Published in final edited form as:
Blood. 2011 January 27; 117(4): 1390–1392. doi:10.1182/blood-2010-08-302703.

Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia

Julie Makani[#1,2], **Stephan Menzel**[#3], **Siana Nkya**[1], **Sharon E. Cox**[1,4], **Emma Drasar**[3,5], **Deogratius Soka**[1], **Albert N. Komba**[1], **Josephine Mgaya**[1], **Helen Rooks**[3], **Nisha Vasavda**[3], **Gregory Fegan**[2,6], **Charles R. Newton**[1,4,6,7], **Martin Farrall**[8], and **Swee Lay Thein**[3,5]

[1]Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania

[2]University of Oxford, Oxford, United Kingdom

[3]King's College London School of Medicine, Molecular Haematology, London, United Kingdom

[4]London School of Hygiene and Tropical Medicine, London, United Kingdom

[5]King's College Hospital, Haematological Medicine, Denmark Hill, London, United Kingdom

[6]Centre for Geographic Medical Research–Coast, Kenya Medical Research Institute, Kilifi, Kenya

[7]Institute of Child Health, University College London, London, United Kingdom

[8]University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom

[#] These authors contributed equally to this work.

Abstract

Fetal hemoglobin (HbF, $\alpha_2\gamma_2$) is a major contributor to the remarkable phenotypic heterogeneity of sickle cell anemia (SCA). Genetic variation at 3 principal loci (*HBB* cluster on chromosome 11p, *HBS1L-MYB* region on chromosome 6q, and *BCL11A* on chromosome 2p) have been shown to influence HbF levels and disease severity in (β-thalassemia and SCA. Previous studies in SCA, however, have been restricted to populations from the African diaspora, which include multiple genealogies. We have investigated the influence of these 3 loci on HbF levels in sickle cell patients from Tanzania and in a small group of African British sickle patients. All 3 loci have a significant impact on the trait in both patient groups. The results suggest the presence of *HBS1L-MYB* variants affecting HbF in patients who are not tracked well by European-derived markers, such as *rs9399137*. Additional loci may be identified through independent genome-wide association studies in African populations.

Correspondence: Swee Lay Thein, King's College London School of Medicine, James Black Centre, 125 Coldharbour Lane, London, SE5 9NU, United Kingdom; sl.thein@kcl.ac.uk; or Stephan Menzel, King's College London School of Medicine, James Black Centre, 125 Coldharbour Lane, London, SE5 9NU, United Kingdom; stephan.menzel@kcl.ac.uk; or Julie Makani, King's College London School of Medicine, James Black Centre, 125 Coldharbour Lane, London, SE5 9NU, United Kingdom; jmakani@muhas.ac.tz.
The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Authorship

Contribution: J. Makani, S.M., and S.L.T. designed the research and wrote the manuscript; J. Makani, D.S., A.N.K., J. Mgaya, E.D., N.V., G.F., C.R.N., and S.L.T. collected data; S.N., S.E.C., and H.R. performed genotyping and analyzed results; J. Makani, S.M., H.R., S.L.T., and M.F. analyzed and interpreted results; and all authors commented on drafts of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Introduction

Sickle cell anemia (SCA) is a monogenic disease caused by a single mutation (*HBB* glu6val) within the gene encoding the β-subunit of adult hemoglobin (HbA, $\alpha_2\beta_2$), but remarkable clinical variability is introduced through additional genetic and nongenetic factors.1,2 A major ameliorating factor is an inherent ability to produce fetal hemoglobin (HbF, $\alpha_2\gamma_2$); elevated levels are correlated with reduced morbidity and mortality in patients with SCA.3,4 Genetic variants at 3 principal loci have been shown to contribute to the interindividual HbF variation in sickle patients,5–7 the region on chromosome 11p that contains the *HBB* and olfactory receptor gene clusters8 and 2 hematopoietic regulator loci: one on chromosome 6q (*HBS1L-MYB* intergenic polymorphism, *HMIP*) and one on chromosome 2p (*BCL11A*).

The sickle mutation is prevalent in Sub-Saharan Africa; 80% of the global 300 000 annual affected newborns occur in Africa, with one of the highest incidence rates in Tanzania (~ 8000 births per year).9,10 So far, studies of the modifier loci have been restricted to populations from the African diaspora, which include multiple genealogies with brief (< 15-30 generations) coalescent times, which can generate unusual linkage disequilibrium patterns.11 The sickle mutation exists in Africa on diverse genetic backgrounds,12 and each distinctive African population studied has the potential to offer unique clues about genes and other factors that might alleviate disease. Before we venture into the unknown, though, and search for new genes, it is prudent to first evaluate the presence and impact of the 3 known major loci on the HbF trait. Here we present such data for a cohort of SCA patients from East Africa compared with an African British SCA population.

