PHENOTYPE–GENOTYPE RELATIONSHIPS IN MONOGENIC DISEASE: LESSONS FROM THE THALASSAEMIAS

D. J. Weatherall

The remarkable phenotypic diversity of the β-thalassaemias reflects the heterogeneity of mutations at the β-globin locus, the action of many secondary and tertiary modifiers, and a wide range of environmental factors. It is likely that phenotype–genotype relationships will be equally complex in the case of many monogenic diseases. These findings highlight the problems that might be encountered in defining the relationship between the genome and the environment in multifactorial disorders, in which the degree of heritability might be relatively low and several environmental agents are involved. They also emphasize the value of an understanding of phenotype–genotype relationships in designing approaches to gene therapy.

THALASSAEMIA
Inherited disorder caused by the abnormal production of haemoglobin.

Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK.
Correspondence to D.J.W.

A central problem for the medical sciences in general, and for clinical genetics in particular, is the extent to which it will be possible to relate findings at the molecular level to clinical phenotypes. This question will touch on every aspect of medical practice and research in the post-genome period. It will dominate predictive genetics and genetic counselling, particularly at a time when much effort is being directed at extensive population studies to find the genes that are involved in common, multifactorial diseases, most of which have been shown to have a relatively small inherited component.

In estimating the magnitude of this problem it is helpful to ask to what extent it has been possible to relate the molecular pathology of simple monogenic diseases to their associated clinical phenotypes. Here, I review progress that has been made in the case of the inherited disorders of haemoglobin, notably the THALASSAEMIAS, which are the most common genetic diseases and among the first to be analysed at the molecular level[1,2].

Inherited disorders of haemoglobin

The normal human haemoglobins. The structure of human haemoglobin (Hb) changes during development[3,4]. All the normal haemoglobins are tetramers of two pairs of unalike globin chains. Adult (HbA) and fetal (HbF) haemoglobins have α-chains that are combined with β- (HbA, $\alpha_2\beta_2$), δ- (HbA$_2$, $\alpha_2\delta_2$) or γ-chains (HbF, $\alpha_2\gamma_2$), whereas in the embryo, α-like chains called ζ-chains combine with γ- (Hb Portland, $\zeta_2\gamma_2$) or ε-chains (Hb Gower 1, $\zeta_2\varepsilon_2$), and α- and ε-chains form Hb Gower 2 ($\alpha_2\varepsilon_2$). Embryonic haemoglobin is confined to the yolk-sac stage of development and thereafter is replaced by HbF until shortly before term. After birth, HbF is replaced by HbA and HbA$_2$ over the first year of life, although in normal adults small amounts of HbF, constituting ~1% of the total haemoglobin, continue to be produced (FIG. 1).

The α-like genes are encoded on chromosome 16 in the order 5′-ζ2-ψζ1-ψα2-ψα1-α2-α1-θ-3′, whereas the β-like genes form a cluster on chromosome 11 as 5′-ε-^{G}γ-^{A}γ-ψβ-δ-β-3′. The β-like genes contain two introns of 122–130 and 850–900 base pairs (bp), between codons 30 and 31, and 104 and 105, respectively. The α- and ζ-genes contain similar, although smaller, introns. As well as typical promoter and enhancer sequences, each globin gene cluster has an

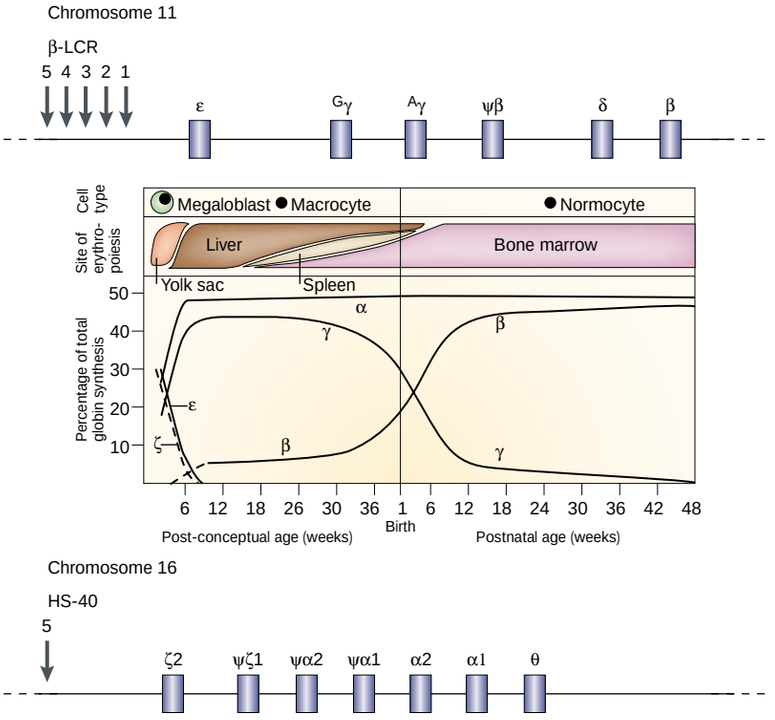


Figure 1 | **The α-globin gene cluster on chromosome 16 and the β-globin gene cluster on chromosome 11.** Vertical arrows indicate the location of DNaseI hypersensitive sites that are thought to be involved in globin gene regulation. The products of the $^{G}\gamma$- and $^{A}\gamma$-genes are γ-chains with either glycine ($^{G}\gamma$) or alanine ($^{A}\gamma$) at position 136. The insert shows the sequential activation of the embryonic, fetal and adult globins. A megaloblast is a large red-cell precursor, a macrocyte is a large red cell and a normocyte is a normal-sized red cell. (LCR, locus control region.)

important upstream regulatory region. In the β-gene cluster, this is called the locus control region (LCR), whereas in the α-gene cluster it is designated HS-40. The β-globin LCR establishes a transcriptionally active domain that spans the entire β-globin gene cluster. Each cluster contains various binding sites for both erythroid-specific and more ubiquitous DNA-binding proteins. The developmental regulation of the globin genes reflects their sequential activation in a 5′–3′ direction; the way in which these developmental switches is controlled in globin gene expression is still not fully understood[4].

The inherited disorders of haemoglobin. It has been estimated that ~7% of the world's population are carriers for different inherited disorders of haemoglobin, making them the commonest human monogenic diseases[5]. They are divided into two main groups, the structural haemoglobin variants and the thalassaemias, which result from defective synthesis of the globin chains (BOX 1). There is a third family comprising conditions in which there is a defect in the normal switch from fetal to adult haemoglobin production that is called hereditary persistence of fetal haemoglobin (HPFH). Although of no clinical importance *per se*, the co-inheritance of some forms of HPFH can modify the phenotypes associated with the structural haemoglobin variants or thalassaemias.

Over 700 structural haemoglobin variants have been identified[6], but only three, sickle haemoglobin (HbS), HbC and HbE, occur at a high frequency in different populations. The gene for HbS is distributed throughout sub-Saharan Africa, parts of the Mediterranean region, the Middle East and certain regions of India, whereas HbC is restricted to West Africa and parts of the Mediterranean region. HbE, the commonest structural haemoglobin variant, occurs at a very high frequency in parts of India, Myanmar and throughout Southeast Asia.

The thalassaemias are classified into α-, β-, δβ- and εγδβ-thalassaemias, on the basis of the particular globin chain, or chains, that is ineffectively synthesized[1]. Each of these forms of thalassaemia is extremely heterogeneous. There are two main varieties of α-thalassaemia, α^{+}- and α^{o}-thalassaemia, designated in this way because they reflect either a partial or complete defect in α-globin synthesis from the affected chromosome, respectively[7]. Normal individuals have two α-globin genes per haploid genome and so their genotype can be written αα/αα. In the α^{+}-thalassaemias, one of the linked α-globin genes is lost by deletion, –α/αα, or inactivated by a point mutation, $\alpha^{T}\alpha/\alpha\alpha$. In the α^{o}-thalassaemias, both of the linked α-globin genes are lost, most commonly by deletions that involve part or all of the α-globin gene cluster; the heterozygous genotype is expressed as ––/αα. The β-thalassaemias are similarly subdivided into the β^{o}-thalassaemias, in which there is no β-globin chain production, and the β^{+}- or β^{++}-thalassaemias, in which there is a severe or mild reduction in the output of β-globin chains, respectively. The δβ- and εγδβ-thalassaemias, which are quite rare, result from a series of deletions involving the β-globin gene cluster that remove either the δ- and β-genes or all the genes of the cluster. Similarly, many forms of HPFH result from deletions of this cluster, or from point mutations in the γ-gene promoters[4].

