

# Intergenic variants of *HBS1L-MYB* are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults

Swee Lay Thein<sup>\*†‡</sup>, Stephan Menzel<sup>\*</sup>, Xu Peng<sup>§</sup>, Steve Best<sup>\*</sup>, Jie Jiang<sup>\*</sup>, James Close<sup>\*¶</sup>, Nicholas Silver<sup>\*</sup>, Ageliki Gerovasilli<sup>\*</sup>, Chen Ping<sup>§</sup>, Masao Yamaguchi<sup>§</sup>, Karin Wahlberg<sup>\*</sup>, Pinar Ulug<sup>\*</sup>, Tim D. Spector<sup>¶</sup>, Chad Garner<sup>\*\*</sup>, Fumihiko Matsuda<sup>§</sup>, Martin Farrall<sup>††</sup>, and Mark Lathrop<sup>§</sup>

<sup>\*</sup>King's College London School of Medicine, Division of Gene and Cell Based Therapy, King's College Hospital, London SE5 9PJ, United Kingdom; <sup>†</sup>Department of Haematological Medicine, King's College Hospital, Denmark Hill, London SE5 9PJ, United Kingdom; <sup>§</sup>Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique, 91057 Evry, France; <sup>¶</sup>SANE/POWIC, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford OX3 7JX, United Kingdom; <sup>¶¶</sup>Twin Research and Epidemiology Unit, King's College London, St. Thomas' Hospital Campus, Lambeth Palace Road, London SE1 7EH, United Kingdom; <sup>\*\*</sup>Epidemiology Division, Department of Medicine, University of California, Irvine, CA 92697-7550; and <sup>††</sup>Department of Cardiovascular Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom

Edited by Yuet Wai Kan, University of California School of Medicine, San Francisco, CA, and approved May 22, 2007 (received for review December 21, 2006)

**Individual variation in fetal hemoglobin (HbF,  $\alpha_2\gamma_2$ ) response underlies the remarkable diversity in phenotypic severity of sickle cell disease and  $\beta$  thalassemia. HbF levels and HbF-associated quantitative traits (e.g., F cell levels) are highly heritable. We have previously mapped a major quantitative trait locus (QTL) controlling F cell levels in an extended Asian-Indian kindred with  $\beta$  thalassemia to a 1.5-Mb interval on chromosome 6q23, but the causative gene(s) are not known. The QTL encompasses several genes including *HBS1L*, a member of the GTP-binding protein family that is expressed in erythroid progenitor cells. In this high-resolution association study, we have identified multiple genetic variants within and 5' to *HBS1L* at 6q23 that are strongly associated with F cell levels in families of Northern European ancestry ( $P = 10^{-75}$ ). The region accounts for 17.6% of the F cell variance in northern Europeans. Although mRNA levels of *HBS1L* and *MYB* in erythroid precursors grown *in vitro* are positively correlated, only *HBS1L* expression correlates with high F cell alleles. The results support a key role for the *HBS1L*-related genetic variants in HbF control and illustrate the biological complexity of the mechanism of 6q QTL as a modifier of fetal hemoglobin levels in the  $\beta$  hemoglobinopathies.**

$\beta$  hemoglobinopathies | complex trait | F cells

Sickle cell disease and  $\beta$  thalassemia are among the most common genetic diseases worldwide and have a major impact on global health and mortality (1). Both these hemoglobinopathies display a remarkable diversity in their disease severity. A major ameliorating factor is an innate ability to produce fetal hemoglobin (HbF,  $\alpha_2\gamma_2$ ). HbF levels vary considerably, not only in patients with these  $\beta$  hemoglobin disorders, but also in healthy normal adults. The distribution of HbF and F cells (FCs, erythrocytes that contain measurable HbF) in healthy adults is continuous and positively skewed. Although the majority of adults have HbF of less than 0.6% of total hemoglobin, 10–15% of individuals have increases ranging from 0.8% to 5% (2). The latter individuals are considered to have heterocellular hereditary persistence of fetal hemoglobin (hHPFH), which refers to the uneven distribution of HbF among the erythrocytes. When coinherited with  $\beta$  thalassemia or sickle cell disease, hHPFH can increase HbF output to levels that are clinically beneficial (3, 4).

