Mtatio, SN; Mgaya, J; Singh, T; Mariki, H; Rooks, H; Soka, D; Mmbando, B; Thein, SL; Barrett, JC; Makani, J; Cox, SE; Menzel, S (2015) 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 medical genetics, 16 (1). p. 4. ISSN 1471-2350 DOI: https://doi.org/10.1186/s12881-015-0148-3

Downloaded from: http://researchonline.lshtm.ac.uk/2162925/

DOI: 10.1186/s12881-015-0148-3

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
Genetic association of fetal-hemoglobin levels in individuals with sickle cell disease in Tanzania maps to conserved regulatory elements within the MYB core enhancer

Siana N Mtatiro1,2,3,7*, Josephine Mgaya1, Tarjinder Singh4, Harvest Mariki1, Helen Rooks3, Deogratius Soka1, Bruno Mmbando1,5, Swee Lay Thein3, Jeffrey C Barrett4, Julie Makani1, Sharon E Cox1,6 and Stephan Menzel3,8*

Abstract

Background: Common genetic variants residing near upstream regulatory elements for MYB, the gene encoding transcription factor cMYB, promote the persistence of fetal hemoglobin (HbF) into adulthood. While they have no consequences in healthy individuals, high HbF levels have major clinical benefits in patients with sickle cell disease (SCD) or β thalassemia. Here, we present our detailed investigation of HBS1L-MYB intergenic polymorphism block 2 (HMIP-2), the central component of the complex quantitative-trait locus upstream of MYB, in 1,022 individuals with SCD in Tanzania.

Methods: We have looked at 1022 individuals with HbSS or HbS/β0 in Tanzania. In order to achieve a detailed analysis of HMIP-2, we performed targeted genotyping for a total of 10 SNPs and extracted additional 528 SNPs information from a genome wide scan involving the same population. Using MACH, we utilized the existing YRI data from 1000 genomes to impute 54 SNPs situated within HMIP-2.

Results: Seven HbF-increasing, low-frequency variants (β > 0.3, p < 10^-5, f ≤ 0.05) were located in two partially-independent sub-loci, HMIP-2A and HMIP-2B. The spectrum of haplotypes carrying such alleles was diverse when compared to European and West African reference populations: we detected one such haplotype at sub-locus HMIP-2A, two at HMIP-2B, and a fourth including high-HbF alleles at both sub-loci (‘Eurasian’ haplotype clade). In the region of HMIP-2A a putative functional variant (a 3-bp indel) has been described previously, but no such candidate causative variant exists at HMIP-2B. Extending our dataset through imputation with 1000 Genomes, whole-genome-sequence data, we have mapped peak association at HMIP-2B to an 11-kb region around rs9494145 and rs9483788, flanked by two conserved regulatory elements for MYB.

Conclusions: Studies in populations from the African continent provide distinct opportunities for mapping disease-modifying genetic loci, especially for conditions that are highly prevalent there, such as SCD. Population-genetic characteristics of our cohort, such as ethnic diversity and the predominance of shorter, African-type haplotypes, can add to the power of such studies. 

* Correspondence: siananikya@muhimbili-wellcome.org; stephan.menzel@kcl.ac.uk

1 Muhimbili Wellcome Programme, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania
3 King’s College London, Molecular Haematology, Division of Cancer Studies, London, UK

© 2015 Mtatiro et al; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Sickle cell disease (SCD) is a hemoglobin disorder caused by the Glu6Val mutation in the β chain of adult hemoglobin. The resulting hemoglobin variant, HbS, is prone to polymerization, disrupting red blood cell shape, function and life span. SCD is prevalent in Sub-Saharan Africa, where it is a significant contributor to childhood mortality [1]. In Tanzania, 8,000-11,000 affected children are born annually [2]. The most common and severe forms of the disease are due to homozygosity for the mutation (HbSS) or compound heterozygosity with β0 thalassemia (HbS/β0thalassemia). Where newborn screening and prophylactic penicillin are available, childhood mortality due to SCD is significantly reduced, but patients nevertheless remain at risk for chronic complications and premature death. The disease is milder in those patients that carry significant amounts of fetal hemoglobin (Hbf) in their circulating red blood cells [3]. Similar to healthy populations, Hbf persistence in patients with SCD is partially genetically controlled, and three Hbf quantitative-trait loci (QTLs) - HBG2 [4,5], BCL1A [6,7] and HBSIL-MYB [8] - have been identified. Knowledge of the genetic factors underlying Hbf persistence is helping to interpret the clinical variability of SCD and has led to the identification of novel molecular targets for the therapeutic reactivation of Hbf.

