Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *The Year in Human and Medical Genetics*

Modifier genes in Mendelian disorders: the example of hemoglobin disorders

Vijay G. Sankaran,[1,2] Guillaume Lettre,[3,4] Stuart H. Orkin,[1,2] and Joel N. Hirschhorn[5,6]

[1]Division of Hematology/Oncology, Children's Hospital Boston, Boston, Massachusetts. [2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. [3]Montreal Heart Institute, Montréal, Québec, Canada. [4]Département de Médecine, Université de Montréal, Succursale Centre-ville, Montréal, Québec, Canada. [5]Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital Boston, Boston, Massachusetts. [6]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts

Addresses for correspondence: Joel N. Hirschhorn, M.D., Ph.D., Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital Boston, Boston, MA 02115; and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142. joelh@broadinstitute.org
Stuart H. Orkin, M.D., Division of Hematology/Oncology, Children's Hospital Boston, MA 02115; and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115. stuart_orkin@dfci.harvard.edu

The disorders of hemoglobin, including sickle cell disease (SCD) and β-thalassemia, are the most common "Mendelian" genetic diseases in the world. Numerous studies have demonstrated the complexity in making genotype–phenotype correlations in both SCD and β-thalassemia. Indeed, patients with exactly the same set of pathogenic globin mutations can have dramatically variable clinical courses. We discuss natural history studies that have attempted to delineate the factors responsible for the variability among the numerous clinical complications noted in these diseases. We then discuss, in depth, two well characterized ameliorating factors in the β-hemoglobin disorders, concomitant α-thalassemia, and elevated levels of fetal hemoglobin (HbF). We use the study of HbF regulation to illustrate how important insights into the genetic modifiers in Mendelian diseases can be achieved through the study of such factors. We finally go on to discuss future avenues of research that may allow us to gain further insight into the poorly understood clinical heterogeneity of this fascinating set of common genetic diseases.

Keywords: sickle cell disease; β-thalassemia; fetal hemoglobin; association studies

Overview

The most common set of monogenic diseases in the world are the inherited hemoglobin disorders. These disorders are thought to have reached such a high prevalence due in large part to the carrier protection afforded against malaria.[1] Approximately 7% of the world's population is a carrier for at least one of these diseases.[2] It is estimated that each year over 500,000 babies with severe forms of these disorders are born.[3] Unfortunately, many children born with these diseases do not receive adequate treatment and therefore death occurs in the first few years of life. The burden of these diseases is undoubtedly going to increase in the coming years, as many developing countries are making an epidemiologic transition as a consequence of the reduction in infant mortality from nutritional and infectious causes.

There are two major categories of the inherited hemoglobin disorders: the structural hemoglobin variants and the thalassemia syndromes. Generally, carriers or heterozygotes for these variants are asymptomatic, whereas those who inherit two copies of these mutations are often affected by this spectrum of diseases. The structural hemoglobin variants result primarily from amino acid substitutions in one of the two polypeptide chains, α- or β-globin, that compose the hemoglobin (Hb) molecule. These mutations can be innocuous or alternatively can structurally destabilize or alter functional properties of the assembled Hb molecule. On a worldwide basis, only three structural variants are widespread. All three of these structural

doi: 10.1111/j.1749-6632.2010.05821.x

variants are due to missense mutations in the β-globin polypeptide and include HbS, HbC, and HbE.[2] The homozygous state of HbS results in sickle cell anemia, whereas the compound heterozygous state for HbS and HbC results in HbSC disease, which shares many features with sickle cell anemia, though it is generally milder.[4,5] HbS may also be found in the compound heterozygous state with β-thalassemia mutations causing HbS-β-thalassemia that also has extensive clinical overlap with sickle cell anemia. Collectively, these disorders are grouped together with the term *sickle cell disease* (SCD). The etiology of the complications in SCD is primarily attributable to the propensity of this sickle Hb to polymerize in the deoxygenated state, which results in the red blood cells becoming deformed into sickle shapes.[4] These misshapen cells then occlude small blood vessels, which leads to attacks of pain in bones from ischemia; sequestration of blood in the spleen, liver, or lungs; and thrombosis of cerebral vessels, causing strokes.[5] In addition to these acute consequences of sickle Hb polymerization, long-term changes can also result. These alterations include red cell membrane damage, chronic hemolysis, altered interactions between sickle red blood cells and vascular endothelial cells, and inflammatory responses from such interactions.[6]