FC levels are strongly correlated with HbF in adults within the normal range (including hHPFH) (2), and F cells are generally used as an indirect measure of HbF within normal individuals because of the poor sensitivity for HbF assay in the lower range (see *Materials and Methods*). A logarithmic transformation of FC removes skewness and leads to a distribution that is approximately normal for a representative population sample (5). The

heritability of HbF and FC is estimated to be 89% (5). Cis-acting variants and rare mutations at the  $\beta$  globin gene locus explain some of the variability (2), but >50% of the variance is unlinked to this locus (6). Our previous study of an Asian-Indian kindred in which  $\beta$  thalassemia and hHPFH were segregating identified a QTL that mapped to a 1.5-Mb interval on chromosome 6q23 with a logarithm of odds (lod) score of 6.3 (7, 8) (Fig. 1*a*). This interval contains five known protein-coding genes (*ALDH8A1*, *HBS1L*, *MYB*, *AH11*, and *PDE7B*), none of which harbored mutations (nonsynonymous variants), and three (*HBS1L*, *MYB*, and *AH11*) are expressed in erythroid progenitor cells (9, 10).

To explore further the role of the 6q23 QTL on HbF control, we studied two panels (824 and 1,217 individuals, respectively) of twin pairs of North European origin. In a high-resolution association study, we identified multiple genetic variants that are strongly associated with FC levels in the Caucasian controls ( $P = 10^{-75}$ ). These genetic variants reside in three linkage disequilibrium (LD) blocks within *HBS1L* and 5' to *HBS1L*, and *MYB* in the intergenic region.

To delineate the functional significance of these genetic variants, we performed an expression profile of *HBS1L* and *MYB* during erythropoiesis. We observed a striking correlation of increased *HBS1L* expression in erythroid progenitor cells with presence of the single nucleotide polymorphisms (SNPs) associated with high trait values in the three LD blocks. The present study illustrates the power of QTL mapping for positional identification of trans-acting genetic variants influencing regulation of HbF levels, a major ameliorating factor of sickle cell disease and  $\beta$  thalassemia.

## Results

We genotyped two panels (824 and 1,217 individuals, respectively; see Table 1) of twin pairs of North European origin

Author contributions: S.L.T. designed research; S.M., X.P., S.B., J.J., J.C., N.S., A.G., C.P., M.Y., K.W., and P.U. performed research; T.D.S. contributed valuable material; S.M., C.G., F.M., M.F., and M.L. analyzed data; and S.L.T., M.F., and M.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: QTL, quantitative trait locus/loci; HbF, fetal hemoglobin; FC, F cells; lod, logarithm of odds; hHPFH, heterocellular hereditary persistence of fetal hemoglobin; LD, linkage disequilibrium; HMIP, *HBS1L MYB* intergenic polymorphism; HPRT, hypoxanthine phosphoribosyltransferase; MZ, monozygotic; DZ, dizygotic.

<sup>†</sup>To whom correspondence should be addressed. E-mail: sl.thein@kcl.ac.uk.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0611393104/DC1](http://www.pnas.org/cgi/content/full/0611393104/DC1).

© 2007 by The National Academy of Sciences of the USA



**Table 2. Significance tests for sex, age, XmnI-G $\gamma$ , and the three selected *HBS1L-MYB* markers**

Fixed effects <i>P</i> values							
Panel	Sex	Age	XmnI-G $\gamma$ Additive	HBS1L-MYB			
				Block 1 (rs52090901)	Block 2 (rs9399137)		Block 3 (rs6929404)
				Additive	Additive	Dominance	Additive
Panel 1	0.0001	0.0005	10 <sup>-14</sup>	0.006	10 <sup>-24</sup>	0.0005	0.01
Panel 2	0.004	0.0005	10 <sup>-22</sup>	0.002	10 <sup>-24</sup>	0.06	0.005
Combined	10 <sup>-5</sup>	10 <sup>-7</sup>	10 <sup>-35</sup>	10 <sup>-5</sup>	10 <sup>-45</sup>	0.0004	0.0002

The significance tests are conditional on the presence of the nontested parameters in the model. For *HBS1L-MYB*, these are different from *P* values for the marginal test statistics in SI Table 4 because of partial LD between the markers. *P* values for dominance are shown only when significant. We employed a stepwise statistical procedure to select the markers shown here. New markers were incorporated into the ANOVA only if they accounted for a significant proportion of the trait variance when more strongly associated markers were already included in the model (thus accounting for linkage disequilibrium with these). We selected the marker with the most significant test statistic to incorporate at each step until no remaining markers gave a significant trait association ( $P > 0.01$ ). We obtained equivalent results using either the markers genotyped in the first twin panel, or the combined data with markers that were characterized in both panels.