Unfortunately, this classification of the inherited haemoglobin disorders is not completely straightforward. A few of the structural haemoglobin variants are synthesized at a reduced rate, or are highly unstable, and result in the phenotype of thalassaemia. For example, the substitution at β-codon 26 (GAG→AAG) that gives rise to HbE, β 26 (Glu→Lys), also activates a cryptic splice site that causes abnormal mRNA processing[8]. So, β^{E}-chains are produced in reduced amounts, which results in a mild β-thalassaemic phenotype. Furthermore, because the thalassaemias and structural haemoglobin variants occur together at a high frequency in many populations, it is not uncommon for an individual to inherit genes for both types of condition. For example, the compound heterozygous state for β-thalassaemia and HbE, HbE β-thalassaemia, is frequently encountered in parts of the Indian subcontinent and throughout Southeast Asia[1].

The β-thalassaemias pose by far the most important global public health problem, and so this review is confined largely to their phenotype–genotype relationships and those of their interactions with HbE. However, I also consider the α-thalassaemias insomuch as they have an important role in modifying the phenotype of the β-thalassaemias.

Box 1 | Genetic disorders of human haemoglobin

Structural variants
Over 700 described. First called by letters of the alphabet (for example, HbC, HbE and HbS), but later by place of discovery. Mainly due to single amino-acid substitutions, although a few have elongated or short globin chains. All have the general structure $\alpha_2\beta_2$, except for Hb Bart's and HbH, which are γ_4- or β_4-homotetramers, respectively, and which are formed when α-chain production is defective in α-thalassaemia.

Thalassaemias
Disorders due to defective and imbalanced globin production. The β-thalassaemias result from over 200 different mutations of β-globin genes. The α-thalassaemias result from more than 80 different deletions or point mutations in the α-globin genes.

Hereditary persistence of fetal haemoglobin
A heterogeneous group of inherited defects in the switch from fetal to adult haemoglobin production, with persistent fetal haemoglobin production.

Phenotypic diversity of β-thalassaemias

The hallmark of the β-thalassaemias is defective β-globin synthesis, which leads to imbalanced globin chain production and an excess of α-chains. The excess chains aggregate in red-cell precursors, and cause abnormal cell maturation and their premature destruction in the bone marrow. There is abundant evidence that the severity of β-thalassaemia is related to the degree of globin-chain imbalance[9]. It is equally clear that its most important complications, such as SPLENOMEGALY, bone disease, and endocrine and cardiac damage, can be related to the degree of ANAEMIA together with the magnitude of iron loading of the tissues that results from the increased absorption of iron and from repeated blood transfusion[9].

The β-thalassaemias have extremely diverse clinical phenotypes[1]. At the severe end of the spectrum, many homozygous or compound heterozygous states are characterized by profound anaemia from early life that, if not treated with regular blood transfusions, leads to death in the first year — a condition known as β-thalassaemia major. Conversely, many patients with the same disease have a milder illness that ranges from being only slightly less severe than the major form, through a spectrum of decreasing severity of anaemia, to one which is symptomless and is ascertained only by routine examination of the blood. This diverse collection of β-thalassaemias of varying severity constitutes the β-thalassaemia intermedias.

Even the heterozygous states for β-thalassaemia show wide phenotypic diversity. Typically, the inheritance of a single β-thalassaemia allele is associated with mild anaemia and characteristic morphological changes of the red cells. However, in some cases, the effect of a single β-thalassaemia allele can be completely silent with no definable haematological abnormalities, whereas in others it might cause a phenotype as severe as the major forms of the illness — that is, a dominantly inherited form of β-thalassaemia[10,11].

Definition of the β-thalassaemia phenotype

In attempting to relate phenotype to genotype for any disease, it is essential to have a consistent definition of the severity of the phenotype. Because nearly all the clinical manifestations and complications of the β-thalassaemias can be related to the degree of anaemia, the steady-state haemoglobin level should be adequate to compare different genotypes. Unfortunately, the situation is not as simple as this. For, although it is easy to define phenotypes at the severe (major) and mild (minor) ends of the spectrum of severity, the forms that lie in between are much more difficult to categorize[1]. The haemoglobin level at presentation is of little value because many babies are first seen during an INTERCURRENT ILLNESS, such as severe infection, which might temporarily exacerbate their anaemia. The phenotype is not static and the haemoglobin level can decline with age owing to many factors, including progressive enlargement of the spleen, nutritional deficiency and, particularly in poorer countries, recurrent infection. Furthermore, the complications of the disease, particularly splenomegaly, bone deformity, OSTEOPOROSIS and iron loading, are not always related to the haemoglobin level.

Because of these difficulties, the relatively few studies that have even attempted a rigorous definition of the β-thalassaemia phenotype in relationship to genotype[12–16] have had to fall back on phenotypic classifications based on the haemoglobin level and transfusion history of the patients, with all the inherent shortcomings. In essence, they have divided patients into those who are transfusion dependent and those with intermediate forms of the illness, subdivided according to their steady-state haemoglobin levels and their requirements for intermittent transfusion over a relatively long period of observation[15,16]. It is against this unsatisfactory background that the relationship between phenotype and genotype in β-thalassaemia must be examined.

Mechanisms underlying phenotypic diversity

The remarkable variability in clinical severity of the β-thalassaemias reflects both genetic and environmental factors. It is becoming apparent that the genetic element involves many loci, some of which are directly involved with the basic defect in globin synthesis, whereas others, which modify the variable complications of the disease, have nothing to do with globin. For this reason, it is convenient to classify these genetic modifiers into the following groups: primary, the many different mutations of the β-globin genes that underlie β-thalassaemia; secondary, loci that are also involved in globin synthesis; and tertiary, loci that are not involved in globin production but that might modify the complications of the disease in many different ways. The latter group includes the many different polymorphisms that have been co-selected with the thalassaemias and that might further modify their phenotypes.

Complications acquired as the result of the primary defect in β-globin synthesis can have a profound effect on the phenotype. Similarly, it is becoming increasingly clear that for at least some forms of the disease, environmental and social factors might also have an important role in modifying individual responses to the different forms of thalassaemia.

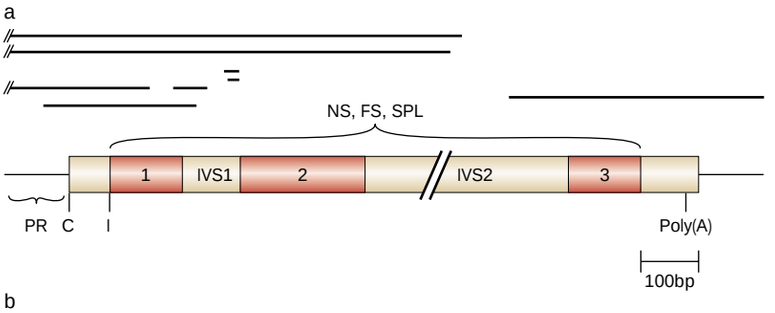
SPLENOMEGALY
Enlargement of the spleen that results in the pooling of red cells and in anaemia.

ANAEMIA
A reduction in the haemoglobin level or red-cell count, which leads to defective tissue oxygenation.

INTERCURRENT ILLNESS
An illness unrelated to the primary disease (for example, infection or malnutrition in a child with thalassaemia).

