Hb F Levels in β-Thalassemia Carriers and Normal Individuals: Known and Unknown Quantitative Trait Loci in the β-Globin Gene Cluster

Licínio Manco, Raquel Santos, Catarina Rocha, Luís Relvas, Celeste Bento, Tabita Maia, Verónica Gomes, António Amorim, and Maria J. Prata

ABSTRACT

In the already identified quantitative trait loci (QTL) modulating Hb F levels are cis-acting haplotypes of the β-globin gene cluster itself, although the single nucleotide polymorphisms (SNPs) accounting more for the association, remain uncertain. In this study, the role in Hb F production of previously reported candidate SNPs within the β-globin gene cluster was reexamined, along with a yet poorly studied variation in the BGLT3 gene. In a sample of β-thalassemia (β-thal) carriers, we succeeded in replicating the significant association between increased Hb F levels and rs7482144 (C>T) (HBG2 XmnI), which is the most well-established variation in the cluster influencing the trait. This SNP was found to be in strong linkage disequilibrium (LD) with a variation in the HBBP1 gene, which consistently revealed a similar association signal. Remarkably, much stronger than the latter associations were those involving both rs968857 (T allele) (3′ HBBP1) and rs7924684 (G allele) (BGLT3), two SNPs that were also in strong LD. As the pattern of LD detected in the β-globin gene cluster does not correlate with a tight linkage between markers, complex interactions between SNPs at the cluster seem to modulate Hb F. Seeing that no such associations were detected in normal subjects, the question can be raised on whether, under erythropoiesis stress, epigenetic mechanisms contribute to change the regulation of the entire β-globin gene cluster. In conclusion, we provide statistical evidence for a new player within the β-globin gene cluster, BGLT3, that in cooperation with other regions influences Hb F levels in β-thal carriers.

ARTICLE HISTORY

Received 11 June 2021
Revised 21 January 2022
Accepted 21 February 2022

KEYWORDS

β-Globin gene cluster; β-thalassemia (β-thal) carriers; Hb F regulation; HBBP1 and BGLT3 genes; hereditary persistence of fetal hemoglobin (HPFH)

Introduction

Hb F (α2γ2) is a major modulator of the high diversity of clinical phenotypes associated to β hemoglobinopathies such as sickle cell disease and β-thalassemia (β-thal) [1]. Thus, knowledge of the genetic factors and pathways that regulate the switch from γ- to adult β-globin in humans has the potential to highlight new therapeutic targets aiming to increase Hb F production levels [1–3]. In recent years, genetic association studies revealed a few γ-globin repressors in adult cells, including the major transcriptional repressors BCL11A and MYB that are encoded by genes where candidate single nucleotide polymorphisms (SNPs) were identified [4–6].

It is also well documented that the causes of unusually high Hb F levels in adults, collectively known as hereditary persistence of fetal hemoglobin (HbF), are deletions or single base substitutions in the β-globin gene cluster (11p15.4) [3,7]. Epidemiological and clinical genetic studies have implicated the XmnI −158 (C>T) [rs7482144 (C>T)] polymorphism, located upstream to the HBG2 gene, as a common variant influencing Hb F levels mainly under conditions of erythropoietic stress [1]. This SNP (also known as the XmnI polymorphism) is one of the shared variations usually examined to infer ββ (Hb S or HBB: c.20A>T) and β-thal haplotypes, which have been widely addressed in the field of hemoglobinopathies because some of them are associated with increased Hb F levels and milder clinical courses. For example, in sickle cell disease, the Senegal and Arab-Indian haplotypes, the two conventional haplotypes that carry the same allele T at the HBG2 XmnI site, relate to high levels of Hb F (>15.0%) and less severe phenotypes, whereas the Bantu haplotype, which carries the alternative allele C at HBG2 XmnI, is related to lower levels of Hb F (<5.0%) and more severe phenotypes [1]. However, whether the variant responsible for the association signal is rs7482144, or instead, is its haplotypic combination with other variants in the cluster, remains uncertain.

Elevated levels of Hb F are commonly found in patients with deletions or point mutations in the intergenic region between genes HBG1 (Cγ) and HBD (δ), indicating that this
region plays an important role in regulating γ-globin expression [8–10]. Notably, in this precise region are located two genes that produce long non coding RNAs (lncRNAs), the Hb subunit βψ1 (HBBP1) and the β-globin locus transcript 3 (BGLT3) [11]. It was demonstrated that the HBBP1 region is required for γ-globin silencing in adult erythrocyte cells, possibly by binding the major γ-globin repressor BCL11A [11]. Moreover, a variant in the second intron of HBBP1, rs10128556 (G>C), was found to be strongly associated with elevated Hb F levels in sickle cell disease [12,13] and also in β-thal [10,14]. This SNP was found to be in LD with the HBG2 XmnI (rs7482144) polymorphism in sickle cell disease patients from India [13], Saudi Arabia [15] and African Americans [12].

With regard to the lncRNA BGLT3 gene, a recent study by Ivaldi et al. [16] showed that, in contrast to the HBBP1 gene, this locus is a positive regulator of γ-globin gene expression, possibly through multiple mechanisms including looping to the γ-globin genes and recruiting coregulators via the BGLT3 transcript itself. Still concerning the HBG1-HBD intergenic region, there is another SNP, rs968857 (T>C), located downstream to the HBBP1 gene, that was also previously identified as a significant contributor to differences in Hb F levels in sickle cell disease patients [9].

