POLYMORPHIC VARIATIONS INFLUENCING FETAL HEMOGLOBIN LEVELS: ASSOCIATION STUDY IN BETA-THALASSEMIA CARRIERS AND IN NORMAL INDIVIDUALS OF PORTUGUESE ORIGIN

Clara Pereira[1], Luís Relvas[2], Celeste Bento[2], Augusto Abade[1], M. Letícia Ribeiro[2], Licínio Manco[1,2]

[1]Research Centre for Anthropology and Health (CIAS), Department of Life Sciences, University of Coimbra, Coimbra, Portugal.

[2]Department of Hematology, Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal.

Key words: HbF; beta-thalassemia carriers; HPFH; BCL11A, HMIP, HBG2 and KLF1 loci.

SHORT TITTLE:
POLYMORPHIC VARIATIONS INFLUENCING HBF LEVELS IN BETA-THALASSEMIA CARRIERS AND IN NORMAL INDIVIDUALS

Corresponding author:
Licínio Manco, PhD
Research Centre for Anthropology and Health
Department of Life Sciences
University of Coimbra
3000 Coimbra (Portugal)
Phone: +351239855210
Fax: +351239855211
E-mail: lmanco@antrop.uc.pt

ABSTRACT

Three major loci have been associated with HbF levels, including -158C/T (XmnI) at *HBG2* promoter region, and several polymorphisms at *BCL11A* intron-2 and *HBS1L-MYB* (HMIP) intergenic region. Mutations in the KLF1 gene were recently associated with increased HbF levels.
This study aims to evaluate whether genetic variability at these loci influence HbF levels in β-thalassemia carriers and in normal individuals of Portuguese origin.
Sixty five β-thalassemia carriers, HbF levels ranging from 0.2% to 9.5%, and 60 individuals with normal haematological parameters, HbF levels ranging from 0.2% to 7.4%, were selected for this study.
In β-thal carriers linear regression models revealed strong statistical significant association for *HBG2* (XmnI) rs7482144 (β=0.45; P=5.85x10^{-7}), and nominal significance for *BCL11A* rs766432 (β=0.21; P=0.02) and *HMIP* rs9399137 (β=0.20; P=0.01). In normal individuals, a case (HbF>2%; n=15) *vs.* control (HbF<1.7%; n=45) model, showed nominal significant associations for *BCL11A* SNPs rs11886868 (OR=4; P=0.001), rs766432 (OR=3.7; P=0.002) and rs7606173 (OR=0.36; P=0.03). *KLF1* rs3817621 was not found associated with HbF levels.
Our results suggest that in Portuguese β-thal carriers the *HBG2* XmnI polymorphism is strongly associated with HbF levels. In normal individuals, *BCL11A* polymorphisms, but not *HMIP* or *HBG2* (XmnI) *loci*, are nominally associated with HbF expression.

INTRODUCTION

Fetal hemoglobin (HbF, α2γ2) levels vary among individuals as a consequence of genetic variation. The identification of modulators responsible for the reactivation of gamma-globin genes has been extensively studied, since high levels of HbF contribute to ameliorate the clinical manifestations and severity of sickle cell anemia (SCA) and β-thalassemia (β-thal) [1,2]. Thus, the understanding of these modulation factors and how the fetal to adult (γ to β) globin chains switch occurs in the course of human ontogeny could be important to identify targets for the development of new therapies to improve the treatment of these hemoglobinopathies [1,2].

Recent genome wide association studies (GWAS), conducted in both healthy nonanemic populations and in patients with sickle cell disease or beta thalassemia of European, African or Asian descent, have identified several single nucleotide polymorphism (SNP) variants along the genome associated with HbF levels (or the highly correlated F-cells number). The three important HbF levels loci determinants identified were: i) SNPs at promoter nucleotide of *HBG2* on chromosome 11p15, including the -158 C/T (rs7482144) variant also known as XmnI; ii) SNP variants located at the *HBS1L-MYB* intergenic region (HMIP) on chromosome 6q23, and iii) SNPs located at the *BCL11A* gene on chromosome 2p16 [1,3,4,5]. The erythroid transcription factor KLF1 activates BCL11A and assists in coordinating the switch from fetal to adult hemoglobin [6]. Individuals with loss-of-function KLF1 mutations were known to have persistently elevated HbF [7-9].

A modest elevation of HbF (1–5%), called heterocellular hereditary persistence of fetal hemoglobin (HPFH) or Swiss type HPFH, is observed in 10-15% of apparently normal individuals in the absence of any hematological disorder [10]. This condition had been regarded as a multifactorial quantitative trait linked to polymorphic variations in regulatory sequences of the γ- and β-globin genes, especially the -158 C>T polymorphism [11-13], or in introns of γ-genes, but also with other loci mapping elsewhere in the genome (Xp 22.2–22.3 region, 6q23 and 8q) [14-16].

