in SCD, biologists, philosophers, anthropologists, ontologists and data scientists. These experts should share a common understanding of existing SCD knowledge, make domain assumptions explicit, discuss the scope of the model through competency questions and define SCD concepts, relations and other axioms to be included (30). However, it was not always easy to converge to a single conceptualization of a domain, and this process was particularly time-consuming as each expert conceptualizes the domain depending on how he/she came to understand it, which is often related to particular experiences.

It is known that an ontology is never complete and always dynamic, and we expect the SCDO to continue to evolve for many years considering the context in which it has been developed. Currently, the SickleInAfrica consortium (https://www.sickleinafrica.org), which includes the Sickle Pan-African Research Consortium (SPARCo), Sickle Africa Data Coordinating Center (SADaCC) and Sickle Pan African Network (SPAN), intends to collaboratively collect large-scale SCD patient datasets, using the SCDO to harmonize these datasets and facilitate data integration, information retrieval and analysis. SCDO should be able to handle the challenges of a potential exponential growth of SCD datasets by keeping SCD knowledge updated, possibly on a daily basis, revealing gaps in the existing knowledge and identifying new hypotheses and research questions. To achieve this, some of the processes involved in ontology update and evolution need to be automated to minimize time and resources invested by curators.

Limitations

Despite the successful international collaborative effort to develop this SCDO, there were some notable limitations. The absence of experts from South Asia (i.e. India) leaves a void of characteristics that might be specific to that region, which has one of the highest prevalence rates of SCD in the world (11). Furthermore, the SCDO in its current iteration is available only in the English language, limiting its applicability to regions where the majority of individuals are non-English speakers. While constructing SCDO has been a successful effort, it has yet to be applied; hence, the full strength of its utility is unknown. Some efforts are in progress to mitigate these limitations.

Future directions

Future efforts include translation of the ontology into French and Portuguese, to make it more accessible to native French and Portuguese speakers in Africa and beyond. In addition, we aim to create a lay person version of the SCDO to make terms accessible to non-medical experts, as has been previously done with the HPO (31). Offering an SCDO layperson-friendly version enables the use of SCDO in patient data collection forms, allowing patients to perform standardized self-phenotyping.

To enable semantic interoperability with other ontologies, future efforts will be made to add logical definitions or equivalence axioms that utilize other OBO Foundry ontologies (32). This will allow for machine readable definitions and allow automated reasoning and inferencing. Moreover, the annotations of SCDO terms will continue to be enriched in numerous ways to ensure a high coverage of existing and future SCD knowledge. An example would be to include terms for measurement units into the ontology and to link diagnostic measurements to their relevant units and information about the measurement normal ranges, amongst many other aspects that will enrich SCDO terms.

An SCDO application is already under way. Using the SCDO design, an SCD-based case report form (SCD-CRF) has been developed. The SCD-CRF provides a standardized tool that can be utilized and adapted as needed in multinational cohort studies. Because it is linked to and derived

from the SCDO, studies utilizing the SCD-CRF will capture data more uniformly in comparison to prior studies for which tools were created *ad hoc* for a single study. Currently, research in rare disease has been hampered by a lack of standardization across studies with individual research groups employing definitions for exposures, measures and outcomes that may differ substantially. Developing a knowledge base and data capture tools for a specific disease will impose greater rigor on these cohort studies as well as better inter-study comparability if standardized, agreed-upon definitions are used. The process herein described for the SCDO offers a model approach for the construction of other genetic disease-specific ontologies.

Conclusion

The SCDO was created via a collaborative and iterative process, and this first SCDO release provides a comprehensive description of clinically relevant aspects of SCD, standardizing common SCD vocabulary. This will facilitate seamless data sharing and collaborations including meta-analysis within the SCD community and support the development and curation of data-basing and clinical informatics in SCD. The ontology will continue to be developed by the SCD community using best practices and guidelines. We hope that the SCDO will prove to be a valuable resource for researchers, clinicians, patients and anyone affected by SCD and facilitate the global unification of SCD knowledge. Currently, the SCDO represents the most comprehensive compendium in the SCD field to our knowledge. We anticipate that the SCDO will drive the expansion of scientific content in this field and enhance information structuring, searching and retrieval and can lead to new hypotheses and discoveries.

Authors' contributions

1. Project leads: N.M., A.W.
2. Project coordinator: V.N.
3. SCDO key curators: J.H., A.G., K.M.
4. SCDO Ontologists and Developers: G.K.M., S.J., K.G., C.B.H., M.H.
5. SCDO Content Providers: K.O.-F., all working group chairs and members
6. SCDO Working Group Chairs: K.A., A.C., F.C., C.C.-L., N.H., J.K.-M., D.M., O.N., B.T., M.T., C.R., S.O.-A., K.O.-F.

Supplementary data

Supplementary data are available at *Database* Online.

Acknowledgements

The development of the SCDO is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number U24HL135600; National Institutes of Health Common Fund grants U24HG006941 (H3ABioNet) and U01HG009716 and Wellcome Trust/AESA Ref: H3A/18/001 and 1U54HG009790-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest. None declared.

Sickle Cell Disease Ontology Working Group in alphabetical order

Adekunle Adekile[1], Kofi A. Anie[2], Cherif Ben Hamda[3], Biobele Brown[4], Daima Bukini[5], Andrew Campbell[6], Melek Chaouch[3], Emile Chimusa[7], Catherine Chunda-Liyoka[8], Jemima Dennis-Antwi[9], Vimal K. Derebail[10], Miriam Flor-Park[11], Amy Geard[12,13], Kais Ghedira[3], Melissa Haendel[14], Neil A. Hanchard[15], Jade Hotchkiss[7], Mario Jonas[7], Muntaser Ibrahim[16], Clair Ingram[7], Baba Inusa[17], Adijat Ozohu Jimoh[18], Simon Jupp[19], Karen Kamga[7,20], Zainab Abimbola Kashim[18], Jennifer Knight-Madden[21], Guida Landouré[22,23], Philomene Lopez-Sall[24], Julie Makani[5], Leonard Malasa[5], Tshepiso Masekoameng[7], Gaston Mazandu[7], Khuthala Mnika[7], Nicola Mulder[25], Nchangwi Syntia Munung[26], Deogratias Munube[27], Liberata Mwita[5], Victoria Nembaware[7], Obiageli Nnodu[28], Solomon Ofori-Acquah[29,30], Kwaku Ohene-Frempong[31], Alex Osei-Akoto[32], Vivian Paintsil[33], Sumir Panji[25], Mohamed Cherif Rahimy[34], Charmaine Royal[35], Raphael Z Sangeda[36], Bamidele Tayo[37], Ines Tiouiri[3], Furahini Tluway[5], Marsha Treadwell[38], Leon Tshilolo[39], Nicole Vasilevsky[40], Kasadhakawo Musa Waiswa[41], Ambroise Wonkam[7]

1. Kuwait University, Kuwait City, Kuwait.
2. Haematology and Sickle Cell Centre, London North West University Healthcare NHS Trust & Imperial College London, London, United Kingdom
3. Laboratory of Bioinformatics, Biomathematics and Biostatistics (LR16IPT09), Institute Pasteur of Tunis, University of Tunis El Manar, Tunis, Tunisia
4. Department of Paediatrics, College of Medicine, University of Ibadan, Ibadan, Nigeria
5. Sickle Cell Programme, Department of Haematology and Blood Transfusion, Muhimbili University of Health and Allied Sciences, Tanzania
6. Division of Hematology, Children's National Medical Center, George Washington School of Medicine and Health Sciences, Washington, DC, USA
7. Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
8. University Teaching Hospitals - Children's Hospital, University of Zambia, Lusaka, Zambia
9. Sickle Cell Foundation of Ghana. East Legon-Accra, Ghana
10. UNC Kidney Center, Division of Nephrology and Hypertension, University of North Carolina at Chapel Hill
11. Onco Hematology Unit, Instituto da Criança, Hospital das Clínicas, Universidade de São Paulo, Brazil
12. Gene Transfer Technology Group, UCL EGA Institute for Women's Health, University College London, London, UK
13. University of the Witwatersrand, School of Pathology, Antiviral Gene Therapy Research Unit, Health Sciences Faculty, South Africa
14. Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA

15. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
16. Institute of Endemic Diseases, - University of Khartoum, Sudan
17. Department of Paediatric Haematology, Evelina Children's Hospital, Guy's and St Thomas NHS Trust, London
18. Department of Genetics, Genomics and Bioinformatics - National Biotechnology Development Agency (NABDA), Abuja, Nigeria
19. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, United Kingdom
20. FMBS, University of Yaounde
21. The Sickle Cell Unit, Caribbean Institute for Health Research, the University of the West Indies
22. Faculté de Médecine et d'Odontostomatologie, USTTB, Mali
23. Service de Neurologie, CHU du Point "G", Bamako, Mali
24. Department of Pharmacy, Biochemistry Unit, Cheikh Anta Diop University, Dakar, Senegal.
25. Computational Biology Division, IDM, CIDRI-Africa Wellcome Trust Centre, University of Cape Town Faculty of Health Sciences Anzio Road, Observatory, Cape Town, South Africa,
26. Department of Medicine, University of Cape Town, South Africa
27. Department of Paediatrics and Child Health, Makerere University/Mulago National Referral Hospital, Kampala, Uganda
28. Department of Haematology & Blood Transfusion, College of Health Sciences, University of Abuja, Abuja, Nigeria
29. Dean, School of Biomedical and Allied Health Sciences
30. Director, SickleGenAfrica:Sickle Cell Disease Genomics Network of Africa
31. President, Sickle Cell Foundation of Ghana
32. Dept. of Child Health, School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology/Komfo Anokye Teaching Hospital, Kumasi, Ghana
33. Komfo Anokye Teaching Hospital, Kumasi, Ghana
34. Centre de Prise en Charge Medicale Integree du Nourrisson et de la Femme Enceinte atteints de Drepanocytose, (The National Sickle Cell Disease Institute), Cotonou, Republic of Benin
35. Departments of African & African American Studies, Biology, and Community & Family Medicine, Duke University, Durham, USA
36. Department of Pharmaceutical Microbiology, Muhimbili University of Health and Allied Sciences, Tanzania
37. Department of Public Health Sciences, Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA
38. UCSF Benioff Children's Hospital Oakland, Department of Hematology/Oncology 747 52nd Street Oakland, CA USA 94609
39. Centre de Formation et d'Appui sanitaire/MONKOLE, Kinshasa, DR Congo
40. Oregon Clinical & Translational Research Institute, Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, USA
41. Department of Medicine, Makerere University College of Health sciences, University college of Health sciences