The present study intends to add novel insights into the relationship between Hb F levels and polymorphic variations within the β-globin gene cluster, by analyzing two samples of Portuguese subjects, one of β-thal carriers and the second of normal individuals with Hb F ranging from standard levels to common forms of HPFH. We have examined a total of seven SNPs in the β-globin gene cluster (Figure 1) that were analyzed either individually or in the context of the inferred haplotypic combinations. Six of the SNPs define the distinct background haplotypes of the HBB mutation underlying sickle cell disease, representing a set that includes the three SNPs within the β-globin gene cluster that, individually, have accumulated the strongest evidence of influencing Hb F levels [the afore mentioned rs7482144 (C>T), rs968857 (T>C) and rs10128556 (G>A) SNPs]. The seventh SNP is a variation in BGLT3 [rs7924684 (G>A)], here analyzed for the first time as a modulator of Hb F levels in β-thal carriers, extending the coverage of the HBG1-HBD intergenic region to three polymorphic variations.

### Material and methods

#### Populations

We studied 71 Portuguese β-thal carriers (36 males, 35 females), aged between 2–77 years old (mean 32.76 years), with Hb F levels ranging from 0.2–8.6% (mean 1.95%). The hematological parameters of the study population are shown in Table 1. This study population is an enlarged sample of that previously analyzed by Pereira et al. [17]. Hb A2 levels were above the normal level, ranging from 3.5–6.1% (mean 4.78%). Hb A2 and Hb F levels were determined by high performance liquid chromatography (HPLC) (VARIANT II™, Bio-Rad Laboratories Inc., Hercules, CA, USA). Seven different HBB mutations were identified in heterozygous β-thal minor subjects: IVS-1-6 (T>C) (HBB: c.92+6T>C) 27 subjects; codon 39 (CAG>TAG) (HBB: c.118C>T) 22 subjects; codon 15 (TGG>TGA) (HBB: c.48G>A) eight subjects; IVS-I-1 (G>A) (HBB: c.92+1G>A) seven subjects; codon 15 (TGG>TAG) (HBB: c.47G>A) three subjects; IVS-I-110 (G>A) (HBB: c.93-21G>A) two subjects, and codon 6 (−A (HBB: c.20delA) two subjects.

A second population was also analyzed, consisting of 51 healthy subjects (33 females; 18 males), aged 2–75 years old (mean 32.02 years), with normal Hb A2 levels (ranging from 2.0–3.4%; mean 2.66%). This population, herein referred to as ‘normal population,’ consisted of a control sub-group of 36 subjects with normal Hb F levels, ranging from 0.1–1.6% (mean 0.697%), and another sub-group of 15 subjects having

### Table 1. Description of demographic and hematological data (mean ± SD) in the two analyzed populations of β-thalassemia carriers and healthy individuals.

| Parameters | β-Thal Carriers | Normal Population |
|------------|----------------|------------------|
| Age (years) | 32.76 ± 20.59 | 32.02 ± 20.58 |
| Age range (years) | 2–77 | 2–75 |
| Gender (n) | M: 36; F: 35 | M: 18; F: 33 |
| Hb F (%) | 1.95 ± 1.80 | 1.56 ± 1.87 |
| Hb F (C) | 0.2–8.6 | 0.1–9.1 |
| Hb A2 (%) | 4.78 ± 0.66 | 2.66 ± 0.31 |
| Hb A2 (range) | 3.5–6.1 | 2.0–3.4 |
| Hb (g/dL) | 11.66 ± 1.14 | 13.21 ± 2.01 |
| MCV (fl) | 67.56 ± 5.87 | 84.33 ± 10.11 |
| MCH (pg) | 21.25 ± 2.14 | 27.41 ± 2.42 |
| MCHC (g/dL) | 31.93 ± 1.76 | 32.43 ± 2.73 |
| RDW (%) | 15.83 ± 2.57 | 14.11 ± 1.77 |

M: Males; F: females; Hb: hemoglobin; MCV: mean corpuscular volume; MCH: mean corpuscular Hb; MCHC: mean corpuscular Hb concentration; RDW: red blood cell distribution width.

Figure 1. Schematic representation of the β-globin gene cluster showing the variations examined and the two pairs of SNPs that revealed the stronger levels of LD. Black rectangle: protein coding gene; gray rectangle: non coding gene.
elevated levels of Hb F, ranging from 1.7–9.1% (mean 3.65%), herein classified as 'with HPFH.' This second population is a sub-set of the ‘normal subjects,’ also previously studied by Pereira et al. [17].

All individuals from the two study populations had been enrolled in molecular diagnosis of hemoglobinopathies at the Haematology Unit from the Coimbra Hospital and University Centre (CHUC), Coimbra, Portugal. Written informed consent was obtained from all participants in this study.