In beta-thalassemia, the synthesis of the β-globin polypeptide is reduced as a result of mutations in this gene, leading to an excess of alpha-globin chains which precipitate in red blood cell precursors, contributing to dyserythropoiesis and premature cells death [17]. Ineffective production of mature red blood cells in homozygotes or compound heterozygotes for β-globin mutations leads to anemia with severely clinical manifestations (β-thalassemia major and β-thalassemia intermedia) [17]. Increased production of γ-globin chains can modulate the disease severity by reducing the number of free α-chains, which decreases the dyserythropoiesis [18,19]. Carriers of β-thalassemia (β-thal minor) have a mild asymptomatic anemia with no

clinical symptoms. Several genetic association studies have been performed in β-thal carriers across different populations and have identified sequence variants associated with the HbF levels [20-23].

Nevertheless, results regarding the genetic association basis for both conditions, β-thal carriers and HPFH, are not entirely clear, and some loci previously identified in patients with severe hemoglobinopathies have not yet been assessed in such individuals. Thus, the aim of this study was to evaluate whether genetic variability in *loci BCL11A*, *HMIP*, *HBG2* and *KLF1* are associated with HbF levels in normal subjects with common forms of HPFH and in β-thal carriers of Portuguese origin.

MATERIAL AND METHODS

Study population

A total of 125 unrelated subjects of Portuguese origin were recruited for this study, including heterozygous for beta thalassemia mutations and *HBB* wild-type.

Sixty five (35 females and 30 males) were beta thalassemia carriers, aged 2-77 years (mean age 34.3 y; median age 35 y), with HbF levels ranging from 0.2% to 9.5% and HbA2 levels raised between 3.4% and 6.8% (mean 4.78%). These subjects classified as β-thal *minor* were heterozygous for one of the following mutations in the *HBB* gene [c.118C>T (p.Glu40term); c.48G>A (p.Trp16term); c.92+6T>C; c.126_129 delCTTT; c.92+1G>A; cd.92+110G>A].

A second population group of 60 subjects (34 females; 26 males), aged 2-75 years (mean age 29.87 y; median age 30 y), with normal hematological parameters, HbF levels ranging from 0.2% to 7.4% and normal HbA2, was selected: a control group of randomly recruited subjects (n=35) with normal HbF levels ranging from 0.2% to 1.6%, and 15 subjects with HbF levels >2%, classified as HPFH.

Informed consent was provided by all the participants. HbF and HbA2 levels were determined by high performance liquid chromatography (HPLC) using Variant 2 (Bio-Rad, CA, USA).

Genotyping

Seven SNPs were chosen for this study based on the recent publications in PubMed reporting genetic variants most strongly associated with increased HbF levels: loci *BCL11A* (rs11886868, rs766432 and rs7606173), HBS1L-MYB (*HMIP)* (rs9399137 and rs6934903), *HBG2* (rs7482144), and *KLF1* (rs3817621). SNPs rs11886868, rs7606173, rs9399137 were genotyped by TaqMan assay using pre-designed probes (Applied Biosystems, Foster City, USA). Allelic discrimination was performed according to the manufacturer's instruction on a BioRad RT-PCR system (MiniOpticon, BioRad, CA, USA). SNPs rs766432, rs6934903, rs7482144 and rs3817621 were genotyped by PCR-RFLP using the restriction enzymes AccI, HpyCH4III, XmnI, and AciI, respectively. To assess genotyping reproducibility, 10% of random samples were and re-genotyped for all SNPs with 100% concordance.

A commercially available kit (SALSA MLPA kit P102-B2 HBB), using multiplex ligation-dependent probe amplification (MLPA), was used to screen for deletions in the beta-globin cluster for individuals with HbF levels >5%.

Statistical analysis

The allelic and genotypic frequencies of all polymorphisms were estimated by direct counting and Hardy-Weinberg equilibrium probability values were achieved using an exact test. Associations of SNPs with HbF levels, after logarithmic transformation to near normalize the quantitative trait distribution, were performed: i) in the β-thal *minor* group by linear regression under an additive genetic model; ii) in the normal subjects by a case-control model (subjects with HPFH *vs.* subjects with normal HbF), using 2% HbF as cutoff-point, estimating P values, odds ratio (OR), and 95% confidence intervals (CI), crude and with age and sex as covariates. All these statistical analyses were done using the set-based tests implemented on PLINK software v.1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) [24].

Graphical analyses, normality of the data assessed by the Kolmogorov–Smirnov test and comparisons of HbF levels between genotypes by using non-parametric (Mann-Whitney U for β-thal carriers) or parametric (one way ANOVA, followed by post hoc Bonferroni test for normal individuals) tests according to the distribution of data, were performed with the SPSS software, version 20.

A significant p-value was considered below 7.1×10^{-3} (0.05/7) by applying a Bonferroni correction for multiple testing, and a p-value below 0.05 was considered significant for individual SNPs.