Polymorphisms were identified by resequencing *MYB* and  $\approx 78$  kb of the *HBS1L-MYB* intergenic region. We identified 184 markers for which the minor allele had 5% or greater frequency, 94 of which were selected for genotyping based on their positions, linkage disequilibrium patterns and intermediate association results. We added 27 markers from public databases to provide additional coverage particularly in the 3' flanking regions of *MYB* and *HBS1L*. Altogether, 121 markers were genotyped with average spacing of 4.4 kb, and higher density (1.8 kb average) in the *HBS1L-MYB* intergenic region [supporting information (SI) Table 3].

Twenty-eight markers provided very strong evidence of association ( $P < 10^{-8}$ ) in the first panel, with the most significant results concentrated at sites between *HBS1L* and *MYB* (Fig. 1b). In particular, a 24-kb segment starting 33 kb upstream of *HBS1L* contained 12 markers showing very strong association (*P* values between 10<sup>-28</sup> and 10<sup>-39</sup>, block 2 in Fig. 1a) whereas the other 13 markers from within this segment are less significantly associated with the trait (SI Table 4). Strikingly, the 12 markers with the strongest trait association have similar allele frequencies and are in complete linkage disequilibrium (except for haplotypes with frequency <2%), whereas the others exhibit different linkage disequilibrium patterns (SI Tables 5–7). We confirmed the association by characterizing a subset of 75 markers from the *HBS1L-MYB* interval in the second twin panel (Fig. 1b). The twelve markers with the strongest trait association in the first panel are also the most strongly associated in the second panel (Fig. 1b and SI Table 4). When the data from the two panels were combined, these markers had association *P* values of 10<sup>-50</sup> to 10<sup>-75</sup>.

Markers outside of the 24-kb interval also showed consistent evidence of association in the two panels. In some instances, linkage disequilibrium between trait-associated markers appeared weak. We hypothesized that more than one variant contributes to the QTL. A stepwise statistical selection procedure led to the identification of three markers that accounted independently for a significant proportion of the trait variance even with the other markers included in the ANOVA (Table 2). The three markers were selected in the following order (with *P* values from the combined data for the significance calculated with previously selected markers included): rs9399137 ( $P = 10^{-75}$ ), rs52090901 ( $P = 10^{-10}$ ) and rs6929404 ( $P = 0.0002$ ). The first of these SNPs (rs9399137) is 1 of 12 markers in *HBS1L-MYB* intergenic polymorphism (HMIP) block 2 (so-labeled because of its physical position) with the strongest trait associations. We identified multiple markers in two other trait-associated blocks in strong linkage disequilibrium with rs52090901 (HMIP block 1) and rs6929404 (HMIP block 3) (Fig. 1a). Minor differences in

the association statistics for markers in the same block could be attributed to rare haplotypes and/or a small amount of missing genotype data.

**A Novel Transcript of *HBS1L*.** As part of our characterization of the *HBS1L-MYB* intergenic region, we confirmed by RT-PCR and sequence analysis the existence of a previously uncharacterized transcript of *HBS1L* that is expressed in thymus, Jurkat cells, peripheral leukocytes, and at minimal levels in erythroid progenitors. The transcript was deduced from the sequence of a thymus cDNA clone deposited in a public database (Japanese Database of Transcriptional Start Sites; DBTSS; <http://dbtss.hgc.jp/>; GenBank ID DB114698). This transcript contains an alternative 119-bp first exon (denoted exon 1a), which starts  $\approx 45$  kb upstream of the previously described first exon of the gene (Fig. 1a and SI Fig. 2). A 102-bp repeat-free segment that starts 129 bp upstream of the initiation codon has marked nucleotide homology with other mammals and contains binding site motifs for a putative TATA box and three members of the GATA family of transcription factors (GATA-1, -2, and -3) that regulate gene expression in hematopoietic tissue during both development and adult life (17).