Methods

Tanzanian patients

The Muhimbili Sickle Cell Collaborative Program was established in 2004 in Dar-es-Salaam, Tanzania. Patients were recruited from the hematology clinic in Muhimbili National Hospital, Dar-es-Salaam. Written informed consent was obtained from patients or parents/guardians of children in accordance with the Declaration of Helsinki. Ethical approval was given by the Muhimbili University Research and Publications Committee (no. MU/RP/AEC/VOL XI/33). Hematologic data were included from nontransfused state only and from patients 5 years of age or older. Pairs of phenotype (HbF) and genotype data could be assembled for 1045 patients (Table 1).

British patients

We have included data on a population of 151 British patients (146 with HbSS and 5 with HbS/β^0 sickle genotype) of African-Caribbean (Jamaican, Trinidanian) or West African (Nigerian, Ghanaian, Sierra Leonean) descent from King's College Hospital, London, United Kingdom (Table 1). The patients were recruited through the specialist clinic in the Hematology Outpatient Unit (King's College Hospital Local Research Ethics Committee, protocol no. 01-083). At the time of study, patients ranged from 11 to 64 years of age (median, 29 years) had not been transfused within 120 days and were not receiving hydroxycarbamide.

Hemoglobin profiling by high performance liquid chromatography (Variant II Hemoglobin Testing System) is part of the routine clinical procedures for all patients seen at the clinics in Dar-es-Salaam and London. Genomic DNA isolated from EDTA blood samples was used for genotyping. DNA from Tanzanian samples were first genome-wide amplified by random primer amplification.13

Single nucleotide polymorphism (SNP) genotyping was performed by the TaqMan procedure (Applied Biosystems) at King's College London, as described,7 or by MassARRAY procedure (Sequenom) at The Wellcome Trust Center for Human Genetics, Oxford, United Kingdom. SNP *rs748214* resides within the promoter of the $^{G}\gamma$ globin gene (*HBG2*), which is very similar to that of *HBG1* ($^{A}\gamma$). Typing of *rs748214* involves an initial stage of specific amplification of the *HBG2* promoter encompassing *rs748214* by polymerase chain reaction14 followed by TaqMan genotyping of the polymerase chain reaction product.

Statistical analysis

For quality control, a Hardy-Weinberg test was performed on all genotype results. The HbF percentage (of total hemoglobin) values were natural log-transformed, and the extreme low tail of the distribution was trimmed to create a nearly normally distributed quantitative trait. Genetic association of this trait with SNP alleles was analyzed through multiple linear regression (SPSS, Version 12, IBM), with age and sex included as covariates. Dominance was tested for, but no significant ($P < .05$) effects were detected.

Results and discussion

All 3 principal HbF loci have a significant impact in Tanzanian patients with SCA (Table 2), the strongest association being seen at the *BCL11A* locus on chromosome 2. The considerable effect of alleles at this locus (−0.406 for *rs11886868* and −0.412 *rs4671393*) results in a marked depression of mean HbF values for genotypes containing the minor allele (eg, for *rs4671393*: 3.7% for G/G and 5.4% for A/G compared with 8.1% for A/A). Together with a high prevalence of the minor alleles (26% and 30% for *rs11886868* and *rs4671393*, respectively) in the Tanzanian population, this leads to an important influence of this locus on the overall phenotype (ie, 12.8% of the trait variance can be explained by genetic variation at *rs4671393* alone; Table 2). A similar impact is seen in the African British patient population (Table 2) and has been reported for African American and African Brazilian patients.6–8,15,16 To date, the functional variant causing this strong association signal across most human populations has not been identified.

The largest allelic effect (0.668, Table 2) in the Tanzanian patients was detected at the *HMIP* locus on chromosome 6, specifically sub-locus *HMIP-2* (*rs9399137*), leading to mean HbF values of 8.8% for C/T versus 4.5% for T/T (C homozygotes were not found; data not shown in Table 2). Variant alleles for this marker are rare in the Tanzanian patients, though, as they are in the African British and most patients or healthy persons of African descent.6,7 Therefore, its overall impact is small (1.6% of the trait variance). Hence, *rs9399137*, which acts as tagging SNP for the *HMIP-2* sub-locus in European populations,17,18 does not track

the causative sequence variant at *HMIP-2* on African chromosomes very well7 because of its low frequency on the African chromosomes.

In a conditional regression analysis, there is evidence for a more extensive association signal at *HMIP-2* (Table 2) that is only partially tracked by *rs9399137* and independent of *rs9399137.* The importance of the *HMIP* locus in African populations might therefore have been underestimated by this and other datasets using markers tailored to European studies. Our findings also support the presence of shorter haplotype blocks at *HMIP* in the Tanzanians, which might include the biologically active variable sites, but not *rs9399137*. Shorter blocks would greatly aid further fine-mapping efforts at this important locus.