The thalassemia syndromes are attributable to defective production of the globin polypeptides. This can either be due to impaired production of the α- or β-globin chains, which causes α- or β-thalassemia, respectively. In the case of the α-thalassemia syndromes, unpaired β-globin chains can assemble into an unstable Hb molecule known as HbH, which results in hemolytic anemia. There is a great deal of heterogeneity in this disease, which has been discussed in other review articles.[2,7] In the case of β-thalassemia, unbalanced globin chain synthesis from a reduced production of β-globin chains results in the presence of highly unstable α-globin polypeptides. These polypeptides precipitate in precursors of mature red blood cells, resulting in toxicity, death, and ineffective maturation into red blood cells.[8,9] This causes anemia because of ineffective production of red blood cells (known as ineffective erythropoiesis) and results in an expansion of the erythroid precursor populations within the bone marrow, spleen, as well as in other sites within the body. At the molecular level, it is important to recognize that the mutations causing β-thalassemia are extremely heterogeneous, with over 200 different mutations described in the literature.[2,9] Generally, a few mutations predominate in a particular region of the world. The missense mutation in HbE is structurally innocuous, but it does create a novel splice site resulting in defective production of this Hb.[10] When HbE is found in the compound heterozygous state with a β-thalassemia mutation, a clinically heterogeneous disease known as HbE-β-thalassemia can occur.[1,11] This disease ranges in severity from being essentially asymptomatic, with the presence of a moderate anemia, all the way to a transfusion-dependent form. HbE-β-thalassemia is considered to be part of the spectrum of the β-thalassemia syndromes.

Although many children with the various β-thalassemia syndromes cannot survive without regular blood transfusions, a major source of morbidity and mortality in patients who either are transfused regularly or can survive without transfusions is due to iron overload. This is attributable to both the iron burden resulting from regular transfusions, as well as from the increased physiologic tendency of these patients to absorb iron.[12,13] Indeed, many patients with transfusion-independent forms of thalassemia (thalassemia intermedia) still succumb to the consequences of increased iron overload from deregulated iron absorption.

Natural history

Both SCD and β-thalassemia were among the first single-gene disorders whose molecular etiology was unraveled by using the tools of molecular biology.[1] However, although the field of molecular medicine began to burgeon soon after the elucidation of these initial "molecular" diseases, many years passed before natural history studies began to reveal the actual extent of phenotypic variability in these diseases. As a result of this work, it became evident that these diseases were far from the seemingly simple Mendelian disorders that many workers in the field had come to believe; among patients with known causal mutations, the clinical course can be extremely variable.

The understanding of the natural history of SCD stems from a rather limited number of studies. The majority of children with SCD in natural history studies in the developing world have died early.[14,15] Older studies suggested that the majority of these

deaths were attributable to increased susceptibility to infectious diseases among these children,[15,16] likely secondary to the functional loss of the spleen in this disease. In the developed world, a series of important studies showed that this increased susceptibility was attributable to specific bacterial infections among children with SCD,[2,5,17] leading to arguably the most important intervention to date in patients with SCD, the prophylactic use of penicillin to prevent bacteremia.[18,19] However, it was not known whether the same causes of early death were present in SCD patients in the developing world, until the recent ground-breaking work from Williams *et al.*[20] showing that the spectrum of bacteremia in these children with SCD in Africa mimics what has been observed in other settings.[21]

As a result of the above early interventions, coupled with newborn screening in some countries, the life expectancy for patients with SCD has increased considerably.[2] However, these patients continue to face a number of complications of this disease. Cohort studies in countries including the United States and Jamaica examined the causes and risk factors for various SCD-related complications, including vaso-occlusive crises, acute chest syndrome, strokes, death, and others.[5,22–25] An important common theme of all of these studies is the extent of diversity among patients for these various complications. For example, although a subset of SCD patients may present with frequent vaso-occlusive crises, many patients rarely or almost never suffer from this complication.[23] It is also apparent that other populations with SCD may have a great diversity of clinical manifestations.[26,27] Moreover, most studies have been performed in the developed world and few studies exist to describe the clinical manifestations of SCD in the developing world, where the majority of patients with this disease live.[21]

A similar heterogeneity in the natural history of disease is present for the β-thalassemia syndromes.[2,28] In the majority of cases, patients with homozygous or compound heterozygous β-thalassemia mutations require regular transfusions to maintain Hb levels compatible with survival. However, it is apparent that there are occasional patients, even with mutations normally considered to create transfusion-dependent forms of thalassemia, who have a much milder clinical course and who may not require regular transfusions.[29] The precise reasons for this heterogeneity are not well understood in many instances,[30] although certain modifiers, such as concomitant α-thalassemia mutations or increased fetal hemoglobin (HbF) production (see below), play a role in this phenotypic variability.[31,32] When children with β-thalassemia are given adequate transfusions along with appropriate iron chelation, development can occur in a nearly normal fashion and survival into adulthood is possible.[2] Nonetheless, an important degree of heterogeneity is observed even among patients who receive regular treatment. For example, β-thalassemia patients with a similar degree of apparent iron burden (as assessed by either serum ferritin or liver iron measurements) may have extremely heterogeneous manifestations of iron overload.[33,34] For example, some patients with a similar degree of iron burden may develop endocrine complications such as diabetes or impaired growth, whereas others manifest primarily with cardiac or other complications of iron overload.