**Expression Profile of *HBS1L* and *MYB* During Erythropoiesis.** To investigate the functional significance of the trait-associated genetic variants, we used real-time quantitative RT-PCR to study the expression levels of *HBS1L* and *MYB* during erythropoiesis. As *HBS1L-1a* was expressed at very low levels in erythroid progenitors, it was excluded from the study. Erythroid cells obtained from 35 individuals (23 from the twin-pair panels, 2 from the Asian-Indian pedigree and 10 from other Caucasian volunteers) were cultured by using a two-phase liquid system as described (10), and RT-PCR was performed with total RNA obtained from erythroid progenitor cells on days 0 and 3 phase II erythroid culture for each individual included in the study. We hypothesized that contrasts between the extreme genotypes would be the most informative to detect effects on expression, so individuals who were homozygous at the trait-associated sites within block 2 were chosen for these studies. Alleles associated with high trait values for a block are denoted as “H” and the alleles associated with low trait values for a block as “L”. The genotype status was usually equivalent for all of the markers within a block because of the strong linkage disequilibrium. In a few instances when the genotypes were not equivalent, we classified individuals according to the predominant pattern (see legend to SI Fig. 3).

HbF and FC levels were significantly associated with geno-

types in the three blocks in the samples selected for the expression study, as expected. We observed a striking relationship of increased *HBSIL* expression measured at day 0 associated with the presence of the H genotype in the three trait associated blocks, and a statistically less significant relationship for day 3 expression (SI Fig. 3). These results strongly suggest that the biological effects of genetic variants in one or more of these blocks include modulation of *HBSIL* expression.

## Discussion

Our study has identified the principal genetic variants that account for the chromosome 6q QTL for F cells/HbF. These variants are distributed within three LD blocks, which we refer to as *HBSIL* *MYB* Intergenic Polymorphism (HMIP) blocks 1, 2 and 3. HMIP blocks 1, 2 and 3 span a nearly contiguous segment  $\approx 79$  kb long, starting 188 bp upstream from *HBSIL* exon 1 and ending 45 kb upstream of *MYB* (Fig. 1a). Among the 12 markers exhibiting the strongest evidence of association, one, rs52090909, is located in the 5' UTR of exon 1a of *HBSIL*. The other strongly associated markers in HMIP block 2, are either in intron 1a (rs9376090, rs9399137, rs9402685 and rs11759553), or directly upstream of the 5' UTR of *HBSIL* exon 1a (rs4895440, rs4895441, rs9376092, rs9389269, rs9402686, rs11154792 and rs9483788). HMIP block 1 is also located within intron 1a of *HBSIL* whereas HMIP block 3 is located between exon 1a of *HBSIL* and the first exon of *MYB*. Whereas markers within each of the trait-associated blocks are in strong linkage disequilibrium, there is less linkage disequilibrium and a greater diversity of frequent haplotypes between markers in different blocks (SI Table 8). The markers interspersed within a trait-associated block that are less significantly associated with the trait have lower linkage disequilibrium with the block markers (SI Tables 5–7). Each of the trait-associated blocks contains at least one marker that had also been characterized in the HapMap data set (18). As we found no significant linkage disequilibrium with HapMap markers outside of the region studied here, we concluded that the trait-associated blocks were confined to the *HBSIL*-*MYB* segment. A test of linkage in the European dizygotic (DZ) twins showed that the 6q23 QTL is completely accounted for by the markers in the three trait-associated blocks (unadjusted lod = 1.79,  $P = 0.002$ ; lod adjusted for three markers that identify the trait-associated blocks = 0.0).

Based on measured haplotype analysis (SI Tables 8 and 9), we estimate that 17.6% of the trait variance is attributed to the markers in the three *HBSIL*-*MYB* blocks. An additional 11.6% of the trait variance is influenced by the *Xmn* I variant on chromosome 11. As the overall heritability of the FC trait in Europeans is 89% (5), it is suggested that additional genetic or other familial factors contribute substantially (residual heritability = 59.8%) to the trait variation. The genetic variants that are associated with high F cell levels are also strongly correlated to increased expression of *HBSIL* in cultured erythroid cells.