OSTEOPOROSIS
Reduction in the amount of bone without a change in its composition. Associated with bone pain and fractures.

a



b

Mild β-thalassaemia alleles			'Silent' β-thalassaemia alleles
–90 C→T	–30 T→A	CD 26 G→A (Hb E)	–92 C→T
–88 C→A	–30 T→C	CD 27 G→T Knossos	–101 C→T
–88 C→T	–29 A→G	IVS1 6 T→C	5′ UTR +10 –T
–87 C→A	–28 A→G (mild in blacks;	3′ UTR 47 C→G	5′ UTR +33 C→G
–87 C→G	severe in Chinese)	PolyA AATAAA→AACAAA	IVS2 844 C→G
–87 C→T	5′ UTR +22 G→A	PolyA AATAAA→AATUAA	CAP +1 A→C
–86 C→G	CD 19 A→G Malay	PolyA AATAAA→AATAGA	3′ UTR +6 C→G
–31 A→G	CD 24 T→A	PolyA AATAAA→AATAAC	

Figure 2 | **Human β-globin mutations. a** | A schematic representation of the human β-globin gene with the main classes of mutations that cause β-thalassaemia. Exons are indicated in red and non-coding regions, such as introns (IVS1 and IVS2), are in yellow. Deletions associated with thalassaemia are shown above the gene. Various point mutations have been identified in coding and non-coding parts of the gene: PR, promoter; C, CAP site; I, initiation codon; NS, nonsense; FS, frameshift; SPL, splicing. **b** | Mild and silent β-thalassaemia mutations. The mild mutations are shown on the left and the silent mutations on the right. Those named Knossis, Malay and HbE are all due to splice mutations that also produce a structural haemoglobin variant at a reduced level, which results in the phenotype of mild β-thalassaemia. (UTR, untranslated region; IVS, intervening sequence (intron); CD, coding region (exon).)

Primary modifiers: β-thalassaemia alleles

Over 200 different mutations have been identified in the β-globin genes of patients with β-thalassaemia[1,17]. With the exception of a few deletions, the bulk of them consist of point mutations or the loss of one or two bases, which interferes with gene function either at the transcriptional, translational or post-translational levels (FIG. 2). The resulting phenotypes reflect the effects of the β^0-thalassaemias, in which there is no β-globin gene product, and the β^+- or β^{++}-thalassaemias, in which there is a marked or mild reduction in the output of β-chains, respectively[1].

Some of the clinical heterogeneity of the β-thalassaemias can be explained by the differing severity of particular alleles. Clearly, the β^0-thalassaemias should be associated with a severe phenotype although, as we shall see later, this is not always the case. The β^+- and β^{++}-thalassaemia alleles are remarkably diverse in their effect on the output of β-globin chains. They are most easily described by their phenotypic effects in heterozygotes.

A few β-thalassaemia mutations are completely 'silent': they have no demonstrable effects in carriers and have usually been ascertained by finding individuals with intermediate forms of β-thalassaemia in whom one parent has typical β-thalassaemia traits and the other seems to be normal[18–20]. Overall, they are uncommon except for the –101 C→T mutation, which has been observed frequently in the Mediterranean region. There, it interacts with a variety of more severe β-thalassaemia alleles to produce mild forms of β-thalassaemia intermedia[19,20].

The mild β-thalassaemia alleles are listed in FIG. 2b. The majority result from mutations in the promoter elements of the β-globin genes or in the poly(A) cleavage sites; a few involve mutations at cryptic splice sites in exons or consensus sequences in introns[17]. Phenotypically, they all result in milder, although clearly definable, changes in the red cells in heterozygotes, and disorders of intermediate severity in homozygotes. Overall, their interactions with severe alleles result in transfusion-dependent disorders or intermediate forms of β-thalassaemia at the more severe end of the spectrum[1].

The world distribution of the different β-thalassaemia alleles is shown in FIG. 3. The only common mild alleles are the promoter element mutations in African populations[21], βIVS1 6 T→C in the Mediterranean region[22,23] and βCD 26 G→A (which gives rise to HbE[8] and is widely distributed throughout the Indian subcontinent and Southeast Asia[1]). Clearly therefore, because mild β-thalassaemia alleles, with the exception of HbE, are relatively uncommon (FIG. 3) in many high-frequency populations, mild alleles cannot account for many of the less severe forms of the disease.

From recessive to dominant inheritance. A form of β-thalassaemia with a dominant mode of inheritance was first identified in an Irish family in which several members had a moderately severe form of β-thalassaemia that was clearly inherited as a Mendelian dominant[10]. Subsequently, it was found that the underlying mutation of the β-globin gene involves two deletions of 4 and 11 bp in exon 3 (interrupted by an insertion of 5 bp), which give rise to a frameshift and the predicted synthesis of an elongated β-chain variant with an abnormal carboxyl terminus[24]. It was suggested that heterozygotes for this condition are more severely affected than those for other forms of β-thalassaemia because, as well as producing an excess of α-chains, they synthesize highly unstable β-chain products that bind haem and precipitate in the red-cell precursors; more recent studies of the constitution of the inclusion bodies in the bone marrow in this condition have confirmed that this is the case[25]. Since this first description, numerous families with dominantly inherited β-thalassaemia have been described, arising from a heterogeneous series of mutations that include missense mutations, minor deletions leading to the loss of intact codons and frameshifts[26–28].

As shown in FIG. 4, most in-phase chain-termination mutations that result in dominantly inherited β-thalassaemia are in exon 3 or beyond, whereas those that are recessively inherited lead to termination in exons 1 or 2. In the latter case, very little abnormal β-globin mRNA is found in the cytoplasm of red-cell precursors. It has been suggested that the effects of these premature termination codons on the accumulation of mRNA (nonsense-mediated RNA decay) might reflect a surveillance mechanism to prevent mRNAs coding for truncated peptides[29–31]. Conversely, the exon 3 mutations that cause dominant β-thalassaemia are

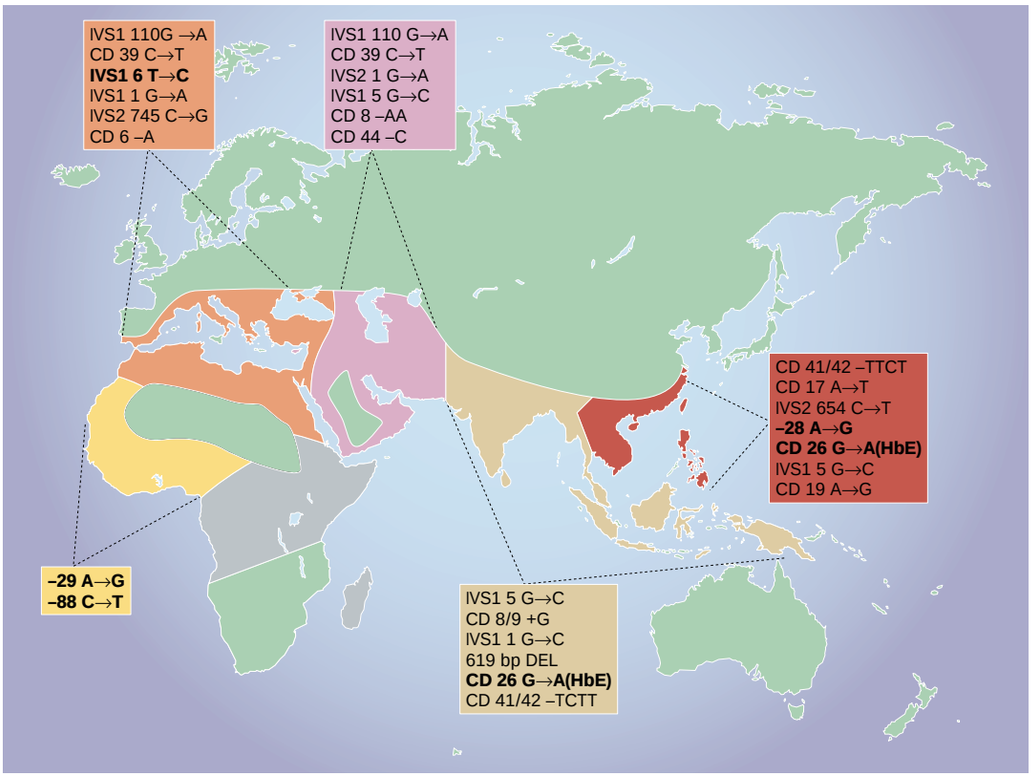


Figure 3 | **The global distribution of the β-thalassaemia mutations.** The common mild mutations are shown in bold. β-thalassaemia also occurs in the regions shaded in grey, but little is known about its molecular pathology in these areas.

associated with substantial amounts of abnormal cytoplasmic mRNA, leading to the synthesis of β-chain products[24,28] that are unstable and hence that also act in a dominant-negative fashion and damage red-cell precursors.

Variation in the degree of instability of these abnormal β-globin gene products provides a further basis for phenotypic variation. Although highly unstable products precipitate in the red-cell precursors and produce dominant β-thalassaemia, less unstable products are able to combine with α-globin subunits to produce a haemoglobin tetramer that survives through the different stages of red-cell maturation, only to precipitate in the mature red cell in the peripheral blood. In this case, red-cell production is relatively normal and the phenotype is that of a HAEMOLYTIC ANAEMIA associated with inclusion bodies in the red blood cells. Indeed, there is a wide range of phenotypes that extends from typical thalassaemic disorders, through conditions in which there is defective ERYTHROPOIESIS and severe haemolysis, to pure haemolytic anaemias[27,28].