This situation is further complicated in the case of β-thalassemia syndromes that are characterized as being intermediate in severity. The most notable and common group of these disorders are the HbE-β-thalassemia syndromes. As discussed above, HbE-β-thalassemia can range in severity, from being asymptomatic to being a transfusion-dependent disease. Until recently, the full extent of this diversity was not well appreciated.[11,35] It is apparent that many patients who were once thought to require transfusions for survival can actually do extremely well in the absence of such treatment.[11] However, the study of this clinical heterogeneity is complicated by the important observation that even a single patient can have a dramatically variable clinical course over his or her lifetime.[1,11] Such observations stress the important role that nongenetic factors, such as environmental exposures, may have upon the clinical course of Hb diseases, and also highlight an important complexity that must be considered when attempting to develop scoring systems for severity in these diseases.[1,35]

Hb disease modifiers

As is apparent from our discussion on the natural history of these diseases, we currently have a rather limited grasp on the etiology of heterogeneity in these disorders, even among patients who acquire the exact same set of Mendelian mutations. However, there are a few cases in which clearly identified modifiers of clinical severity in

the β-hemoglobin disorders are known. These well-characterized modifiers primarily act directly to alter the known pathophysiology of SCD or β-thalassemia. In the case of SCD, factors that reduce HbS deoxygenation and/or polymerization appear to be important and primary modifiers of the disease. In the case of β-thalassemia, any factor that reduces the imbalance between α- and β-globin polypeptide chain production is an important modifier of these diseases.

α-Thalassemia

A major well characterized modifier of severity in both SCD and β-thalassemia is the concomitant presence of α-thalassemia. Higgs *et al.*[36] and Embury *et al.*[37] first described this important clinical effect in SCD, where a milder clinical course was observed with a greater reduction in α-globin production. In SCD, the reduction in α-globin polypeptides results in a reduced concentration of HbS within a red blood cell, which then reduces the likelihood that HbS will become deoxygenated and undergo polymerization.[38] As a result, there is a reduced severity of certain SCD-related complications, particularly those associated with hemolysis, although compound heterozygous patients can still be severely affected by many of the complications of SCD. In the context of β-thalassemia, a reduced concentration of α-globin chains reduces the globin chain imbalance and therefore allows for more effective erythropoiesis.[1,39,40] Although this interaction among α- and β-thalassemia has been appreciated for many years, the critical nature of globin chain balance has been highlighted by the finding that the presence of α-globin duplications can interact with heterozygous β-thalassemia to cause a syndrome resembling homozygous β-thalassemia.[41,42]

Fetal hemoglobin

In 1948, Janet Watson[43] hypothesized that infants with SCD were protected from the clinical manifestations of their disease secondary to the presence of a unique form of Hb found in the fetus and in early infancy. Similar hypotheses were put forth to explain the delayed onset of symptoms in patients with β-thalassemia.[29] Furthermore, observations in certain populations with Hb disorders, such as SCD patients from Saudi Arabia[27,44,45] and India,[26] indicated that higher steady-state levels of HbF tend to result in less severe disease. The conclusions drawn from these disparate observations were subsequently confirmed in larger epidemiological studies, which demonstrated the quantitative partial ameliorating effect of HbF levels in SCD[22–24,46] and the β-thalassemia syndromes.[1,8,9,11,28]

With the identification and characterization of different Hb isoforms, it became evident that a unique β-like globin subunit, γ-globin, is expressed in much of fetal life and declines soon after birth.[1] Together with two α-globin chains, two γ-globin polypeptides combine with a heme moiety to form HbF. γ-Globin polypeptides compensate for reduced production of β-globin chains in β-thalassemia, thus ameliorating the symptoms of this disease. In SCD, HbF inhibits polymerization of HbS and therefore serves as a powerful antisickling agent.

In the following sections, we use the regulation of HbF production to illustrate the benefit of modern human genetics of complex traits in understanding clinical heterogeneity in "simple" monogenic diseases. Specifically, although HbF levels themselves have a strong genetic influence, it has been possible to apply modern methods to study the genetic determinants of HbF levels, and thereby identify modifiers of disease severity.

Genetic regulation of HbF

Although the level of HbF falls promptly after birth (to reach <1% in adults), the eventual level maintained in adults varies among individuals. This interindividual variation in HbF levels is highly heritable ($h^2 \sim 0.60–0.89$),[47,48] suggesting that genetic investigation could identify factors that control HbF production. In fact, early DNA mapping and resequencing at the β-globin gene cluster had identified deletions and rare point mutations that prevent globin gene switching and are characterized by very high HbF levels (10–40%) throughout adulthood. This Mendelian form of high γ-globin gene expression and HbF production, termed *hereditary persistence of fetal hemoglobin* (HPFH), is rare and cannot explain the variation in HbF levels observed in the general population. However, HPFH provided the first evidence that HbF production is genetically controlled, and suggested that common DNA polymorphisms at the β-globin locus could act in *cis* to modulate HbF levels in normal individuals. This

prediction was supported by identification of the *Xmn*I polymorphism (rs7482144) in the proximal promoter of HBG2 ($^{G}\gamma$-globin).[45,49,50] The T allele at rs7482144 is associated with increased HbF levels, and is frequent in individuals of Northern European (HapMap CEU T allele frequency: 28%) and west African (HapMap YRI T allele frequency: 5%) ancestry. The *Xmn*I-rs7482144 polymorphism explains 2–10% of the variation in HbF levels, depending on the frequency of the T allele in the different populations considered.[51–53] Despite 25 years of research, it is still unclear whether the *Xmn*I-rs7482144 itself, or a marker in linkage disequilibrium (LD) with it controls HbF levels at the β-globin locus. Analysis of the five classic sickle cell haplotypes also suggested the presence of additional DNA sequence variants at the β-globin locus that influence HbF production, whereas both the Senegal and Arab–Indian haplotypes carry the *Xmn*I-rs7482144 T allele; SCD patients with the Arab–Indian haplotype have much higher HbF levels.[27,44,45] Extensive DNA resequencing in Arab–Indian SCD populations could therefore yield additional DNA polymorphisms associated with HbF levels.