Interestingly, however, FC levels and *HBSIL* expression were not significantly correlated in this sample set despite the association of both traits with the same genetic variants. Examination of the samples showed that it was principally due to the inclusion of two individuals with high FC values who harbor the LL genotype and exhibit low *HBSIL* expression. The presence of such samples is not unexpected given the selection on genotype, and the fact that most of the FC trait variance (82%) is not accounted for by the *HBSIL*-*MYB* locus.

In a previous study of 26 individuals selected to have high or low HbF, we found a negative correlation between FC levels and *HBSIL* expression (10). The previous sample partially overlaps with the present data set, but it contains 13 (50%) individuals with the block 2 H/L genotype, and only 13 with H/H or L/L genotypes. In an attempt to reconcile the results in these two data sets, we reexamined *HBSIL* expression by repeating all of

the RT-PCR experiments. Using the new data from all 47 individuals in the combined sample set, we found significant association of block 2 genotypes with FC levels ( $P = 0.007$ ) and with *HBSIL* expression (day 0:  $P = 0.01$ ; day 3:  $P = 0.03$ ). After adjustment for genotype effects under an additive model, the residual FC trait and *HBSIL* expressions values were negatively correlated (day 0:  $\rho = -0.31$ ,  $P = 0.04$ ; day 3:  $\rho = -0.39$ ,  $P = 0.01$ ) as reported in the original subset. We conclude that multiple factors affect both the FC trait and *HBSIL* expression, and that these factors include, but are not limited to the genetic variants within *HBSIL*-*MYB* region. The sampling scheme used for ascertainment (e.g., selection on genotype or phenotype) may impact the magnitude and the direction of the observed relationships.

The biological complexity underlying gene regulation in this region is further illustrated through analysis of *MYB* expression. Although *MYB* expression was not significantly related to the genotype status (SI Fig. 4) or to FC levels in the block 2 H/H vs. L/L samples, *MYB* expression at day 3 was positively correlated to *HBSIL* expression (SI Table 10). Moreover, significant correlation remained after adjustment of *HBSIL* for the associated *HBSIL*-*MYB* genotypes. Thus, it would seem that the correlation of *HBSIL* and *MYB* expression is principally due to factors outside of the *HBSIL*-*MYB* locus.

The location of the most significantly associated variants and their correlation with *HBSIL* expression implicate *HBSIL* in the F Cell QTL. *HBSIL* (16), is a putative member of the “GTPases” superfamily (19), which bears a close relationship to the eEF-1A (eukaryotic elongation factor 1A, or EF 1 $\alpha$ ) and eRF3 (eukaryotic release factor 3) families (16, 20). GTPases, which bind and hydrolyze GTP, are involved in regulating a variety of critical cellular processes, including protein synthesis, cytoskeleton assembly, protein trafficking and signal transduction (19). Recently it has been shown that another GTP-binding protein, the secretion-associated and RAS-related (SAR) protein may be a key molecule in the induction of  $\gamma$ -globin expression by hydroxyurea (21). The role of *HBSIL* on FC levels is not immediately apparent and could be manifested indirectly through its effect on the expression of various cytokines and transcription factors that impact erythroid cell growth (15).

The present study illustrates how genetic approaches can contribute new knowledge to the regulation of human hemoglobin through dissection of the quantitative genetic variation. The identification of novel transacting genetic variants that are associated with modulation of HbF and FC levels is a key step toward resolving some of the outstanding biological questions in the field and has the potential for novel diagnostic and therapeutic applications.

## Materials and Methods

**Subjects and Phenotyping.** Study participants consisted of monozygotic and same-sex DZ twin pairs of North European descent. The study participants were phenotyped for F-cell levels and genotyped for the *Xmn*I-G $\gamma$  site and 121 other markers. The twin pairs who were not selected for HbF or F-cell levels or any disease or trait, were recruited from the TwinsUK Adult Twin Registry ([www.twinsuk.ac.uk](http://www.twinsuk.ac.uk)) (11). The average age of the participants was 47.6 years of age, ranging from 18 to 79 years of age. The average FC level of the sample was 4.06% of total erythrocytes (SD 3.15%; range 0.23% to 36.7%).