References

1. Williams,T.N. and Weatherall,D.J. (2012) World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb. Perspect. Med*, **2**, a011692.
2. Rees,D.C., Williams,T.N. and Gladwin,M.T. (2010) Sickle-cell disease. *Lancet.*, **376**, 2018–2031.
3. Piel,F.B., Patil,A.P., Howes,R.E. *et al.* (2013) Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet.*, **381**, 142–151.
4. Makani,J., Williams,T.N. and Marsh,K. (2007) Sickle cell disease in Africa: burden and research prioritie. *Ann. Trop. Med. Parasitol.*, **101**, 3–14.
5. Diallo,D.A. and Guindo,A. (2014) Sickle cell disease in sub-Saharan Africa: stakes and strategies for control of the disease. *Curr. Opin. Hematol.*, **21**, 210–214.
6. Stephen,N., Nden,N., Gusen,N.J. *et al.* (2018) Prevalence of sickle cell disease among children attending plateau specialist hospital, Jos, Nigeria. *Acta Medica International*, **5**, 20–23.
7. Kauf,T.L., Coates,T.D., Huazhi,L. *et al.* (2009) The cost of health care for children and adults with sickle cell disease. *Am. J. Hematol*, **84**, 323–327.
8. Shriner,D. and Rotimi,C.N. (2018) Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the Holocene wet phase. *Am. J. Hum. Genet.*, **102**, 547–556.
9. Piel,F.B., Hay,S.I., Gupta,S. *et al.* (2013) Global burden of sickle cell anaemia in children under five, 2010-2050: modelling based on demographics, excess mortality, and interventions. *PLoS Med.*, **10**, e1001484.
10. Piel,F.B., Steinberg,M.H. and Rees,D.C. (2017) Sickle cell disease. *N. Engl. J. Med.*, **377**, 305.
11. Makani,J., Ofori-Acquah,S.F., Nnodu,O. *et al.* (2013) Sickle cell disease: new opportunities and challenges in Africa. *Scientific-WorldJournal*, **2013**, 193252.
12. Vichinsky,E., Hoppe,C.C., Ataga,K.I. *et al.* (2019) A phase 3 randomized trial of voxelotor in sickle cell disease. *N. Engl. J. Med*, **381**, 509–519.
13. Ataga,K.I., Kutlar,A., Kanter,J. *et al.* (2017) Crizanlizumab for the prevention of pain crises in sickle cell disease. *N. Engl. J. Med*, **376**, 429–439.
14. Niihara,Y., Miller,S.T., Kanter,J. *et al.* (2018) Investigators of the phase 3 trial of l-glutamine in sickle cell disease. A phase 3 trial of l-glutamine in sickle cell disease. *N. Engl. J. Med*, **379**, 226–235.
15. Bolaños-Meade,J., Cooke,K.R., Gamper,C.J. *et al.* (2019) Effect of increased dose of total body irradiation on graft failure associated with HLA-haploidentical transplantation in patients with severe haemoglobinopathies: a prospective clinical trial. *Lancet Haematol*, **6**, e183–e193.
16. Bernaudin,F., Verlhac,S., Peffault de Latour,R. *et al.* (2019) Association of matched sibling donor hematopoietic stem cell transplantation with transcranial Doppler velocities in children with sickle cell anemia. *JAMA*, **321**, 266–276.
17. Ribeil,J.A., Hacein-Bey-Abina,S., Payen,E. *et al.* (2017) Gene therapy in a patient with sickle cell disease. *N. Engl. J. Med*, **376**, 848–855. doi: 10.1056/NEJMoa1609677.
18. Platt,O.S., Brambilla,D.J., Rosse,W.F. *et al.* (1994) Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N. Engl. J. Med.*, **330**, 1639–1644.
19. Mulder,N., Nembaware,V., Adekile,A. *et al.* (2016) Proceedings of a sickle cell disease ontology workshop - towards the first comprehensive ontology for sickle cell disease. *Appl. Transl. Genomi.*, **9**, 23–29.

20. Pesquita,C. and Couto,F.M. (2012) Predicting the extension of biomedical ontologies. *PLoS Comput. Biol.*, **8**, e1002630.
21. Bodenreider,O., Mitchell,J.A. and McCray,A.T. (2005) Biomedical ontologies. *Pac. Symp. Biocomput.*, 76–78.
22. Noy,NF, Mcguinness,DL. (2001) Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March.
23. Köhler,S., Vasilevsky,N.A., Engelstad,M. *et al.* (2017) The human phenotype ontology in 2017. *Nucleic Acids Res.*, **45**, D865–D876.
24. Kibbe,W.A., Arze,C., Felix,V. *et al.* (2015) Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
25. Amberger,J.S., Bocchini,C.A., Schiettecatte,F. *et al.* (2015) OMIM.org: online Mendelian inheritance in man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
26. Smith,C.L. and Eppig,J.T. (2012) The mammalian phenotype ontology as unifying standard for experimental and high-throughput phenotyping data. *Mamm. Genome*, **23**, 653–668.
27. Smith,B., Ceusters,W., Klagges,B. *et al.* (2005) Relations in biomedical ontologies. *Genome Biol.*, **6**, R46.
28. Vasant,D., Neff,F., Gormanns,P. *et al.* (2015) DIAB: an ontology of type 2 diabetes stages and associated phenotypes. *Proceedings of Phenotype Day ISMB*, **2015**, 24–27.
29. Hoehndorf,R., Schofield,P.N. and Gkoutos,G.V. (2015) The role of ontologies in biological and biomedical research: a functional perspective. *Brief Bioinform.*, **16**, 1069–1080.
30. Aranguren,M.E., Antezana,E., Kuiper,M. and Stevens,R. (2008) Ontology design patterns for bio-ontologies: a case study on the cell cycle ontology. *BMC Bioinformatics*, **9**, S1.
31. Vasilevsky,N.A., Foster,E.D., Engelstad,M.E. *et al.* (2018) Plain-language medical vocabulary for precision diagnosis. *Nat. Genet.*, **50**, 474–476.
32. Smith,B., Ashburner,M., Rosse,C. *et al.* (2007) The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, **25**, 1251–1255.

4. Wonkam, Ambroise et al. “Genetic modifiers of long-term survival in sickle cell anemia.” *Clinical and Translational Medicine* vol. 10,4 e152. 12 Aug. 2020, doi:10.1002/ctm2.152

Authors’ contributions

1. Conceived and designed the experiments: AW, EC, AA.
2. Performed the experiments: AW, **Khuthala Mnika**, GP, VJNB.
3. Patient recruitment, samples and clinical data collection and pro- cessing: AW, VJNB, **Khuthala Mnika**, GP.
4. Analyzed the data: AW, EC**, Khuthala Mnika**, GP, AA.
5. Contributed reagents/materials/analysis tools: AW, EC, NM, DS, CNR.
6. Wrote the paper: AW, EC.
7. Revised and approved the manuscript: AW, AA, EC, **Khuthala Mnika**, VJNB, NM, DS, CNR.

Received: 10 July 2020 | Revised: 30 July 2020 | Accepted: 31 July 2020 | Published online: 10 August 2020

DOI: 10.1002/ctm2.152

RESEARCH ARTICLE

CLINICAL AND TRANSLATIONAL MEDICINE Open Access

WILEY

Genetic modifiers of long-term survival in sickle cell anemia

Ambroise Wonkam[1,2] | **Emile R. Chimusa**[1,2] | **Khuthala Mnika**[1] | **Gift Dineo Pule**[1] | **Valentina Josiane Ngo Bitoungui**[1] | **Nicola Mulder**[3] | **Daniel Shriner**[4] | **Charles N. Rotimi**[4] | **Adebowale Adeyemo**[4]

[1] Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[2] Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health SciencesUniversity of Cape Town, Cape Town, South Africa

[3] Computational Biology Division, Department of Integrative Biomedical Sciences, Institute for Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

[4] Center for Research on Genomics and Global Health, National Human Genome Research InstituteNational Institutes of Health, Bethesda, Maryland, USA

Correspondence
Ambroise Wonkam, MD, PhD, Division of Human Genetics, Department of Medicine, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, Cape Town, South Africa.
Email: ambroise.wonkam@uct.ac.za

Funding information
NIH Clinical Center, Grant/Award Number: U24-HL-135600; National Human Genome Research Institute, Grant/Award Number: 1U01HG007459-01; National Institutes of Health, Grant/Award Number: 1ZIAHG200362

Abstract

Background: Sickle cell anemia (SCA) is a clinically heterogeneous, monogenic disorder. Medical care has less-than-optimal impact on clinical outcomes in SCA in Africa due to several factors, including patient accessibility, poor access to resources, and non-availability of specific effective interventions for SCA.

Methods: Against this background, we investigated 192 African participants who underwent whole exome sequencing. Participants included 105 SCA patients spanning variable clinical expression: a "long survivor" group (age over 40 years), a "stroke" group (at least one episode of overt stroke), and a "random" group (patients younger than 40 years without overt cerebrovascular disease). Fifty-eight ethnically matched homozygous hemoglobin A controls were also studied. Findings were validated in an independently recruited sample of 29 SCA patients. Statistical significance of the mutational burden of deleterious and loss-of-function variants per gene against a null model was estimated for each group, and gene-set association tests were conducted to test differences between groups.