Linkage studies

Despite its strong effect on HbF levels, the *Xmn*I-rs7482144 polymorphism is not sufficient to classify individuals in the low or high ends of the HbF level distribution; some individuals have the *Xmn*I-rs7482144 T allele and low HbF levels, and others have high HbF levels but the C allele at *Xmn*I-rs7482144, which is otherwise associated with decreased HbF. This incomplete correlation suggested that HbF is a complex genetic trait, influenced by multiple DNA sequence variants of modest effect, as well as environmental factors. Specifically, because estimates of narrow-sense heritability for HbF are substantial ($h^2 = 0.60$–0.89),[47,48] and are only partially explained by the *Xmn*I-rs7482144 polymorphism, there are likely other *cis* and *trans* genetic factors that regulate HbF production.

Linkage studies analyze genetic markers across the human genome in families to identify chromosomal regions linked to phenotypic variation within the families. Several linkage scans were carried out to find loci linked to HbF, leading to the identification of HbF linkage peaks on chromosomes 6q23,[54] 8q,[55,56] and Xp22.2.[57] Linkage signals on chromosomes 6q23 and 8q have since then been replicated in additional families; replication data are still lacking for the Xp22.2 linkage peak. Because linkage analysis is better suited to find alleles of moderate or strong effect on complex traits and diseases,[58,59] these results, if correct, suggested that variants might exist with moderate or strong effects on HbF levels. These results therefore supported the notion that HbF is a polygenic trait controlled by genetic factors in *trans* with the β-globin locus on chromosome 11.

Candidate–gene association studies

Linkage scans implicate large chromosomal regions that segregate together within families, which are typically on the order of several megabases of DNA, and usually include many genes. Higher resolution mapping can be achieved by association studies, which search for correlations (LD) between specific alleles and traits or diseases of interest; these correlations usually span much shorter genomic regions. Thein *et al.*[52] used association studies in healthy Northern European twins with F cell measurements available (F cells represent the fraction of erythrocytes that contain HbF, and are highly correlated with HbF levels) to fine-map the 6q23 HbF linkage peak to a small intergenic interval (∼80 kb) between HBS1L and MYB. This association has been replicated in healthy Sardinians,[32] Jamaicans,[60] and in African-American and Brazilian patients with SCD,[53] establishing the HBS1L–MYB intergenic region as the second bona fide HbF-associated locus (in addition to the *XmnI* polymorphism at the β-globin locus itself). Interestingly, genetic variants within the HBS1L–MYB region have pleiotropic effects and have been shown to associate not only with HbF levels, but also with many other hematologic traits,[61–63] suggesting a general role for DNA polymorphisms in the HBS1L–MYB intergenic region in hematopoiesis.

HBS1L and MYB encode, respectively, a Guanosine-5′-triphosphate (GTP)-binding protein and a transcriptional regulator of hematopoiesis. It is currently unclear whether HBS1L, MYB, both, or any of these gene products directly regulate HbF production. HbF-associated variants in the HBS1L–MYB intergenic region were suggested to control the expression of HBS1L in erythroid progenitor cells and lymphoblasts.[52,64] However, MYB is a key regulator of erythropoiesis, and its overexpression is associated with inhibition of γ-globin gene expression.[65] Recent functional work in

erythroid precursor cells using DNase I sensitivity assays and chromatin immunoprecipitation suggested the presence of multiple distal regulatory in the HBS1L–MYB intergenic region.[66] Importantly, additional functional studies will be needed to link how HbF-associated single nucleotide polymorphisms (SNPs) in the HBS1L–MYB region influence the activity of these transcriptional regulatory sequences, and how this relates to HBS1L and MYB activity, as well as HbF production.