RESULTS

The demographic characteristics and hematological parameters of the populations under investigation are shown in Table 1.

Beta thalassemia carriers

Regarding the β-thal *minor* group, total genotyping rate was 0.942. The minor allele frequencies (MAF) observed for the seven polymorphisms in the total sample were displayed in Table 2. The genotype distributions were in agreement with the Hardy-Weinberg equilibrium (P>0.05). Deletional mutations in the β-genes cluster were excluded in individuals with HbF levels >5%.

Using a dominant model (genotypes homozygous for the ancestral allele *versus* homozygous and heterozygous for the derived allele), we analysed the distribution of log transformed HbF values according to SNP genotypes by the Mann-Whitney U test (Table S1). Significant differences were observed for SNPs *BCL11A* rs766432 (P=0.034), HMIP rs9399137 (P=0.005) and HBG rs7482144 (P<0.001). Figure 1 shows in box-plots the distribution of HbF levels (log-transformed) within each genotype of the three statistical significant SNPs: individuals who were homozygous for the major ancestral allele showed lower values when compared to those with the derived allele. In accordance, linear regression analysis under an additive genetic model to test the association between SNPs and HbF levels showed a strong statistical significant value for *HBG2* (XmnI) rs7482144 (β=0.46; $P=5.86x10^{-7}$) and nominal significant association with SNPs *BCL11A* rs766432 (β=0.22; P=0.029) and *HMIP* rs9399137 (β=0.21; P=0.011), also using age and sex as covariates (P=0.011, P=0.018 and $P=6.77x10^{-7}$, respectively) (Table 2).

Normal subjects

In relation to the normal subjects, genotyping rate was 0.998. The MAF were presented in Table 3 for all the studied polymorphisms. Genotype distributions were in agreement with Hardy-Weinberg equilibrium (P>0.05). No deletional mutations were detected in the β-genes cluster among subjects with HbF levels >5%.

We analysed the log transformed HbF mean values according to SNP genotypes by one-way ANOVA and statistical significant differences were observed for the *BCL11A* SNPs rs11886868 (P<0.001), rs766432 (P=0.008) and rs7606173 (P=0.006) (Table S2). Individuals homozygous for the derived allele at rs11886868 (CC) and rs766432 (CC) and for the ancestral allele at rs7606173 (GG) have HbF levels about 3 times higher than those homozygous for the different allele (Bonferroni post-hoc tests: P<0.001, P=0.008 and P=0.004, respectively). The distribution of HbF

levels (log-transformed) within each genotype group of the three BCL11A SNPs in box-plots (Figure 2), show the homozygous individuals for the ancestral allele at rs11886868 (TT) and rs766432 (AA) and for the derived allele at rs7606173 (CC) with the lowest HbF values.

We subdivided participants in two groups considering the HbF levels: a group (n = 15) with levels of HbF between 2% and 7.4%, classified with HPFH, and a group (n= 45) with HbF levels in the normal range, below 2% (0.1% to 1.6%). The mean of HbF (4.28% *vs.* 0.67%) differs significantly between the two groups (P<0.001). The case-control study, under an additive model, showed nominal significant associations between *BCL11A* SNPs rs11886868 (OR=4; 95% CI, 1.6-9.6; P=0.001), rs766432 (OR=3.7; 95% CI, 1.5-8.9; P=0.002) and rs7606173 (OR=0.36; 95% CI, 0.14-0.93; P=0.03) and HbF levels (Table 3). Using age and sex as covariates, Fisher's association *p*-values were 0.0015, 0.0034 and 0.041, respectively.

If we look only within the subgroup with normal HbF levels (<2%), linear regression models showed statistical significant associations for SNPs *BCL11A* rs11886868 (P=0.006) and rs7606173 (P=0.009), but not for *BCL11A* rs766432, HMIP, XmnI and KLF1 SNPs (P>0.05).

To examine haplotypes among the three *BCL11A* loci (rs11886868, rs766432 and rs7606173), we performed a linkage disequilibrium (LD) analysis. Value r2 for LD of the three loci was 0.66 (rs11886868 with rs766432), 0.46 (rs11886868 with rs7606173) and 0.31 (rs766432 and rs7606173). Association results of *BCL11A* haplotypes with HbF levels are presented in Table 4. The predominant haplotypes consisted of TAC and CCG with frequencies of 0.397 and 0.308, respectively. Haplotype with the three HbF associated alleles CCG was significantly higher (0.53) in HPFH individuals than in normal group (0.23): P=0.002. On the other hand, the complementary haplotype TAC showed a marginal effect against HbF (P=0.034) (0.23 in HPFH individuals *vs.* 0.45 in controls).

DISCUSSION

Individuals with severe forms of hemoglobinopathies due to beta-globin chain disorders, including beta-thalassemia major and intermedia and sickle cell disease, typically demonstrate clinical phenotypes whose severity is inversely proportional to the degree of preservation of HbF expression. Common polymorphisms at just three loci account for a meaningful fraction of the variation in HbF levels. These loci include the beta-globin gene cluster itself, an intergenic interval between the HBS1L and MYB genes (HMIP) and *BCL11A*.

