

Control of fetal hemoglobin: new insights emerging from genomics and clinical implications

Swee Lay Thein^{1,2,*}, Stephan Menzel¹, Mark Lathrop³ and Chad Garner⁴

¹Division of Gene and Cell Based Therapy, King's College London School of Medicine, Molecular Haematology, Denmark Hill Campus, London SE5 9NU, UK, ²Department of Haematological Medicine, King's College Hospital, Denmark Hill, London SE5 9RS, UK, ³Centre National de Génotypage, Institut Génomique, Commissariat à l'Energie Atomique, 91006 Evry, France and ⁴Department of Epidemiology, University of California, Irvine, 224 Irvine Hall, Irvine, CA 92697-7550, USA

Received August 18, 2009; Revised and Accepted August 18, 2009

Increased levels of fetal hemoglobin (HbF, $\alpha_2\gamma_2$) are of no consequence in healthy adults, but confer major clinical benefits in patients with sickle cell anemia (SCA) and β thalassemia, diseases that represent major public health problems. Inter-individual HbF variation is largely genetically controlled, with one extreme caused by mutations involving the β globin gene (*HBB*) complex, historically referred to as pancellular hereditary persistence of fetal hemoglobin (HPFH). These Mendelian forms of HPFH are rare and do not explain the common form of heterocellular HPFH which represents the upper tail of normal HbF variation, and is clearly inherited as a quantitative genetic trait. Genetic studies have identified three major quantitative trait loci (QTLs) (*Xmn1-HBG2*, *HBS1L-MYB* intergenic region on chromosome 6q23, and *BCL11A* on chromosome 2p16) that account for 20–50% of the common variation in HbF levels in patients with SCA and β thalassemia, and in healthy adults. Two of the major QTLs include oncogenes, emphasizing the importance of cell proliferation and differentiation as an important contribution to the HbF phenotype. The review traces the story of HbF quantitative genetics that uncannily mirrors the changing focus in genetic methodology, from candidate genes through positional cloning, to genome-wide association, that have expedited the dissection of the genetic architecture underlying HbF variability. These genetic results have already provided remarkable insights into molecular mechanisms that underlie the hemoglobin 'switch'.

INTRODUCTION

In humans, a shift from γ - to β -globin gene expression around birth, underlies the switch from fetal to adult hemoglobin (Hb) production such that by 6 months of age, the major Hb is HbA ($\alpha_2\beta_2$) (1). Residual amounts of HbF, however, continue to be synthesized throughout adult life with the majority of adults having <1% HbF. Inappropriately high γ globin expression in adult life is associated with deletions in the *HBB* cluster or with point mutations upstream of the *HBG* genes (2). These hereditary persistence of fetal hemoglobin (HPFH) mutations are characterized by significant elevations of HbF ranging from 10–40% in heterozygotes with the HbF homogeneously distributed among the erythrocytes, prompting the descriptive term, pancellular HPFH. Apart from the isolated increase in HbF levels, carriers for HPFH mutations are

otherwise normal. Although the increases in HbF mirror the molecular diversity of HPFH mutations, within each molecular class a range of HbF levels that do not show clear Mendelian inheritance, has been noted.

Variable increases in HbF levels have also been noted in individuals with sickle cell anemia (SCA) and β thalassemia (3–5), which are caused by mutations affecting the *HBB* gene and inherited as Mendelian recessives. Individuals with SCA (Hb-SS) have HbF levels ranging from 1–30% (6). An indication that additional variation at the β globin cluster is responsible for some of this variability came from the discovery of the 'sickle β haplotypes', and that the β^S gene on certain β^S haplotypes are associated with higher HbF levels and a milder disease (7). HbF response is also variable in β thalassemia (8). Although some of this variability

*To whom correspondence should be addressed at: Molecular Haematology, James Black Centre, 125 Coldharbour Lane, London SE5 9NU, UK. Tel: +44 2078485443; Fax: +44 2078485444; Email: sl.thein@kcl.ac.uk