In summary, the broad spectrum of β-thalassaemia alleles can produce a wide spectrum of different β-thalassaemia phenotypes that ranges from silent carriers to homozygous (recessive), heterozygous (dominant) or compound heterozygous inheritance of the principal forms of the disease. But this is not the whole story.

HAEMOLYTIC ANAEMIA
Anaemia due to reduced red-cell survival.

ERYTHROPOIESIS
Differentiation and maturation of red blood cells.

Secondary modifiers

Sibship and family studies have shown that there is wide phenotypic diversity even among individuals with the same β-thalassaemia genotype. There are two particularly striking examples. First, not every homozygote or compound heterozygote for β^0-thalassaemia, in which there is no output of β-chains, is severely affected[32,33]. Second, studies of compound heterozygotes of Indian or Southeast Asian origin, who have inherited an HbE allele from one parent and the same β-thalassaemia allele from the other, show wide phenotypic variability in the resulting disorder, HbE β-thalassaemia[34]. For example, recent studies in Sri Lanka have shown that individuals with this condition who carry identical β-thalassaemia mutations have phenotypes that range from transfusion-dependent anaemia in early life to a clinically 'silent' condition that is ascertained by chance in middle age[35]. At least some of this remarkable phenotypic diversity can be explained by the action of the products of other loci involved in globin synthesis.

Because the severity of the anaemia of β-thalassaemia reflects defective β-globin chain production, which leads to excess α-chains and their deleterious effects on red-cell production and survival, it follows that anything that modifies the magnitude of the surfeit of α-chains should have an important effect on the phenotype. Variation at two loci that mediate this effect have been identified — the α- and γ-globin loci.

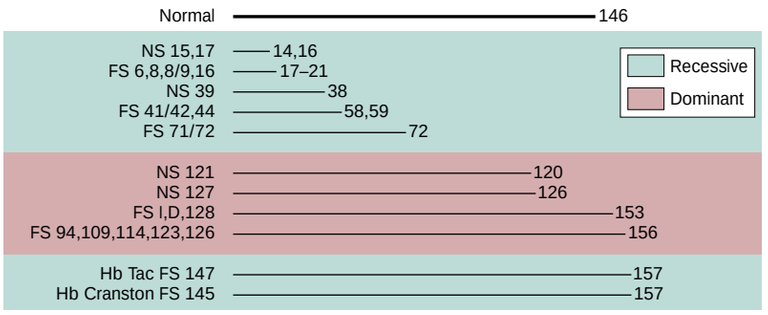


Figure 4 | **Dominant β-thalassaemia.** Nonsense (NS) or frameshift (FS) mutations in exons 1 and 2 of the β-globin gene are associated with recessive β-thalassaemia. Exon 3 mutations result in long and sometimes unstable β-chain products that bind haem and precipitate in red-cell precursors. The lengths (in amino acids) of the abnormal products are indicated, with the wild-type product shown at the top. Note that, some elongated products are associated with recessive forms of β-thalassaemia (for example, Hb Tac and Hb Cranston).

Co-inheritance of α-thalassaemia. Because α-thalassaemia coexists with β-thalassaemia at a high frequency in many populations (FIGS 3, 5), it is not uncommon to inherit both conditions[36]. So, homozygotes or compound heterozygotes for severe β-thalassaemia alleles might also be heterozygous or homozygous for α^+-thalassaemia, or heterozygous for α^o-thalassaemia. From studies of the phenotypes of these remarkable experiments of nature, it is apparent that the co-inheritance of α-thalassaemia can ameliorate the severity of β-thalassaemia. And because there is a wide range of phenotypic expression of the different α-thalassaemia alleles[1,7], this provides a further mechanism for extensive clinical diversity of the β-thalassaemias. These interactions are reviewed in detail elsewhere[1]. In short, the co-inheritance of different α-thalassaemia alleles might reduce the severity of the homozygous or compound heterozygous states for β^o-thalassaemia to some degree[37] and can convert the severe forms of β^+-thalassaemia into milder, non-transfusion-dependent conditions[38].

As well as providing a mechanism for the amelioration of the β-thalassaemia phenotype, the fact that the coexistence of α-thalassaemia can reduce the severity of β-thalassaemia provides clear evidence that the chief pathophysiological mechanism in the β-thalassaemias is imbalanced globin chain production rather than the under-production of haemoglobin. So, although the red cells of individuals who have inherited both types of thalassaemia might be grossly under-haemoglobinized, the anaemia is less severe and consequently the phenotype is milder[39].

Variation in fetal haemoglobin production. When it became clear that some homozygous β^o-thalassaemics have a mild clinical phenotype and are able to maintain a relatively high haemoglobin level, all of which is HbF, it seemed likely that an unusual propensity for the production of HbF after birth might be an important factor in modifying the clinical course of β-thalassaemia. There is now good evidence that this is the case. Normal children and adults produce small amounts of HbF that seem to be confined to particular red-cell populations called F cells[40]. In a patient with β-thalassaemia, the γ-chains bind some of the excess α-chains to produce HbF, and so red-cell precursors that are synthesizing γ-chains come under intense selection, a mechanism that accounts for much of the increased HbF in the blood of β-thalassaemics[41]. Because the number of F cells in normal individuals is under genetic control[42], it is not surprising that there is a variable propensity for producing HbF in patients with β-thalassaemia. However, it is now clear that many different genes must be involved, some in the β-globin gene cluster, others on different chromosomes.

There are several determinants within the β-globin gene cluster that are involved in setting the level of HbF in β-thalassaemia. The conditions that constitute hereditary persistence of fetal haemoglobin result from deletions that involve the β-globin gene cluster or point mutations in the promoters of one or other of the duplicated γ-globin genes. They are all characterized by the persistent production of high levels of HbF into adult life[1,4]. However, these genetic variants are rare and, numerically, play a relatively small part in the modification of the β-thalassaemia phenotype. By contrast, there is a relatively common polymorphism at position –158 in the $^G\gamma$-gene, which involves a C→T change[43]. Although this seems to have little effect in normal people, there is good evidence that homozygous individuals have an increased propensity to produce HbF under conditions of haemopoietic stress, and that this can have the effect of raising fetal haemoglobin levels in patients with β-thalassaemia. As this polymorphism is widespread, it is an important factor in the modification of β-thalassaemia phenotypes, particularly those with β^o-thalassaemia of the intermediate variety[9,37,44,45]. In addition, some β-thalassaemia alleles might themselves favour a higher output of HbF. This is certainly true in the case of promoter mutations[21], or deletions that involve the promoter elements of the β-globin gene[46], an observation that might reflect competition between γ- and β-globin gene promoters for rate-limiting regulatory proteins or for interaction with the LCR. The possible role of γ-globin gene triplication or other structural changes in the β-globin gene complex in the modification of HbF production are reviewed elsewhere[1,17].