Results: In the "long survivor" group, deleterious/loss-of-function variants were enriched in genes including *CLCN6* (a voltage-dependent chloride channel for which rare deleterious variants have been associated with lower blood pressure) and *OGHDL* (important in arginine metabolism, which is a therapeutic target in SCA). In the "stroke" group, significant genes implicated were associated with increased activity of the blood coagulation cascade and increased complement activation, for example, *SERPINC1*, which encodes antithrombin. Oxidative stress and glutamate biosynthesis pathways were enriched in "long survivors" group. Published transcriptomic evidence provides functional support for the role of the identified pathways.

Conclusions: This study provides new gene sets that contribute to variability in clinical expression of SCA. Identified genes and pathways suggest new avenues for other interventions.

KEYWORDS
Africa, genetic modifiers, sickle cell disease, whole exome sequencing

1 | INTRODUCTION

Sickle cell disease (SCD) is a group of blood disorders caused by mutations in *HBB* that promote polymerization of hemoglobin and sickling of erythrocytes. The most common and most clinically severe form of SCD is sickle cell anemia (SCA [MIM: 603903]) and is caused by homozygosity of the sickle mutation HbS. While the estimated age of the sickle mutation is ~7300 years,[1] the mutation has persisted in appreciable frequencies because of the protection that heterozygotes have against malaria. An estimated 305 800 babies are born each year with SCD worldwide, with nearly 75% of the births occurring in sub-Saharan Africa (SSA).[2] Despite this high incidence, most SSA countries lack effective public health programs focused on SCD.[3]

Clinical expression of SCD often shows considerable variation in features such as the severity of anemia, the frequency of painful vaso-occlusive crises, stroke, and mortality. Both genetic and non-genetic factors are known to influence the severity of SCD. For example, high fetal hemoglobin (HbF) levels have long been known to be associated with less severity[4]; HbF levels are under genetic control and are amenable to therapeutic manipulation.[5] Similarly, co-inheritance of α-thalassemia is protective against some SCD-related complications, such as hemolysis, stroke, and kidney disease.[6,7] Availability of appropriate comprehensive medical care for SCD that combines newborn screening, penicillin prophylaxis, screening for stroke risk, transfusion program, and the use of hydroxyurea, in high income countries, usually mitigates morbidity and facilitates longer survival of SCD patients.[8] This is, however, not the case in most of SSA, due to the combined effect of inappropriate care and often severe clinical complications,[7,9] compounded by other factors such as malaria, malnutrition, infectious diseases of childhood, and poverty. Thus, up to 90% of infants with SCD in SSA are believed to die needlessly by 5 years of age.[10]

Different *HBB* (MIM: 141900) haplotypes (Cameroon, Central African Republic, Benin, Senegal, and Arabian/Indian) have also been associated with variable clinical expression of SCD.[2] However, a recent study of sickle *HBB* haplotypes based on sequence data rather than restriction site data led to a reclassification of the classical haplotypes and found sub-structuring of haplotypes that may have confounded previous associations of the haplotypes with clinical severity.[1] To search the genome for protein coding genes that affect disease severity, we utilized a whole exome sequence (WES)-based approach to explore the specific hypothesis that rare gene-associated and function-altering variants with potentially high penetrance are associated with SCA clinical sub-phenotypes,[11] and could lead to the identification of modifiable pathophysiological pathways. We studied two distinct clinical SCA groups, a "long survivor" group and a "stroke" group, which we contrasted with a "random" group to infer critical genes and pathways. We also contrasted the "long survivor" and "stroke" groups as extreme phenotypes that are likely to come from the relatively benign versus the most severe (respectively) ends of the phenotypic spectrum of SCA.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

The study was performed in accordance with the guidelines of the Helsinki Declaration. Ethical approval was given by the National Ethical Committee Ministry of Public Health, Republic of Cameroon (No 033/CNE/DNM/07); and the University of Cape Town, Faculty of Health Sciences Human Research Ethics Committee (HREC RE: 132/2010).

2.2 | Patients and methods

Recruitment for the discovery group was conducted in Cameroon at the Yaoundé Central Hospital and Laquintinie Hospital in Douala, as previously described.[12] Sociodemographic and clinical data including blood counts and hemoglobin (Hb) electrophoresis were obtained on enrolment. The present study was restricted to SCA (ie, HbSS), the most prevalent and severe form of SCD. This study design decision was taken to remove the potential confounding effect of beta-thalassemia and hemoglobin C, both of which are associated with milder forms of SCD. The discovery sample of patients included three groups. (a) A "long survivor group" comprised SCA patients aged over 40 years; this cut-off was based on a life expectancy of 43 years for SCA in the Cooperative Study.[4] (b) A "stroke" group comprised SCA patients with at least one clinical episode of overt stroke, a devastating complication of SCA that is considered to be a proxy of severity[4] and is influenced by genetic modifiers.[13] (c) A "random group"

comprised patients younger than 40 years with no known cerebrovascular disease and randomly selected from among clinically stable patients. No patient was on hydroxyurea, at the time of recruitment. Fifty-eight ethnically matched homozygous HbAA controls were randomly selected from apparently healthy blood donors recruited in Yaoundé.[14] The replication cohort consisted of adult SCA patients (age range 18-51 years, mean age 26.1 years) from the Demographic Republic of Congo (DRC), recruited at the Haematology Clinic, Groote Schuur Hospital in Cape Town, South Africa.

2.3 | Sickle cell disease mutation, β-globin gene cluster haplotypes, and 3.7 kb α-globin gene deletion

DNA was extracted from peripheral blood. Molecular analysis to determine the presence of the sickle mutation was carried out by polymerase chain reaction (PCR), followed by DdeI restriction analysis.[12] Using published primers and methods, five restriction fragment length polymorphism (RFLP) sites in the β-globin gene cluster were amplified to analyze the *HBB* haplotype background.[12] The 3.7 kb α-globin gene deletion was screened using expand-long template PCR.[14]

2.3.1 | Whole exome sequencing

DNA samples underwent sequencing at Omega Bioservices, Omega Bio-tek, Inc, Emory University, USA. The Roche Nimblegen SeqCap EZ MedExome v2.0 (~47 Mb target), which has enhanced coverage of medically relevant genes, was used for sequence capture. Samples passing quality control, library preparation, and exome capture were sequenced on an Illumina HiSeq 4000 sequencer. Read mapping and alignment were performed as described in the Supporting Information Methods. Because different calling methods produce large numbers of differing variants, we adopted an ensemble approach implemented in VariantMetaCaller[15] (see Figure S1 and details in Supporting Information Methods).

2.3.2 | Variant calling quality control, annotation, and prioritization

Joint variant calling was conducted from three independent callers (Figure S1) on each subject group defined above and across all samples (Supporting Information Methods). Before applying the ensemble approach (VariantMetaCaller[15]) across the resulting variant set from three callers (Supporting Information Methods) from each independent subjects' group, we filtered each resulting VCF file using the GATK tool Variant Filtration.[16] Variant filtering procedures and quality control assessment are as described in the Supporting Information Methods. Final call-sets were produced from VariantMetaCaller.[17] We used ANNOVAR to perform gene-based annotation (Supporting Information Methods) on each independent final call-set per subjects' group. First, each resulting functional annotated call set was independently filtered for predicted functional status. We used 21 in silico prediction tools (*SIFT, LRT, MutationTaster, MutationAssessor, FATHMM, fathmm-MKL, RadialSVM, LR, PROVEAN, MetaSVM, MetaLR, CADD, GERP++, DANN, M-CAP, Eigen, GenoCanyon, Polyphen2 HVAR, Polyphen2 HDIV, PhyloP*, and *SiPhy*) to identify variants whose predicted functional status is "deleterious" (D), "probably damaging" (D), "disease_causing_automatic" (A) or "disease_causing" (D). We retained a variant if it had at least 17 predicted functional status D (Supporting Information Methods). Second, the retained variants from each data set were further filtered for rarity, exonic variants, and nonsynonymous mutations, yielding a final candidate list of in silico predicted mutant variants from each subject group. We also reported the aggregated SiPhy score from all identified SNPs within a gene.

2.3.3 | Gene-specific differences in SNP frequencies

To assess gene set differences in frequencies, we first computed SNP-specific allele frequencies. Assuming a population evolving under the Wright-Fisher model under selective neutrality, and with an expected number of mutations, we used a stepwise constant effective population size to (1) compute the allele frequency difference, (2) estimate group pairwise differences, and (3) compute a test statistic for which an excess of large values indicated deviation from the null model (Supporting Information Materials).[18] Second, assuming SNPs in 40 kb upstream and downstream (exon) within a gene are close and possibly in Linkage Disequilibrium (LD), SNP-specific unusual allele frequency summary statistics of SNPs in the defined gene region were aggregated as described in Supporting Information Materials to obtain gene-specific differences in SNP frequencies.

2.3.4 | Burden and rare-variant analyses

Rare variants in a gene or region may influence a phenotype in different directions and with differing magnitude of effect.[11,19,28] Therefore, gene-set analyses were performed to determine the associations between genes and phenotype as defined by the three participant categories. Gene

sets of variants comprised non-synonymous, missense, and stop lost/stop variants (in core *in silico* mutant genes) in the exome datasets and in 40 kb upstream and downstream (exon) within a gene were mapped within the coordinates of the gene in the human genome build hg19. We conducted rare variant analysis using the optimal unified sequence kernel association test (SKAT-O[19]) in the exome dataset for our primary comparisons, namely: (a) Long survivors versus Random SCA (n = 82), (b) Stroke versus Random SCA (n = 79), and (c) Stroke versus Long survivors (n = 59). We also analyzed Stroke versus Controls (n = 81), Long survivors versus Controls (n = 84), and Random SCA versus Controls (n = 114). All analyses used an adaptive minor allele frequency threshold of $1/\sqrt{2n}$ (where n is the number of individuals) for definition of "rare variant."[11,19] A total number of 17 285 gene sets had two or more rare variants and were included in the SKAT-O rare variant analysis The SKAT-O[19] analysis was performed using a linear weighted kernel with a missingness cutoff of 0.9 and adjustments for covariates including age and principal components of population stratification. To account for multiple testing, empirical *P*-values were calculated after 100 000 permutations; therefore, a *P*-value of .05 was considered significant. While permutation testing was used to adjust for multiple testing for each comparison, we also considered the fact that we had three primary hypotheses and adjusting for the number of primary hypotheses led to our considering *P*-values < .017 (ie, .05/3) as significant.