Genome-wide association studies

The extensive repertoire of common human DNA sequence variants, such as the HapMap project,[67,68] and low-cost genotyping arrays allow researchers to move away from candidate–gene-based association studies and perform genome-wide association (GWA) experiments in an unbiased manner.[69] Genome-wide association GWA studies interrogate millions of common SNPs for association to human complex traits. The GWA approach has proven extremely useful to find robust genetic associations, highlighting promising and often unanticipated underlying biology.[70] For HbF variation, two GWA studies were initially carried out: one in 179 healthy Northern Europeans selected from the tails of the F cell distribution (lower 5th and upper 95th percentile) among 5,184 individuals with F cell measures available,[51] and one in 4,305 healthy Sardinians.[32] Both studies replicated the previously known associations between HbF levels and the β-globin locus on chromosome 11 and the HBS1L–MYB intergenic region on chromosome 6. They also identified and replicated a new HbF locus on chromosome 2: SNPs in intron 2 of the zinc-finger protein encoded by BCL11A were strongly associated with HbF levels and explained ∼15% of the phenotypic variation in HbF levels.[32,51] This genetic association to HbF levels, originally described in populations of European ancestry, has since then been replicated in African-American and Brazilian SCD patients,[53] and in Chinese and Thai patients with β-thalassemia or the HbE trait.[71,72] The identification of BCL11A is a clear example of the advantage of GWA as an unbiased discovery tool in human genetics: BCL11A would not be an *a priori* strong candidate gene for modulating HbF levels, and the BCL11A locus did not emerge from any of the HbF linkage scans.

Following up on the observation that genetic variants in the BCL11A gene were significantly associated with HbF levels, we demonstrated that the gene product encoded by BCL11A serves as a direct transcriptional regulator of the fetal to adult Hb switch in humans.[73] In particular, we provided evidence that the variants associated with high HbF levels were associated with lower levels of BCL11A expression. Additionally, BCL11A protein expression varies at different developmental stages, such that stages when the γ-globin genes were robustly expressed had low or absent expression of the full-length forms of this protein. Additionally, short variant forms were noted to be present in these earlier stages of ontogeny. BCL11A proteins interact in erythroid cells with specific transcriptional regulators of erythropoiesis, including GATA-1 and FOG-1. Additionally, BCL11A interacts with a repressor complex, known as the NuRD complex, which may contribute to its repressive activity at the γ-globin genes. Knockdown of BCL11A expression (and hence, protein) in primary human erythroid precursor cells leads to dramatic elevations in the expression of the γ-globin genes without resulting in a major perturbation of erythropoiesis. Moreover, knockout of BCL11A in mice prevents appropriate shutoff of endogenous embryonic β-like globin gene expression, and also failure to properly silence human γ-globin transgenes.[74] In further mechanistic studies, we showed that BCL11A acts at a distance from the γ-globin gene promoters within the β-globin locus and cooperates with other factors such as GATA-1 and the transcription factor SOX6 to repress γ-globin expression.[75]

Together, these recent findings establish BCL11A as the first genetically and biochemically validated regulator of the fetal to adult globin switch in humans. Further work should lead to an improved understanding of the mechanisms by which this transcription factor functions to silence the γ-globin genes. Our studies in the mouse suggest that BCL11A mediates the evolutionarily divergent globin gene switches in other mammals, strongly supporting the notion that this factor is a critical mediator of these processes.[74] These findings also show how genetic studies of disease modifiers can identify new potential targets for therapy because downregulating or inhibiting BCL11A *in vivo* could be a desirable therapeutic approach.

Recent work has extended these findings and revealed further informative connections to the human genetics of HbF regulation. By studying a

family from Malta where elevations in HbF levels were noted in certain family members (ranging from 3–19%), a new locus linked to the HbF elevation was found on chromosome 19.[76] By using mapping and sequencing the genes in the region, a heterozygous nonsense mutation in the KLF1 gene was found in affected family members. By gene expression profiling and chromatin analysis, the investigators found that haploinsufficiency of KLF1 appears to affect HbF expression directly and also appears to act indirectly through the downregulation of BCL11A.[76] These findings were also supported by a companion study that utilized mouse models to show that downregulation of KLF1 affected BCL11A expression and globin gene expression.[77]

HbF-associated SNPs and hemoglobinopathy severity

HbF is a strong modifier of disease severity in β-thalassemia and SCD patients. Because the SNPs identified at the BCL11A, HBS1L–MYB, and β-globin explain a significant fraction of the interindividual variation in HbF levels (20–50%), genotypes at these SNPs could become useful predictive tests. Uda *et al.*[32] genotyped the BCL11A rs11886868 SNPs in β-thalassemia patients, and showed that the C allele, associated with increased HbF levels, was found significantly more often in transfusion-independent patients (β-thalassemia intermediate) than in patients that depend on blood transfusions for survival (β-thalassemia major). Modulation of β-thalassemia severity by HbF-associated SNPs has now been extended to the β-globin and HBS1L–MYB loci.[31,72] One limitation in these studies of β-thalassemia severity is that extreme clinical differences are used, and clinically important subtle differences among patients are ignored for the sake of simplicity. Although these studies are promising, further studies will be necessary to dissect to what extent these genetic variants explain the diversity of clinical manifestations noted in these diseases.

In SCD, one of the best measures of overall morbidity is the pain crisis rate.[23] We tested the effect of SNPs at the three known HbF loci on pain crisis rate in African-American patients from the Cooperative Study of Sickle Cell Disease (CSSCD); in aggregate, these SNPs were associated with pain crisis rate, even after considering HbF measurements.[53] This suggests that HbF-associated SNP genotypes provide additional predictive information, beyond steady-state HbF levels. Validation of these preliminary results in independent cohorts are warranted, as they might lead to insight in the identification and follow-up of hemoglobinopathy patients more at risk of severe complications. This would be particularly useful in young patients (<5 years old) in whom HbF levels have not yet reached a steady-state.