Genotyping

Genomic DNA isolated from peripheral blood using the QIAsymphony® DSP DNA Mini Kit (Qiagen GmbH, Hilden, Germany) was analyzed through a 6-multiplex system by SNAPSHOT reaction described in Borges et al. [18] that was further enlarged to a 7-multiplex system to also cover a SNP in BGLT3. The seven target SNPs implemented in the system were as follows: rs7482144 (C>T), rs113425530 (G>T), rs2070972 (A>C), rs7924684 (G>A), rs10128556 (G>A), rs968857 (T>C) and rs16911905 (G>C), respectively, located in genes/chromosome positions (hg38) 5'-HBBG2 (chr11: 5,254,939), HBBG2 (11: 5,253,490), HBBG2 (11: 5,253,487), BGLT3 (11: 5,245,498), HBBP1 (11: 5,242,453), 3'-HBBP1 (11: 5,239,228) and 5'-HBB (11: 5,228,060). Genotyping conditions were essentially as described in Borges et al. [18]. Briefly, the target fragments were amplified by multiplex polymerase chain reaction (PCR), of 5 µL per reaction, containing 2.5 µL of Qiagen® Multiplex PCR Kit, 0.5 µL of Primer Mix (to 2 µM each primer), 1.5 µL of distilled water and 0.5 µL of DNA (Qiagen GmbH). For the SNAPSHOT strategy, seven single base extension (SBE) primers were designed to hybridize with the sequences immediately adjacent to the target SNPs and the sequencing reaction was performed using the Ready Reaction Mix SNAPSHOT Multiplex Kit (Applied Biosystems, Foster City, CA, USA). The SBE reactions were analyzed using the Applied Biosystems® 3130 Genetic Analyzer (Applied Biosystems) and the GeneMapper Software 6 (Applied Biosystems). A multiplex ligation-dependent probe amplification (MLPA) (MRC-Holland, Amsterdam, The Netherlands), was used to screen for deletions in the β-globin gene cluster for individuals with Hb F levels >5.0% using a commercial kit (SALSA MLPA kit P102-B2 HBB).

Statistical analyses

Allele frequencies in the different SNPs were estimated by direct counting with Arlequin version 3.5 [19] (http://cmpg.unibe.ch/software/arlequin35/Arlequin35.html). The same software was used to test the Hardy-Weinberg equilibrium, to conduct haplotype inference through the expectation maximization/Excoffier-Laval-Balding (EM/ELB) algorithm and to calculate LD parameter r² and associated p values. The software PLINK version 1.07 (https://zzz.bwh.harvard.edu/plink/download.shtml) [20], was used to evaluate the association between individual SNPs and inferred haplotypes with levels of Hb F. Two different statistical approaches were used for the individual genetic variants: (i) simple linear regression, under an additive genetic model, and conditional analysis, to test in β-thal carriers the association between SNPs and Hb F levels, after logarithmic transformation of the Hb F values to fit a normal distribution; (ii) a case-control association test in the normal population using logistic regression, in the additive model, assumed to be subjects with HPFH vs. subjects with normal Hb F levels, using 1.6% Hb F as cutoff. Crude and adjusted p values for age and sex as covariates were obtained. The Bonferroni-adjusted significance level of p < 0.007 (0.05/7) was used to correct for multiple testing with seven SNPs. The program QUANTO, v.1.2.4 power calculator (http://hydra.usc.edu/gxe/) was used to estimate the power of association as a function of the frequency of the minor allele, assuming an additive model [21].

All variants were searched in the VarSome (https://varsome.com/) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases, to predict possible consequences of the nucleotide changes, and the scores for combined annotation dependent depletion (CADD) and genomic evolutionary rate profiling (GERP) were addressed from the genome browser Ensembl (https://wwwensembl.org). The CADD provides a ranking for single nucleotide variants based on scores that as higher they are the more likely to be deleterious, while GERP yields positive scores for highly-conserved positions and negative scores for highly-variable positions.

Results

β-Thalassemia carriers

Genetic data and association results with Hb F levels for the seven SNPs within β-globin gene cluster here examined as well as for the two most common mutations (HBB: c.92+6T>C and HBB: c.118C>T) in β-thal carriers are shown in Table 2. All genotype distributions are in accordance with the Hardy-Weinberg equilibrium after Bonferroni correction (p<0.007).

In order to evaluate the influence of each of the seven SNPs in the Hb F levels of β-thal carriers, basic simple linear regression in the additive genetic model was carried on, which showed significant associations with increased levels of Hb F for the minor allele of SNPs rs7482144 (T) (β=4.096e-4), rs7924684 (G) (BGLT3 (β=0.716; p=1.095e-5), rs10128556 (A) (HBBP1) (β=0.65; p=8.092e-4) and rs968857 (T) (3'-HBBP1) (β=0.70; p=7.091e-6) (Table 2). The β⁰ mutation at HBB: c.118C>T was also found to be strongly associated with increased levels of Hb F (β=0.927; p=1.78e-4); in contrast, the β⁺ mutation at HBB: c.92+6T>C is significantly associated with decreased levels of Hb F (β = -1.094; p=1.63e-6). When adjusting with age and sex as covariates, these significant associations are maintained (Table 2).