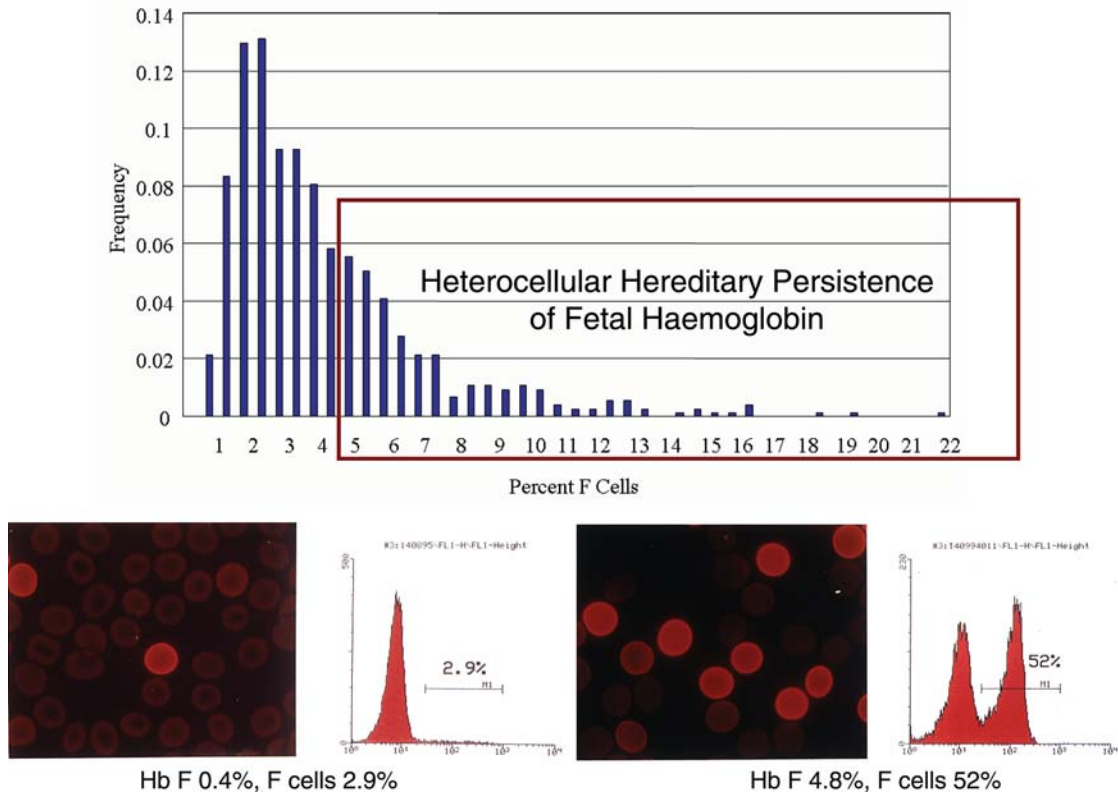


Figure 1. Population distribution of the F cell trait. The main panel shows the positively-skewed, continuous distribution of % F-cell measurements. Approximately 10% of the population at the upper end of the distribution has >4.5% F cells (FC) (corresponding to ~0.6% HbF), and are designated as having heterocellular hereditary persistence of fetal hemoglobin (hHPFH). The panels below indicate F cell values at the two extremes of the distribution. FC were detected using anti- γ globin antibody on fixed erythrocytes smears or by fluorescence-activated cell sorting (FACS) after labeling of the intracellular HbF (24).

can be explained by the specific β thalassemia mutation itself, and the β chromosomal background (9), a substantial proportion of the HbF increase is clearly unlinked to the *HBB* cluster (10–13). In some cases, the levels of HbF increase is sufficient to compensate for the complete lack of HbA, resulting in mild disease, i.e. β^0 thalassemia intermedia and no transfusion dependence (14).

It is now clear that the variable HbF increases in disease occurs on an underlying background of variable HbF persistence that is inherited as a quantitative genetic trait. We describe the discovery of the loci influencing the common HbF variation, their contribution to an emerging understanding of Hb switching, and the genetic architecture underlying HbF control in healthy adults, and in patients with β thalassemia and SCA. Finally, we speculate on the experimental approaches that will be required to elucidate the maximum genetic variance of HbF.

HbF AND F CELLS AS QUANTITATIVE TRAITS

The residual amounts of HbF in adults are distributed unevenly among the red blood cells; those that contain measurable amounts, are termed F cells (FC) (15). The first evidence that HbF persistence shows variability in a general population came in the early 1960s from screening of Swiss army recruits. Thirty-one out of 3000 recruits had HbF

above 0.7% (16,17), outliers for a normal population, and were labeled as having ‘Swiss Type HPFH’. Although heritability was shown, Swiss HPFH was otherwise quite different from classical HPFH: the inheritance did not follow Mendelian patterns, the HbF elevation was modest, and the HbF was unevenly distributed among the erythrocytes, hence the alternative term *heterocellular* HPFH. Other population surveys confirmed that the levels of FC and HbF vary considerably (by more than 20-fold); the distribution is continuous and positively skewed (18–21). Heterocellular HPFH represents ~10% of the population with HbF levels between 0.8 and 5% within the upper tail of this continuous distribution [Fig. 1; (22)].

HbF and FC are similarly distributed and closely correlated traits in normal individuals ($r > 0.9$) (19,21). Twin studies show that genetic factors account for 89% of the variability in F cell levels (23); the remaining 11% of FC variance is accounted for by age and sex (2%), and unknown environmental factors. HbF is usually measured by high-performance liquid chromatography which is highly accurate and reliable. However, the assay is imprecise at values <0.4% of total hemoglobin, which includes a large proportion of the non-anemic population. Thus, in the normal population, FC which can be quantified with high sensitivity by immunofluorescence using an anti- γ globin antibody, provide a much better measure of the trait (24).

IDENTIFICATION OF THREE MAJOR QTLs FOR Hb F AND FC

Association with the *Xmn1-HBG2* polymorphism at the *HBB* locus

In 1985, a polymorphism (C/T at position -158 of *HBG2*, later termed *Xmn1-HBG2* or *rs7482144*) (Fig. 2A) was identified from re-sequencing of the *HBB* genes (25), and shown to promote the expression of *HBG2*, and to contribute to HbF variability. Subsequent independent studies confirmed the association between the *Xmn1-HBG2* T allele and increased HbF and FC, as well as milder disease among individuals with SCA and β thalassemia from different population groups (9,26). The T allele was also shown to be associated with the Swiss-type HPFH (27).

Statistical analyses have identified no evidence for dominance at the locus, suggesting an additive effect of the *Xmn1-HBG2* polymorphism T allele. In a non-anemic North European population, the *Xmn1-HBG2* genotype was estimated to account for 13–32% of the total F-cell phenotypic variation (28). The QTL identified with the *Xmn1-HBG2* does not show Mendelian segregation and the distributions of the trait within each genotype class overlap considerably (20). Presence of the ‘T’ allele does not always dictate the presence of a high HbF phenotype, and high HbF has been associated with β haplotypes that do not include this allele (27,29,30). The quantitative trait is characterized by genetic heterogeneity such that to produce a full high-HbF phenotype, *Xmn1-HBG2* must exist on a genetic background requiring the presence of additional factors. The *Xmn1-HBG2* site achieves its importance through its large impact on the trait variance and its high frequency ($\sim 30\%$) in most population groups, including Europeans, Africans and Asian Indians.