HBSIL-MYB is unique among the Hbf modifier loci because it has marked pleiotropic effects, i.e., in healthy individuals it affects general hematological parameters [9] as well as Hbf. It has been postulated that changes in Hbf levels caused by this locus are secondary to altered kinetics of erythropoiesis [10]. The locus consists of several linkage disequilibrium (LD) blocks of common variants, which affect erythroid traits independently [8]. The most effective of these, termed HMIP-2 (HBSIL-MYB intergenic polymorphism, block 2) has been shown to influence disease severity in patients with SCD [11] and β thalassemia [8,12]. HMIP-2 variants reside within the core enhancer for MYB [13], a key hematopoietic regulator gene [14]. It is divided further into sub-loci HMIP-2A and -2B, which provide independent Hbf association in African populations, including SCD patients [11,15-18]. A 3-bp deletion (rs66650371) at HMIP-2A is suspected to directly cause Hbf variability [19], but is independent of the trait association seen at HMIP-2B. Therefore, causative variants acting at HMIP-2B are still to be discovered.

To better define the Hbf association signal at HMIP-2B, and to identify candidate variants for trait causation, we dissected HMIP-2 and its effect on Hbf persistence in a large SCD patient cohort from Tanzania. The Tanzanian population is well-suited to genetic fine-mapping studies, with a marked ethnic diversity [20,21] and the increased mapping resolution that is characteristic for African chromosomes [22,23].

Methods

Study subjects, sample collection and phenotyping

Only patients with Hb SS or HbS/β0 thalassemia genotype were included in this study. Enrollment of patients, diagnosis and confirmation of sickle phenotypes as well as the quantification of hemoglobin subtypes were performed as previously described [24]. Informed consent was obtained for each patient and ethical approval given by the Muhimbili University Research and Publications Committee (MU/RP/AEC/VOLX1/33). During follow-up clinics, a 2-ml blood sample was collected from non-transfused SCD patients (confirmed Hb SS genotype) who are not on hydroxyurea treatment. This study includes 1,022 individuals with Hbf measured (by HPLC, Variant I, Biorad, Hercules, CA, USA) at the age of 5 years or older. The median age of the SCD population is 11 years; males and females are represented equally. Hbf values vary significantly, with a median of 5.4% (of total hemoglobin).

Genotyping

DNA was extracted from archived buffy coat using the Nucleon BACC II system (GE Healthcare, Little Chalfont, UK). Genotypes for 528 regional SNPs were extracted from a genome-wide SNP set generated at the Wellcome Trust Sanger Institute on the Human Omnichip 2.5 platform (Illumina, La Jolla, CA, USA), as described elsewhere [25]. Targeted genotyping was performed, adding ten markers with known trait association: rs9376090, rs9399137, rs9402686, rs9389269 and rs9494142 by TaqMan procedure [16], rs9389268 and rs9376091 by PCR product sequencing (amplification and sequencing: F: 5'-TGCTTCTTGGCAGT GAATTACCTTTGT-3', R: 5'-AGTTTGGTGCCAAAGG TAGCAGAT-3'), indels rs66650371 and rs11321816 by multiplex PCR fragment sizing (F1: 5'-GTTTGAGTTGTG CAGAAGACAAAAGC-3', R1: 5'-VIC-TAAGTGCTCCTT GAGGAAACC-3', F2: 5'-FAM-TCACCTTTAAAAAGGC CGGTATTG-3', R2: 5'-GTGT-AAGCCTTTGGAAGCAT-3' and rs35786788 by SNpShot procedure (F5'-FAM TCACCTTTAAAAAGGC CGGTATTG-3', R5'-GTGT-AAGC ACTTTGGCAAGCAT-3', extension: 5'-ACTATATCTGT GCACAGAAATACAG-3'). All assays were performed under supplier-recommended conditions (Applied Biosystems, Foster City, CA, US), including the fragment sizing, which used the Taq Gold (Applied Biosystems) microsatellite genotyping protocol. Fragment sizes and SNpShot products were evaluated by capillary electrophoresis (3130 Genetic analyzer, Applied Biosystems), with subsequent allele scoring using GeneMarker v1.95 (SoftGenetics, State College, PA, USA). Marker quality control consisted of Hardy-Weinberg equilibrium testing and call rate evaluation (cut off >80%). Imputation with MACH 1.0 [26,27] was used to fill in missing genotypes.
Statistical analysis
Phased variant call files from the 1000 Genomes project [28] for the YRI population sample were accessed on 24/4/2013 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr6.phase1_release.v3.20110123.snps_indels.genotypes.vcf.gz) using the ‘Data Slicer’ tool at http://browser.1000genomes.org. haplotype files were derived, purged of non-informative variants (monomorphic and singletons) and used to impute 54 non-genotyped variants in the target area, using MACH 1.0.16 [29,30]. GWAS data was processed with the PLINK software package (http://pngu.mgh.harvard.edu/purcell/plink/). Test for genetic association with ln[\%HbF], including conditional analysis, was performed with STATA v12 (Stata Corp, College Station, TX) using multiple linear regression with age and sex as covariates. Haplotype relative effects were estimated using multifactor ANOVA in R (http://www.r-project.org/), correcting for pair-wise comparison using Tukey’s method, including age and sex as covariates.