Blood samples were collected in EDTA, and F cells were enumerated by flow cytometry of 20,000 cells by using a monoclonal anti- $\gamma$  globin chain antibody conjugated with fluorescein isothiocyanate (FITC) (22). Current methods of quantifying HbF are not sensitive enough for measuring levels in the 0–1% range, the range usually encountered in normal subjects. Therefore, in normal subjects, the trait is represented by F cells

measured by using a monoclonal antibody against  $\gamma$  chains of HbF ( $\alpha_2\gamma_2$ ).

The study was approved by the local Research Ethics Committee (LREC No: 01-332 and LREC No: 01-083) of King's College Hospital, London. XmnI-G $\gamma$  genotyping was performed on genomic DNA as described (23).

**SNP Discovery.** A systematic investigation of genetic variants between *HBSIL-MYB* was made by resequencing this 125-kb region by using DNA from 32 European control subjects. The genomic sequence encompassing the region (NT\_025741.13, 39,480,452–39,606,881, 126,430 bp) was excised with 1-kb each of adjacent sequences at both ends. PCR primers were designed by PRIMER3 to generate a total of 139 PCR amplicons (ranging from 759 bp to 1,725 bp, with an average length of 1,208 bp) with an overlap of >160 bp between adjacent amplicons. In addition, 428 internal primers were also used for sequencing. Resequencing of the human *MYB* gene was performed with 50 PCR amplicons generated by PRIMER3 to cover the 15 exons and parts of the introns. PCR was undertaken in 15- $\mu$ l reaction volume by using 1 unit of ExTaq DNA polymerase (TaKaRa Biomedicals, Paris, France) and 25 ng of genomic DNA. The PCR profile consisted of an initial melting step of 5 min at 94°C, followed by 35 cycles of 5 s at 98°C, 30 s at 60°C, and 2 min at 72°C; and a final elongation step of 10 min at 72°C. PCR products were purified by using Bio-gel P100 Gel (Bio-Rad Inc, Hercules, CA). PCR products were sequenced by using the Bigdye Terminator cycle sequencing chemistry method. Reactions were purified by using Sephadex G-50 Superfine (Amersham Biosciences, Uppsala, Sweden) before applying to the ABI 3730 DNA Analyzers. Detection of genetic variants was performed with in-house software (the Genalys program available at <http://www.cng.fr>).

**Erythroid Cell Cultures and Expression Analysis of *HBS1L* and *MYB* by Quantitative Real-Time PCR.** Erythroid cells were cultured by using a two-phase liquid system [modified from Fibach *et al.* (24)]. Mononuclear cells were isolated from peripheral blood by centrifugation on a gradient of Ficoll-Hypaque and cultured for 7 days in phase I medium which consisted of serum-free StemSpan (Stem Cell Technologies, U.K.) supplemented with 1  $\mu$ g/ml cyclosporin A, 25 ng/ml interleukin-3 (IL-3), 50 ng/ml human stem cell factor (Sigma, U.K.), and 0.01% BSA. Cells were incubated at 37°C, 5% CO<sub>2</sub>. After 7 days, nonadherent cells were collected and reseeded at a concentration of  $2.5 \times 10^5$  cells/ml in phase II medium [StemSpan supplemented with  $10^{-7}$  M dexamethasone (Sigma, U.K.), 50 ng/ml stem cell factor, and 2 units/ml human recombinant erythropoietin (EPO; Sigma, U.K.). The cultures were diluted once or twice to maintain the cell concentration lower than  $1 \times 10^6$  cells per ml in phase II. Cell samples were collected from phase II cultures on days 0 and 3.

Total RNA was isolated from erythroid cells by using Tri-reagent (Sigma, U.K.) and quantified by absorbance at 260 nm. cDNA was synthesized by using SuperScript III reverse transcriptase (Invitrogen, U.K.) from 1  $\mu$ g of total RNA. Primers and probes were designed by using Primer Express 2.0 program and synthesized by Applied Biosystems. Quantitative RT-PCR was carried out in an ABI 7900 HT Sequence Detection System by using TaqMan master mix and the protocol of the manufacturer (Applied Biosystems). Sequences of the primers and probes were as follows: *MYB* probe, 6-FAM-TGCTACCAACACAGAACCACACATGCA-TAMRA; *MYB* forward primer, 5'-ATGATGAAGACCCCTGAGAAG-GAAA-3'; *MYB* reverse primer, 5'-AACAGGTGCACTGTCTC-CATGA-3'; *HBS1L* probe, 6-FAM-CTATAACTACGATGAA-GATTTT-TAMRA; *HBS1L* forward primer, 5'-TCTAC-AGACTGGCCGTAGAGATCA-3' (in exon 2); *HBS1L* reverse primer, 5'-CCCGCATCGGAATGTT-3' (in exon 1).