Positional cloning and the discovery of the *HMIP* locus on chromosome 6q23.3

Early studies suggested that high HbF determinants segregated independently of the *HBB* in some families with β thalassemia and SCA; these families were often discovered through the probands who had unexpectedly mild disease (11,31,32). In one such extended family of Asian-Indian origin, complex segregation analysis showed strong evidence for a major HbF gene that was inherited independently of the *HBB* cluster (11,33). Using a regressive model that included the major QTL genotype and the effects of age, β thalassemia and the *Xmn1-HBG2* site, a genome-wide linkage analysis of the Asian-Indian kindred identified a major locus on chromosome 6q23–q24 (34). Follow-up haplotype mapping narrowed the candidate interval to 1.5 Mb (35) but re-sequencing of the five known genes in the region (*ALDH8A1*, *HBSIL*, *MYB*, *AH11* and *PDE7B*) revealed no mutations that could be implicated based on functional significance and segregation within the family (36).

High resolution association mapping was then carried out on a sample of North European ancestry (37). A set of common SNPs spanning a nearly contiguous segment of ~ 79 kb within *HBSIL*, the intergenic region 5' to *HBSIL* and the *MYB* oncogene, showed very strong association with

F cell levels (p-value $\sim 10^{-75}$ at the most significantly associated SNP) (37). The SNPs were distributed in three linkage disequilibrium blocks referred to as *HBSIL-MYB* intergenic polymorphism (*HMIP*) blocks 1, 2 and 3. Common alleles within the three haplotype blocks completely account for the FC variance attributed to the 6q23 QTL in the European sample, with block 2 (24 kb) (Fig. 2B) showing the strongest effect and accounting for the majority of the variance. In Northern Europeans, the 6q locus accounts for $\sim 19\%$ of the population trait variance.

Genome-wide association studies and *BCL11A* on chromosome 2

By early 2006, developments in genetic tools and genotyping platforms expedited by the International Human HapMap Project, led to a dramatic extension of the scope of genetic association studies, resulting in the genome-wide association study (GWAS) replacing the genome-wide linkage study as the most popular agnostic approach to whole-genome analysis (38–40). Two GWAS of HbF/FC have been reported. The first GWAS utilized a selected genotyping study design, targeting 179 individuals with contrasting extreme FC values (FC values above 95th or below fifth percentile points), chosen from a phenotyped cohort of 5184 individuals (41). Not only were the γ globin gene region (with the strongest signal at the *Xmn1-HBG2* site) and the chromosome 6 locus identified, but the study also found a new F-cell locus in intron 2 of the oncogene *BCL11A* on chromosome 2p16 (Fig. 2C). All SNPs implicated at this locus are common polymorphisms (with $>10\%$ minor allele frequency); the *BCL11A* locus accounts for 15.1%; 6q, 19.4% and γ globin region, 10.2% of the F-cell variability in Northern Europeans (41).

Statistical genetic theory predicts that association studies using the selected genotyping approach will be more powerful and cost effective than a design based on unselected individuals (42). A criticism of the study design has been that it will be powerful for identifying genotypes that underlie the rare and extreme trait values, which are different from the genotypes responsible for normal variation in the trait. This criticism was put to rest by the second GWAS of 4000 individuals from Sardinia with unselected HbF, which replicated significant association at the same three loci (43). The fact that both approaches, using FC or HbF as a quantitative trait, have identified the same set of three major loci is an additional argument that HbF and F cell are closely related traits.

MECHANISM OF HbF VARIATION

Influence of the *Xmn1-HBG2* variant (*rs7482144*) could be explained by a direct effect on *HBG2* gene expression (supported by relative increase of G γ globin in carriers) but, unlike the non-deletional variants associated with the Mendelian forms of HPFH, *in-vitro* functional studies have been inconclusive.

Of the three major QTLs, SNPs within the 14 kb intron 2 of *BCL11A* correlate most strongly with HbF expression (41,43–45). The *BCL11A* genotype that is associated with high HbF is associated with reduced *BCL11A* expression (46). *BCL11A* is