The rs7482144 C/T polymorphism (XmnI), located at the promoter region of the *HBG2* gene, has been associated with an enhancement of HbF levels under erythropoietic stress in homozygous β^+-thalassemia [25] or sickle cell anemia [4,23], but also with a modest elevation of HbF (1–5%) in normal individuals in several association studies [11-13]. In a group of Portuguese β-thal carriers we have found a strong association between increased HbF levels and the rs7482144 T allele, which confirms previously described associations of the *HBG2* -158T allele (XmnI Gγ+) with increased levels of HbF in β-thal carriers of Portuguese [23], Italian [20] and Indian [22] descent. However, our study did not show statistical significant correlations in normal subjects with HPFH regarding the effect of the *HBG2* XmnI polymorphism on the synthesis of HbF levels. Although this result is in contrast with other studies in normal individuals from Algerian [12] and Tunisian [13] populations, our findings are in accordance to previous reports for minor or absence of consequences on HbF levels in normal individuals of Italian [20] and Indian [22] origin. These inconsistent results regarding normal subjects may depend on the different statistical models used in the analysis or to the lower cutoff point (~1%) used to distinguish normal *versus* raised HbF levels in case-control studies [12,13]. We have used a 2% cutoff as a limit for the normal HbF to overcome eventual discrepancies in the lower quantifications by the HPLC method. Moreover, linear regression models to assess the correlation of polymorphisms in individuals with HbF levels between 0.1 and 1.7 (i.e., below the considered cutoff point of 2%) did not show significant association with HbF for the XmnI polymorphism.

BCL11A expression is important for gamma to beta globin gene switching and the increase in HbF is associated with reduced BCL11A expression [26]. Several studies validated BCL11A as a transcriptional repressor acting in the HBG promoter to form a repressor complex [27,28]. Xu et al [29] using a sickle cell disease mouse model demonstrate that interference with BCL11A function increases fetal hemoglobin levels and, thereby, reduces the severity of the disease. These data illustrates the potential of the transcription factor BCL11A as a therapeutic target for the treatment of sickle cell disease and beta-thalassemia. Variation at *BCL11A* gene is estimated

to explain ~15% of the variance in HbF levels and several SNPs have been identified within 20 kb in *BCL11A* intron-2 as the most highly associated with the HbF, including rs11886868 [18], rs766432 [5] and rs7606173 [2]. These SNPs lie within an erythroid enhancer and act combinatorial to influence BCL11A regulation [30]. Our studies in the Portuguese normal subjects showed that these three BCL11A SNPs, rs11886868, rs766432 and rs7606173 are nominally associated with HbF values. The haplotype analysis showed haplotype CCG, with the three HbF associated alleles, significantly associated with HbF (P=0.002).This is in concordance with previous data at BCL11A showing that the rs11886868 genotype distribution was markedly different between Sardinian individuals with normal HbF values and those with HPFH [18]. In regarding to β-thal carriers, our data show nominal significant association between *BCL11A* rs766432 and HbF levels, as previous found in Chinese adult β-thal heterozygotes [21].

Regarding the *HBS1L-MYB* intergenic region (HMIP), the most significant associated SNP in beta-thalassemia and sickle cell anemia among Chinese, European, and African patients, was rs9399137, located at the haplotype block 2 [1]. Overexpression of *MYB*, a critical regulator of erythroid development, but not *HBS1L*, in human K562 erythroid cells was shown to decrease *HBG* expression [31]; and down-regulation of *MYB* expression in adult erythroid cell cultures by shRNAs was reported to result in increased *HBG* expression [32]. More recently, functional studies using human erythroid progenitors (HEPs) showed that several *HBS1L-MYB* intergenic variants located in regulatory elements may reduce transcription factor bindings, affecting long-range interactions with *MYB*; this results in decreased *MYB* expression and consequently elevated HbF levels [33]. Our data in β-thal carriers showed nominal statistically significant associations between HMIP rs9399137 and HbF levels, replicating previous findings in Chinese adult β-thal heterozygotes [21]. However, no significant correlation with HbF levels was found in our study sample of normal individuals for the *HMIP* variants. This result is different from previous findings regarding the association of rs9399137 with HbF levels in normal subjects of northern European origin [3].