There are also genetic determinants responsible for increasing the output of HbF in some patients with β-thalassaemia that are not encoded in the β-globin gene cluster. For example, in families with milder forms of β-thalassaemia owing to increased HbF production, unusually high levels of HbF are sometimes found in one of the heterozygous parents, or one or more unaffected relatives have slightly increased levels of HbF[16]. In studies of several generations of a large family in which a gene of this type segregated independently from the β-globin gene cluster, the locus involved has been assigned to chromosome 6 (REFS 47,48). However, analyses of similar families indicate that there are genetic determinants involved in increased HbF production in β-thalassaemia that are not linked to the β-globin gene cluster or chromosome 6 (REF. 49). There is also a locus on the X chromosome that seems to have an effect on the numbers of F cells in adults[50], although its role, if any, in determining the level of HbF in β-thalassaemia is not yet clear. In short, it is apparent that

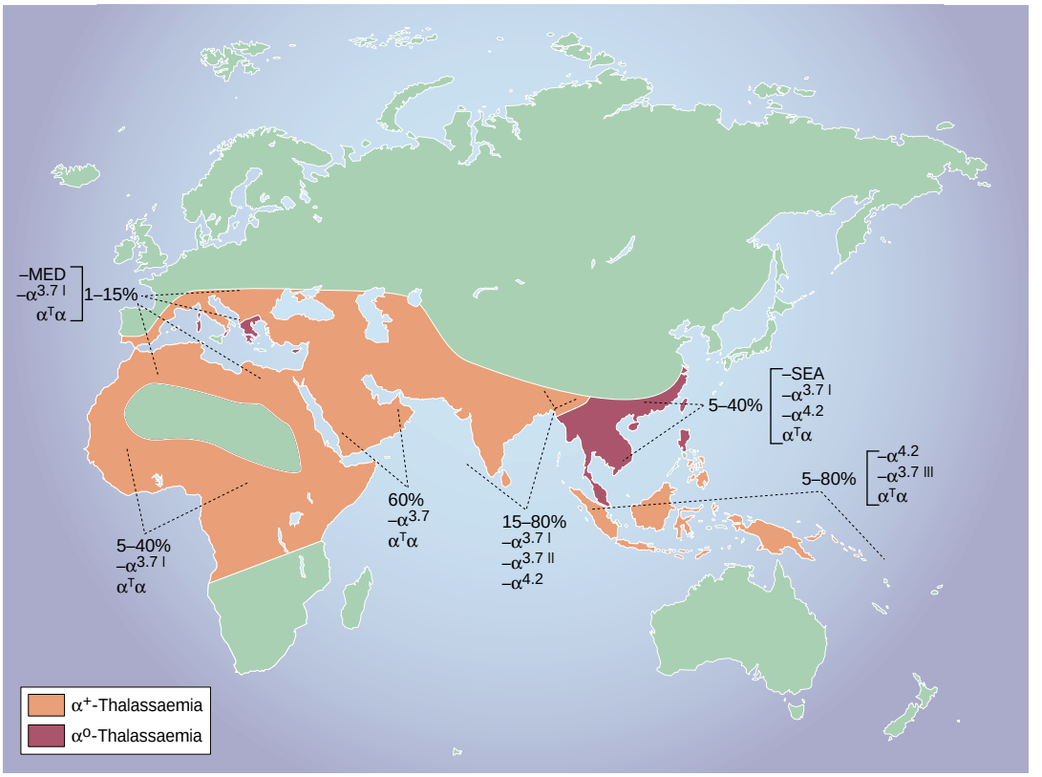


Figure 5 | **The global distribution of the α-thalassaemias.** The α^+-thalassaemias result from deletions of 3.7 kb or 4.2 kb, which remove a single α-globin gene. There are three subvarieties of $-\alpha^{3.7}$, designated $-\alpha^{3.7\,I}$, $-\alpha^{3.7\,II}$ and $-\alpha^{3.7\,III}$, depending on the site of the crossover event that underlies the deletion. Non-deletion forms of α-thalassaemia are written α^T. In some cases, they are associated with the production of a haemoglobin variant. For example, α^{CS} refers to the α-globin chain-termination mutant, Hb Constant Spring. This mutation downregulates the α2-globin gene and is also associated with the production of an elongated α-chain variant. The α^0-thalassaemias result from deletions of both of the linked α-globin genes, and are further characterized by their length and place of discovery. (MED, Mediterranean; SEA, Southeast Asia.) The small fork in northern India represents a localized population with an extremely high frequency of α-thalassaemia. Note that the α-thalassaemia distribution is not as well charted as the thalassaemia distribution shown in FIG. 3, and it is therefore not possible to be as precise about the ranges of the various alleles.

there are several genes that are not linked to the β-globin gene complex that can fine tune the level of HbF, both in normal adults and in those with β-thalassaemia. Presumably they encode transcription factors that are involved with the activation or repression of γ-chain synthesis or in modulating the kinetics of haemopoietic-cell development to make γ-chain synthesis more likely in conditions of haemopoietic expansion.

Increasing the severity of β-thalassaemia. Just as the α-thalassaemias, or a genetically determined increase in HbF production, can ameliorate the phenotype of β-thalassaemia, variability at the α-chain loci can also have the opposite effect. Instead of the duplicated α-globin gene arrangement, αα, some individuals are heterozygous or even homozygous for triplicated or quadruplicated α-globin gene arrangements, ααα or αααα[51,52]. They are found in most populations, although the frequency of chromosomes that contain additional α-globin genes is not known in detail. They have no phenotypic effect in normal people — presumably the small excess of α-chains that is synthesized can be dealt with by proteolysis — but this is not the case in individuals with β-thalassaemia. The consequences on the phenotype of β-thalassaemia of the inheritance of additional numbers of α-globin genes have been best defined in heterozygotes. For example, β-thalassaemia carriers who are heterozygous or homozygous for triplicated α-globin gene arrangements have a β-thalassaemia disorder of intermediate severity[53,54]; a similar effect is seen in those who have inherited a chromosome with the quadruplicated α-globin gene arrangement[55,56].

Tertiary modifiers

Particularly now that patients with β-thalassaemia are living longer, it is becoming apparent that variability at loci that have nothing to do with globin chain production might have important phenotypic effects, related particularly to the complications of the disease. Although so far there are only limited data about these tertiary modifying genes, it seems likely that they will become of increasing importance, particularly if the polymorphisms that affect their function are common in populations in which β-thalassaemia occurs at a high frequency.

Bilirubin metabolism. Because of the rapid turnover of red-cell precursors in patients with β-thalassaemia and the resulting breakdown of haem products, many of those with more severe forms of the disease are mildly jaundiced and have a propensity to gallstone formation and gall bladder disease. It has been found that the level of BILIRUBIN in β-thalassaemia heterozygotes is related to a polymorphism in the promoter of the gene that is involved in the hepatic glucuronidation of bilirubin, UDP-glucuronosyltransferase (*UGT1*). In normal individuals, the promoter has a run of six TA repeats $(TA)_6$. Individuals who are homozygous for an additional repeat, $(TA)_7$, tend to have mild hyperbilirubinaemia; β-thalassaemia heterozygotes with the $(TA)_7$ arrangement can have more persistent jaundice[57,58]. Recently, it has been found that the $(TA)_7$ arrangement is extremely common in Sri Lanka and that patients with HbE β-thalassaemia, who are homozygous for $(TA)_7$, have unusually high bilirubin levels and a significantly increased likelihood of developing gallstones (A. Premawardhena, manuscript in preparation).

Iron metabolism. Cardiac disease, hepatic disease and diabetes are important complications of β-thalassaemia that reflect tissue damage from iron loading, not only from transfusion but also from increased intestinal absorption. Although there have been few studies to date, preliminary data indicate that the common mutation of *HFE* that causes hereditary haemochromatosis, C282Y, might be involved in the variability of iron loading in some patients with the intermediate forms of β-thalassaemia[59]. Furthermore, there is recent evidence that the β-thalassaemia trait favours higher rates of iron loading in C282Y homozygotes[60]. However, this mutation is rare in parts of the world in which β-thalassaemia is common and so it will probably have only a small role in iron loading in the more severe forms of β-thalassaemia. By contrast, the *HFE* polymorphism H63D, the functional significance of which is still being evaluated, occurs commonly throughout many of the populations affected by β-thalassaemia[61]. The further study of genetic variability in the rate of iron loading in the thalassaemias will be of considerable importance, because polymorphisms that result in more effective iron absorption are likely to have had a selective value in the past, and because there are now so many candidate genes that are involved in iron homeostasis[62].

Bone disease. Another increasingly common problem in young adults with β-thalassaemia is progressive osteoporosis, associated with bone pain and fractures that might, in part, be related to secondary HYPOGONADISM due to iron-mediated damage to the hypothalamic–pituitary axis. There is increasing evidence that this complication might be modified by polymorphisms at loci that are involved in bone metabolism, including the genes for the vitamin D receptor, collagen, and the oestrogen receptor[62–66].

Phenotypic variation due to co-selection. There is strong evidence that the high frequency of the α-thalassaemias[67,68], and almost certainly the β-thalassaemias[69], is a reflection of heterozygote advantage against *P. falciparum* malaria. The fact that, as shown in FIG. 3, every population has its own unique set of β-thalassaemia mutations indicates that, in evolutionary terms, the exposure of these high-frequency populations to malaria might have been fairly recent. In other words, individual mutations have arisen locally and have been expanded by heterozygote advantage. Recent studies indicate that exposure to malaria has not simply expanded mutations at the haemoglobin loci, but that varying susceptibility to malaria is also reflected by polymorphisms at many other loci, including the major histocompatibility complex loci *HLA-DR* (REF. 70), tumour-necrosis factor-α (*TNF*)[71], intercellular adhesion molecule 1 (*ICAM1*)[72], and others[69]. Just as in the case of the thalassaemias, the malaria-related polymorphisms of these systems vary greatly between different racial groups, again reflecting fairly recent exposure to malaria. Because these systems have an important role in defence mechanisms against many infectious agents, it follows that children with thalassaemia from different parts of the world might have quite different responses to infection, an important complication in this disease[73].