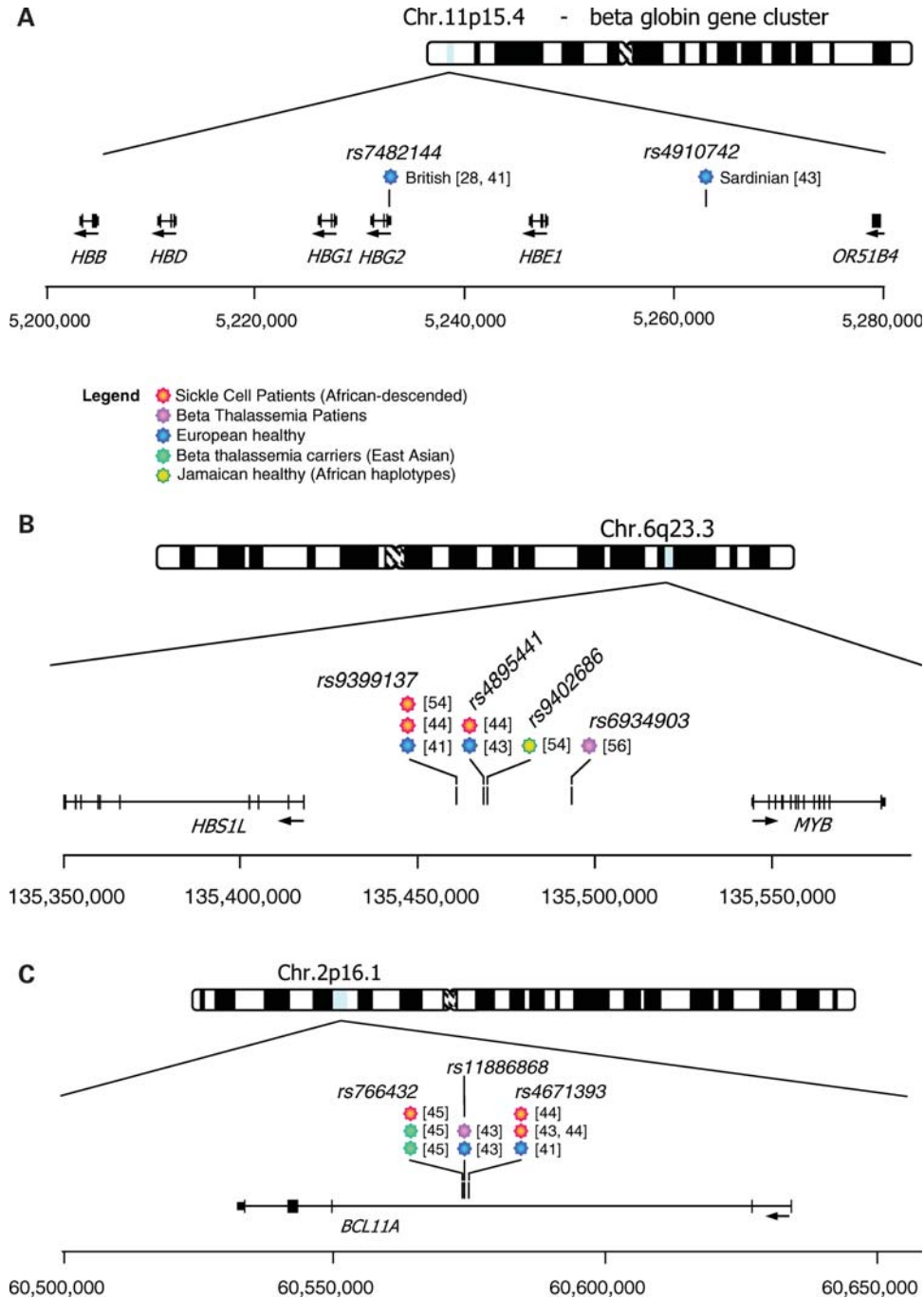


Figure 2. Large-scale association mapping of the three main HbF/FC QTLs in patients and population samples. Shown are the SNPs representing the strongest association signal in each study as referenced. For the chromosome 2 (C) and 6 (B) QTLs, the signals cluster well, and the implicated SNP is highly dependent on the study design (e.g. GWAS or saturation genotyping). In the *HBS1L*-*MYB* region on chromosome 6, tight linkage disequilibrium complicates high-resolution mapping. For the *HBB* cluster (A), the Sardinian population seems to present a genetic scenario different from Northern Europeans such as the British twins shown here. Evidence from smaller-scale studies (data not shown) points to an effect of *rs7482144* (the *Xmn1*-*HBG2* site) in sickle cell patients as well. Genome distances and feature sizes are from the UCSC genome browser (NCBI 36.1), the ideograms, with permission, from Michelle M Le Beau, Chicago.

highly expressed in erythroid progenitors, the shorter isoforms appeared to be restricted to primitive erythroblasts, and the full-length isoforms to adult-stage erythroblasts (46). More recent studies using human β globin locus YAC transgene and *bcl11a*^{-/-} mouse, showed that in the absence of BCL11A, developmental silencing of the human γ globin

genes is markedly impaired in the definitive erythroid lineage. Relaxation of *HBG* silencing in *bcl11a*^{+/-} mouse suggests that not only is *BCL11A* a developmental-stage specific repressor but the effect is quantitative (47). The exact molecular mechanisms by which *BCL11A* represses *HBB* expression in humans are still not clear, but studies in

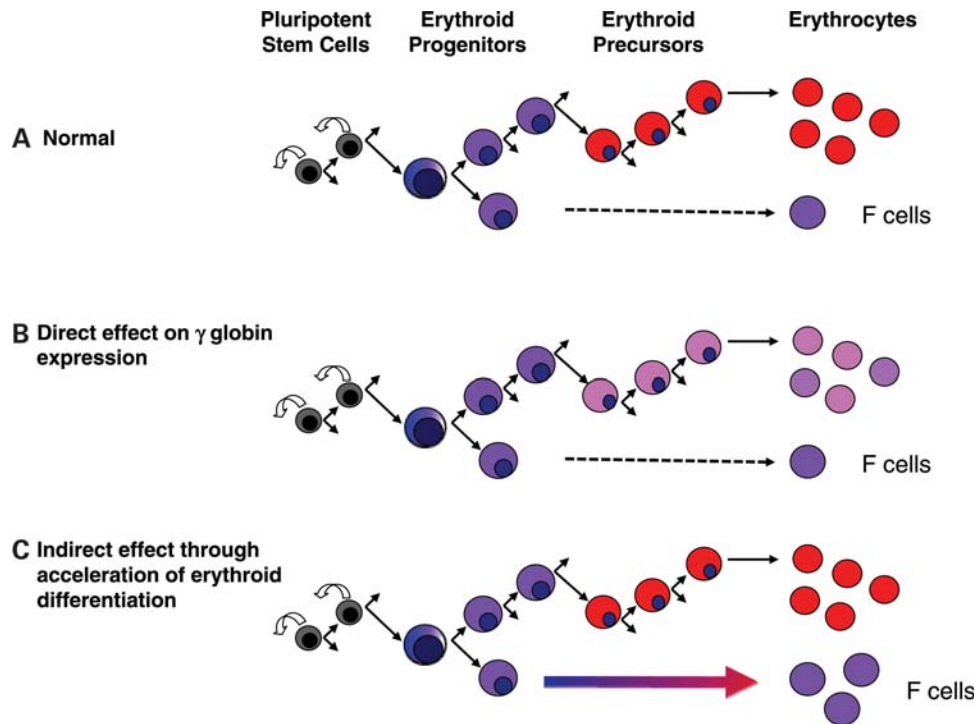


Figure 3. Proposed mechanisms of fetal hemoglobin modulation (adapted from Stamatoyannopoulos, G. *Experimental Hematology* 2005; **33**: 259–271). (A) Normal fetal and adult hemoglobin are produced by cells from the same lineage, i.e. hemoglobin switching occurs through a change in transcription program within the same cell. A change in transcription program occurs with erythropoietic differentiation, from one that predominantly supports γ globin expression to one that supports only β globin expression. FC represents the progeny of erythroid progenitor cells that terminate early before the change in transcription program occurs. Two plausible mechanisms of HbF modulation have emerged from biological studies of the QTLs. (B) Direct effect on γ globin gene expression (*BCL11A* is proposed as a developmental stage-specific repressor); and (C) Indirect via perturbation of kinetics of erythropoiesis—*HMIP 2* contains a distal regulatory locus that controls *MYB* expression, whose levels affect differentiation of erythropoiesis. An accelerated erythropoiesis (mimicking stress erythropoiesis) leads to the release of more progenitor cells that are still synthesizing significant amounts of HbF.