2.3.5 | Network and enrichment analysis

From the gene lists obtained from the analyses of deleterious/loss-of-function variants and unusual allele frequency spectra, we next sought to identify networks of interactions using a comprehensive human protein-protein interaction (PPI) network.[20] We examined how the genes in these networks were associated with phenotypes, pathways, biological processes, and molecular functions. Pathways and networks were drawn from various bioinformatics databases, including KEGG, Panther, Biocarta, Reactome, and the Gene Ontology (GO) Consortium database (Supporting Information Materials). Enrichment analysis was performed using a custom script as well as the DAVID and PANTHER tools.

3 | RESULTS

3.1 | Characterization of participants

Clinical characteristics of the 105 Cameroonian SCA patients are shown in Table 1. The median ages were 44.0, 20.5, and 16.5 years, for the "long survivor," "stroke," and "random" SCA groups, respectively. A total of five of 23 (21.7%) patients had stroke before the age of 16 years. "Long survivor" SCA patients had significantly lower leucocyte counts and health care utilization rates. Two *HBB* haplotypes (Benin and Cameroon) were observed in the sample. The overall distribution of haplotypes was not significantly different among the three groups, although it should be noted that a greater proportion of the "stroke" group carried at least one Cameroon haplotype when compared to the other two groups (38.5% vs 9.1% and 15.4%, Table 1). The distribution of the 3.7 kb α-globin gene deletion genotypes and mean HbF levels did not differ significantly among the three groups.

A total number of 8 458 386 variants were called in the whole exome sequence dataset, of which 80 226 were exonic, distributed as 0.4% stop loss, 0.3% stop gain, 4.5% synonymous, and 94% non-synonymous or splice site variants (Figures 1A,B; Figure S2). Principal component analysis (PCA) of the study sample with the African populations (AFR) from the 1000 Genomes Project showed that the study samples clustered separately from the other Africans (Supporting Information Materials), which is not surprising because samples from Cameroon were not included in the 1000 Genomes Project. PCA plots (Figure 1A; Figure S3) showed no global population differences among the SCA patients and control groups, that is, cases and controls clustered together (Supporting Information Materials). The replication SCA sample from DRC had a median age of 26 years and 13.8% (4/29) had a clinically overt stroke.

3.1.1 | Mutational burden of genes in discovery samples

Among SCA patients in the discovery sample, we detected significant differences in the burden of non-synonymous, function-altering variants (Figure 2A) in a total of 49 genes (Table 2; Tables S1 and S5): 17 genes in the "long survivor" group (Figure 3; Figure S4A), 19 genes in the "stroke" group (Figure 3; Figure S5a), and 39 genes in the "random" group (Figure 3; Figure S6A). Five genes were found only in the "long survivor" group (*ATP2B4 [MIM: 108732], CLCN6 [MIM: 602726], OGDHL [MIM: 617513], ESR2 [MIM: 601663]*, and *SLC7A8 [MIM: 604235]*), and a different set of five genes were found only in the "stroke" group (*SLC22A5 [MIM: 603377], HGF [MIM: 142409], IVD [MIM: 607036], ABCC1 [MIM: 158343]*, and *SNTB1 [MIM: 600026]*) (Table S1). No genes overlapped between the "stroke" and "long survivor" groups while also absent from the "random" group (Table S1; Figure 2A). Six genes (*CPS1 [MIM: 608307]*, *PYGB [MIM: 138550]*, *MARCH10 [MIM: 613337]*, *SLC4A5 [MIM: 609802]*, *NADSYN1 [MIM:*

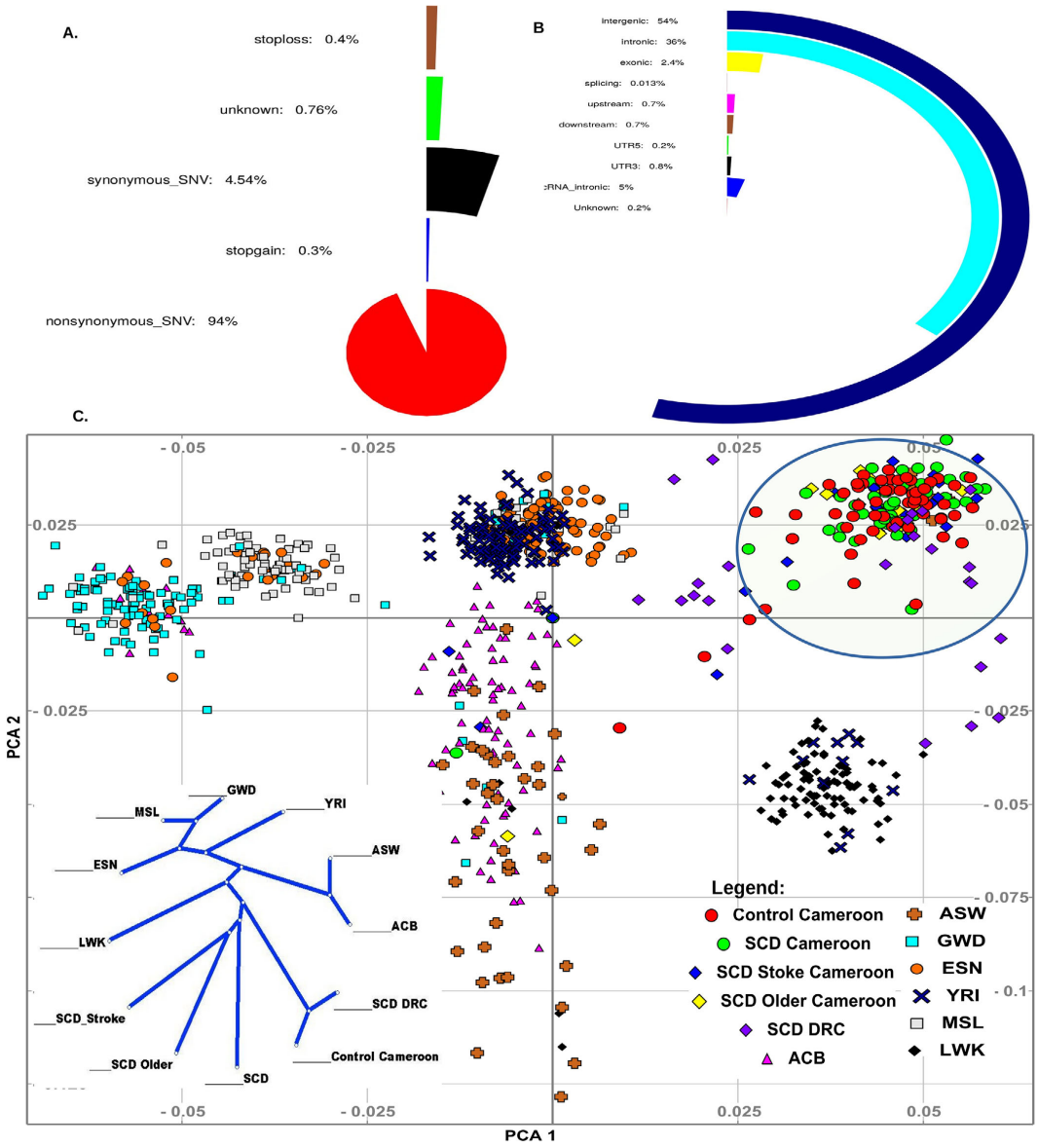


FIGURE 1 SCA exome map characteristics. A), Percentages of functions covered from 80 226 exonic variants. B, Overall percentage of variant functions from 8 458 386 variants in discovery and replication cohorts. C) Principal component analysis (PCA) plot of the three Cameroonian SCA sub-groups ("random," "stroke," and "long survivor"), and African ancestry samples from 1000 Genomes phase 3 release, indicating the Cameroonian SCD patients and controls are relatively homogeneous and from similar background. Date from DRC participants is scattered in the convex of Western and Eastern African populations, supporting a bantu migration route from west to south Africa[21]

TABLE 1 Characteristics of the 105 SCA patients that underwent WES displayed as median (25th to 75th percentiles) or percent (%)

Variables		Random SCA(N = 56)	SCA with overt stroke (N = 23)	Long survivor SCA (N = 26)	*P*-values*
Age (years)		16.5 (9.2-25.7)	20.5 (16.25-25.75)	44 (41-49.5)	**<.0001**
Gender	F/M (50/46)	28/28	12/11	15/12	.854
Haematological indices	RBC (10źš/L)	2.7 (2.2-3.2)	2.9 (2.4-3.5)	3.4 (2.6-3.9)	.101
	Hb (g/dL)	7.8 (7.1-8.8)	8.2 (7.3-8.8)	8.2(6.8-9.8)	.462
	MCV (Fl)	84 (78-92.5)	84 (73.2-93.7)	81 (74-89.5)	.824
	MCHC (g/dL)	33.6 (31.0-36.6)	34.1 (31.7-36.5)	32.8 (30.4-35.1)	.420
	WBC (10^9/L)	12.4 (10.3-36.6)	14.4 (10.3-17.8)	9.4 (8.23-12.2)	**.026**
	Lymphocytes (10^9/L)	5.2 (3.9-6.9)	5.5 (3.5-7.9)	3.9 (2.7-4.8)	**.005**
	Monocytes (10^9/L)	1.3 (1.0-1.9)	1.3 (0.8-2.1)	0.9 (0.8-1.3)	**.022**
	Platelets (10^9/L)	342.5 (289.2-342.5)	359.5 (283-440.5)	329 (228.5-452.5)	.583
	HbA_2 (%)	3.7 (3.2-4.2)	4.0 (3.2-4.8)	3.6 (3.2-4.9)	.325
	HbF (%)	13.1 (2.7-17.3)	10.6 (5.2-13.7)	9.4 (3-14.3)	.231
Clinical events	Vaso-occlusive crisis (n/year)	0.6 (1-5)	2 (1-3)	2 (1-3.5)	.172
	Consultations (n/year)	0.5 (0-5)	1 (0-6)	0 (0-0)	**.030**
	Hospitalization (n/year)	0.3 (0.5-2)	1 (1-2)	0 (0-1.5)	.606
	Blood transfusion (%)	75.9	85.0	70.0	.315
	Stroke (%)	0	100	5.0	**<.0001**
3.7 α-globin gene deletion genotypes	αα / αα	57.8%	58.8%	46.7%	.652
	αα/ α3.7	31.1%	29.4%	26.7%	
	α3.7/ α3.7	11.1%	11.8%	26.6%	
***HBB* Haplotype**[a]	Benin/Benin	90.9%	61.5%	84.6%	.834
	Benin/Cameroon	3.0%	38.5%	7.7%	
	Cameroon/Cameroon	6.1%	0.0%	7.7%	

Abbreviations: RBC, red blood cell counts; Hb, hemoglobin; MCV, mean corpuscular volume; MCHC, mean corpuscular hemoglobin concentration; WBC, white blood cell counts;

[a]Percentage of individuals not chromosomes.