BCL11A controls the *HBB* cluster in association with the erythroid KLF1 transcription factor. The KLF1 transcription factor occupies the BCL11A promoter and it is known that knockdown of KLF1 leads to decreased BCL11A in human erythroid progenitors, derepressing γ-globin chains synthesis [6,7]. Several point mutations in the KLF1 gene were found to be associated with HPFH. These KLF1 mutations are predicted to affect gene function, and include missense mutations, frameshift mutations, deletions, splice site and nonsense mutations [7,8]. Moreover, a first KLF1 gene promoter mutation (-148G>A) was also associated with increased HbF levels in an adult female subject of Serbian origin [9]. This mutation is located within a Sp1 binding site, resulting in drastically reduced CAT reporter gene expression in K562 cells and in reduced KLF1 gene

expression in vivo [9]. In our work, trying to identify additional KLF1 gene promoter variations that could be associated with HbF levels, we investigated a common SNP G/C (rs3817621), located 102 bp upstream the -148G>A variant, which has a minor allele frequency of 25% in the European population (1000 genomes project EUR population, http://www.ensembl.org/). The no significant association between rs3817621 and HbF levels, observed both in β-thal carriers and wild-type individuals, may be explained by the high population frequency of this variant, in contrast to the rare KLF1 -148G>A variation (rs79334031), with a 4% frequency for the minor A-allele in the European population (1000 genomes project EUR population, http://www.ensembl.org/), and located in a functional element of the KLF1 promoter region.

In conclusion, our results suggest that the increase of HbF levels in Portuguese individuals with normal hematological parameters is nominally associated with *BCL11A* polymorphisms, but not with *HMIP* or the *HBG2* XmnI polymorphisms. On the other hand, the increase of HbF levels in Portuguese β-thal carriers is strongly associated with the *HBG2* XmnI polymorphism, but also with *BCL11A* rs766432 and *HMIP* rs9399137. These results are consistent with previous findings in European and other populations.

Acknowledgements

This study was performed with the support of *Fundação para a Ciência e Tecnologia* (FCT) (PEst-OE/SADG/UI0283/2013) and Forum Hematologico, Serviço de Hematologia, CHUC.

REFERENCES

[1] Thein SL, Menzel S, Lathrop M, et al., Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. Hum Mol Genet 18 (2009) 216–223.

[2] Bauer DE, Orkin SH, Update on fetal hemoglobin gene regulation in Hemoglobinopathies. Curr Opin Pediatr 23 (2011) 1–8.

[3] 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 haemoglobin levels in adults. Proc Natl Acad Sci USA 104 (2007) 11346–11351.

[4] 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 105 (2008) 1869–11874.

[5] Bhatnagar P, Purvis S, Barron-Casella E, et al., Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. J Hum Genet 56 (2011) 316–323.

[6] Zhou D, Liu K, Sun CW, et al., KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat Genet 42 (2010) 742-744. doi: 10.1038/ng.637.

[7] Borg J, Papadopoulos P, Georgitsi M, et al., Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat Genet 42 (2010) 801–805.

[8] Gallienne AE, Dréau HM, Schuh A, Old JM, Henderson S., Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. Haematologica 97 (2012) 340-343. doi: 10.3324/haematol.2011.055442.

[9] Radmilovic M, Zukic B, Petrovic MS, et al., Functional analysis of a novel KLF1 gene promoter variation associated with hereditary persistence of fetal hemoglobin. Ann Hematol 92 (2013) 53-58. doi: 10.1007/s00277-012-1625-9.

[10] Thein SL, Menzel S., Discovering the genetics underlying foetal haemoglobin production in adults. Br J Haematol 145 (2009) 455-67. doi: 10.1111/j.1365-2141.2009.07650.x.

[11] Sampietro M, Thein SL, Contreras M, Pazmany L, Variation of HbF and F-cell number with the Gc Xmn I (C-T) polymorphism in normal individuals. Blood 79 (1992) 832–833.

[12] Zertal-Zidani S, Ducrocq R, Sahbatou M, Satta D, Krishnamoorthy R. Foetal haemoglobin in normal healthy adults: relationship with polymorphic sequences cis to the beta globin gene. Eur J Hum Genet. 10 (2002) 320-326.

[13] Jouini L, Bibi A, Ouali F, et al., Contribution of β-globin cluster polymorphisms to raise fetal hemoglobin levels in normal adults. Mol Biol Rep. 39 (2012) 4619-4625. doi: 10.1007/s11033-011-1253-9.

[14] 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 4 (1) (2009) e4218.

[15] Dover G-J, Smith K-D, Chang Y-C, Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. Blood 80 (1992) 816–824.

[16] Craig JE, Rochette J, Fisher CA, et al., Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. Nat Genet 12 (1996) 58–64.

[17] Weatherall DJ, Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2 (2001) 245–255.

[18] 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 USA 105 (2008) 1620–1625.

[19] Steinberg MH, Sebastiani S, Genetic modifiers of sickle cell disease, Am J Hematol 87 (2012) 795–803. http://dx.doi.org/10.1002/ajh.23232.

[20] Guida V, Cappabianca MP, Colosimo A, et al., Influence of Ggamma-158C --> and beta-(AT)x(T)y globin gene polymorphisms on HbF levels in Italian beta-thalassemia carriers and wild-type subjects. Haematologica 91 (2006) 1275-1276.