Next, we performed conditional analyses between these four common variants highly associated with Hb F levels and we observed that the two SNPs rs7924684 (A>G) (BGLT3) and rs968857 (C>T) (3'-HBBP1) remained...
Table 4. Four of the 11 obtained haplotypes with a frequency of common mutations, undertaken, combining all seven SNPs plus the two most under study, a haplotype-based association analysis was then HBB with Hb F levels (H1, H2, H4 and H7). The haplotypes H1 (c.118C>T) and H2 (c.92+6T>C), which includes the minor alleles at SNPs (CGAGGTGTT), which includes the minor alleles at SNPs rs7924684 and rs968857 T together with the mutated levels of Hb F, and just differ by the presence or absence of levels of Hb F (p = 0.0089 and p = 0.00168, respectively).

**Table 3.** Conditional analyses of the individually significant associations between single nucleotide polymorphisms and Hb F levels in -thalassemia carriers of Portuguese origin.

| Chr: Chromosome position; SNP: single nucleotide polymorphism; MAF: minor allele frequency; Alleles (A1:A2); A1-minor, A2-major; p HWE: p values for Hardy-Weinberg equilibrium; β: regression coefficient; (SE): standard error; r²: explained variance; p values (Wald test asymptotic) for the log-transformed Hb F values, using a linear regression model. *p Value using age and sex as covariates.

**Table 4.** Frequencies of haplotypes combining the seven analyzed single nucleotide polymorphisms and the two most common β mutations [IVS-1 (T>C) and codon 39 (C>T)] and association with Hb F levels in the 71 -thalassemia carriers of Portuguese origin.

| Haplotype | Frequency | β (SE) | p Value |
|-----------|-----------|--------|---------|
| H1 | TGAGATGTC | 0.125 | 0.681 | 0.095 |
| H2 | TGAGATGTT | 0.0491 | 1.295 | 0.132 |
| H3 | CGAGGGT GCCG | 0.0862 | -0.272 | 0.0097 |
| H4 | CGAGGGT GCCG | 0.1347 | -1.11 | 0.258 |
| H5 | CGAGGGT GCCG | 0.0146 | 0.317 | 0.0027 |
| H6 | CGAGGGT GCCG | 0.323 | -0.209 | 0.0138 |
| H7 | CGAGGT GGT T | 0.0706 | 0.718 | 0.0583 |
| H8 | CGAGGGT GCCG | 0.0216 | -0.674 | 0.0185 |
| H9 | CGAGGGT GCCG | 0.0141 | 0.361 | 0.0306 |
| H10 | CGAGGGT GCCG | 0.0221 | 0.0317 | 0.0161 |
| H11 | CGAGGGT GCCG | 0.0614 | 0.607 | 0.0387 |

**Table 2.** Association between Hb F levels and single nucleotide polymorphisms within the β-globin gene cluster and the two most common β mutations in 71 -thalassemia carriers of Portuguese origin.

| Chr: Chromosome position; Gene SNP | Alleles (A1:A2) | MAF (total) | p HWE | β (SE) | r² | p Value (unadjusted) |
|-----------------------------------|-----------------|-------------|--------|--------|----|---------------------|
| 115:254,939 S'HBB2 (Xmn1) rs7482144 T:C | 0.197 | 0.447 | 0.685 (0.185) | 0.167 | 4.096e-4a | 6.499e-4b |
| 115:253,490 HBB2 rs11345530 T:G | 0.007 | 1.0 | -0.752 (0.103) | 0.008 | 0.460 | 0.457 |
| 115:253,487 HBB2 rs2070972 C:A | 0.345 | 0.007 | -0.258 (0.217) | 0.02 | 0.239 | 0.134 |
| 115:245,498 BGLT3 rs7924684 G:A | 0.472 | 0.814 | 0.716 (0.151) | 0.246 | 1.095e-5b | 8.25e-6b |
| 115:242,453 HBBP1 rs10128556 A:G | 0.204 | 0.465 | 0.650 (0.185) | 0.151 | 8.092e-4b | 1.845e-3b |
| 115:239,228 3'HBBP1 rs968857 T:C | 0.430 | 0.636 | 0.700 (0.144) | 0.255 | 7.091e-6b | 5.342e-6b |
| 115:228,060 HBB2 rs16911905 C:G | 0.092 | 1.0 | 0.374 (0.307) | 0.021 | 0.277 | 0.085 |
| 115:226,924 HBB rs35724775 T:C | 0.190 | 0.059 | -0.109 (0.208) | 0.285 | 1.631e-6b | 1.628e-6b |
| 115:226,774 HBB rs11594907 T:C | 0.155 | 0.197 | 0.923 (0.234) | 0.185 | 1.778e-4 | 1.533e-4 |

**SNP:** Single nucleotide polymorphism.

*Significant association *p* values (<0.05) are bold.