Similar to *HMIP*, the β-globin cluster had a muted effect on HbF in the Tanzanian patients, presumably because of low allele frequencies and lack of power. Alleles at *rs7482144* exert a strong effect on HbF, but the A allele (also referred to as Xmn1 $^{G}\gamma$, +) is absent in the Central African Republic or Bantu β^{S} haplotype, which seems prevalent in Tanzania. 12,19,20 Other African populations with the Senegal β^{S} haplotype that contains the *rs7482144* SNP would be better suited to study the effects of this variant. Such differences in haplotype and allele frequencies between populations in Africa provide a strong argument for the necessity of genome-wide association studies carried out in individual African populations.

A strong signal adjacent to the *HBB* cluster, recently detected in African American patients, 8 is significant ($P = .024$, Table 2) in SCA patients from Tanzania but disappears ($P = .14$) when linkage disequilibrium with *rs7482144* is taken into account. In conclusion, this additional HbF locus seems absent in Tanzanians.

The number of British patients studied here is too small to detect more than the strongest markers and to statistically compare findings with those from Tanzanians. The results show similar impacts of the HbF loci (Table 2), with the exception of *rs7482144* at *HBG2*, where no association was seen in British patients. *rs7482144* failed quality control in this group (Hardy-Weinberg test, $P = .02$) because of an excess of homozygote (A/A) genotypes. The fact that patients with this genotype were subsequently found to originate mostly from Sierra Leone illustrates the potentially confounding influence of hidden heterogeneity or admixture.

To uncover new loci and variants controlling HbF in populations where SCA is endemic, genetic studies focused on individual African populations may be more informative, and patient resources across Africa, such as the Muhimbili Sickle Cell Collaborative Program, can make important contributions toward this goal.

Acknowledgments

The authors thank the patients and staff of Muhimbili National Hospital, Muhimbili University of Health and Allied Sciences and Hematology Outpatient Unit, King's College Hospital; Dr Kirk Rockett, The Wellcome Trust Center for Human Genetics in Oxford for crucial advice and help with DNA handling, primer-extension preamplification (PEP), and genotyping procedures; Professor Kevin Marsh (Kenya Medical Research Institute, Kelifi, Kenya) for support; and Claire Steward for help in preparation of the manuscript.

This work was supported by The Wellcome Trust, United Kingdom (JKM 072064; project grant 080025; strategic award 084538), Kenya Medical Research Institute, Center for Geographic Medicine Research–Coast, and Medical Research Council, United Kingdom (G000111, ID51640) (S.L.T.).

References

1. Sebastiani P, Solovieff N, Hartley SW, et al. Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. Am J Hematol. 2010; 85(1):29–35. [PubMed: 20029952]
2. Thein SL. Genetic modifiers of the beta-haemoglobinopathies. Br J Haematol. 2008; 141(3):357–366. [PubMed: 18410570]
3. Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in sickle cell disease: life expectancy and risk factors for early death. N Engl J Med. 1994; 330(23):1639–1644. [PubMed: 7993409]
4. Platt OS, Thorington BD, Brambilla DJ, et al. Pain in sickle cell disease: rates and risk factor. N Engl J Med. 1991; 325(1):11–16. [PubMed: 1710777]
5. Thein SL, Menzel S, Lathrop M, Garner C. Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. Hum Mol Genet. 2009; 18(R2):R216–R223. [PubMed: 19808799]
6. 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 U S A. 2008; 105(33):11869–11874. [PubMed: 18667698]
7. Creary LE, Ulug P, Menzel S, et al. Genetic variation on chromosome 6 influences F cell levels in healthy individuals of African descent and HbF levels in sickle cell patients. PLoS ONE. 2009; 4(1):e4218. [PubMed: 19148297]
8. Solovieff N, Milton JN, Hartley SW, et al. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. Blood. 2010; 115(9):1815–1822. [PubMed: 20018918]
9. Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. Blood. 2010; 115(22):4331–4336. [PubMed: 20233970]
10. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ. 2008; 86(6):480–487. [PubMed: 18568278]
11. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. Science. 2009; 324(5930):1035–1044. [PubMed: 19407144]
12. Nagel, RL., Steinberg, MH. Genetics of the β^S gene: origins, genetic epidemiology, and epistasis in sickle cell anemia. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. Steinberg, MH.Forget, BG.Higgs, DR., Nagel, RL., editors. Cambridge, United Kingdom: Cambridge University Press; 2001. p. 711-755.
13. Zhang L, Cui X, Schmitt K, et al. Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci U S A. 1992; 89(13):5847–5851. [PubMed: 1631067]
14. Craig JE, Sheerin SM, Barnetson R, Thein SL. The molecular basis of HPFH in a British family identified by heteroduplex formation. Br J Haematol. 1993; 84(1):106–110. [PubMed: 7687855]
15. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc Natl Acad Sci U S A. 2008; 105(5):1620–1625. [PubMed: 18245381]
16. Sedgewick AE, Timofeev N, Sebastiani P, et al. BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. Blood Cells Mol Dis. 2008; 41(3):255–258. [PubMed: 18691915]
17. Menzel S, Garner C, Gut I, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat Genet. 2007; 39(10):1197–1199. [PubMed: 17767159]
18. 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 U S A. 2007; 104(27):11346–11351. [PubMed: 17592125]