*P-value in three-way comparisons; significant *P*-values are **bolded**.

608285], and *CACNA1H [MIM: 607904]*) were common to all three SCA groups (Table S1). Notably, none of the genes identified among SCA patients (Table S1) were significant in the HbAA control samples.

Among key candidate genes, with known role in SCA-related pathophysiology, only *NADSYN1*, which is involved in glutamine pathways, has significantly common mutations in all three subgroups (Table S1). While *SERPINE1* and *SERPINA1* involved in coagulation pathways, had recurrent mutations in the random group only, but not in the two "extreme" groups. However, gene-set rare-variant association analyses showed significance difference in *NADSYN1* gene between "Stroke" versus "Random" SCA (Table 3).

3.1.2 | Replication of mutational burden of genes in an independent sample

The characteristics of the replication cohort (including clinical events and hematological indices) are shown in Table S2. In the replication exome dataset, we identified 35 genes with mutations with functional impact as previously defined (Table S3). Twelve genes from the discovery analysis were replicated (Figure 2A; Table 2). Three genes (*HGF, SNTB1*, and *SERPINC1 [MIM: 107300]*) were found in the "stroke" group but not in the "long survivor" group (Figure 2A). Conversely, five genes (*CLCN6*, *OGDHL*, *COL6A3 [MIM: 120250]*, *INSR [MIM: 147670]*, and *NOS3 [MIM: 163729]*) were found in the "long survivor" group but not in

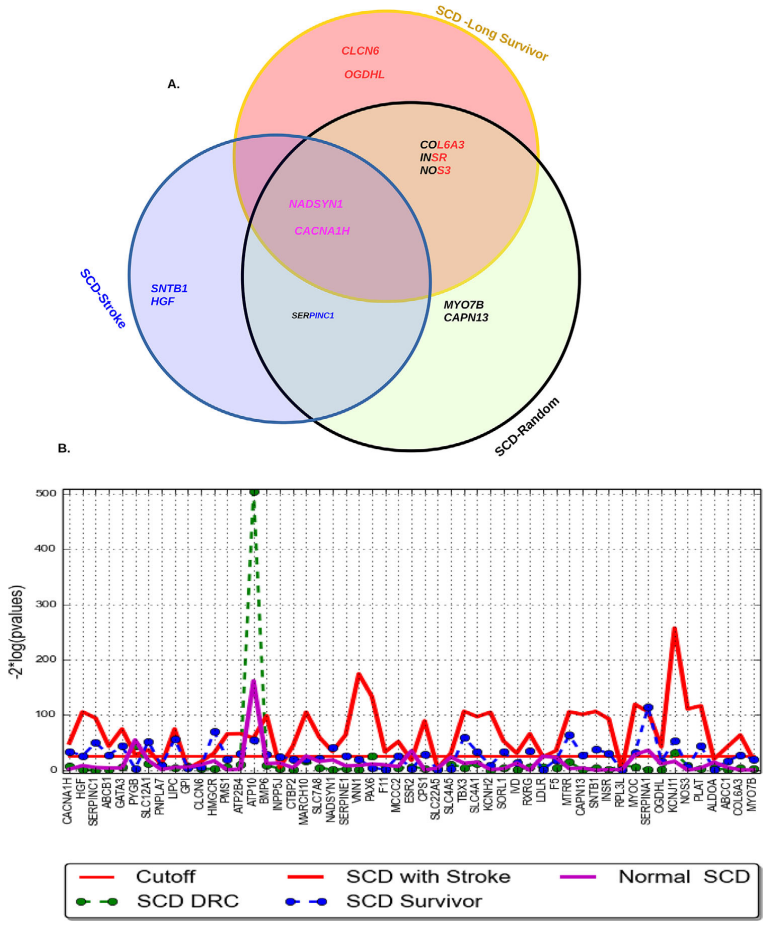


FIGURE 2 Gene mutations in three Cameroonian SCA sub-groups (only the 12 genes replicated in patients from Congo are represented here). A, The Venn diagram shows the overlap of the replicated candidate mutations in the three SCA groups. B, Gene-specific genetic differentiation based on identified 49 mutant genes identified (Table S1), among 58 HbAA Cameroonian controls, each SCA patient sub-group, and the replication cohort of SCA patients from Congo

the "stroke" group. Two genes (*NADSYN1* and *CACNA1H*) were common to all three groups (Figure 2A).

3.1.3 | Pathways and biological processes associated with genes with high mutational burdens

The PPI network formed from 17 genes containing high mutational burdens among the "long survivor" group was enriched for glutamate metabolism ($P = .0035$; Figure S4) and clustered with the fetal liver cell type ($P = .0087$) but had no association with any human disease/disorder. The PPI network of 19 genes in the "stroke" group was enriched for the arginine biosynthesis ($P = .00078$; Figure S5B), showed an association with hypertension ($P = .00781$), and clustered with monocytes ($P = .029$). The set of 39 genes with mutations identified from the "random" group (including the genes in common with the "long survivor" and "stroke" groups) was enriched for complement and coagulation cascades ($P = 1.129 \times 10^{-6}$; Figure S6B), associated with cholesterol level, diabetes mellitus, and thrombophilia ($P = .00103$, .0072, and .0095, respectively), and clustered with the adrenal cortex ($P = .00018$). The three sets of mutations from all 105 Cameroonian SCA patients (Figure 3A; Table S1)

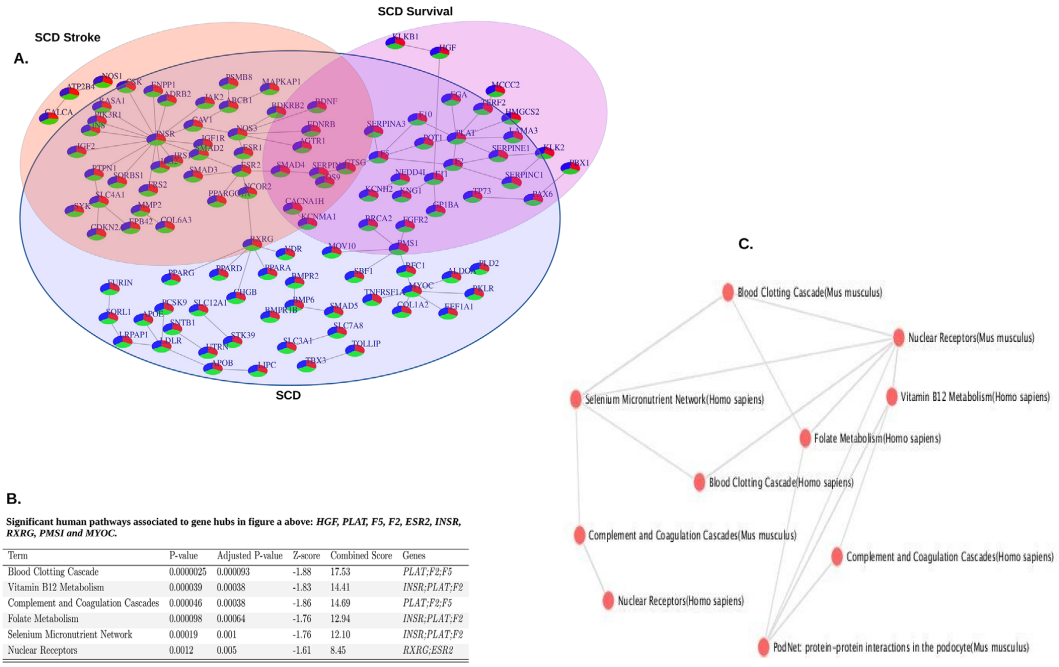
TABLE 2 Genes with high burdens of deleterious and loss-of-function mutations in SCA patients in both discovery and replication samples

Gene Max # SNPs[1]	Gene name	cDNA change[£]	Protein change	ExAC AFR	ExAC EUR	*Z*-scores[¥] "Random" SCA	"Stroke" SCA	"Long survivor" SCA	Replication samples	
NADSYN1	12	NAD synthetase 1	c.G175A	p.E59K	0	0	15.15	15.15	15.15	16.21
CACNA1H	4	Calcium channel, voltage-dependent, T type, alpha 1H subunit	c.C1538T	p.S513L	0	0	15.05	15.18	15.05	17.02
SERPINC1	5	Serpin peptidase inhibitor, clade C, member 1	c.G973C	p.A325P	0	0	19.75	19.75	–	18.54
INSR	3	Insulin receptor	c.G3311A	p.R1104H	0	0	15.83	–	15.83	20.82
NOS3	9	Nitric oxide synthase 3	c.G1585A	p.G529S	0	0	15.73	-	15.73	18.74
COL6A3	9	Collagen, type VI, alpha 3	c.T7463C	p.I2488T	0	0	19.472	–	16.95	17.01
HGF	1	Hepatocyte growth factor	c.C1595T	p.A532V	0	0	–	19.91	–	18.93
SNTB1	4	Syntrophin, beta 1	c.A173G	p.N58S	0	0	–	19.27	–	16.98
CLCN6	5	Chloride channel 6	c.G992C	p.C331S	0	0	–	–	19.058	18.84
OGDHL	5	Oxoglutarate dehydrogena like	c.C632T	p.T211M	0	0	–	–	19.95	15.97
CAPN13	4	Calpain 13	c.C336G	p.I112M	0	0	18.074	–	–	15.76
MYO7B	26	Myosin VIIB	c.G77A	p.G26D	0.0006	0	17.660	–	–	20.14

[¥]The z-scores are obtained from aggregating the SiPhy (29-way) score based on identified mutants SNPs within genes (See details in Table S5 of all the mutations found).