Results
Genetic association with HbF in the HBS1L-MYB intergenic region
In 1,022 SCD patients, we scanned the HBS1L-MYB intergenic region (chr6:135,318,635-135,518,635, Figure 1) for genetic association with HbF levels (ln[\%HbF]), evaluating 538 SNP (single nucleotide polymorphism) markers from a combination of genome-wide chip array and targeted genotyping. Confirming our own previous findings [9,15] and replicating what is seen across human populations [4,6-13], we detected trait association over the entire region, with the peak signal (P < 10^{-7}) located at the HMIP-2 locus (Figure 1, Table 1). Strong trait association at sub-locus HMIP-2A was detected with rs66650371 (the 3-bp indel proposed to be functional [19]), rs35786788 and rs9399137, and at HMIP-2B with rs9494145, rs9389269 and rs9402686 (Figure 1, Table 1). The partial independence of the two sub-loci was confirmed in our dataset by conditional analysis (Table 1).

African-specific trait association at HMIP-2B
In contrast to HMIP-2A, a good candidate variant for causing the HbF association at HMIP-2B has so far not been identified. We therefore studied HbF association at this sub-locus in more detail. We included additional variants from 1000 Genomes project sequence data [28] through imputation with YRI (Yoruba, West African) haplotypes, which are well-matched with our population regarding the allele frequencies of regional markers [18]. The YRI dataset harbors 51 polymorphic variants within the HMIP-2B region (chr6:135,426,573-135,435,501), and another 17 such variants in an additional 5-kb fragment we decided to include on the 3’ side of the locus, adjacent to MYB. 15 patients carrying European-type high-HbF haplotypes (tagged by rs1376090 [18]) were excluded from imputation and analysis to make full use of the increased mapping resolution of the shorter African haplotypes, with the goal to localize the QTL more precisely. Association analysis subsequent to imputation identified an additional SNP with strong HbF association, rs9483788 (Figure 2). The area of significant association (p < 0.01) thus extends between chr6:135,427,000 and 135,438,000, with the peak between MYB enhancer elements −71 and −63 (Figure 2).

HMIP-2 haplotypes
To dissect the haplotype architecture underlying the trait association pattern at HMIP-2, we phase-aligned genotypes for the seven strongly-associated markers. To relate our data to findings in other populations, we also included rs9376090, which tags European and Asian high-HbF alleles, and rs4895441, which is part of the HMIP-2B sub-locus in other SCD patient populations.

The most prevalent haplotype carried the reference (low-HbF) allele in all positions, representing the global ancestral situation, universally associated with low HbF levels [18]. The other haplotypes contained high-HbF associated alleles in at least one position (Figure 3). An Asian-type haplotype [18] (high-HbF alleles at all seven positions, including the ancestry-informative marker rs9376090) was detected at low frequency, but with high HbF levels (9.3%, median, Figure 4). The remaining haplotypes belonged to two minor clades more common in African populations, one with HbF-promoting alleles at HMIP-2A only (termed clade A-b’), and one with HbF-promoting alleles at HMIP-2B only (clade a-B’). In Tanzanian patients, A-b’ was represented by a single haplotype, whereas the a-B’ clade was more diverse (haplotypes a-B1’, a-B2’, a-B3’). A-b’, a-B2’ and a-B3’ haplotypes all significantly boost HbF levels, compared to the ancestral haplotype a-B’ (Table 2), while a-B1’ does not. a-B1’ contains only a single non-ancestral allele, rs4895441-G, which is not strongly HbF-associated in Tanzanian patients. a-B3’, which contains two high-HbF associated alleles, rs9494145-C and rs9483788-C, significantly increases HbF levels over the ancestral haplotype. A stronger effect is detected when all four HbF-boosting alleles, rs9389269-C, rs9402686-A, rs9494145-C and rs9483788-C, are present (in a-B2’, Table 2).