Given the observed pattern of LD between the SNPs under study, a haplotype-based association analysis was then undertaken, combining all seven SNPs plus the two most common mutations, HBB: c.92+6T>C (rs35724775) and HBB: c.118C>T (rs11549407), whose results are detailed in Table 4. Four of the 11 obtained haplotypes with a frequency >0.01, were revealed to be significantly associated with Hb F levels (H1, H2, H4 and H7). The haplotypes H1 (TGAGATGTC) and H2 (TGAGATGTT), depicting the minor allele T for the rs7482144 (HBB2 Xmn1) polymorphism, showed the most significant association with elevated levels of Hb F (p = 0.0089 and p = 0.00168, respectively). However, these two haplotypes also displayed the minor alleles at the three SNPs rs7924684 G, rs10128556 A and rs968857 T, that are individually associated with increased levels of Hb F, and just differ by the presence or absence of the mutated codon 39 T allele. Another haplotype, H7 (CGAGGT GGT T), which includes the minor alleles at SNPs rs7924684 G and rs968857 T together with the mutated codon 39 T allele, only showed a marginal nominal association with high Hb F levels (p = 0.04). Finally, H4 (CGAGGG T GCCG), the most common haplotype in the
sample, included the major alleles at SNPs rs7482144 (HBG2 XmnI), rs7924684 (BGLT3), rs10128556 (HBBP1) and rs968857 (3'-HBBP1) in combination with the β+ mutation IVS-1-6 (T>C), revealed strong association with the decreased Hb F levels (β = −1.11; p = 6.07e-6).

**Normal population**

In the sample of healthy subjects, genotype distributions were also in agreement with Hardy-Weinberg equilibrium (p>0.05) (Table 5). The participants were arranged in two groups according to levels of Hb F: a control group of 36 subjects with normal Hb F levels (≤1.6%) and a case group of 15 subjects with higher Hb F levels (≥1.7%) herein assumed to carry HPFH. To investigate the relationship between the seven SNPs and levels of Hb F, a case-control study using logistic regression was undertaken in the additive model. The analysis yielded a nominal significant association for the rs2070972 minor C allele (OR = 2.162; CI 95% 0.889-5.255; p = 0.089) (Table 5). Using sex and age as covariates, near nominal associations were obtained for rs113425530 (p = 0.076) and rs2070972 (p = 0.086) (Table 5). Both SNPs are located in the same intronic region of the HBG2 gene, and despite being extremely rare to each other (2 bp apart), they were not in LD (r² = 0.0004) (Supplementary Table 1).

Accordingly, the haplotype analysis showed that haplotype H5 (GTGGTG), which contains the two minor alleles at rs113425530 (T) and rs2070972 (C), presents a frequency significantly higher in individuals with HPFH than in those with normal Hb F (0.133 vs. 0.015, respectively; p = 0.015) (Table 6). In contrast, the haplotype H6 (GGAGGTTG), with the two complementary alleles, rs113425530G and rs2070972A, was not observed in HPFH individuals and attained a 0.162 frequency in individuals with normal Hb F levels (p = 0.019) (Table 6). No other significant associations with levels of Hb F were detected for the remaining SNPs or haplotypes in the sample of healthy individuals.

**In silico analysis**

Table 7 shows the in silico analyses for the examined variations based on VarSome and ClinVar databases, and the CADD and GERP scores addressed from the genome browser Ensembl. The analysis showed that all seven SNPs located outside HBB were identified as benign by VarSome, and none was reported in ClinVar except the SNP rs7482144 (HBG2 XmnI) which appeared with ‘uncertain significance.’ In contrast, the two mutations, HBB: c.118C>T (codon 39) and HBB: c.92+6T>C (IVS-1-6), were identified in ClinVar or VarSome as ‘pathogenic.’ The seven variants that lie outside HBB are associated with low CADD scores and negative GERP scores, again in opposition with the above two β mutations, which have high CADD scores (40.0 and 21.7, respectively) and positive GERP scores (2.26 for both).

**Discussion**

In the present study, strong associations with elevated levels of Hb F were identified in β-thal carriers implying four SNPs within the β-globin gene cluster, namely rs7482144 (HBG2 XmnI) (p = 4.096e-4), rs7924684 (BGLT3) (p = 1.095e-5), rs10128556 (HBBP1) (p = 8.092e-4) and rs968857 (3'-HBBP1) (p = 7.091e-6), as well as the common β0 mutation (HBB: c.118C>T (p = 1.78e-4) (Table 2). The strongest signal was in rs968857, located downstream to the HBBP1 gene. This variant is located in a DNA fragment known not only to be involved in the Corfu deletion (a 7.2 kb deletion between the HBBP1 and HBD, typically

Table 5. Association between Hb F levels and single nucleotide polymorphisms at the β-globin gene cluster in 51 healthy individuals of Portuguese origin.