[£]Exonic. nonsynonymous variants that were considered damaging according to 21 different functional scores from the annotation databases, including SIFT, LRT, MutationTaster, MutationAssessor, FATHMM, fathmm-MKL, RadialSVM, LR, PROVEAN, MetaSVM, MetaLR, CADD, GERP++, DANN, M-CAP, Eigen, GenoCanyon, Polyphen2 HVAR, Polyphen2 HDIV, PhyloP, and SiPhy, as previously reported.[8] Abbreviations: Max #SNPs: Maximum number of nonsynonymous variants observed among the three SCD groups; SNP: Single Nucleotide Polymorphism; ExAC: Exome Aggregation Consortium; AFR: African; EUR: European. The reported cDNA change, Protein change, and all ExAc frequencies are from the putative deleterious variants with top SiPhy (29-way) score.

formed a network (Figure 3B) through gene hubs including *HGF*, *PLAT [MIM: 173370]*, *F5 [MIM: 612309]*, *F2 [MIM: 176930]*, *ESR2*, *INSR*, *RXRG [MIM: 180247]*, *PMS1 [MIM: 600258]*, and *MYOC [MIM: 601652]*. These gene hubs were associated with blood clotting cascade, vitamin B12 metabolism, and thrombophilia ($P = .0000025$, .000039, and .00029, respectively; Figure 3B,C). In the replication sample, we found enrichment of genes in glutamate metabolism, response to oxidative stress, complement/coagulation, and hemoglobin synthesis (Figure S7**)**. In the replication sample, novel findings that were not seen in the discovery samples were enrichment of genes in the following pathways: focal adhesion, angiogenesis, immune response and inflammation, hemoglobin production, longevity, nitric oxide, calcium signaling, and heme metabolism.



Significant human pathways associated to gene hubs in figure a above: *HGF, PLAT, F5, F2, ESR2, INSR, RXRG, PMSI and MYOC.*

Term	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Blood Clotting Cascade	0.0000025	0.000093	-1.88	17.53	*PLAT;F2;F5*
Vitamin B12 Metabolism	0.000039	0.00038	-1.83	14.41	*INSR;PLAT;F2*
Complement and Coagulation Cascades	0.000046	0.00038	-1.86	14.69	*PLAT;F2;F5*
Folate Metabolism	0.000098	0.00064	-1.76	12.94	*INSR;PLAT;F2*
Selenium Micronutrient Network	0.00019	0.001	-1.76	12.10	*INSR;PLAT;F2*
Nuclear Receptors	0.0012	0.005	-1.61	8.45	*RXRG;ESR2*

FIGURE 3 Overlapping biological networks of identified candidate mutations in three Cameroonian SCA sub-groups. A, Overlap of three networks of the identified candidate mutations (Table S1). B, Significant pathways associated with gene hubs. C, Interaction network of top associated pathways of gene hubs

3.1.4 | Gene-set allele frequency differentiation among groups of SCD patients and controls

We tested for gene-specific differences in SNP frequencies among the three SCA groups (Figure 2B, Figures S8 A, B, and C). Comparing the "long survivor" and "stroke" groups, we found six genes exhibiting significant differentiation, including *VKORC1* ([MIM: 608547] $p = 5.44 \times 10^{-7}$), *FGA* ([MIM: 134820] $P = 3.40 \times 10^{-6}$), *FGFR3* ([MIM: 134934] $P = 3.45 \times 10^{-6}$), *PIGG* ([MIM: 616918] $P = 8.64 \times 10^{-6}$), *HFE2* ([MIM: 608374] $P = 1.16 \times 10^{-5}$), and *P1BA* ($P = 1.83 \times 10^{-5}$) (Figure S8C). These genes were clustered in complement and coagulation cascades ($P = .00007$; Figure 2A) and expressed in the liver ($P = .0054$). Comparing the "long survivor" group and the rest of the SCA patients (ie, "stroke" and "random"), three genes exhibited unusual differentiation (Figure S8B), including *CYP21A2*, *P2RY2 [MIM: 600041]*, and *PLAT* ($P = 6.28 \times 10^{-6}$, 1.22×10^{-5}, and 1.97×10^{-5}, respectively), enriched for blood coagulation ($P = .0013$) (Figures S9A, and S9C), associated with thrombophilia ($P = .006818$), and expressed in the liver ($P = .0053$). These genes are also implicated in fibrinolysis and the response to oxygen levels (GO:0042730 and GO:0070482, respectively; Figure S9C).

Comparing the "stroke" group and the rest of the SCA patients, we found frequency differences in *ITGA2B [MIM: 607759]*, *CALCA [MIM: 114130]*, *LTC4S [MIM: 246530]*, *EMILIN1 [MIM: 130660]*, *SERPINC1*, *NPHS1 [MIM: 602716]*, *ACD [MIM: 609377]*, and *USP37* ($P = 3.46 \times 10^{-7}$, 5.36×10^{-7}, 3.96×10^{-6}, 4.96×10^{-6}, 6.22×10^{-6}, 5.85×10^{-6}, 1.86×10^{-5}, and 1.48×10^{-5}, respectively; Figure S8). These genes were enriched for blood coagulation ($P = .00131$), and glycosylphosphatidylinositol anchor biosynthesis ($P = .0014$), associated with rheumatoid arthritis ($P = .046$) and Fanconi anemia ($P = .0467$), and expressed in the liver ($P = .0053$). In the comparison of the controls against each patient group including "random," "stroke," and "long survivor" groups, we found *PLAT* and *SLC19A2* [MIM: 603941] to be consistently significant in gene-specific differences in allele frequencies across all the comparisons (Table S4). A total of 29, 27, and 23 genes exhibited significantly unusual gene-specific differences in allele frequency with nominal *P*-values ranging from 2.5×10^{-5} to .05, in "random," "stroke," and "long survivor" patients' groups against the Cameroon controls,

TABLE 3 Significant genes from gene-set rare-variant association analyses

Region	Gene	Min. #Rare Variants tested	"Stroke" Versus"Random" SCA	"Long Survivor" Versus"Random" SCA	"Stroke" Versus"Long Survivor"	"Random" SCAVersus"Controls"
1p36.22	***CLCN6***	19	0.4686	**0.0080**	0.0527	1.0
1p21.1	***COL11A1***	69	**0.006066**	0.6627	0.1376	1.0
1q24.2	*F5*	12	0.5863	0.5849	0.0287	0.912
1q25.1	***SERPINC1***	21	0.0716	0.033	**0.017**	1.0
1q32.1	***ATP2B4***	13	**0.0082**	**0.0045**	0.0554	0.04
2p23.1	*CAPN13*	11	0.4759	0.0223	0.1827	0.812
2p13.1	***SLC4A5***	87	0.0462	**0.0061**	0.0505	0.023
2q34	***CPS1***	5	0.0737	0.7004	0.0767	0.021
2q37.3	***COL6A3***	9	**0.0104**	0.6201	0.0398	0.001
5p15.31	***MTRR***	3	0.2059	**0.0119**	0.02263	1.0
5q13.3	*HMGCR*	9	0.3361	0.4565	0.0453	0.062
5q31.1	***SLC22A5***	17	0.0478	0.6778	**0.0041**	0.011
7q21.11	***HGF***	45	0.0257	**0.0063**	0.0174	0.041
7q36.1	*NOS3*	39	0.342	0.0324	0.6525	0.76
8p11.21	***PLAT***	21	0.4267	**0.012**	0.4917	0.058
8q24.12	*SNTB1*	23	0.0228	0.354	0.2357	0.049
10p14	*GATA3*	12	0.2278	0.0356	0.3434	1.0
11q13.4	***NADSYN1***	130	**0.0172**	0.045	0.0247	0.017
14q11.2	***SLC7A8***	34	0.568	**0.0061**	0.05819	0.719
14q32.13	*SERPINA1*	78	0.0392	0.2231	0.0665	0.033
15q21.1	*SLC12A1*	97	0.0372	0.0848	0.5735	0.013
16p13.3	***CACNA1H***	134	0.0534	0.0229	**0.0139**	0.012
16p13.11	***ABCC1***	209	**0.0083**	**0.0155**	0.653	0.0246
17p12	***COX10***	49	0.3804	**0.0143**	0.3224	1.0
17q21.31	***SLC4A1***	156	0.5027	0.472	**0.007**	0.091
17q23.2	*MARCH10*	329	0.0375	0.0706	0.0287	0.028
19p13.3	***ABCA7***	278	0.194	**0.0151**	0.0884	0.078
19p13.2	***INSR***	178	**0.0106**	0.4731	0.0261	0.0191
20p11.21	*PYGB*	376	0.0257	0.043	0.0282	0.027

Table 3 shows all genes significant at a permutation $P < .05$ for at least one of the primary comparisons. Genes and *P*-values significant at adjusted $P < .017$ are in bold. Numbers of individuals are: (a) "Stroke' versus "Random" SCA (n = 79), (b) "Long Survivor" versus" Random" SCA (n = 82), (c) "Stroke" versus" Long Survivor" (n = 59). "Random" SCA versus "Controls" (n = 114) included as comparison

respectively (Table S4). These genes are enriched with several genes identified with recurrent putative deleterious variants (Table 2 and Table S1) and those that clustered in hubs of an interaction network (Figure 2; Figures S4, S5, and S6). Specifically, *HBG2* [MIM: 142250] featured significant differentiation in allele frequencies with all SCD patients ($P = 1.31 \times 10^{-5}$, Table S4).