Discussion
Performing trait-association for SNPs across the HBS1L-MYB intergenic region on chromosome 6q24.3 in SCD patients from Tanzania, we detected significant association with HbF levels at HMIP-2, a globally-prevalent HbF QTL [7-9,11,15,17,19,31-35] residing within the
Figure 1 Association of common genetic variation with HbF levels across the HBS1L-MYB intergenic region on chromosome 6q23.3. Shown are LOP (−log_{10} (P-value)) scores for 1,022 patients, tested for association of ln[%HbF] with 109 common variants from a combination of Illumina Human Omnichip 2.5 data and PCR-based genotyping. A: Genetic association is present over the entire interval, but reaches genome-wide significance only at HMIP-2, whereas the other two LD blocks detected in Europeans (HMIP-1 and HMIP-3, [8]) display only low-level association in our dataset. B: HbF association at HMIP-2. The six SNPs showing strong association are indicated. In addition to un-conditioned analysis (black dots), the presence two partially independent association signals (sub-loci HMIP-2A and HMIP-2B) is shown by conditioning analysis on rs66650371 (tagging HMIP-2A, black diamonds) and rs9494145 (tagging HMIP-2B, open diamonds, see also Table 1).
Table 1 Association of HMIP-2 variants with fetal-hemoglobin levels (ln[HbF%]) in Tanzanian patients with SCA