| Chrposition (hg38) | Gene | SNP | Alleles (A1:A2) | MAF (total) | p HWE | MAF (Hb F>1.6%) | OR (95% CI) | p Value | p Value (unadjusted) | p Value (or adjusted) |
|-------------------|------|-----|----------------|-------------|--------|----------------|-------------|---------|---------------------|----------------------|
| 11:5,254,939      | 5’HBG2 XmnI | rs7482144 | T:C           | 0.033       | 0.126  | 0.267          | 0.361       | 0.563    | (0.191–1.652)       | 0.296                | 0.342                |
| 11:5,253,490      | HBG2   | rs113425530 | T:G           | 0.118       | 0.518  | 0.233          | 0.069       | 4.068    | (0.103–15.00)       | 0.035*                | 0.086                |
| 11:5,253,487      | HBG2   | rs2070972  | C:A           | 0.431       | 0.779  | 0.567          | 0.375       | 2.162    | (0.889–5.255)       | 0.089                | 0.076                |
| 11:5,254,498      | BGLT3  | rs7924684 | A:G           | 0.353       | 0.545  | 0.367          | 0.347       | 1.101    | (1.427–2.832)       | 0.84                | 0.7                  |
| 11:5,242,453      | HBBP1  | rs10128556 | A:G           | 0.373       | 1.0    | 0.3            | 0.403       | 0.631    | (0.249–1.595)       | 0.330                | 0.399                |
| 11:5,239,228      | 3’HBBP1| rs968857  | C:T           | 0.467       | 0.198  | 0.467          | 0.404       | 1.419    | (0.520–3.866)       | 0.495                | 0.408                |
| 11:5,228,060      | HBB    | rs11691905 | C:G           | 0.031       | 0.015  | 0.031          | 0.083       | 0.441    | (0.057–3.478)       | 0.431                | 0.385                |

Chr: Chromosome position; SNP: single nucleotide polymorphism; Alleles (A1:A2): A1-minor:A2-major; MAF: minor allele frequency; p HWE: p values for Hardy-Weinberg equilibrium; OR (95% CI): odds ratio (95% confidence interval).

Association was tested under a case-control model using a cutoff of 1.6% for Hb F level. The OR shows the minor allele, 95% CI and p values for allelic association obtained with logistic regression under the additive model.

*Significant association p values (<0.05) is bold.
associated with elevated Hb F levels) [22,23] but also to contain binding sites for the major γ-globin repressor BCL11A [8]. A finer examination of sequence encompassing the Corfu deletion in patients with sickle cell disease also led to pinpoint rs968857, showing a different distribution in patients with high Hb F compared with patients with low Hb F levels [9]. Furthermore, in silico analysis revealed that the C>T change at rs968857 eliminates binding sites for the transcription factors NF-E2 and AP-1, two transcription factors known to participate in the Hb switch mechanisms [9]. In the Portuguese β-thal carriers, this SNP was found in high LD (r² = 0.799) with rs7924684 located on the BGLT3 gene, and consistently, the latter SNP showed the second strongest signal of association with increased Hb F levels. The BGLT3 gene has recently been implicated in γ-globin expression [16], and this is the first report in β-thal carriers showing evidence that SNP rs7924684 modulates Hb F levels.

In the present study, the rs7482144 (HBG2 XmnI) T allele was also found to be strongly associated with increased Hb F levels (p = 4.096e⁻⁴), replying the association previously reported in our study based on a subset of these β-thal carriers [17]. Here, another SNP revealed an association of the same order of magnitude as the HBG2 XmnI polymorphism, the rs10128556 lying in the HBBP1, in which the minor allele (A) was also found to account significantly for increased Hb F levels (p = 8.092e⁻⁴). Despite these two SNPs being separated by about 12.5 kb, they are in strong LD (r² = 0.873) in Portuguese β-thal carriers, similarly to that observed in sickle cell disease patients from different populations [12,13,15]. Both rs7482144 and rs10128556 are among the SNPs usually analyzed to assess the frequencies of the three remaining SNPs significantly associated with Hb F, including rs10128556 (Table 3). Similarly, rs10128556 lost significance upon conditioning on the same three SNPs, including rs7482144. In contrast, and rather unexpectedly, rs7924684 (in BGLT3) and rs968857 (in 3’-HBPP1), both retain a strongly significant association with Hb F after conditioning on rs7482144 and rs10128556. This result indicates that these two SNPs located in the non coding regions of the β-globin gene cluster (rs7924684 and rs968857) are important modulators of Hb F levels, disclosing greater signals for this trait association than previously described for HBG2 XmnI rs7482144, which up to now was the variation within the cluster considered to account for the more elevated proportion of the variability in Hb F levels.

Focusing now on the haplotype distribution in β-thal carriers (Table 4), we stress that the two haplotypes (H1 and H2) depicting the minor alleles for the four SNPs individually associated with increased levels of Hb F [rs7482144 (HBG2 XmnI), rs10128556 (HBPP1), rs7924684 (BGLT3) and rs968857 (3’-HBPP1)], revealed the most significant association with elevated Hb F. As these two haplotypes differ only by the presence or absence of the mutated codon 39 allele (rs11549407), it seems that this β⁰ mutation (defining H2), amplifies the role of the remaining haplotypic background in enhancing levels of Hb F (individually, rs11549407 also showed a significant positive trait association). Looking upon the three haplotypes with the major alleles at the four former SNPs (H4, H5 and H6), only H4 was significantly associated with Hb F, but this time with decreased levels. This is the most common haplotype containing the HBB: c.92+6T>C β⁺ mutation, which was the HBB mutation associated with lower levels of Hb F in β-thal carriers. Haplotype H7 was found to be marginally associated with increased Hb F levels (p = 0.04), a result that might have been prompted by the presence of the mutated codon 39 T allele in the haplotype.