19. 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 U S A. 1985; 82(7):2111–2114. [PubMed: 2580306]
20. Nagel RL, Fabry ME, Pagnier J, et al. Hematologically and genetically distinct forms of sickle cell anemia in Africa: the Senegal type and the Benin type. N Engl J Med. 1985; 312(14):880–884. [PubMed: 2579336]
21. [Accessed March 2006] University of California Santa Cruz genome browser. http://genome.uscs.edu/cgi-bin/hgGateway

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Summary data for both patient populations

		Tanzanian patients	British patients
N	HbS/HbS	529 female, 516 male	82 female, 64 male
	HbS/β^0	0	2 female, 3 male
Age, y	Median	13	29
	Range	5-45	11-64
HbF	Geometric mean, %	4.40	5.60
	Interquartile range, %	2.7-7.8	3.2-10.3

Europe PMC Funders Author Manuscripts

Blood. Author manuscript; available in PMC 2017 August 14.

Table 2

Association of the 3 principal loci (BCL11A, HMIP, and the β-globin gene cluster) with HbF levels

Genetic variants tested				Tanzanian HbSS patients				British HbSS patients			
Locus	SNP name	Position* on chromosome	Allele change	Minor allele frequency	Effect† of allele change	*P*	n	Minor allele frequency	Effect† of allele change	*P*‡	n
Chromosome 2											
BCL11A	*rs11886868*	60 573 750	C → T	0.26	−0.406	3.0×10^{-30}	1000	0.29	−0.302	8.5×10^{-4}	146
BCL11A	*rs4671393*	60 574 455	A → G	0.30	−0.412	3.9×10^{-28}	845	0.28	−0.440	3.7×10^{-6}	145
Chromosome 6											
HMIP-1	*rs28384513*	135 417 902	A → C	0.21	−0.146	1.9×10^{-4}§	1021	0.21	+0.103	.36	142
HMIP-2	*rs9376090*	135 452 920	T → C	0.01	+0.471	.016	1021	0.02	+0.581	.058	147
HMIP-2	*rs9399137*	135 460 710	T → C	0.01	+0.668	8.3×10^{-6}	975	0.07	+0.529	.006	126
HMIP-2	*rs9389269*	135 468 851	T → C	0.03	+0.400	1.4×10^{-5}§	1016	0.04	+0.461	.031	146
HMIP-2	*rs9402686*	135 469 509	G → A	0.06	+0.342	1.6×10^{-4}§	1013	0.04	+0.460	.031	146
HMIP-2	*rs9494142*	135 473 333	T → C	0.13	+0.085	.060	1014	0.10	+0.218	.127	144
Chromosome 11											
HBG2	*rs7482144*‖	5 232 745	G → A	0.01	+0.562	1.6×10^{-4}	991	0.07	+0.136	.43	147
OR51B5/6	*rs5006884*	5 329 827	C → T	0.05	+0.164	.024	957	0.13	+0.217	.15	145

Representative SNPs for *BCL11A* on chromosome 2, the *HBS1L-MYB* intergenic polymorphism (*HMIP*) on chromosome 6, and *HBG2* on chromosome 11, which encodes G_γ-globin are shown. *rs7482144* has been extensively studied as the *Xmn1* G_γ*-polymorphism, a component of the classic sickle haplotypes, where the alternative allele A is detected as positive (site present, Senegal and Arab/Indian haplotypes). Allele frequencies in the patient population were derived from all genotyped individuals (n > 1300).*

* The chromosomal position given here is based on the public human genome assembly National Center for Biotechnology Information 36.1 (University of California Santa Cruz, genome browser21).

† "+" or "−" indicates whether the allele change elevates or depresses HbF (expressed as ln[% HbF]). The value reported here (the regression coefficient β) indicates by how much the trait value changes, on average, when one of the alleles in a genotype is changed. This is also referred to as the additive allelic effect.

‡ Although the allelic effect sizes (β) are often comparable between the 2 groups, *P* values are much larger in the British patients because of the smaller number of subjects studied.

§ Association with these makers remains significant at $P < .001$ after linkage disequilibrium with *rs9399137* is taken into account (ie, these markers display HbF association independent of *rs9399137*).

‖ *rs7482144* is also referred to as the *Xmn*1-G_γ site, where "+" indicates the alternative allele A.