K562 cells suggest *BCL11A* binds to a core motif (GGCCGG at positions -56 to -51) in the *HBG* promoter to form a repressor complex (48).

The variants most strongly associated in the 6q QTL reside in a 24 kb non-protein coding region between *HBS1L* and *MYB* oncogene (*HMIP* block 2) (37). Recent studies show that *HMIP 2* contains a distal regulatory locus evidenced by the presence of several prominent erythroid-specific GATA-1 signals that coincided with DNaseI hypersensitive sites, and the presence of intergenic transcripts in erythroid precursor cells (49). It is suggested that the *HMIP 2* regulatory elements distally control *MYB* expression, that in turn influences erythroid differentiation, and indirectly, the control of HbF levels. *MYB* is a quantitative trait gene (50); erythroid precursor cells from individuals with higher HbF and higher F cell levels have lower *MYB* expression that was also associated with lower erythrocyte count but higher erythrocyte volume, and higher platelet counts. Further, genotype variability at *HMIP 2* has a pleiotropic impact on several types of peripheral blood cells: erythrocytes counts and volume, hemoglobin concentration, platelet and monocyte counts in healthy individuals of European ancestry (51).

Thus, the biological effect of the QTLs on HbF expression includes two plausible mechanisms: (1) direct effect on *HBG* expression (activation or repression of *HBG* transcription) thereby increasing or decreasing the amount of HbF per cell; and (2) alteration of the kinetics of erythroid maturation and

differentiation, mimicking a situation encountered in stress erythropoiesis that results in accelerated erythropoiesis with the release of more erythroid progenitors that synthesize predominantly HbF i.e. FC, leading to an increase in circulating HbF [Fig. 3; (52,53)].

IMPACT OF THE HbF QTLs

Influence of the *HBB* locus and the *Xmn1-HBG2* site on HbF levels in SCA and β thalassemia has been validated by many studies in several populations.

In the Asian Indian family in which the chromosome 6q23.3 locus was first identified, the QTL affects individuals with and without β thalassemia (33). Among β thalassemia heterozygotes, those who were homozygous for the 6q23.3 high F cell QTL allele had trait values ranging from 10 to 24% HbF compared with a range of 0.3–3.6% in the individuals that were homozygous for the alternative allele. In family members without the β thalassemia mutation, high F cell 6q23.3 QTL homozygotes had 1.1–3.0% HbF compared with 0.1–1.0% HbF in homozygotes for the alternative allele. Systematic studies of other patient groups have subsequently shown that the 6q QTL is also important in healthy individuals and in African American (44) and British sickle cell patients with predominantly African admixed ancestry (54) and sickle cell patients from Brazil (44).

The chromosome 6q23.3 QTL contributes 3–7% of the trait variance in these populations (44). The comparatively small contribution of the QTL to the trait variance in sickle patients compared with northern Europeans could be attributed to lower allele frequencies in African populations (55). Strong association with the locus was shown in Chinese β thalassemia heterozygotes, despite the confounding influence of the variable underlying ineffective erythropoiesis due to the wide spectrum of thalassemia mutations (56).

The *BCL11A* QTL has shown the strongest effect on HbF/F cell levels to date. In patients with SCA, the QTL accounts for 7–12% (44,45) of the trait variance. The locus has also been shown to influence HbF or F cell levels in individuals of Chinese and Thai descent with β thalassemia or HbE (45). In the Sardinian population the C alleles of the *rs11886868* SNP at the *BCL11A* locus is strongly associated with high HbF levels and it is significantly more frequent in patients with milder β thalassemia disease (thalassemia intermedia) compared with the transfusion-dependent thalassemia major patients (43). Both groups have identical β and α genotypes, the Hb pattern comprised only HbF and trace amount of HbA₂ with no HbA, suggesting that variation at the *BCL11A* locus can alleviate disease severity through raising HbF levels.

In SCA, the three known Hb/F cell loci contribute >20% to the HbF variance with a corresponding reduction in the frequency of acute pain associated with sickling (44).

High HbF expression clearly has ameliorating effects on sickle cell disease and β thalassemia, however, it is unknown if the effects translate to a selective advantage for carriers of the high HbF alleles in regions with a high incidence of the diseases. Genetic studies have shown that individuals with hemoglobinopathies concurrent with high expression of HbF can maintain normal fitness levels that would otherwise be severely limited by the debilitating consequences of their disease. The hemoglobinopathies are not present in the ancestral European population yet the alleles at the known major genes remain at high frequency and variation in the HbF expression persists. Strong evidence for natural selection has not been reported around the three major HbF/F cell QTLs (57) and similar allele frequencies have been observed for the QTL in different ethnic groups, suggesting that a selective sweep favoring genes for HbF persistence has not occurred in the relatively recent past. It is possible that the pleiotropic effect of the major genes for HbF expression (for example, *HMIP* effect on other hematological variables) (51) ensure the persistence of the alleles in the absence of strong selection from the hemoglobinopathies. After the transition to expression of adult hemoglobin, there is no known biological role for HbF. Genetic studies of HbF/FC in varied populations will allow for more powerful analysis of the evolutionary history of the alleles at the known QTLs and a better understanding of their biological importance.