All data were normalized by using the endogenous hypoxanthine phosphoribosyltransferase (HPRT) control. Assays for HPRT are available from the Applied Biosystem database. To quantify gene expression, a relative standard method was used. The quantities of targets and of the endogenous HPRT were determined from the appropriate standard curves. The target amount was then divided by the HPRT amount to obtain a normalized value. One of the experimental samples on day 0 (HPRT normalized) was designated as the calibrator, and given a relative value of 1.0. All quantities (HPRT normalized) were expressed as *n*-fold relative to the calibrator.

**RNA Analysis.** RNA was obtained from Clontech-Europe, U.K. or prepared from cultured cells by using Tri Reagent (Sigma, U.K.) according to manufacturer's instructions. One microgram of total RNA was reverse transcribed by using SuperScript III RT (Invitrogen, U.K.) and oligo(dT) primers. One hundred nanograms of cDNA was then used in a 25- $\mu$ l PCR containing TaqGold (Applied Biosystems, U.K.) at 2.5 mM MgCl<sub>2</sub> and 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s.

**Genotyping.** Markers in the target region were selected for genotyping from the dbSNP ([www.ncbi.nlm.nih.gov/SNP/](http://www.ncbi.nlm.nih.gov/SNP/)) and HapMap ([www.hapmap.org/](http://www.hapmap.org/)) databases, or from the sequencing experiments described above. Most markers were genotyped by Taqman (Applied Biosystems, Foster City, CA). Taqman reactions were performed according to the manufacturer's instructions by using 5.0 ng of purified and quantified genomic DNA. Plate reading was conducted on ABI Prism 7900HT sequence Detection System, and analysis was undertaken with SDS 2.0 software. A small number of markers were genotyped by direct sequencing with techniques as described above, or by using the tetra primer ARMS method (25). The T homopolymer upstream of *MYB* (Bpil) was genotyped on a microsatellite genotyping platform from Applied Biosystems, by using an ABI Prism 3100 Genetic Analyzer.

**Statistical Methods.** The relationship of the quantitative trait with age, sex, and marker genotypes was evaluated by using the mixed-model ANOVA procedure (PROC MIXED) from SAS version 8.2 (SAS Institute Inc., Cary, NC) with restricted maximum likelihood estimation. Monozygotic (MZ) and DZ twins were assumed to have distinct trait variances and covariances. The combined data from Panel 1 and Panel 2 were analyzed assuming common trait variances and equal covariances for MZ and DZ twins in the two panels. Age, sex and marker genotypes were incorporated as fixed effects for analysis. Likelihood ratio tests were used to evaluate hypotheses involving equality of the variances and covariances in different subsets of the data, and to test the fit of the additive genetic model. Haplotype estimates were obtained with the MERLIN and fugue programs (26) and haploview programs (27).

PAP (version 4.2; <http://hasstedt.genetics.utah.edu/>) was used to estimate effects and to obtain likelihood ratio test statistics in the measured haplotype analysis by modifying the measured genotype procedure (qmlprmv). Briefly, the phenotype trait was simultaneously adjusted for age, sex and the XmnI-G $\gamma$  marker whilst fitting a measured genotype model. The variance, correlations for DZ and MZ twins, haplotype means and dominance terms were estimated by maximum likelihood conditional on the observed genotypes at the sites included in the model, the adjusted trait phenotype and the family structure. MZ twins were constrained to be identical-by-descent at the *HBSIL-MYB* locus by inclusion of a completely linked and fully informative indicator marker. The mean associated with the combination of two haplotypes,  $H_i$  and  $H_j$ , was written as  $M_i + M_j$ , except when considering dominance. In the latter case, the mean was ex-

pressed as  $M_i + M_j + D_S$  for haplotype combinations with presence of hypothesized dominant allele at site S. Under the between-site additive model, the haplotype mean was written as the sum of means associated with the alleles at each site, plus a site-specific dominance term when it was included in the model. Likelihood ratio tests were used to test specific hypotheses involving nested models. A variance-components linkage analysis of FC levels in the DZ twins was performed with the MERLIN program (26) allowing for linkage disequilibrium

between markers (28). Tests of population stratification (admixture) were performed with the QTDT program (29).