Genetics and genomics: beyond HbF

Most of the current efforts to identify modifiers of the major hemoglobinopathies have been focused around the regulation and reactivation of HbF production. There are other important modifiers of severity for β-thalassemia and SCD, including red blood cell counts, which are genetically controlled; identifying loci that control these modifiers could highlight new therapeutic approaches. Moreover, β-thalassemia and SCD are extremely heterogeneous diseases, with some patients presenting severe complications and others not.[78] The identification of genetic variation that protects some hemoglobinopathy patients from most severe complications should improve risk stratification and treatment strategies. Fortunately, large GWA studies in well-phenotyped cohorts, such as the CSSCD, are ongoing and should yield new exciting insights into the clinical heterogeneity of these diseases. As we move to the next step, hemoglobinopathy modifier genetics should also benefit from second-generation DNA resequencing technologies, allowing investigators to consider not only common genetic variation, but also all forms (common, rare, and structural) of genetic variation, and their contribution to disease heterogeneity.

It is expected that the major limitation in our search to understand how genetic modifiers influence severity in β-thalassemia and SCD patients will come more from a lack of appropriately large and well-phenotyped cohorts rather than genetic or technical challenges. As such, and given the healthcare burden that hemoglobinopathies represent worldwide,[2,29] priorities need to be set to create new large prospective patient cohorts to allow continued genetic research. A major challenge that remains will be to define the range of clinical phenotypes in these diseases in large enough samples to provide adequate power, and to better understand the full extent of variability that

exists in these disorders. This will be a critically needed step before further and more comprehensive studies on the genetic basis for variation in these model Mendelian disorders can be more fully appreciated. Clearly, the future appears to be quite fruitful and will require an integrated approach by clinicians caring for patients with these disorders, epidemiologists studying the global burden of these diseases, geneticist interested in understanding this variability, clinical researchers designing studies to better understand the nature of this variation, and basic scientists who will study the mechanisms that underlie the poorly understood variability of the Hb disorders.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Weatherall, D.J. 2001. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* **2:** 245–255.
2. Weatherall, D., O. Akinyanju, S. Fucharoen, *et al*. 2006. Inherited disorders of hemoglobin. In *Disease Control Priorities in Developing Countries* 2nd ed.: 663–680. Oxford University Press. New York.
3. Weatherall, D.J. 2010. The inherited diseases of hemoglobin are an emerging global health burden. *Blood* **115:** 4331–4336.
4. Bunn, H.F. 1997. Pathogenesis and treatment of sickle cell disease. *N. Engl. J. Med.* **337:** 762–769.
5. Serjeant, G.R. 1997. Sickle-cell disease. *Lancet* **350:** 725–730.
6. Stuart, M.J. & R.L. Nagel. 2004. Sickle-cell disease. *Lancet* **364:** 1343–1360.
7. Fucharoen, S. & V. Viprakasit. 2009. Hb H disease: clinical course and disease modifiers. *Hematol. Am. Soc. Hematol. Educ. Prog.* 26–34.
8. Olivieri, N.F. 1999. The beta-thalassemias. *N. Engl. J. Med.* **341:** 99–109.
9. Rund, D. & E. Rachmilewitz. 2005. Beta-thalassemia. *N. Engl. J. Med.* **353:** 1135–1146.
10. Orkin, S.H., H.H. Kazazian Jr. S.E. Antonarakis, *et al*. 1982. Abnormal RNA processing due to the exon mutation of beta E-globin gene. *Nature* **300:** 768–769.
11. Premawardhena, A., C.A. Fisher, N.F. Olivieri, *et al*. 2005. Haemoglobin E beta-thalassaemia in Sri Lanka. *Lancet* **366:** 1467–1470.
12. Tanno, T., N.V. Bhanu, P.A. Oneal, *et al*. 2007. High levels of GDF15 in thalassemia suppress expression of the iron regulatory protein hepcidin. *Nat. Med.* **13:** 1096–1101.
13. Nathan, D.G. 2005. Thalassemia: the continued challenge. *Ann. N.Y. Acad. Sci.* **1054:** 1–10.
14. Ibidapo, M.O. & O.O. Akinyanju. 2000. Acute sickle cell syndromes in Nigerian adults. *Clin. Lab. Haematol.* **22:** 151–155.
15. Fleming, A.F., J. Storey, L. Molineaux, *et al*. 1979. Abnormal haemoglobins in the Sudan savanna of Nigeria. I. Prevalence of haemoglobins and relationships between sickle cell trait, malaria and survival. *Ann. Trop. Med. Parasitol.* **73:** 161–172.
16. Molineaux, L., A.F. Fleming, R. Cornille-Brogger, *et al*. 1979. Abnormal haemoglobins in the Sudan savanna of Nigeria. III. Malaria, immunoglobulins, and antimalarial antibodies in sickle cell disease. *Ann. Trop. Med. Parasitol.* **73:** 301–310.
17. Powars, D., G. Overturf, J. Weiss, *et al*. 1981. Pneumococcal septicemia in children with sickle cell anemia. Changing trend of survival. *JAMA* **245:** 1839–1842.
18. Gaston, M.H., J.I. Verter, G. Woods, *et al*. 1986. Prophylaxis with oral penicillin in children with sickle cell anemia. A randomized trial. *N. Engl. J. Med.* **314:** 1593–1599.
19. Falletta, J.M., G.M. Woods, J.I. Verter, *et al*. 1995. Discontinuing penicillin prophylaxis in children with sickle cell anemia. Prophylactic Penicillin Study II. *J. Pediatr.* **127:** 685–690.
20. Williams, T.N., S. Uyoga, A. Macharia, *et al*. 2009. Bacteraemia in Kenyan children with sickle-cell anaemia: a retrospective cohort and case-control study. *Lancet* **374:** 1364–1370.
21. Hankins, J. & R.E. Ware. 2009. Sickle-cell disease: an ounce of prevention, a pound of cure. *Lancet* **374:** 1308–1310.
22. Platt, O.S., D.J. Brambilla, W.F. Rosse, *et al*. 1994. Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N. Engl. J. Med.* **330:** 1639–1644.
23. Platt, O.S., B.D. Thorington, D.J. Brambilla, *et al*. 1991. Pain in sickle cell disease. Rates and risk factors. *N. Engl. J. Med.* **325:** 11–16.
24. Castro, O., D.J. Brambilla, B. Thorington, *et al*. 1994. The acute chest syndrome in sickle cell disease: incidence and risk factors. The cooperative study of sickle cell disease. *Blood* **84:** 643–649.
25. Adams, R., V. McKie, F. Nichols, *et al*. 1992. The use of transcranial ultrasonography to predict stroke in sickle cell disease. *N. Engl. J. Med.* **326:** 605–610.
26. Kar, B.C., R.K. Satapathy, A.E. Kulozik, *et al*. 1986. Sickle cell disease in Orissa State, India. *Lancet* **2:** 1198–1201.
27. Perrine, R.P., M.J. Brown, J.B. Clegg, *et al*. 1972. Benign sickle-cell anaemia. *Lancet* **2:** 1163–1167.
28. Cao, A. & R. Galanello. 2010. Beta-thalassemia. *Genet. Med.* **12:** 61–76.
29. Weatherall, D.J. & J.B. Clegg. 2001. Inherited haemoglobin disorders: an increasing global health problem. *Bull. World Health Organ.* **79:** 704–712.
30. Taher, A.T., K.M. Musallam, A. El-Beshlawy, *et al*. 2010. Age-related complications in treatment-naive patients with thalassaemia intermedia. *Br. J. Haematol.* **150:** 486–489.
31. Galanello, R., S. Sanna, L. Perseu, *et al*. 2009. Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood* **114:** 3935–3937.
32. Uda, M., R. Galanello, S. Sanna, *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.
33. Taher, A., C. Hershko, M.D. Cappellini. 2009. Iron overload in thalassaemia intermedia: reassessment of iron chelation strategies. *Br. J. Haematol.* **147:** 634–640.