From the data herein obtained in β-thal carriers, it is difficult to assess the real role of each of the SNPs in the modulation of Hb F levels, mainly due to the complex pattern of LD encompassing the entire β-globin gene cluster, and in particular the unusual LD relationships between the four SNPs that were individually found significantly associated with Hb F. The functional importance of rs7482144 (HBG2 XmnI) in raising Hb F levels in β-thal or sickle cell disease was recently clarified by disruption studies using CRISPR-Cas9, evidencing that this region is a binding domain for a repressive transcription factor or alternatively contains a DNA motif for a still unknown transcriptional activator [24]. Paradoxically, HBPP1, despite of being a pseudogene,

---

**Table 7. In silico analysis of the seven single nucleotide polymorphisms within the β-globin gene cluster examined in this study, and the two most common β mutations found in the β-thalassemia carriers.**

| Chr:position (hg38) | Gene | SNP | cDNA HBB: | Frequency EUR (gnomAD) | Impact | ClinVar | VarSome | CADD Score | GERP |
|---------------------|------|-----|-----------|------------------------|--------|---------|---------|-----------|------|
| 11:5,254,939 S' HBG2 XmnI | rs7482144 | c.-158C>T | – | 0.263 | NC | US | B | 5.421 | -3.85 |
| 11:5,253,487 HBG2 | rs2070972 | c.316-82T>G | – | 0.471 | NC | NR | B | 1.007 | -3.61 |
| 11:5,242,498 BGLT3 | rs7924684 | n.49G>A¹ | – | 0.469 | NC | NR | B | 2.666 | -0.26 |
| 11:5,242,453 HBPP1 | rs10128556 | c.-29 +997G>A | – | 0.321 | NC | NR | B | 0.147 | -1.24 |
| 11:5,239,228 HBPP1 | rs968857 | c.-29 +4222>G>A | – | 0.489 | NC | NR | B | 2.91 | -1.83 |
| 11:5,228,060 HBG | rs16911190 | c.-86-481C>G | – | 0.056 | NC | NR | B | 0.20 | -2.94 |
| 11:5,226,924 HBG | rs35724775 | c.92 +6T>C [IVS-I-6 (T>C)] | 0.0002 | NC | P | P | 21.70 | 2.26 |
| 11:5,236,774 HBG | rs11549407 | c.118C>T [codon 39 (C>T)] | 0.00067 | nonsense | P | P | 40.00 | 2.26 |

Chr: Chromosome position; SNP: single nucleotide polymorphism; EUR: European (non Finnish); NC: non coding; US: uncertain significance; B: benign; NR: not reported; P: pathogenic; CADD: combined annotation dependence depletion; GERP: genomic evolutionary rate profiling.

¹Human Genome Variant Society nomenclature: NC_000007.3:g.52118G>C-A.
was demonstrated to be evolutionarily highly conserved [25], and concerning BGLT3, which codes for a long non-coding RNA of function not yet fully clarified, our own data based on direct sequencing analysis of the entire BGLT3 in the β-thal carriers indicate that it is highly invariable except for rs7924684 (results not shown).

In the normal population, only rs113425530 and rs2070972, 2 bp apart on the HBG2 gene but without showing LD between each other, appeared to be involved in Hb F regulation (Table 5). Indeed, near significant associations with Hb F levels were found for both polymorphisms and fittingly the unique haplotype with the two minor alleles at rs113425530 (T) and rs2070972 (C) was significantly more frequent in HPFH individuals than in those with normal Hb F values (Table 6). Until now, scarce support exists on the influence of these HBG2 polymorphisms in Hb F regulation, although a study involving patients from Thailand with Hb E (HBB: c.79G>A)/β0-thal has reported rs2070972 as being associated with high Hb F levels [26].

The post-hoc statistical power of the study was evaluated, and for the β-thal carriers, the estimated power of association was between 12.0 and 99.0% for the seven analyzed SNPs, showing that the four associated SNPs (5’HBG2 rs7482144, BGLT3 rs7924684, HBOBP1 rs10128556, and 3’HBOBP1 rs968857) achieve adequate statistical power (e.g., >80.0%) to detect true evidence for an association. In regard to the normal population, the estimated power ranged from 6.0 to 77.0%, and at least for the HBG2 rs113425530 polymorphism, which is associated with Hb F, the obtained power was close to 80.0%.

The in silico analyses for the examined variations showed that all the seven SNPs located outside HBB must have no functional or clinical significance, in contrast with the two mutations HBB: c.118C>T (codon 39) and HBB: c.92+6T>C (IVS-1-6), which in ClinVar or VarSome databases, appear as ‘pathogenic’ (Table 7). These two mutations were associated with much higher CADD scores then any of the other seven SNPs, indicating that both are predicted to be functionally relevant. In addition, their GERP scores were positive, suggesting they are under strong evolutionary constraint, also in opposition with the negative GERP scores for the other seven SNPs, indicating that both are predicted to be functionally relevant. In conclusion, this study reinforces the complex pattern of SNP interactions that modulate Hb F levels, consistent with the consensual idea that multiple functional SNPs within the β-globin gene cluster act cooperatively to influence Hb F regulation, possibly by changing the binding sites for transcription factors or repressors related to the γ- to β-globin gene switching process. Given that different patterns of association were detected between β-thal carriers and normal subjects, the question can be raised on whether, under erythropoiesis stress, epigenetic mechanisms contribute to change the regulation of the entire β-globin gene cluster. Of note, BGLT3 was incorporated in the list of regions carrying alleles with Hb F regulatory consequences, which in combination with alleles in other positions within the cluster can trigger the recruitment of positive or negative coregulators of Hb F. Knowing more about what SNPs influence Hb F levels is crucial to anticipate disease severity associated with β hemoglobinopathies, such as sickle cell disease and β-thal, and has the potential to be translated in the future into new therapeutic approaches for Hb F reactivation.