[21] Farrell JJ, Sherva RM, Chen ZY, et al., A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. Blood 117 (2011) 4935-4945. doi: 10.1182/blood-2010-11-317081.

[22] Dabke P, Colah R, Ghosh K, Nadkarni A, Effect of Cis Acting Potential Regulators in the β Globin Gene Cluster on the Production of HbF in Thalassemia Patients. Mediterr J Hematol Infect Dis 5 (2013) e2013012. doi: 10.4084/MJHID.2013.012.

[23] Gonçalves I, Ducrocq R, Lavinha J, et al., Combined effect of two different polymorphic sequences within the beta globin gene cluster on the level of HbF. Am J Hematol 57 (1998) 269-1276.

[24] 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 81 (2007) 559–575.

[25] Camaschella C, Mazza U, Roetto A, et al., Genetic interactions in thalassemia intermedia: analysis of b mutations, a genotype, c promoters, and b-LCR hypersensitive sites 2 and 4 in Italian patients. Am J Hematol 48 (1995) 82–87.

[26] Sankaran VG, Menne TF, Xu J, et al., Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A, Science 322 (2008) 1839–1842. doi:10.1126/science.1165409.

[27] Xu J, Sankaran VG, Ni M, ET AL., Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev 24 (2010) 783–798 doi:10.1101/gad.1897310

[28] Sankaran VG, Xu J, Ragoczy T, et al., Developmental and species-divergent globin switching are driven by BCL11A. Nature 460 (2009) 1093–1097. doi:10.1038/nature08243

[29] Xu J, Peng C, Sankaran VG, et al., Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science 334 (2011) 993–996. doi:10.1126/science.1211053

[30] Bauer DE, Kamran SC, Lessard S, et al., An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342 (2013) 253-257. doi: 10.1126/science.1242088.

[31] Jiang J, Best S, Menzel S, et al. cMYB is involved in the regulation of fetal hemoglobin production in adults. Blood 108 (2006) 1077-1083. DOI 10.1182/blood-2006-01-008912.

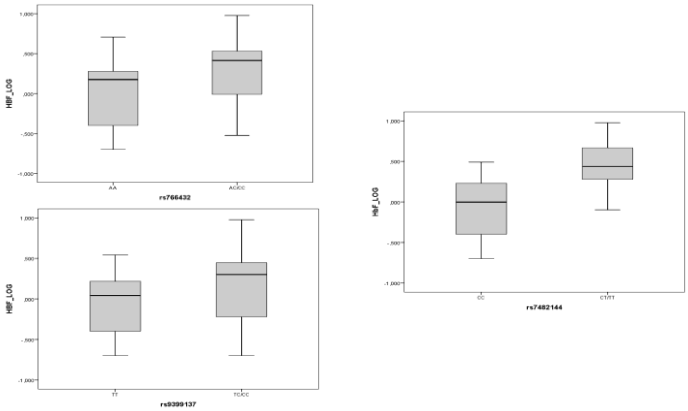
[32] Sankaran VG, Menne TF, Scepanovic D, et al. MicroRNA-15a and -16–1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc Natl Acad Sci USA 108 (2011) 1519-1524. doi: 10.1073/pnas.1018384108.

[33] Stadhouders R, Aktuna S, Thongjuea S, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. E.J Clin Invest 124 (2014):1699-1710. doi: 10.1172/JCI71520.

Legends

Figure 1: Box plots showing the distribution of log-transformed HbF levels within genotypes of the SNPs *BCL11A* rs766432, HMIP rs9399137 and HBG rs7482144 in Portuguese individuals with beta-thal minor. Each rectangle represents the data between the 25th and 75th quartiles, and the bar within each rectangle is the median value for HbF.

Figure 2: Box plots showing the distribution of log-transformed HbF levels within genotypes of the *BCL11A* SNPs rs11886868, rs766432 and rs7606173 in Portuguese individuals with normal hematological parameters. Each rectangle represents the data between the 25th and 75th quartiles, and the bar within each rectangle is the median value for HbF.



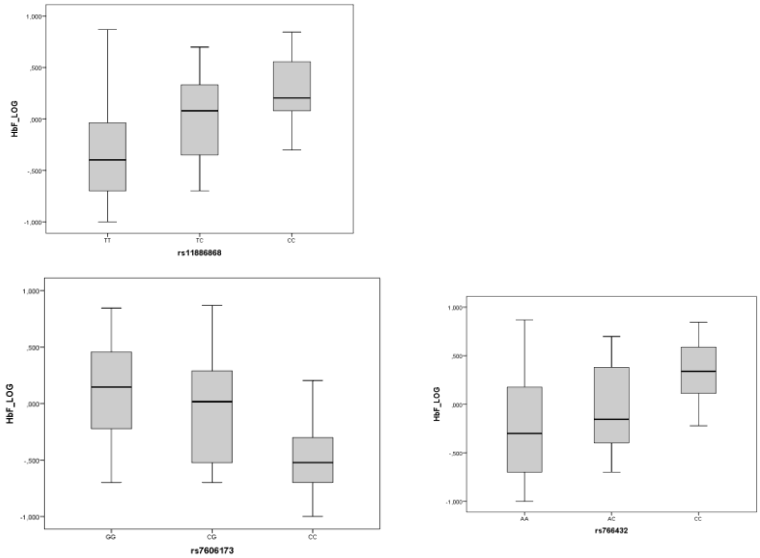


Table 1: Description of demographic and hematological data (mean ± SD) in the two population groups studied: beta-thalassemia carriers and subjects with normal hematological parameters.