Acquired and environmental factors

Although it is self-evident that environmental factors must have some effect on the phenotype of the thalassaemias, it is not possible in the absence of any twin data or similar studies to calculate the relative importance of genetic compared with environmental factors in modifying the clinical phenotype. Acquired complications, such as progressive enlargement of the spleen, folic-acid deficiency and recurrent infection, can undoubtedly modify the clinical course of β-thalassaemia[1]. Much less is known about the overall role of the environment. Preliminary studies to compare the clinical course of patients with HbE β-thalassaemia, who carry identical β-thalassaemia mutations and who live in tropical countries or more developed Western societies, indicate that the environment might have an important role in modifying the phenotype of this disease[34,74]. The mechanisms involved are not clear. It seems likely that the pattern of infectious illnesses and the frequency of recurrent fever, which might shorten the survival of HbE (REF. 75), might be important factors; the deleterious effect of high body temperatures on HbH, which is the β_4-molecule that reflects defective α-chain synthesis in α-thalassaemia, has been well documented[76].

Finally, there are various ethnological and cultural factors that have an important role in modifying patients' responses to the severe forms of thalassaemia. What little is known about this neglected aspect and the phenotypic diversity of the disease has been summarized recently[1,73].

BILIRUBIN
A principal metabolic product of haemoglobin breakdown.

HYPOGONADISM
Reduction in ovarian or testicular function. This might be primary, due to disease of the ovaries or testes, or secondary due to disease of the hypothalamic–pituitary axis.

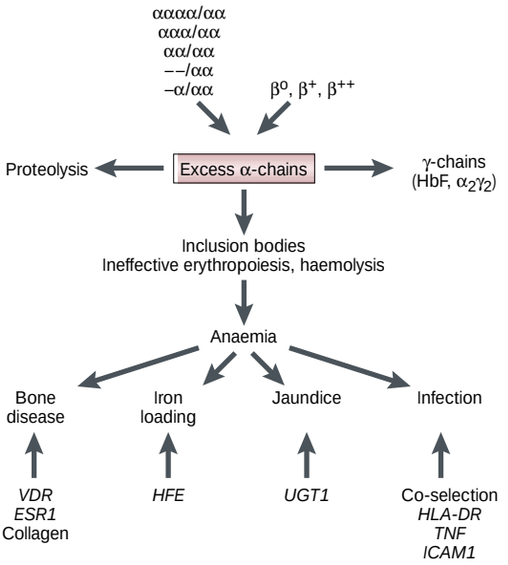


Figure 6 | **A summary of the main genetic mechanisms that contribute to the phenotypic diversity of the β-thalassaemias.** The secondary modifiers, that is the α- and γ-globin genes, are shown at the top of the figure, as they affect the magnitude of the excess of α-chains. The tertiary modifiers are shown at the bottom of the figure: *VDR*, vitamin D receptor; *ESR1*, oestrogen receptor; collagen, several genes determined in collagen synthesis; *HFE*, the locus for hereditary haemochromatosis; *UGT1*, UGT glucuronyltransferase involved in bilirubin glucuronidation; *HLA-DR*, major histocompatibility complex locus; *TNF*, tumour-necrosis factor-α; *ICAM1*, intercellular adhesion molecule 1.

Conclusion

The picture that is emerging is that the phenotypic diversity of the β-thalassaemias is determined by layer upon layer of complexity: a wide variety of primary mutations at the β-globin genes; the action of two well-defined secondary modifying loci; and several less well-characterized tertiary modifiers. It also reflects the effects of co-evolution and the so far neglected but clearly important role of the environment (FIG. 6). Even with this background of knowledge, it is still difficult to give a precise prognosis for a baby with β-thalassaemia, which bears out this extremely complex interaction between the genome and the environment. And so far we only have a limited picture of phenotypic variation that is mediated by individual modifiers; it will be necessary to extend these studies into larger populations to determine how they interact with each other and the environment to produce these remarkably diverse phenotypes.

There is abundant evidence that other monogenic diseases have equally variable clinical pictures, even when they result from the same mutations. There has been less progress in identifying genetic modifiers for these conditions, although in a few cases a pattern is emerging that is similar to the β-thalassaemias[77–79]. The principal lesson from the thalassaemia field is that, before it is possible to start making real sense of phenotypic diversity, it is very important to establish the precise pathophysiological mechanism of the disease. So, although it might be possible to identify modifiers by sophisticated genome searches, without an understanding of the cellular pathology of a disease it will be very difficult to determine the significance of putative loci that are ascertained in this way.

These observations provide some indication of the difficulties that will be encountered when trying to identify the major genes involved in susceptibility to common multifactorial disorders, such as heart disease, hypertension, diabetes, the common psychoses and so on. If this still primitive understanding of the mechanisms for the phenotypic diversity of the β-thalassaemias has told us anything, it is that it will be absolutely vital accurately to define phenotypes and broad pathological mechanisms at the same time as embarking on a search of the genome for susceptibility loci. If this is not done, many candidate genes that are no more than very minor phenotypic modifiers might be mistaken for important players in the basic pathogenesis of these complex diseases.

Finally, it is becoming apparent that, at least from what has been learnt from the human haemoglobin field, a better understanding of the mechanisms of phenotypic modification might provide important information for work directed at the partial or complete correction of monogenic diseases by somatic-cell gene therapy. For example, the observations that the basic pathophysiological mechanism of the anaemia of β-thalassaemia is imbalanced globin production, which leads to abnormal red-cell maturation, and its modification by the action of at least three different gene loci, provides several options for novel therapeutic strategies; β- or γ-globin synthesis could be augmented, or a way of selectively reducing α-globin production could be designed.

Knowledge of the pathophysiology of phenotype–genotype relationships also provides some indication of the degree to which the augmentation of a defective gene might be required to control a particular disease. Because it is unlikely that preliminary efforts to increase β-globin chain production by gene-replacement therapy will result in complete restoration of normal output from the β-globin locus, a central question is how much gene product would be required to control the disease. Although a relatively high output might be required to control the more severe forms of β-thalassaemia, studies of the milder varieties, notably HbE β-thalassaemia, indicate that growth and development might be restored by raising the steady-state haemoglobin level by as little as 1–2 gm dl^{-1} (REFS 1,34). Conversely, a similar result in the case of sickle-cell anaemia would be unlikely to ameliorate the disease and might even exacerbate it; the production of a genetically engineered cell population that raised the haemoglobin level by only a small degree would still leave the number of sickleable cells well above 30% of the red-cell population. At this level of sickleable cells, or above, vaso-occlusive episodes are common[80]; the introduction of the normal cell line would simply increase the viscosity of the blood, further increasing the likelihood of such events.

The type of haemoglobin that is produced by such manipulations, and its intercellular distribution, might be equally important. Because HbF inhibits sickling, the stimulation of its production across the red-cell population might be advantageous, but the introduction of a small population of cells expressing HbF might have the opposite effect — it would raise the haemoglobin level but leave a large population of sickleable cells. Similarly, because HbF has a high oxygen affinity, it is far from clear whether its minor elevation would benefit patients with severe forms of β-thalassaemia intermedia.

In conclusion, a better knowledge of the mechanisms of phenotypic diversity and of pathophysiological mechanisms in monogenic disorders not only will inform efforts to identify the genetic components of complex disease, but also is likely to provide important information for devising somatic gene therapy treatments for monogenic disorders[81].