We thank C. Steward for help in preparation of the manuscript, and M. Foglio, S. Heath, D. Zelenika, A. Boland, and C. Julier for valuable advice and technical assistance. This work was supported by Medical Research Council Grant G0000111, ID51640 (to S.L.T.) and the French Ministry of Higher Education and Research (M.L.). TwinsUK is supported by the Wellcome Trust and Framework V European Union Grant GenomEU Twin.

- Weatherall DJ, Clegg JB (2001) *Bull WHO* 79:704–712.
- Thein SL, Craig JE (1998) *Hemoglobin* 22:401–414.
- Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP (1994) *N Engl J Med* 330:1639–1644.
- Ho PJ, Hall GW, Luo LY, Weatherall DJ, Thein SL (1998) *Br J Haematol* 100:70–78.
- Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL (2000) *Blood* 95:342–346.
- Garner C, Tatu T, Game L, Cardon LR, Spector TD, Farrall M, Thein SL (2000) *GeneScreen* 1:9–14.
- Craig JE, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, Demenais F, Thein SL (1996) *Nat Genet* 12:58–64.
- Garner C, Mitchell J, Hatzis T, Reittie J, Farrell M, Thein SL (1998) *Am J Hum Genet* 62:1468–1474.
- Close J, Game L, Clark BE, Bergounioux J, Gerovassili A, Thein SL (2004) *BMC Genomics* 5:33.
- Jiang J, Best S, Menzel S, Silver N, Lai MI, Surdulescu GL, Spector TD, Thein SL (2006) *Blood* 108:1077–1083.
- Spector TD, MacGregor AJ (2002) *Twin Res* 5:440–443.
- Sampietro M, Thein SL, Contreras M, Pazmany L (1992) *Blood* 79:832–833.
- Emambokus N, Vegiopoulos A, Harman B, Jenkinson E, Anderson G, Frampton J (2003) *EMBO J* 22:4478–4488.
- Oh IH, Reddy EP (1999) *Oncogene* 18:3017–3033.
- Cantor AB, Orkin SH (2002) *Oncogene* 21:3368–3376.
- Wallrapp C, Verrier S-B, Zhouravleva G, Philippe H, Philippe M, Gress TM, Jean-Jean O (1998) *FEBS Lett* 440:387–392.
- Ko LJ, Engel JD (1993) *Mol Cell Biol* 13:4011–4022.
- The International HapMap Consortium, Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) *Nature* 437:1299–1320.
- Bourne HR, Sanders DA, McCormick F (1990) *Nature* 348:125–132.
- Inge-Vechtormov S, Zhouravleva G, Philippe M (2003) *Biol Cell* 95:195–209.
- Tang DC, Zhu J, Liu W, Chin K, Sun J, Chen L, Hanover JA, Rodgers GP (2005) *Blood* 106:3256–3263.
- Thorpe SJ, Thein SL, Sampietro M, Craig JE, Mahon B, Huehns ER (1994) *Br J Haematol* 87:125–132.
- Craig JE, Sheerin SM, Barnetson R, Thein SL (1993) *Br J Haematol* 84:106–110.
- Fibach E, Manor D, Oppenheim A, Rachmilewitz EA (1989) *Blood* 73:100–103.
- Ye S, Dhillon S, Ke X, Collins AR, Day IN (2001) *Nucleic Acids Res* 29:E88–8.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) *Nat Genet* 30:97–101.
- Barrett JC, Fry B, Maller J, Daly MJ (2005) *Bioinformatics* 21:263–265.
- Abecasis GR, Wigginton JE (2005) *Am J Hum Genet* 77:754–767.
- Abecasis GR, Cardon LR, Cookson WO (2000) *Am J Hum Genet* 66:279–292.