3.1.5 | Gene-set rare-variant association analyses

The main findings from the rare variant association tests are summarized in Table 3. We found 19 gene sets that were significantly associated at an adjusted $P < .017$ in the comparison of "stroke" versus "random," "long survivor" versus "random," and/or "long survivor" versus "stroke" (Table 3, Figure 3). Gene sets significantly associated with stroke included *ATP2B4, COLL11A1, COL6A3, NADSYN1, ABCC1 [MIM: 158343]*, and *INSR* while the gene sets significantly associated with long survival included *CLNC6, SLC24A5 [MIM: 609802], MTRR [MIM: 602568], HGF, PLAT, SLC7A8, ABCC1, COX10 [MIM: 602125]*, and *ABCA7 [MIM: 605414]*. One gene, *ATP2B4*, was associated with both "stroke" and "long survivor" groups. For the extreme contrast of "stroke" versus "long survivor", only four gene sets – *SERPINC1, SLC22A5, CACNA1H*, and *SLC4A1* [MIM: 109270] – were significant. Notably, these

TABLE 4 Significant pathways in this study that are also significant in global transcriptomic studies of SCD[21,22]

Pathway identified in the present study	Transcriptomic evidence			
	Lowest *P*-values in the present study	Associated phenotypes	*P*-values in original studies	Source
Starch and sucrose metabolism	2.6×10^{-9}	molecular risk profile	1.1871	21
One carbon pool by folate/folate pathway	5.8×10^{-6}	molecular risk profile	3.3263×10^{-8}	21
Complement and coagulation cascades	3.9×10^{-8}	molecular risk profile	2.73693×10^{-10}	21
Complement and coagulation cascades	3.9×10^{-8}	acute crisis in children	.016	22
Oxidative Stress	1.8×10^{-6}	top severity score in children	.00442	22
Oxidative Stress	1.8×10^{-6}	acute crisis in children	5.6×10^{-4}	22
Heme biosynthesis	1.7×10^{-10}	top severity score in children	.00925	22
Heme biosynthesis	1.7×10^{-10}	acute crisis in children	.005	22
Regulation of cellular response to stress	1×10^{-09}	top severity score in children	.00008	22
Colorectal cancer/ DNA repair system	.0002	molecular risk profile	1.51002×10^{-12}	21

four genes were not significant in the stroke association or long survival association tests, indicating that additional information was gained from this comparison.

The significant gene sets in the rare variant association analyses included most of the genes found to harbor recurrent deleterious variants in both the discovery and replication cohort (Table 2) and/or showing unusual allele frequency distributions (Table S4).

3.1.6 | Comparison with GWAS of sickle cell anemia and related traits

To contextualize our findings against the results of GWAS that have been conducted for sickle cell anemia and related traits, we queried the NHGRI-EBI GWAS Catalog for all associations related to "sickle cell anemia" (EFO ID: Orphanet_232) and "haemoglobin F" (EFO ID: EFO_0004576). None of the genes significant in the present study (Tables S3 and S4) was significant in the relevant GWAS and vice versa. Therefore, our results represent findings unique to the exome space and not found by GWAS. We also note that there is no overlap between the genes in this study and genes significantly associated with HbF by GWAS. This observation implies that the effect of the genes we found to be significantly associated with long survival or stroke in this study are not mediated via HbF levels (one of the strongest and most consistently associated SCA modifiers).

3.1.7 | Functional support for identified pathways from transcriptomic studies

Gene expression provides functional evidence that a pathway is dysregulated in a disorder. To this end, we queried the significant pathways identified by sequence analysis in the present study against the most differentially dysregulated pathways in the two largest studies of global gene expression profiles in relation to SCD severity.[22] Our findings show that most of our significant pathways also show significant transcriptomic differences in relation to SCD severity (Table 4), thus providing supportive evidence that these pathways are important in SCD pathophysiology.

4 | DISCUSSION

Our study addresses the issue of genetic modifiers of clinical variation in SCA in SSA using a whole-exome sequencing approach. We utilized a design that included "long survivors" (representing patients surviving to the fifth decade despite the harsh environment and lack of state-of-the-art medical care) and overt stroke patients (representing one of the most severe complications of SCA). By including a "random" comparison group, we provided a reference comparison group of the "average" SCA patient. This phenotype grouping that explicitly recognizes that some complications take time to emerge (eg, stroke- or age-related complications/mortality) is a methodologic approach that,

perhaps, could be considered innovative and a strength of this manuscript. Given that there were no significant findings among the groups for fetal hemoglobin levels, classical sickle haplotypes, and α-thalassemia, our study implicitly controlled for these known factors for clinical heterogeneity in SCD. However, the lack of significance of HbF levels, a well-established and strongest known modifier of SCD childhood complications and that is influenced by genomic variations and therapeutic interventions, probably reflects the fact that the sample studied have all survived past the "under-5-year-old" mortality hazards (malaria, bacterial sepsis, diarrheal disease, or splenic sequestration, etc.).

The main findings point to different gene sets that are enriched for deleterious and loss-of-function mutations in phenotypically defined groups of patients and with evidence of genetic association with different phenotypes, providing support for the complexity of the genetic architecture of SCD phenotypic variability. Notably, pathways represented by these genes point to relevant pathophysiological mechanisms, including some that are already therapeutic targets. Our findings of the involvement of glutamine (*NADSYN1*) and arginine (*OGDHL* and *NOS3*) are novel and noteworthy. Decreased erythrocyte glutamine levels contribute to alterations in the erythrocyte redox environment and hemolysis and play a role in the pathogenesis of pulmonary hypertension in SCD.[23] L-Glutamine was recently approved by the US FDA as a medication for SCD.[24] A non-synonymous variant in the arginine-fifty homeobox gene (*ARGFX*) was previously associated with stroke in SCD.[13] Low-dose supplementation with L-arginine improved liver function, oxidative stress, nitric oxide metabolite levels, cardiovascular dysfunction, and sickle cell-related pain.[25] *CACNA1H* is associated with hypertension and is therapeutically targetable by calcium channel blockers.[26] Thus, our findings using sequence analysis of African SCA patients has support from prior studies. Moreover, the findings of vascular and NO signaling variants are consistent with the clinical observations that long term survival among African Americans patients are dependent on vasculopathic complications.[27]

A major finding of this study is the observation that "long survivor" group was characterized by mutational burdens in *CLCN6* and *OGHDL*. Rare, deleterious mutations in *CLCN6* (a voltage-dependent chloride channel) have been associated with lower blood pressure.[28] Given that increased blood pressure is a major risk factor for stroke in SCD,[29] this suggests that SCD patients with *CLCN6* mutations live longer due to a reduced risk of stroke. *OGHDL* is important in arginine metabolism, which is a key factor in the hemolysis-endothelial dysfunction observed in SCD and has become a target for therapeutic interventions as noted above. Interestingly, in a recent trial in patients with SCD, the median number of pain crises over 48 weeks was lower among those who received oral therapy with L-glutamine.[30] Variants in genes involved in complement and coagulation cascade and fibrinolysis appear to be important for all SCD patients and particularly for susceptibility to stroke. It is well known that SCD involves a hypercoagulable state.[3,13] *SERPINC1* encodes antithrombin, implicating loss of SERPINC1 activity with increased blood coagulation. Transcriptomic expression of complement and coagulation components in circulation was increased in a cluster of African American SCD patients with higher severity and mortality rate.[21,22] Moreover, there is evidence of increased complement activation in older patients with SCD,[31] and a complement gene *C5* mutation was associated with stroke in SCD patients.[13] In summary, the evidence from this and other studies suggest that long survival is characterized by mutations that confer protection for adverse phenotypes (notably stroke) given that these genes influence intermediate phenotypes for stroke (including blood pressure and endothelial function) while the overt stroke phenotype in SCA is associated with mutations in genes that are involved in the complement and coagulation cascade. Identification of genes such as *LTC4S*, that displayed specific signal of unusual difference in SNPs frequencies among patients with stoke (Figure S8A), will deserve future investigations with appropriate experimental design to explore other genes involved in leukotriene synthesis pathways, that have long been associated with SCD pain, airway hyperresponsiveness, and hospitalization rates.[32,33] Specifically, future studies should investigate leukotriene levels as potential marker for stroke, in as much as leukotriene antagonists are being tested in Phase 2 trial for SCD-related comorbidities in SCD.[34]

Long survival in SCA is a composite phenotype that includes factors that decrease mortality-causing events (eg, strokes) and/or are associated with improved indices of health (eg, favorable blood pressure and lipid profiles), which in turn have their own risk factors. While the present study was not designed to measure most of these intermediate phenotypes, we note that the genes with high mutational burden found in the present study were annotated to functional pathways related to some of the intermediate phenotypes and also show association with severity phenotypes in independent transcriptomic studies (as shown in Table 4). Further studies are needed to investigate the mechanisms by which these mutational changes influence the phenotype. We also recommend that future longitudinal studies with larger sample sizes, from diverse population groups and settings in Africa, Europe, and America, should include HbF levels, blood pressure, markers of oxidative stress, arginine/ glutamine

levels, blood coagulation markers, and markers of heme pathways/hemolysis as important clinical variables for genotype-phenotype association, in relation to long-term survival.

Involvement of genes in the pathways of vitamin B12 and folate metabolism is expected as these pathways are important for regulation of erythropoiesis. Subjects with SCD are at higher risk of cobalamin deficiency, justifying supplementation in clinical practice.[35] Identification of genes involved in mitotic check-point and DNA repair, starch, and sucrose metabolism, and solute carriers require further study to explore their roles in modifying the SCD phenotype. Similarly, future studies on heme pathways can explore if hemolysis metabolism or susceptibility are responsible for findings in this signaling hub in the replication samples.