Disclosure statement

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

Funding

This study was partially supported by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the COMPETE 2020 Operational Program for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT (Fundaçao para a Ciência e a Tecnologia), in the framework of the Project [POCI-01-0145-FEDER-007274] to i3S and Project [UIDB/00283/2020] to CIAS. V. Gomes was supported by FCT under the program contract provided in Decree-Law #57/2016 of 29 August 2018.

ORCID

Licinio Manco http://orcid.org/0000-0002-2636-0288
Celeste Bento http://orcid.org/0000-0003-1080-411X
Verónica Gomes http://orcid.org/0000-0003-2815-6376
António Amorim http://orcid.org/0000-0002-7518-6247
Maria J. Prata http://orcid.org/0000-0002-0583-1028

References

[1] Thein SL. Genetic basis and genetic modifiers of β-thalassemia and sickle cell disease. Adv Exp Med Biol. 2017;1013:27–57.
[2] Thein SL. Molecular basis of β thalassemia and potential therapeutic targets. Blood Cells Mol Dis. 2018;70:54–65.
[3] Venkatesan V, Srinivasan S, Babu P, et al. Manipulation of developmental gamma-globin gene expression: an approach for healing hemoglobinopathies. Mol Cell Biol. 2020;41(1):e00253–20.
[4] 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. 2011;108(27):11346–11351.
[5] 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 β-thalassemia. Proc Natl Acad Sci USA. 2008;105(5):1620–1625.
[6] 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.
[7] Steinberg MH. Fetal hemoglobin in sickle hemoglobinopathies: high HbF genotypes and phenotypes. JCM. 2020;9(11):3782.
[8] Sankaran VG, Xu J, Byron R, et al. A functional element necessary for fetal hemoglobin silencing. N Engl J Med. 2011;365(9):807–814.
[9] Akinsheye I, Alsdutan A, Solovieff N, et al. Fetal hemoglobin in sickle cell anemia. Blood. 2011;118(1):19–27.
[10] Hu L, Huang L, Han Y, et al. Association of polymorphisms in the HBG1-HBD intergenic region with HbF levels. J Clin Lab Anal. 2020;34(6):e23243.
[11] Huang P, Keller CA, Giardine B, et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel
fetal hemoglobin regulatory element. Genes Dev. 2017;31(16):1704–1713.

[12] Galanneau 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.

[13] Bhanushali AA, Himani K, Patra PK, et al. Hb F levels in Indian sickle cell patients and association with the HBB locus variant rs10128556 (C>T) and the HBG Xmn1 (Arab-Indian) variant. Hemoglobin. 2017;41(4-6):317–320.

[14] Nuinoo M, Makarasara W, Mushiroda T, et al. A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E. Hum Genet. 2010;127(3):303–314.

[15] Vathipadiekal V, Alsultan A, Baltrusaitis K, et al. Homozygosity for a haplotype in the HBG2-OR51B4 region is exclusive to Arab-Indian haplotype sickle cell anemia. Am J Hematol. 2016;91(6):E308–E311.

[16] Ivaldi MS, Diaz LF, Chakalova L, et al. Fetal γ-globin genes are regulated by the BGLT3 long noncoding RNA locus. Blood. 2018;132(18):1963–1973.

[17] Pereira C, Relvas L, Bento C, et al. Polymorphic variations influencing fetal hemoglobin levels: association study in betathalassemia carriers and in normal individuals of Portuguese origin. Blood Cells Mol Dis. 2015;54(4):315–320.

[18] Borges E, Tchonhi C, Couto CSB, et al. Unusual β-globin haplotype distribution in newborns from Bengo, Angola. Angola. Hemoglobin. 2019;43(3):149–154.

[19] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564–567.

[20] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–575.

[21] Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med. 2002;21(1):35–50.

[22] Galanello R, Melis MA, Podda A, et al. Deletion delta-thalassemia: the 7.2 kb deletion of Corfu delta beta-thalassemia in a non-beta-thalassemia chromosome. Blood. 1996;75(8):1747–1749.

[23] Kulozik AE, Yarwood N, Jones RW. The Corfu delta β zero thalassemia: a small deletion acts at a distance to selectively abolish β globin gene expression. Blood. 1988;71(2):457–462.

[24] Weber L, Frati G, Felix T, et al. Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. Sci Adv. 2020;6(7):eaay9392.

[25] Moleirinho A, Seixas S, Lopes AM, et al. Evolutionary constraints in the β-globin cluster: the signature of purifying selection at the δ-globin (HBD) locus and its role in developmental gene regulation. Genome Biol Evol. 2013;5(3):559–571.

[26] Sherva R, Sripichai O, Abel K, et al. Genetic modifiers of Hb E/beta0 thalassemia identified by a two-stage genome-wide association study. BMC Med Genet. 2010;11(1):51.