34. Viprakasit, V., C. Lee-Lee, Q.T. Chong, *et al*. 2009. Iron chelation therapy in the management of thalassemia: the Asian perspectives. *Int. J. Hematol.* **90:** 435–445.
35. Sripichai, O., W. Makarasara , T. Munkongdee, *et al*. 2008. A scoring system for the classification of beta-thalassemia/HbE disease severity. *Am. J. Hematol.* **83:** 482–484.
36. Higgs, D.R., B.E. Aldridge, J. Lamb, *et al*. 1982. The interaction of alpha-thalassemia and homozygous sickle-cell disease. *N. Engl. J. Med.* **306:** 1441–1446.
37. Embury, S.H., A.M. Dozy, J. Miller, *et al*. 1982. Concurrent sickle-cell anemia and alpha-thalassemia: effect on severity of anemia. *N. Engl. J. Med.* **306:** 270–274.
38. Embury, S.H., M.R. Clark, G. Monroy & N. Mohandas. 1984. Concurrent sickle cell anemia and alpha-thalassemia. Effect on pathological properties of sickle erythrocytes. *J. Clin. Invest.* **73:** 116–123.
39. Kan, Y.W. & D.G. Nathan. 1970. Mild thalassemia: the result of interactions of alpha and beta thalassemia genes. *J. Clin. Invest.* **49:** 635–642.
40. Higgs, D.R. 2004. Gene regulation in hematopoiesis: new lessons from thalassemia. *Hematol. Am. Soc. Hematol. Educ. Prog.* 1–13.
41. Premawardhena, A., C.A. Fisher, N.F. Olivieri, *et al*. 2005. A novel molecular basis for beta thalassemia intermedia poses new questions about its pathophysiology. *Blood* **106:** 3251–3255.
42. Faa, V., M. Masala, A. Cao, *et al*. 2010. Alpha globin gene duplications in beta thalassemia patients with intact beta globin gene. *Blood Cells Mol. Dis.* **44:** 156–158.
43. Watson J. 1948. The significance of the paucity of sickle cells in newborn Negro infants. *Am. J. Med. Sci.* **215:** 419–423.
44. Pembrey, M.E., W.G. Wood, D.J. Weatherall & R.P. Perrine. 1978. Fetal haemoglobin production and the sickle gene in the oases of Eastern Saudi Arabia. *Br. J. Haematol.* **40:** 415–429.
45. Miller, B.A., N. Olivieri, M. Salameh, *et al*. 1987. Molecular analysis of the high-hemoglobin-F phenotype in Saudi Arabian sickle cell anemia. *N. Engl. J. Med.* **316:** 244–250.
46. Higgs, D.R. & W.G. Wood. 2008. Genetic complexity in sickle cell disease. *Proc. Natl. Acad. Sci. USA* **105:** 11595–11596.
47. Garner, C., T. Tatu, J.E. Reittie, *et al*. 2000. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood* **95:** 342–346.
48. Pilia, G., W.M. Chen, A. Scuteri, *et al*. 2006. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2:** e132.
49. Labie, D., O. Dunda-Belkhodja, F. Rouabhi, *et al*. 1985. The -158 site 5' to the G gamma gene and G gamma expression. *Blood* **66:** 1463–1465.
50. Labie, D., J. Pagnier, C. Lapoumeroulie, *et al*. 1985. Common haplotype dependency of high G gamma-globin gene expression and high HbF levels in beta-thalassemia and sickle cell anemia patients. *Proc. Natl. Acad. Sci. USA* **82:** 2111–2114.
51. Menzel, S., C. Garner, I. Gut, *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.
52. Thein, S.L., S. Menzel, X. Peng, *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.
53. Lettre, G., V.G. Sankaran, M.A. Bezerra, *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.
54. Garner, C., J. Mitchell, T. Hatzis, *et al*. 1998. Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production, on chromosome 6q23. *Am. J. Hum. Genet.* 62: 1468–1474.
55. Garner, C., N. Silver, S. Best, *et al*. 2004. Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin. *Blood* **104:** 2184–2186.
56. Garner, C.P., T. Tatu, S. Best, *et al*. 2002. Evidence of genetic interaction between the beta-globin complex and chromosome 8q in the expression of fetal hemoglobin. *Am. J. Hum. Genet.* **70:** 793–799.
57. Dover, G.J., K.D. Smith, Y.C. Chang, *et al*. 1992. Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. *Blood* **80:** 816–824.
58. Risch, N.J. 2000. Searching for genetic determinants in the new millennium. *Nature* **405:** 847–856.
59. Risch, N. & K. Merikangas. 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516–1517.
60. Creary, L.E., P. Ulug, S. Menzel, *et al*. 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.
61. Menzel, S., J. Jiang, N. Silver, *et al*. 2007. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood* **110:** 3624–3626.
62. Ganesh, S.K., N.A. Zakai, F.J. van Rooij, *et al*. 2009. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41:** 1191–1198.
63. Soranzo, N., T.D. Spector, M. Mangino, *et al*. 2009. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41:** 1182–1190.
64. Dixon, A.L., L. Liang, M.F. Moffatt, *et al*. 2007. A genome-wide association study of global gene expression. *Nat. Genet.* **39:** 1202–1207.
65. Jiang, J., S. Best, S. Menzel, *et al*. 2006. cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood* **108:** 1077–1083.
66. Wahlberg, K., J. Jiang, H. Rooks, *et al*. 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.
67. Altshuler, D., L.D. Brooks, A. Chakravarti, *et al*. 2005. A haplotype map of the human genome. *Nature* **437:** 1299–1320.