Characteristics	**β-thal minor**	**Normal individuals**
Age (mean ± SD) (years old)	33.25 (±20.6)	29.87 (± 20.8)
Age (range) (years old)	2-77	2-75
Males (n)	30	26
Females (n)	35	34
Hematological parameters		
HbF (%)	1.35 (± 1.02)	1.36 (± 1.61)
HbF (range; %)	0.2-9.5	0.2-7.4
HbA2 (%)	4.76 (± 0.76)	2.65 (± 0.29)
HbA2 (range; %)	3.4-6.8	2.0-3.5
RBC ($x10^{12}$) *	5.54 (± 0.74)	4.79 (± 0.75)
Hb (g/dL) *	11.8 (± 1.46)	13.03(± 2.00)
MCV (fL) *	67.54 (± 6.77)	84.92 (± 9.30)
MCH (pg) *	21.44 (± 2.26)	27.53 (± 3.86)
MCHC (g/dL) *	31.78 (± 1.86)	32.37 (± 2.62)
RDW (CV %) *	15.78 (± 2.82)	14.40 (± 1.98)
Reticulocytes (% RBC) **	2.09 (± 1.25)	1.97 (± 1.55)
WBC ($x10^{3}$) *	8.08 (± 3.22)	7.23 (± 2.53)
Platelets ($x10^{6}$) *	283.28 (± 121.71)	286.58 (± 105.91)

Hematological parameters on 53 (*) and 39 (**) subjects with β-thal *minor*, and

29 (*) and 22 (**) normal subjects.

Table 2. HbF association results for SNPs associated with genes BCL11A, HBS1L-MYB (HMIP), HBG2 and KLF1 in individuals with beta-thalassemia minor of Portuguese origin.

Chr:position	Gene	SNP	N	Alleles (1:2)	MAF	P-HWE	β (SE)	Variance (%)	P	P*
2:60493111	BCL11A	rs11886868	64	C:T	0.250	0.739	0.084 (0.087)	1.49	0.336	0.271
2:60492835	BCL11A	rs766432	63	C:A	0.135	0.296	0.215 (0.097)	7.49	**0.029**	**0.011**
2:60498316	BCL11A	rs7606173	62	C:G	0.459	0.797	-0.086 (0.073)	2.23	0.247	0.163
6:135097880	HMIP	rs9399137	64	C:T	0.258	1	0.209 (0.079)	10.03	**0.011**	**0.018**
6:135130426	HMIP	rs6934903	52	A:T	0.135	1	0.151 (0.121)	3.03	0.217	0.381
11:5254939	HBG2	rs7482144	64	T:C	0.146	0.594	0.455 (0.082)	33.35	**$5.858x10^{-7}$**	**$6.773x10^{-7}$**
19:12887391	KLF1	rs3817621	62	C:G	0.218	0.263	0.172 (0.097)	5.04	0.079	0.110

Abbreviations: MAF: Minor allele frequency; Alleles: 1-minor, 2-major; N: number of samples; p-HWE: p-value for Hardy-Weinberg Equilibrium. P: p-value unadjusted; P*: p-value using age and sex as covariates.
The table includes the effect sizes of the minor allele (regression coefficient beta, β), standard error (SE) and *p*-values (Wald test asymptotic p-value) for the HbF levels log-transformed using a linear regression model.
Chromosome position (Chr:position) is according Ensembl.
Significant association p-values (P<0.05) are in bold.

Table 3. HbF association results for SNPs associated with genes BCL11A, HBS1L-MYB (HMIP), HBG2 and KLF1 in individuals with normal hematological parameters of Portuguese origin.