Links

DATABASE LINKS HbF | β-globin | hereditary persistence of fetal haemoglobin | HbE | α-globin | *UGT1* | *HFE* | vitamin D receptor | oestrogen receptor | *P. falciparum* | *HLA-DR* | *TNF* | *ICAM1* | sickle-cell anaemia

ENCYCLOPEDIA OF LIFE SCIENCES Globin synthesis | Thalassaemias | Selection and common monogenic diseases

1. Weatherall, D. J. & Clegg, J. B. *The Thalassaemia Syndromes* 4th edn (Blackwell Science, Oxford, 2001). **The principal monograph on thalassaemia: the fourth edition covers every aspect of the field and contains over 3,000 references.**
2. Weatherall, D. J. & Clegg, J. B. Thalassaemia — a global public health problem. *Nature Med.* **2**, 847–849 (1996).
3. Weatherall, D. J., Clegg, J. B., Higgs, D. R. & Wood, W. G. in *The Metabolic and Molecular Bases of Inherited Disease* (ed. Scriver, C. R.) 4571–4636 (McGraw–Hill, New York, 2001).
4. Stamatoyannopoulos, G. & Grosveld, F. in *The Molecular Basis of Blood Disease* (eds Stamatoyannopoulos, G., Majerus, P. W., Perimutter, R. M. & Varmus, H.) 135–182 (Saunders, New York, 2001). **An excellent up-to-date review on what is known about the mechanisms of the differential expression of the globin genes during development.**
5. World Health Organization. Guidelines for the control of haemoglobin disorders. Report of the VIth Annual Meeting of the WHO Working Group on Haemoglobinopathies, Cagliari, Sardinia, 8–9 April, 1989 (1994).
6. Huisman, T. H. J., Carver, M. F. H. & Efremov, G. D. *A Syllabus of Human Hemoglobin Variants* (Sickle Cell Foundation, Augusta, Georgia, 1998).
7. Higgs, D. R. in *Baillière's Clinical Haematology. International Practice and Research: The Haemoglobinopathies* (eds Higgs, D. R. & Weatherall, D. J.) 117–150 (Baillière Tindall, London, 1993).
8. Orkin, S. H. *et al.* Abnormal RNA processing due to the exon mutation of β^E-globin gene. *Nature* **300**, 768–769 (1982).
9. Weatherall, D. J. Pathophysiology of thalassaemia. *Clin. Haematol.* **11**, 127–146 (1998).
10. Weatherall, D. J. *et al.* A genetically determined disorder with features both of thalassaemia and congenital dyserythropoietic anaemia. *Br. J. Haematol.* **24**, 681–702 (1973).
11. Stamatoyannopoulos, G., Woodson, R., Papayannopoulou, T., Heywood, D. & Kurachi, M. S. Inclusion-body β-thalassemia trait. A form of β thalassemia producing clinical manifestations in simple heterozygotes. *N. Engl. J. Med.* **290**, 939–943 (1974).
12. Wainscoat, J. S., Thein, S. L. & Weatherall, D. J. Thalassaemia intermedia. *Blood Rev.* **1**, 273–279 (1987).
13. Fiorelli, G., Sampietro, M., Romano, M., Albano, M. & Cappellini, M. D. Clinical features of thalassemia intermedia in Italy. *Birth Defects: Original Articles Ser.* **23**, 287–295 (1988).
14. Cao, A., Gasperini, D., Podda, A. & Galanello, R. Molecular pathology of thalassemia intermedia. *Eur. J. Int. Med.* **1**, 227–236 (1990).
15. Rund, D. *et al.* Genetic analysis of β-thalassemia intermedia in Israel: diversity of mechanisms and unpredictability of phenotype. *Am. J. Hematol.* **54**, 16–22 (1997).
16. Ho, P. J., Hall, G. W., Luo, L. Y., Weatherall, D. J. & Thein, S. L. β thalassaemia intermedia: is it possible to predict phenotype from genotype? *Br. J. Haematol.* **100**, 70–78 (1998). **The most complete analysis so far of the molecular mechanisms for the milder forms of β-thalassaemia.**
17. Huisman, T. H. J., Carver, M. F. H. & Baysal, E. *A Syllabus of Thalassemia Mutations* 1–309 (The Sickle Cell Anemia Foundation, Augusta, Georgia, 1997). **A complete and detailed catalogue of all the published thalassaemia mutations and their geographical distribution up to 1997.**
18. Schwartz, E. The silent carrier of β thalassemia. *N. Engl. J. Med.* **281**, 1327–1333 (1969).
19. Gonzalez-Redondo, J. H. *et al.* A C∅T substitution at nt-101 in a conserved DNA sequence of the promoter region of the β-globin gene is associated with 'silent' β-thalassemia. *Blood* **73**, 1705–1711 (1989).
20. Bianco, I. *et al.* Silent thalassemias: genotypes and phenotypes. *Haematologica* **82**, 269–280 (1997).
21. Gonzalez-Redondo, J. H. *et al.* Clinical and genetic heterogeneity in black patients with homozygous β-thalassemia from the Southeastern United States. *Blood* **72**, 1007–1014 (1988).
22. Tamagnini, G. P., Lopes, M. C., Castanheira, M. E., Wainscoat, J. S. & Wood, W. G. β^+ thalassaemia — Portuguese type: clinical, haematological and molecular studies of a newly defined form of β thalassaemia. *Br. J. Haematol.* **54**, 189–200 (1983).
23. Efremov, D. *et al.* Possible factors influencing the haemoglobin and fetal haemoglobin levels in patients with β-thalassaemia due to a homozygosity for the IVS-1-6 (T-C) mutation. *Br. J. Haematol.* **86**, 824–830 (1994).
24. Thein, S. L. *et al.* Molecular basis for dominantly inherited inclusion body β thalassemia. *Proc. Natl Acad. Sci. USA* **87**, 3924–3928 (1990).
25. Ho, P. J. *et al.* Erythroblastic inclusions in dominantly inherited β thalassaemias. *Blood* **89**, 322–328 (1997).
26. Kazazian, H. H., Dowling, C. E., Hurwitz, R. L., Coleman, M. & Adams, J. G. I. Thalassemia mutations in exon 3 of the β-globin gene often cause a dominant form of thalassemia and show no predilection for malarial-endemic regions of the world. *Am. J. Hum. Genet.* **45**, A242 (1989).
27. Thein, S. L. Dominant β thalassaemia: molecular basis and pathophysiology. *Br. J. Haematol.* **80**, 273–277 (1992).
28. Thein, S. L. Is it dominantly inherited β thalassaemia or just a β-chain variant that is highly unstable? *Br. J. Haematol.* **107**, 12–21 (1999). **An extensive review of the pathophysiology of the dominant β-thalassaemias.**
29. Sachs, A. B. Messenger RNA degradation in eukaryotes. *Cell* **74**, 413–421 (1993).
30. Thermann, R. *et al.* Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* **17**, 3484–3494 (1998).
31. Zhang, J., Sun, X., Qian, Y. & Maquat, L. E. Intron function in the nonsense-mediated decay of β-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA* **4**, 801–815 (1998).
32. Cividalli, G., Kerem, H., Execkiel, E. & Rachmilewitz, E. A. β^o-thalassemia intermedia. *Blood* **52**, 345–349 (1978).
33. Weatherall, D. J. *et al.* The clinical and molecular heterogeneity of the thalassaemia syndromes. *Ann. NY Acad. Sci.* **344**, 83–100 (1980).
34. Rees, D. C., Styles, J., Vichinsky, E. P., Clegg, J. B. & Weatherall, D. J. The hemoglobin E syndromes. *Ann. NY Acad. Sci.* **850**, 334–343 (1998).
35. De Silva, S. *et al.* Thalassaemia in Sri Lanka: implications for the future health burden of Asian populations. *Lancet* **355**, 786–791 (2000).
36. Kan, Y. W. & Nathan, D. G. Mild thalassemia: the result of interactions of α and β thalassemia genes. *J. Clin. Invest.* **49**, 635–642 (1970).
37. Galanello, R. *et al.* Molecular analysis of β^o-thalassemia intermedia in Sardinia. *Blood* **74**, 823–827 (1989).
38. Weatherall, D. J., Pressley, L., Wood, W. G., Higgs, D. R. & Clegg, J. B. The molecular basis for mild forms of homozygous β thalassaemia. *Lancet* **i**, 527–529 (1981).
39. Knox-Macaulay, H. H. M., Weatherall, D. J., Clegg, J. B., Bradley, J. & Brown, M. J. Clinical and biosynthetic characterization of αβ-thalassaemia. *Br. J. Haematol.* **22**, 497–512 (1972).
40. Boyer, S. H., Belding, T. K., Margolet, L. & Noyes, A. N. Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. *Science* **188**, 361–363 (1975).
41. Rees, D. C., Porter, J. B., Clegg, J. B. & Weatherall, D. J. Why are hemoglobin F levels increased in Hb E/β thalassemia? *Blood* **94**, 3199–3204 (1999).
42. Garner, C. *et al.* Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood* **95**, 342–346 (2000). **The most extensive study so far on the genetic factors that modify the small amounts of fetal haemaglobin that are produced in normal adults.**
43. Gilman, J. G. & Huisman, T. H. J. DNA sequence variation associated with elevated fetal $^G\gamma$ globin production. *Blood* **66**, 783–787 (1985).
44. Labie, D. *et al.* Common haplotype dependency of high $^G\gamma$-globin gene expression and high Hb F levels in β-thalassemia and sickle cell anemia patients. *Proc. Natl Acad. Sci. USA* **82**, 2111–2114 (1985).
45. Thein, S. L. *et al.* Association of thalassaemia intermedia with a β-globin gene haplotype. *Br. J. Haematol.* **65**, 370–373 (1987).
46. Thein, S. L. in *Baillière's Clinical Haematology. International Practice and Research: The Haemoglobinopathies* (eds Higgs, D. R. & Weatherall, D. J.) 151–176 (Baillière Tindall, London, 1993).
47. Thein, S. L. & Weatherall, D. J. in *Hemoglobin Switching. B. Cellular and Molecular Mechanisms* (eds Stamatoyannopoulos, G. & Nienhuis, A. W.) 97–112 (Alan R. Liss, New York, 1989).
48. Craig, J. E. *et al.* Haemoglobin switch: dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nature Genet.* **12**, 58–64 (1996). **A model approach to dissecting the loci that modify fetal haemaglobin production in the haemoglobin disorders.**
49. Craig, J. E. *et al.* Genetic heterogeneity in heterocellular hereditary persistence of fetal hemoglobin. *Blood* **90**, 428–434 (1997).
50. Dover, G. J. *et al.* 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 (1992).
51. Higgs, D. R., Old, J. M., Pressley, L., Clegg, J. B. & Weatherall, D. J. A novel α-globin gene arrangement in man. *Nature* **284**, 632–635 (1980).
52. Goossens, M. *et al.* Triplicated α-globin loci in humans. *Proc. Natl Acad. Sci. USA* **77**, 518–521 (1980).
53. Kanavakis, E., Metaxatou-Mavromati, A., Kattamis, C., Wainscoat, J. S. & Wood, W. G. The triplicated α gene locus and β thalassaemia. *Br. J. Haematol.* **54**, 201–207 (1983).
54. Galanello, R. *et al.* A family with segregating triplicated α globin loci and β thalassemia. *Blood* **62**, 1035–1040 (1983).
55. Thompson, C. C., Ali, M. A. & Vacovsky, M. The interaction of anti 3.7 type quadruplicated α-globin genes and heterozygous β-thalassemia. *Hemoglobin* **13**, 125–135 (1989).