<table>
<thead>
<tr>
<th>Marker</th>
<th>Sub-locus</th>
<th>Chr 6 position</th>
<th>allele change</th>
<th>MAF</th>
<th>β</th>
<th>E</th>
<th>P*</th>
<th>E</th>
<th>P**</th>
<th>E</th>
<th>P***</th>
<th>E</th>
<th>P****</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9376090</td>
<td>135,411,228</td>
<td>T → C</td>
<td>0.01</td>
<td>0.78</td>
<td>2.02x10^{-4}</td>
<td>0.17</td>
<td>0.52</td>
<td>0.51</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs41294858</td>
<td>135,412,636</td>
<td>T → C</td>
<td>0.18</td>
<td>-0.16</td>
<td>3.74x10^{-4}</td>
<td>-0.14</td>
<td>1.28x10^{-3}</td>
<td>-0.13</td>
<td>3.35x10^{-3}</td>
<td>-0.12</td>
<td>6.07x10^{-3}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs55731938</td>
<td>135,414,850</td>
<td>G → A</td>
<td>0.21</td>
<td>-0.16</td>
<td>1.65x10^{-4}</td>
<td>-0.15</td>
<td>4.85x10^{-4}</td>
<td>-0.13</td>
<td>1.79x10^{-3}</td>
<td>-0.13</td>
<td>2.68x10^{-3}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7745098</td>
<td>135,415,004</td>
<td>C → T</td>
<td>0.19</td>
<td>-0.16</td>
<td>1.87x10^{-4}</td>
<td>-0.15</td>
<td>7.52x10^{-4}</td>
<td>-0.14</td>
<td>1.78x10^{-3}</td>
<td>-0.13</td>
<td>3.58x10^{-3}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs76267242</td>
<td>135,417,460</td>
<td>G → T</td>
<td>0.19</td>
<td>-0.16</td>
<td>1.87x10^{-4}</td>
<td>-0.15</td>
<td>7.52x10^{-4}</td>
<td>-0.14</td>
<td>1.78x10^{-3}</td>
<td>-0.13</td>
<td>3.58x10^{-3}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs66650371</td>
<td>135,418,635</td>
<td>I → D</td>
<td>0.02</td>
<td>0.69</td>
<td>4.06x10^{-8}</td>
<td>0.59</td>
<td>3.94x10^{-6}</td>
<td>0.13</td>
<td>0.67</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9399137</td>
<td>135,419,018</td>
<td>T → C</td>
<td>0.02</td>
<td>0.63</td>
<td>3.65x10^{-7}</td>
<td>0.14</td>
<td>0.6</td>
<td>0.54</td>
<td>1.87x10^{-5}</td>
<td>0.13</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11321816</td>
<td>135,419,038</td>
<td>I → D</td>
<td>0.14</td>
<td>0.12</td>
<td>1.36x10^{-2}</td>
<td>0.05</td>
<td>0.36</td>
<td>0.12</td>
<td>0.02</td>
<td>0.06</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs35786788</td>
<td>135,419,042</td>
<td>G → A</td>
<td>0.02</td>
<td>0.69</td>
<td>5.07x10^{-8}</td>
<td>0.35</td>
<td>0.19</td>
<td>0.61</td>
<td>2.44x10^{-6}</td>
<td>0.37</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1074849</td>
<td>135,423,412</td>
<td>G → A</td>
<td>0.13</td>
<td>-0.19</td>
<td>3.09x10^{-4}</td>
<td>-0.18</td>
<td>5.22x10^{-4}</td>
<td>-0.17</td>
<td>1.11x10^{-3}</td>
<td>-0.16</td>
<td>1.37x10^{-3}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4895441</td>
<td>135,426,573</td>
<td>A → G</td>
<td>0.05</td>
<td>0.30</td>
<td>1.53x10^{-4}</td>
<td>0.23</td>
<td>4.06x10^{-3}</td>
<td>0.13</td>
<td>0.16</td>
<td>0.09</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9389269</td>
<td>135,427,159</td>
<td>T → C</td>
<td>0.03</td>
<td>0.48</td>
<td>1.11x10^{-6}</td>
<td>0.37</td>
<td>2.76x10^{-4}</td>
<td>0.27</td>
<td>0.04</td>
<td>0.16</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9402686</td>
<td>135,427,817</td>
<td>G → A</td>
<td>0.03</td>
<td>0.49</td>
<td>1.57x10^{-6}</td>
<td>0.36</td>
<td>6.51x10^{-4}</td>
<td>0.27</td>
<td>0.04</td>
<td>0.14</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1411919</td>
<td>135,432,061</td>
<td>A → G</td>
<td>0.34</td>
<td>-0.10</td>
<td>8.67x10^{-3}</td>
<td>-0.09</td>
<td>0.01</td>
<td>-0.06</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9494145</td>
<td>135,432,552</td>
<td>T → C</td>
<td>0.05</td>
<td>0.38</td>
<td>4.97x10^{-7}</td>
<td>0.31</td>
<td>3.45x10^{-5}</td>
<td>0.31</td>
<td>3.45x10^{-5}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2223385</td>
<td>135,435,171</td>
<td>G → A</td>
<td>0.39</td>
<td>-0.11</td>
<td>1.74x10^{-3}</td>
<td>-0.11</td>
<td>2.77x10^{-3}</td>
<td>-0.08</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1320963</td>
<td>135,443,212</td>
<td>A → G</td>
<td>0.33</td>
<td>0.09</td>
<td>1.44x10^{-2}</td>
<td>0.07</td>
<td>0.05</td>
<td>0.03</td>
<td>0.44</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1569534</td>
<td>135,451,580</td>
<td>C → T</td>
<td>0.15</td>
<td>-0.12</td>
<td>1.25x10^{-2}</td>
<td>-0.11</td>
<td>0.02</td>
<td>-0.10</td>
<td>0.04</td>
<td>-0.09</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6929404</td>
<td>135,454,027</td>
<td>C → A</td>
<td>0.49</td>
<td>0.11</td>
<td>2.35x10^{-3}</td>
<td>0.09</td>
<td>6.64x10^{-3}</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear regression analysis: *conditioned for rs66650371, **conditioned on rs9494145, ***conditioned on rs66650371 and rs9494145. (Residual association extends upstream of HMIP-2 and overlaps with the previously reported (Thein et al., PNAS 2007) HMIP-1 signal.)

The chromosomal position is given in hg19 coordinates, derived using UCSC Genome Browser version February 2009.

MAF: Minor allele frequency of the patient cohort; I: inserted allele, D: deleted allele (Δ’TAY’). ß, the un-standardized regression co-efficient is given as a measure of the effect of the allele change on ln[HbF].

Shown are all directly genotyped markers with nominally significant association at p < 0.01.
MYB enhancer region [13]. Some of these variants have been also associated with white blood counts, mean cell volume and mean cell hemoglobin in our population [36]. Our interest was focused on sub-locus HMIP-2B, where a causative variant has not yet been identified. After excluding patients with longer, ‘Eurasian’-type [18] high-HbF haplotypes and including imputed variants from the YRI (Yoruba, 1000 Genomes sequence data) population, we determined the most likely map location of HMIP-2B as an 11-kb segment including the enhancer core element −71 and the interval between elements −71 and −63 (Figure 2), where peak association (rs9494145, rs9483788) was detected.