Chr:position	Gene	SNP	N	Alleles (1:2)	MAF (total)	P-HWE	MAF HbF≥2%) (n=15)	MAF (HbF≤1.6%) (n=45)	OR (CI 95%)	P	P*
2:60493111	BCL11A	rs11886868	60	C:T	0.417	0.187	0.667	0.333	4 (1.66 - 9.61)	**0.001**	**0.0015**
2:60492835	BCL11A	rs766432	60	C:A	0.308	0.222	0.533	0.233	3.75 (1.57 - 8.94)	**0.002**	**0.0034**
2:60498316	BCL11A	rs7606173	59	C:G	0.398	0.175	0.233	0.454	0.36 (0.14 - 0.93)	**0.032**	**0.041**
6:135097880	HMIP	rs9399137	60	C:T	0.233	1	0.300	0.211	1.60 (0.63 - 4.06)	0.318	0.437
6:135130426	HMIP	rs6934903	60	A:T	0.142	0.319	0.200	0.122	1.79 (0.60 - 5.36)	0.290	0.231
11:5254939	HBG2	rs7482144	60	T:C	0.300	0.125	0.300	0.300	1 (0.40 - 2.46)	1	0.815

19:12887391	KLF1	rs3817621	60	C:G	0.233	0.716	0.133	0.267	0.42 (0.13 - 1.33)	0.134	0.298

Abbreviations: OR: Odds Ratio; CI, confidence interval; MAF: Minor allele frequency; Alleles: 1-minor, 2-major; N: number of samples; P-HWE: p-value for Hardy-Weinberg Equilibrium. P: p-value unadjusted; P*: Fisher's p-value using age and sex as covariates.
Association was tested under a case-control model considering a cut-off of 2% for HbF level. The table includes the OR (shown for the minor allele), CI 95% and p-values for allelic association obtained with Fisher's test.
Chromosome position (Chr:position) is according Ensembl.
Nominal significant association p-values (P<0.05) are in bold.

Table S1: Associations of the HbF levels (log-transformed) with genotypes of the studied SNPs in beta-thalassemia carriers of Portuguese origin (n=64).

Loci	SNP	Genotypes HbF Mean (SE)		P
BCL11A	rs11886868	TT (n=35) 1.44 (0.19)	TC+CC (n=29) 1.95 (0.37)	0.380
	rs766432	AA (n=48) 1.38 (0.15)	AC+CC (n=15) 2.71 (0.62)	**0.034**
	rs7606173	GG (n=17) 2.08 (0.32)	GC+CC (n=45) 1.40 (0.17)	0.054
HMIP	rs9399137	TT (n=35) 1.64 (0.14)	TC+CC (n=29) 2.30 (0.37)	**0.005**
	rs6934903	TT (n=39) 1.59 (0.28)	TA+AA (n=13) 2.08 (0.42)	0.135
HBG2	rs7482144	CC (n=48) 1.11 (0.11)	CT+TT (n=16) 3.34 (0.53)	**<0.001**
KLF1	rs3817621	GG (n=36) 1.38 (0.17)	GC+CC (n=26) 2.17 (0.40)	0.119

Abbreviations:
n: number of subjects within each genotype (genotypes heterozygous and homozygous with the derived allele were combined).
P: p-value obtained with Mann-Whitney U test.
HbF levels shown are not-transformed.
Significant p-values (P<0.05) are in bold.

Table S2: Associations of the HbF levels (log-transformed) with genotypes of the studied SNPs in individuals with normal haematological parameters of Portuguese origin (n=60).

Loci	SNP	Genotypes HbF Mean (SD)			P**
BCL11A	rs11886868	**TT** (n=23) 0.88 (0.31)	**TC** (n=24) 1.64 (0.31)	**CC** (n=13) 2.39 (0.52)	**<0.001**
	P*		**0.009**	**<0.001**	
	rs766432	**AA** (n=31) 1.03 (0.23)	**AC** (n=21) 1.64 (0.31)	**CC** (n=8) 2.8 (0.75)	**0.008**
	P*		0.266	**0.008**	
	rs7606173	**GG** (n=24) 1.89 (0.35)	**CG** (n=24) 1.58 (0.37)	**CC** (n=11) 0.53 (0.16)	**0.006**
	P*		0.591	**0.004**	
HMIP	rs9399137	**TT** (n=35) 1.51 (0.27)	**TC** (n=22) 1.37 (0.36)	**CC** (n=3) 2.63 (1.06)	0.336
	P*		1	0.690	
	rs6934903	**TT** (n=45) 1.54 (0.27)	**TA** (n=13) 1.42 (0.34)	**AA** (n=2) 1.4 (0.80)	0.914
	P*		1	1	
HBG2	rs7482144	**CC** (n=32) 1.62 (0.29)	**CT** (n=20) 1.36 (0.37)	**TT** (n=8) 1.43 (0.62)	0.462
	P*		0.431	0.893	
KLF1	rs3817621	**GG** (n=36) 1.77 (0.31)	**CG** (n=20) 1.21 (0.27)	**CC** (n=4) 0.67 (0.31)	0.325
	P*		0.864	0.648	

Abbreviations:
n: number of subjects within each genotype.
P*: Bonferroni post-hoc p-values for multiple comparisons, between homozygous genotypes for the ancestral allele and heterozygous and homozygous genotypes with the derived allele.
P**: P-value obtained with ANOVA.
HbF levels shown are not-transformed.
Significant P-values (P<0.05) are in bold.