56. Beris, P. *et al.* Severe inclusion body β-thalassaemia with haemolysis in a patient double heterozygous for β^o-thalassaemia and quadruplicated α-globin gene arrangement of the anti-4.2 type. *Br. J. Haematol.* **105**, 1074–1080 (1999).
57. Galanello, R. *et al.* Hyperbilirubinaemia in heterozygous β-thalassaemia is related to co-inherited Gilbert's syndrome. *Br. J. Haematol.* **99**, 433–436 (1997).
58. Sampietro, M. *et al.* The expression of uridine diphosphate glucuronosyltransferase gene is a major determinant of bilirubin level in heterozygous β-thalassaemia and in glucose-6-phosphate dehydrogenase deficiency. *Br. J. Haematol.* **99**, 437–439 (1997).
59. Rees, D. C., Singh, B. M., Luo, L. Y., Wickramasinghe, S. & Thein, S. L. Nontransfusional iron overload in thalassemia: association with hereditary hemochromatosis. *Ann. NY Acad. Sci.* **850**, 490–494 (1998).
60. Piperno, A. *et al.* Haemochromatosis in patients with β-thalassaemia trait. *Br. J. Haematol.* **111**, 908–914 (2000).
61. Merryweather-Clarke, A. T., Pointon, J. J., Shearman, J. D. & Robson, K. J. H. Global prevalence of putative haemochromatosis mutations. *J. Med. Genet.* **34**, 275–278 (1997).
62. Andrews, N. C. Iron homeostasis: insights from genetics and animal models. *Nature Rev. Genet.* **1**, 208–217 (2000).
An extensive review of new insights into iron homeostasis that discusses the many gene candidates for modifiers of iron loading.
63. Rees, D. C. *et al.* Genetic influences on bone disease in thalassemia. *Blood* **92**, S532a (1998).
64. Wonke, B. Annotation: bone disease in β-thalassaemia major. *Br. J. Haematol.* **103**, 897–901 (1998).
An up-to-date outline of the acquired and potential genetic factors that might modify osteoporosis in β-thalassaemia.
65. Perrota, S. *et al.* Osteoporosis in β thalassaemia major patients: analysis of the genetic background. *Br. J. Haematol.* **111**, 461–466 (2000).
66. Dresner Pollak, R., Rachmilewitz, E., Blumenfeld, A., Idelson, M. & Goldfarb, A. W. Bone mineral metabolism in adults with β-thalassaemia major and intermedia. *Br. J. Haematol.* **111**, 902–907 (2000).
67. Williams, T. N. *et al.* High incidence of malaria in α-thalassaemic children. *Nature* **383**, 522–525 (1996).
68. Allen, S. J. *et al.* α^+-thalassaemia protects children against disease due to malaria and other infections. *Proc. Natl Acad. Sci. USA* **94**, 14736–14741 (1997).
69. Weatherall, D. J., Clegg, J. B. & Kwiatkowski, D. The role of genomics in studying genetic susceptibility to infectious disease. *Genome Res.* **7**, 967–973 (1997).
An extensive review of the malaria-related genetic polymorphisms and polymorphisms that alter susceptibility to other infectious agents.
70. Hill, A. V. S., Allsopp, C. E. M. & Kwiatkowski, D. Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595–600 (1991).
71. McGuire, W., Hill, A. V. S., Allsopp, C. E. M., Greenwood, B. M. & Kwiatkowski, D. Variation in the TNF-α promoter region associated with susceptibility to cerebral malaria. *Nature* **371**, 508–511 (1994).
72. Fernandez-Reyes, D. *et al.* A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum. Mol. Genet.* **6**, 1357–1360 (1997).
73. Weatherall, D. J. Thalassemia in the next millennium. *Ann. NY Acad. Sci.* **850**, 1–9 (1998).
74. Fouladi, M. *et al.* Hemoglobin E/β thalassemia: the Canadian experience. *Ann. NY Acad. Sci.* **850**, 410–411 (1998).
75. Rees, D. C., Clegg, J. B. & Weatherall, D. J. Is hemoglobin instability important in the interaction between hemoglobin E and β thalassemia? *Blood* **92**, 2141–2146 (1998).
76. Chinprasertsuk, S., Wanachiwanawin, W. & Piankijagum, A. Effect of pyrexia in the formation of intraerythrocytic inclusion bodies and vacuoles in haemolytic crisis of haemoglobin H disease. *Eur. J. Haematol.* **52**, 87–91 (1994).
77. Weatherall, D. J. From genotype to phenotype: genetics and medical practice in the new millennium. *Phil. Trans. R. Soc. Lond.* B **354**, 1995–2010 (1999).
An extensive review of phenotype–genotype relationships in monogenic disease, including those other than the haemoglobin disorders.
78. Summers, K. M. Relationship between genotype and phenotype in monogenic diseases: relevance to polygenic diseases. *Hum. Mutat.* **7**, 283–293 (1996).
79. Wolf, U. Identical mutations and phenotypic variation. *Hum. Genet.* **100**, 305–321 (1997).
A valuable review of the phenotype–genotype relationships for monogenic disease.
80. Bunn, H. F. Pathogenesis and treatment of sickle cell disease. *N. Engl. J. Med.* **337**, 762–769 (1997).
81. Weatherall, D. J. Gene therapy: repairing haemoglobin disorders with ribozymes. *Curr. Biol.* **8**, R696–R698 (1998).

Acknowledgements

This work was supported by the Medical Research Council (MRC) and the Wellcome Trust. I thank my former colleagues in the MRC Molecular Haematology Unit for their help and support.