ACCOUNTING FOR THE REMAINING GENETIC VARIANCE IN THE HbF/F CELL TRAIT

Identification of the exact same set of three major loci in two independent GWASs of Europeans suggests that we have identified the principal QTLs with frequent alleles affecting

HbF production in the general European population. The genetic variance that remains unaccounted for may be due to many common and/or rare alleles with small effects on the trait, i.e. a polygenic component (58). Indeed, what differentiates the genetics of HbF expression from other quantitative traits that have been equally studied (such as height, blood pressure and BMI) are the existence of common major genes and with those genes now identified, Hb F expression is likely to be as complex and difficult to define genetically as other quantitative traits (39,59–61). Identifying the common alleles will require large population samples for replication, and possibly denser maps (62). Identification of rare alleles will also require next-generation DNA sequencing technologies and statistical analysis methods that enable genome-wide sequence analysis in large samples (63). Identification of the chromosome 6q23.3 QTL in the Asian Indian kindred (34) illustrates the power of large genome-wide linkage studies that could uncover rare alleles with moderately high penetrance that would not otherwise be identified in association studies due to their low frequencies. Functional characterization of the known QTLs should lead to discoveries of biological pathways underlying Hb F expression and new candidates for genetic analysis. With next-generation DNA sequencing methods coming into widespread use, deep sequencing of the QTL regions in targeted samples could ease the complexity of defining the causal variants (quantitative trait nucleotides, QTNs). No statistical interactions between the known QTLs have been found, however, it is possible that genetic interaction does exist between the known QTLs with a biological mechanism that is not detectable by statistical interaction models. Statistical evidence for interaction between the *Xmn1-HGB2* locus and a locus on chromosome 8q has been reported but validation and replication of the finding has been elusive (64). If genetic interactions between currently unknown loci with weak marginal effects are significantly affecting HbF expression then traditional linkage and association studies are unlikely to be productive without greater knowledge of the underlying biology. The possibility remains that major QTLs for HbF trait could exist in non-European populations that have not been the subject to genome-wide analysis. Given that a primary motivation for studying the trait is its clinical impact on β thalassemia and sick cell disease, genome-wide association and linkage studies should be undertaken in populations that are stricken with a high prevalence of these disorders.

CONCLUSION

The story of HbF genetics serves as a paradigm for understanding quantitative traits in humans, their genetics and biology and involvement in disease. Inappropriately high HbF ($\alpha^2\gamma^2$) expression caused by rare mutations affecting the *HBB* cluster indicate that the γ globin gene can be reactivated but does not explain the inter-individual variation in HbF, a major contributing factor in the severity of sickle cell disease and β thalassemia.

Common variations in HbF has all the characteristics of a quantitative genetic trait, associated with multiple QTLs,

some linked and others unlinked to the *HBB* cluster. The first known QTL, *Xmn1-HBG2*, has long been implicated through family and population studies and subsequently validated by genetic studies. The second QTL, the *HBS1L-MYB* intergenic region (*HMIP*) on chromosome 6q, was discovered through genetic linkage studies in an extensive kindred, and subsequently validated by genetic association studies. These two QTLs have now been joined by *BCL11A* discovered through GWAS. GWAS not only 'rediscovered' the known QTLs, but highlighted *BCL11A*, previously known as an oncogene involved in leukemogenesis but whose relevance to HbF trait was previously unsuspected. Biological studies of these recently identified QTLs provide data supporting long-held views on mechanisms of hemoglobin switching—changes in *trans*-acting factor environment and perturbation of the kinetics of erythropoiesis. These mechanisms, however, may not be mutually exclusive. The challenge now is to delineate the causal variants or QTNs and the physiological pathways to facilitate identification of novel targets for therapeutic reactivation of HbF. True understanding of how the QTLs and QTNs function require integration of these genetic mapping results with expression patterns in the erythroid cell and knowledge of the causal network of interacting genes. Unlike other complex traits and diseases, only three QTLs explain a relatively large proportion of the underlying genetic contribution to HbF variation, an immediate impact of these findings may be the application of the SNPs in the three QTLs to improve the prediction of one's ability to produce HbF in response to disease which will have implications for genetic counseling and prenatal diagnosis of these hemoglobin disorders.

ACKNOWLEDGEMENTS

We thank Claire Steward for help in preparation of the manuscript.

Conflict of Interest statement. None declared.

FUNDING

S.L.T. received funding from NIH-NHLBI (R01-HL69259-03), and the Medical Research Council, UK (MRC UK G0000111) for this work.