To our knowledge, this is the first investigation of clinical variation in SCD in Africa using a whole-exome sequencing approach. Strengths of the study include well-defined clinical groups, inclusion of an independent replication sample, study sites where treatment is unlikely to confound outcomes, use of several different but complementary analytical approaches and linking the identified genes and pathways to published transcriptomic and therapeutic data. Nonetheless, the study has some limitations. The stroke group consisted of patients with overt stroke and would therefore not have captured patients with silent cerebrovascular events. In addition, we focused on overt stroke, without brain imaging than could have further sub-stratified this phenotype as ischemic or hemorrhagic, and identified subclinical infarcts, that are also found in SCD in children in Africa.[36,37] These sub-classifications could have allowed differential exploration of genetic protective and pathophysiologic risk factors, for example, variants in genes in hemolysis pathways for ischemic stroke, and variants in genes of vasculopathy, hypertension, and connective tissues pathways for hemorrhagic stroke. However, it should be noted that our focus is on a severe event (overt stroke), not on silent subclinical events, which, by definition, do not represent severe clinical events. The ideal study design for outcomes in SCA is a longitudinal study. In its absence, we have used a "random group" in the present study to enable us to distinguish between genes with similar mutation burden in all SCD patients irrespective of the clinical severity versus those genes that exhibit such characteristics in specific extreme groups, such as "long survivor" and "stroke" patients. Future studies with a longitudinal design will indeed provide a better comparison than what is possible with a cross-sectional study. Also, the sample sizes are relatively modest and larger sample sizes would probably yield more findings, as illustrated by the finding of additional genes and pathways found in the replication group but not in the discovery group. Nonetheless, the total number of exomes sequenced in the study represents one of the largest (and only available from Africa) such datasets for SCA severity to date.

In summary, we reported a WES study on clinical phenotypes of SCA in Africa. We generated a catalogue of candidate modifier genes that clustered in pathophysiological pathways important in SCA and with implications for therapeutic intervention. This study fills an important gap in knowledge by using a WES approach focusing on deleterious coding variants important in two specific clinical categories of SCA patients (long survival and overt stroke), in contrast to most other studies that used a GWAS approach and often used fetal hemoglobin levels as a proxy of severity. This study thus makes significant contributions to present knowledge of the natural history and clinical heterogeneity of SCA in SSA, with the potential for informing the design of new therapeutics.

5 | WEB RESOURCES

Online Mendelian Inheritance in Man: http://www.omim.org

SKAT: SNP-Set (Sequence) Kernel Association Test: https://cran.r-project.org/web/packages/SKAT/index.html

ACKNOWLEDGEMENTS

Conceived and designed the experiments: AW, EC, AA. Performed the experiments: AW, KN, GP, VJNB. Patient recruitment, samples and clinical data collection and processing: AW, VJNB, KM, GP. Analyzed the data: AW, EC, KM, GP, AA. Contributed reagents/materials/analysis tools: AW, EC, NM, DS, CNR. Wrote the paper: AW, EC. Revised and approved the manuscript: AW, AA, EC, KM, VJNB, NM, DS, CNR. The molecular experiments of the study were funded by the National Health Laboratory Services (NHLS), South Africa and the National Institutes of Health, USA grants 1U01HG007459-01 and U24-HL-135600 to AW. CNR, AA, and DS are supported by the Intramural Research Program of the Center for Research on Genomics and Global Health (CRGGH). CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The corresponding author confirms that he has full access to all the data in the study and has final responsibility for the decision to submit for publication.

DATA AVAILABILITY STATEMENT

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

CONFLICT OF INTEREST

The authors declare no competing interests. The authors alone are responsible for the content and writing of this article.

ORCID

Ambroise Wonkam https://orcid.org/0000-0003-1420-9051

REFERENCES

1. Shriner D, Rotimi CN. Whole genome sequence-based haplotypes reveal single origin of the sickle allele during the Holocene Wet Phase. *Am J Hum Genet*. 2017;102:547-556.
2. Piel FB, Hay SI, Gupta S, Weatherall DJ, Williams TN. Global burden of sickle cell anaemia in children under five, 2010–2050: modelling based on demographics, excess mortality, and interventions. *PLoS Med*. 2013;10:e1001484.
3. McGann PT, Hernandez AG, Ware RE. Sickle cell anemia in sub-Saharan Africa: advancing the clinical paradigm through partnerships and research. *Blood*. 2017;129:155-161.
4. 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;330:1639-1644.
5. Xu J, Peng C, Sankaran VG, et al. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science*. 2011;334:993-996.
6. Saraf SL, Akingbola TS, Shah BN, et al. Associations of α-thalassemia and BCL11A with stroke in Nigerian, United States, and United Kingdom sickle cell anemia cohorts. *Blood Advance*. 2017;1:693-698.
7. Geard A, Pule GD, Chetcha Chemegni B, et al. Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *Br J Haematol*. 2017;178:629-639.
8. Maitra P, Caughey M, Robinson L, et al. Risk factors for mortality in adult patients with sickle cell disease: a meta-analysis of studies in North America and Europe. *Haematologica*. 2017;102:626-636.
9. Wonkam A, Mnika K, Ngo Bitoungui VJ, et al. Clinical and genetic factors are associated with pain and hospitalisation rates in sickle cell anaemia in Cameroon. *Br J Haematol*. 2018;180:134-146.
10. Grosse SD, Odame I, Atrash HK, Amendah DD, Piel FB, Williams TN. Sickle cell disease in Africa: a neglected cause of early childhood mortality. *Am J Prev Med*. 2011;41:S398-S405.
11. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177-1186.
12. Wonkam A, Bitoungui VJN, Vorster AA, et al. et al. Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLoS ONE*. 2014;9:e92506.
13. Flanagan JM, Sheehan V, Linder H, et al. Genetic mapping and exome sequencing identify 2 mutations associated with stroke protection in pediatric patients with sickle cell anemia. *Blood*. 2013;121:3237-3245.
14. Rumaney MB, Bitoungui VJN, Vorster AA, et al. The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLoS ONE*. 2014;9:e100516.
15. Gézsi A, Bolgár B, Marx P, Sarkozy P, Szalai C, Antal P. Variant MetaCaller: automated fusion of variant calling pipelines for quantitative, precision-based filtering. *BMC Genom*. 2015;16:875.
16. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297-1303.
17. Sifrim A, et al. eXtasy: variant prioritization by genomic data fusion. *Nat Methods*. 2013;10:1083.
18. Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants. *PLOS Genet*. 2011;7:e1001322.
19. Ionita-Laza I, Lee S, Makarov V, Buxbaum J, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Hum Genet*. 2013;92:841-853.
20. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. ancGWAS: a post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. *Bioinformatics*. 2015;32:549-556.
21. Tishkoff SA, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009;324:1035-1044.
22. Desai AA, Lei Z, Bahroos N, et al. Association of circulating transcriptomic profiles with mortality in sickle cell disease. *Blood*. 2017;129:3009-3016.
23. Hounkpe BW, Fiusa MML, Colella MP, et al. Role of innate immunity-triggered pathways in the pathogenesis of Sickle Cell Disease: a meta-analysis of gene expression studies. *Sci Rep*. 2015;5:17822.
24. Morris CR, Kuypers FA, Lavrisha L, et al. A randomized, placebo-controlled trial of arginine therapy for the treatment of children with sickle cell disease hospitalized with vaso-occlusive pain episodes. *Haematologica*. 2013;98:1375-1382.
25. Wilmore DW. Food and Drug Administration approval of glutamine for sickle cell disease: success and precautions in glutamine research. *J Parenter Enteral Nutr*. 2017;41:912-917.
26. Shilo NR, Morris CR. Pathways to pulmonary hypertension in sickle cell disease: the search for prevention and early intervention. *Expert Rev of Hematol*. 2017;10:875-890.
27. Dl Hamideh, Alvarez O. Sickle cell disease related mortality in the United States (1999-2009). *Pediatr Blood Cancer*. 2013;60:1482-1486.
28. Adeyemo A, Gerry N, Chen G, et al. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet*. 2009;5:e1000564.
29. Cheng YC, Cole JW, Kittner SJ, Mitchell BD. Genetics of ischemic stroke in young adults. *Circulation: Cardiovascular Genetics*. 2014;7:383-392.
30. Niihara Y, Miller ST, Kanter J, et al. A phase 3 trial of l-glutamine in sickle cell disease. *N Engl J Med*. 2018;379(3):226-235.

31. Gavriilaki E, Mainou M, Christodoulou I, et al. In vitro evidence of complement activation in patients with sickle cell disease. *Haematologica*. 2017;102:e481-e482.
32. Eiymo Mwa Mpollo MS, Brandt EB, Shanmukhappa SK, et al. Placenta growth factor augments airway hyperresponsiveness via leukotrienes and IL-13. *J Clin Invest*. 2016;126:571-584.
33. Setty BN, Stuart MJ, Dampier C, Brodecki D. Allen JL Hypoxaemia in sickle cell disease: biomarker modulation and relevance to pathophysiology. *Lancet*. 2003;362:1450-1455.
34. Field JJ, Kassim A, Brandow A, et al. Phase 2 trial of montelukast for prevention of pain in sickle cell disease. *Blood Adv*. 2020;4:1159-1165.
35. Dixit R, Nettem S, Madan SS, et al. Folate supplementation in people with sickle cell disease. *Cochrane Database Syst Rev*. 2018;3:CD011130.
36. Kija EN, Saunders DE, Munubhi E, Darekar A, et al. Transcranial Doppler and magnetic resonance in Tanzanian children with sickle cell disease. *Stroke*. 2019;50:1719-1726.
37. Green NS, Munube D, Bangirana P, et al. Burden of neurological and neurocognitive impairment in pediatric sickle cell anemia in Uganda (BRAIN SAFE): a cross-sectional study. *BMC Pediatr*. 2019;19:381.

SUPPORTING INFORMATION

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

How to cite this article: Wonkam A, Chimusa ER, Mnika K, et al. Genetic modifiers of long-term survival in sickle cell anemia. *Clin Transl Med*. 2020;10:e152. https://doi.org/10.1002/ctm2.152