The two HbF-boosting haplotypes underlying this association peak, ‘a – B1’ and ‘a – B3’, share rs9494145-C and rs9483788-C (Figure 3), ‘a - B1’, which contains all four HbF-boosting alleles (rs9389269-C, rs9402686-A, rs9494145-C and rs9483788-C), has the stronger effect of the two. This means that none of the four strongly trait-associated SNPs detected at HMIP-2B in Tanzanian patients appears to fulfill the conditions for being the singular causative variant, i.e. both, being necessary to show a significant effect and sufficient to produce the maximum genetic effect originating from this sub-locus. Thus, additional variants, not present in the 1000 Genomes dataset, might contribute to trait variability.

Long, ‘Eurasian-type’ (with high-HbF associated alleles across all of HMIP-2 [18]), high-HbF haplotypes were present in the patient cohort at a low frequency. These haplotypes are tagged by the ancestry-informative allele rs9376090-C (Figure 3). 24% of individuals with such haplotypes reported Arabic parental ethnicity, compared to 2% in the general cohort. The high HbF levels we observed in such patients (a median of 9.3% in ‘A-B’/a-b’ heterozygotes, Figure 4) are likely due to the presence of

### Table 2: Relative effects of HMIP-2 haplotypes on the In[HbF%] trait

<table>
<thead>
<tr>
<th>Compared haplotypes</th>
<th>Difference</th>
<th>lower</th>
<th>upper</th>
<th>p adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-B2 vs. a-B1</td>
<td>0.029</td>
<td>-0.35</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>a-B1 vs. A-B</td>
<td>0.431</td>
<td>0.08</td>
<td>0.78</td>
<td>7 x 10^-3</td>
</tr>
<tr>
<td>a-B3 vs. a-b</td>
<td>0.332</td>
<td>0.01</td>
<td>0.66</td>
<td>0.04</td>
</tr>
<tr>
<td>A-b vs. a-b</td>
<td>0.636</td>
<td>0.10</td>
<td>1.18</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\*Difference represents the difference in effect size between the two haplotypes on log-transformed HbF levels; negative difference values indicates that the effect size of the second haplotype is bigger than that of the first. ‘Lower’ and ‘upper’ represent boundaries for family-wise 95% confidence intervals. p adjusted is the P-value adjusted for age, sex as well as multiple testing.

Three haplotypes (shown in bold) have significant HbF-boosting effects at the 0.05 level.
the 3-bp deleted allele at HMIP-2A and possibly another functional allele at HMIP-2B. Population stratification might also contribute to higher levels of HbF: Arab/Indian sickle mutation haplotypes on chromosome 11 are known to result in milder disease and high HbF levels [37]. We also observed a residual association after conditioning on HMIP-2A and HMIP-2B. We suspect that these are part of a group of linked SNPs that overlaps the physical location of HMIP-1, a HbF QTL detected upstream of HMIP-2 (Figure 1A) in the European population [8]. However we didn’t feel we have the power to investigate this further with the present dataset.

Conclusions

We have localized HMIP-2B, a QTL for fetal-hemoglobin persistence, to an 11-kb region within the core enhancer for MYB. So far, we have not identified a likely functional variant within or at this locus. Further studies will involve extended sequence analysis in groups of patients carrying a-B2 and a-B3 haplotypes.

Competing interest

The authors declare that they have no competing interests.

Authors’ contribution

Contribution: SLT, SM, SNM, JC6 and J Makani designed the study. DS, HM and J Mgaya collected the samples and phenotype data. SM, HR and SNM commented on the drafts of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors thank the patients and staff of Muhimbili National Hospital, Muhimbili University of Health and Allied Sciences, Tanzania, Haematology Outpatient Unit and staff of King’s College Hospital, London, and members of Professor Thein’s Molecular Hematology group, King’s College London. This work was supported by Wellcome Trust (Grant no: 095009, 095737, 090025 & 084538) and Commonwealth split-side fellowship (TZCN-2012-361). The sponsors of this study are nonprofit organizations that support science in developing countries. 2nd ed. Washington (DC): World Bank; 2006. p. 663–80.

References