REFERENCES

- Stamatoyannopoulos, G. (2001) Molecular and cellular basis of hemeoglobin switching. In Steinberg, M.H., Forget, B.G., Higgs, D.R. and Nagel, R.L. (eds), *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, UK, pp. 131–145.
- Wood, W.G. (2001) Hereditary persistence of fetal hemoglobin and delta beta thalassemia. In Steinberg, M.H., Forget, B.G., Higgs, D.R. and Nagel, R.L. (eds), *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, UK, pp. 356–388.
- Weatherall, D.J. (2001) Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.*, **2**, 245–255.
- Stuart, M.J. and Nagel, R.L. (2004) Sickle-cell disease. *Lancet*, **364**, 1343–1360.
- Thein, S.L. (2008) Genetic modifiers of the beta-haemoglobinopathies. *Br. J. Haematol.*, **141**, 357–366.
- Nagel, R.L. and Platt, O.S. (2001) General pathophysiology of sickle cell anemia. In Steinberg, M.H., Forget, B.G., Higgs, D.R. and Nagel, R.L. (eds), *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, UK, pp. 494–526.
- Labie, D., Pagnier, J., Lapoumeroulie, C., Rouabhi, F., Dunda-Belkhdja, O., Chardin, P., Beldjord, C., Wajzman, H., Fabry, M.E. and Nagel, R.L. (1985) Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. *Proc. Natl. Acad. Sci. USA*, **82**, 2111–2114.
- Ho, P.J., Hall, G.W., Luo, L.Y., Weatherall, D.J. and Thein, S.L. (1998) Beta thalassemia intermedia: is it possible to consistently predict phenotype from genotype? *Br. J. Haematol.*, **100**, 70–78.
- Thein, S.L., Wainscoat, J.S., Sampietro, M., Old, J.M., Cappellini, D., Fiorelli, G., Modell, B. and Weatherall, D.J. (1987) Association of thalassaemia intermedia with a beta-globin gene haplotype. *Br. J. Haematol.*, **65**, 367–373.
- Gianni, A.M., Bregni, M., Cappellini, M.D., Fiorelli, G., Taramelli, R., Giglioli, B., Comi, P. and Ottolenghi, S. (1983) A gene controlling fetal hemoglobin expression in adults is not linked to the non-alpha globin cluster. *EMBO J.*, **2**, 921–925.
- Thein, S.L. and Weatherall, D.J. (1989) A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the beta-globin gene complex. In Stamatoyannopoulos, G. and Nienhuis, A.W. (eds), *Hemoglobin Switching, Part B: Cellular and Molecular Mechanisms*. Alan R Liss, Inc, New York, pp. 97–111.
- Martinez, G., Novelletto, A., Di Rienzo, A., Felicetti, L. and Colombo, B. (1989) A case of hereditary persistence of fetal hemoglobin caused by a gene not linked to the β -globin cluster. *Hum. Genet.*, **82**, 335–337.
- Giampaolo, A., Mavilio, F., Sposi, N.M., Care, A., Massa, A., Cianetti, L., Petri, M., Russo, R., Capellini, M.D. and Marinucci, M. (1984) Heterocellular HPFH: molecular mechanisms of abnormal γ gene expression in association with β thalassemia and linkage relationship with the β globin gene cluster. *Hum. Genet.*, **66**, 151–156.
- Galanello, R., Dessi, E., Melis, M.A., Addis, M., Sanna, M.A., Rosatelli, C., Argioli, F., Giagu, N., Turco, M.P., Cacace, E. *et al.* (1989) Molecular analysis of beta zero-thalassemia intermedia in Sardinia. *Blood*, **74**, 823–827.
- Boyer, S.H., Belding, T.K., Margolet, L. and Noyes, A.N. (1975) Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. *Science*, **188**, 361–363.
- Marti, H.R. (1963) *Normale und anormale menschliche Haemoglobine*. Springer Verlag, Berlin.
- Marti, H.R. and Butler, R. (1961) Hämoglobin F- und Hämoglobin A₂-Vermehrung bei der Schweizer Bevölkerung. *Acta Haematol. (Basel)*, **26**, 65–74.
- Economou, E.P., Antonarakis, S.E., Kazazian, H.H.J., Serjeant, G.R. and Dover, G.J. (1991) Variation in hemoglobin F production among normal and sickle cell adults is not related to nucleotide substitutions in the Gamma promoter regions. *Blood*, **77**, 174–177.
- Miyoshi, K., Kaneto, Y., Kawai, H., Ohchi, H., Niki, S., Hasegawa, K., Shirakami, A. and Yamano, T. (1988) X-linked dominant control of F-cells in normal adult life: Characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome. *Blood*, **72**, 1854–1860.
- Sampietro, M., Thein, S.L., Contreras, M. and Pazmany, L. (1992) Variation of HbF and F-cell number with the G-gamma *Xmn* I (C-T) polymorphism in normal individuals. *Blood*, **79**, 832–833.
- Zago, M.A., Wood, W.G., Clegg, J.B., Weatherall, D.J., O'Sullivan, M. and Gunson, H. (1979) Genetic control of F cells in human adults. *Blood*, **53**, 977–986.
- Thein, S.L. and Craig, J.E. (1998) Genetics of Hb F/F cell variance in adults and heterocellular hereditary persistence of fetal hemoglobin. *Hemoglobin*, **22**, 401–414.
- Garner, C., Tatu, T., Reittie, J.E., Littlewood, T., Darley, J., Cervino, S., Farrall, M., Kelly, P., Spector, T.D. and Thein, S.L. (2000) Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*, **95**, 342–346.
- Thorpe, S.J., Thein, S.L., Sampietro, M., Craig, J.E., Mahon, B. and Huehns, E.R. (1994) Immunochemical estimation of haemoglobin types in red blood cells by FACS analysis. *Br. J. Haematol.*, **87**, 125–132.