68. Frazer, K.A., D.G. Ballinger, D.R. Cox, *et al.* 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449:** 851–861.
69. Hirschhorn, J.N. & M.J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6:** 95–108.
70. Hindorff, L.A., P. Sethupathy, H.A. Junkins, *et al.* 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106:** 9362–9367.
71. Sedgewick, A.E., N. Timofeev, P. Sebastiani, *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.
72. Nuinoon, M., W. Makarasara, T. Mushiroda, *et al.* 2009. A genome-wide association identified the common genetic variants influence disease severity in beta(0)-thalassemia/hemoglobin E. *Hum. Genet.* **127:** 303–314.
73. Sankaran, V.G., T.F. Menne, J. Xu, *et al.* 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322:** 1839–1842.
74. Sankaran, V.G., J. Xu, T. Ragoczy, *et al.* 2009. Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460:** 1093–1097.
75. Xu, J., V.G. Sankaran, M. Ni, *et al.* 2010. Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev.* **24:** 783–798.
76. Borg, J., P. Papadopoulos, M. Georgitsi, *et al.* 2010. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat. Genet.* **42:** 801–805.
77. Zhou, D., K. Liu, C.W. Sun, *et al.* 2010. KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat. Genet.* **42:** 742–744.
78. Serjeant, G.R., D.R. Higgs & I.R. Hambleton. 2007. Elderly survivors with homozygous sickle cell disease. *N. Engl. J. Med.* **356:** 642–643.