25. Gilman, J.G. and Huisman, T.H.J. (1985) DNA sequence variation associated with elevated fetal γ globin production. *Blood*, **66**, 783–787.
26. Labie, D., Dunda-Belkhdja, O., Rouabhi, F., Pagnier, J., Ragusa, A. and Nagel, R.L. (1985) The -158 site 5' to the γ gene and γ expression. *Blood*, **66**, 1463–1465.
27. Efremov, G.D., Gjorgovski, I., Stojanovski, N., Diaz-Chico, J.C., Harano, T., Kutlar, F. and Huisman, T.H.J. (1987) One haplotype is associated with the Swiss type of hereditary persistence of fetal hemoglobin in the Yugoslavian population. *Hum. Genet.*, **77**, 132–136.
28. Garner, C., Tatu, T., Game, L., Cardon, L.R., Spector, T.D., Farrall, M. and Thein, S.L. (2000) A candidate gene study of F cell levels in sibling pairs using a joint linkage and association analysis. *GeneScreen*, **1**, 9–14.
29. Kulozik, A.E., Kar, B.C., Satapathy, R.K., Serjeant, B.E., Serjeant, G.R. and Weatherall, D.J. (1987) Fetal hemoglobin levels and β^S globin haplotypes in an Indian population with sickle cell disease. *Blood*, **69**, 1742–1746.
30. Pissard, S. and Beuzard, Y. (1994) A potential regulatory region for the expression of fetal hemoglobin in sickle cell disease. *Blood*, **84**, 331–338.
31. Cappellini, M.D., Fiorelli, G. and Bernini, L.F. (1981) Interaction between homozygous β^0 thalassaemia and the Swiss type of hereditary persistence of fetal haemoglobin. *Br. J. Haematol.*, **48**, 561–572.
32. Old, J.M., Ayyub, H., Wood, W.G., Clegg, J.B. and Weatherall, D.J. (1982) Linkage analysis of nondeletion hereditary persistence of fetal hemoglobin. *Science*, **215**, 981–982.
33. Thein, S.L., Sampietro, M., Rohde, K., Rochette, J., Weatherall, D.J., Lathrop, G.M. and Demenais, F. (1994) Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. *Am. J. Hum. Genet.*, **54**, 214–228.
34. Craig, J.E., Rochette, J., Fisher, C.A., Weatherall, D.J., Marc, S., Lathrop, G.M., Demenais, F. and Thein, S.L. (1996) Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat. Genet.*, **12**, 58–64.
35. Garner, C., Mitchell, J., Hatzis, T., Reittie, J., Farrell, M. and Thein, S.L. (1998) Haplotype mapping of a major QTL for fetal hemoglobin production on chromosome 6q23. *Am. J. Hum. Genet.*, **62**, 1468–1474.
36. Close, J., Game, L., Clark, B.E., Bergounioux, J., Gerovasilli, A. and Thein, S.L. (2004) Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults. *BMC Genomics*, **5**, 33.
37. Thein, S.L., Menzel, S., Peng, X., Best, S., Jiang, J., Silver, N., Gerovasilli, A., Ping, C., Yamaguchi, M. *et al.* (2007) 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. USA*, **104**, 11346–11351.
38. Manolio, T.A., Brooks, L.D. and Collins, F.S. (2008) A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.*, **118**, 1590–1605.
39. Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease. *Science*, **322**, 881–888.
40. Hardy, J. and Singleton, A. (2009) Genomewide association studies and human disease. *N. Engl. J. Med.*, **360**, 1759–1768.
41. Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H. *et al.* (2007) A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.*, **39**, 1197–1199.
42. Schork, J.J., Nath, S.K., Fallin, D. and Chakravarti, A. (2000) Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. *Am. J. Hum. Genet.*, **67**, 1208–1218.
43. Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G. *et al.* (2008) 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**, 1620–1625.
44. Lettre, G., Sankaran, V.G., Bezerra, M.A., Araujo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N. *et al.* (2008) 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**, 11869–11874.
45. Sedgewick, A.E., Timofeev, N., Sebastiani, P., So, J.C., Ma, E.S., Chan, L.C., Fucharoen, G., Fucharoen, S., Barbosa, C.G., Vardarajan, B.N. *et al.* (2008) BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. *Blood Cells Mol. Dis.*, **41**, 255–258.
46. Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B. and Orkin, S.H. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science*, **322**, 1839–1842.
47. Sankaran, V.G., Xu, J., Ragoczy, T., Ippolito, G.C., Walkley, C.R., Maika, S.D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M.A. *et al.* (2009) Developmental and species-divergent globin switching are driven by BCL11A. *Nature online*, 1–5. doi:10.1038/nature08248.
48. Chen, Z., Luo, H.Y., Steinberg, M.H. and Chui, D.H. (2009) BCL11A represses HBG transcription in K562 cells. *Blood Cells Mol. Dis.*, **42**, 144–149.
49. Wahlberg, K., Jiang, J., Rooks, H., Jawaid, K., Matsuda, F., Yamaguchi, M., Lathrop, M., Thein, S.L. and Best, S. (2009) The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. *Blood*, **114**, 1254–1262.
50. Jiang, J., Best, S., Menzel, S., Silver, N., Lai, M.I., Surdulescu, G.L., Spector, T.D. and Thein, S.L. (2006) cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood*, **108**, 1077–1083.
51. Menzel, S., Jiang, J., Silver, N., Gallagher, J., Cunningham, J., Surdulescu, G., Lathrop, M., Farrall, M., Spector, T.D. and Thein, S.L. (2007) The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood*, **110**, 3624–3626.
52. Stamatoyannopoulos, G. (2005) Control of globin gene expression during development and erythroid differentiation. *Exp. Hematol.*, **33**, 259–271.
53. Bank, A. (2006) Regulation of human fetal hemoglobin: new players, new complexities. *Blood*, **107**, 435–443.
54. Creary, L.E., Ulug, P., Menzel, S., McKenzie, C.A., Hanchard, N.A., Taylor, V., Farrall, M., Forrester, T.E. and Thein, S.L. (2009) 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**, e4218.
55. Creary, L.E., McKenzie, C.A., Menzel, S., Hanchard, N.A., Taylor, V., Hambleton, I., Spector, T.D., Forrester, T.E. and Thein, S.L. (2009) Ethnic differences in F cell levels in Jamaica: a potential tool for identifying new genetic loci controlling fetal haemoglobin. *Br. J. Haematol.*, **144**, 954–960.
56. So, C.C., Song, Y.Q., Tsang, S.T., Tang, L.F., Chan, A.Y., Ma, E.S. and Chan, L.C. (2008) The HBS1L-MYB intergenic region on chromosome 6q23 is a quantitative trait locus controlling fetal hemoglobin level in carriers of beta-thalassaemia. *J. Med. Genet.*, **45**, 745–751.
57. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W. *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.*, **19**, 826–837.
58. Bodmer, W. and Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.*, **40**, 695–701.
59. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
60. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.*, **41**, 56–65.
61. Lettre, G. and Rioux, J.D. (2008) Autoimmune diseases: insights from genome-wide association studies. *Hum. Mol. Genet.*, **17**, R116–R121.
62. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. and Hirschhorn, J.N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356–369.
63. Goldstein, D.B. (2009) Common genetic variation and human traits. *N. Engl. J. Med.*, **360**, 1696–1698.
64. Garner, C., Silver, N., Best, S., Menzel, S., Martin, C., Spector, T.D. and Thein, S.L. (2004) Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin. *Blood*, **104**, 2184–2186.