
From genotype to phenotype: genetics and medical practice in the new millennium

Sir David Weatherall, FRS

Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK

The completion of the human genome project will provide a vast amount of information about human genetic diversity. One of the major challenges for the medical sciences will be to relate genotype to phenotype. Over recent years considerable progress has been made in relating the molecular pathology of monogenic diseases to the associated clinical phenotypes. Studies of the inherited disorders of haemoglobin, notably the thalassaemias, have shown how even in these, the simplest of monogenic diseases, there is remarkable complexity with respect to their phenotypic expression. Although studies of other monogenic diseases are less far advanced, it is clear that the same level of complexity will exist. This information provides some indication of the difficulties that will be met when trying to define the genes that are involved in common multigenic disorders and, in particular, in trying to relate disease phenotypes to the complex interactions between many genes and multiple environmental factors.

Keywords: phenotype–genotype; thalassaemia; haemoglobinopathies; malaria hypothesis; population genetics

1. INTRODUCTION

Following the revolution in the biological sciences resulting from the development of recombinant DNA technology, the role of genetics in the pathogenesis of human disease is dominating biomedical research at the turn of the century. With the promise of the completion of the human genome project in a few years time, and the widespread belief that its fruits have the capacity to completely change medical practice in the future, it seems likely that genetics will continue to play a central role in medical research and practice in the new millennium.

A great deal of progress has already been made in the application of molecular and cell biology to the study of disease. Through a combination of the analysis of candidate genes and defining disease loci by positional cloning, the molecular basis of many monogenic diseases has been determined. Work of this kind has already found wide clinical application in genetic counselling and prenatal diagnosis. The study of monogenic disease at the molecular level, together with major improvements in cytogenetics, has enabled a start to be made in understanding the basis for at least some forms of congenital malformation and mental retardation. The amalgamation of studies of the structure and function of oncogenes, also aided by improvements in cytogenetic analysis, has yielded hitherto undreamt of insights into the mechanisms of malignant transformation. And, again by the use of candidate genes—in this case supported by an increasing ability to identify loci by genome searching—it has been possible to at least make a tentative start towards the identification of some of the genes that may play an important role in some of the common, refractory diseases of the richer

societies, such as coronary artery disease, stroke, diabetes, psychiatric disorders and others.

There is every sign that the rapidly evolving technology that will become available in the post-genome era will facilitate an analysis of the function of the human genome. Automated sequencing methods will continue to improve, and microchip technology will allow the study of genetic variability in large populations, as well as the rapid analysis of gene expression during development and in response to changing physiological and pathological conditions. Similarly, improvements in structural biology should enable the more rapid and effective definition of the function of particular gene products, and the development of sophisticated mathematical modelling and computer programming will undoubtedly help us to start to understand how their action is orchestrated, both in individual cells and in entire organisms.

The central question for the medical sciences, however, is, given the undoubted complexity of the interactions of the 50 000 to 100 000 genes that make up the human genome, both between themselves and with the environment, to what extent will it be possible to relate events at the molecular level with the clinical phenotypes of patients with particular diseases? This problem will permeate every aspect of medical research and practice in the future. Not only will it dominate predictive genetics and genetic counselling, but it will also be of major importance for clinical decision making as new and novel approaches to the treatment of disease become available, particularly if they involve genetic manipulation. Although some of these problems may be solved by mathematical modelling, they will also require careful analysis of the relationship between the molecular pathology of disease and its variation in clinical phenotype in large numbers of patients.

Theoretically, monogenic diseases should be the simplest models for asking to what extent it is possible to predict phenotypes from genotypes. Focusing mainly on the genetic disorders of haemoglobin, since they are the commonest monogenic diseases in man and were the first to be defined at the molecular level, I will try to assess how much progress has been made in our ability to predict phenotype from genotype. I will then briefly review what progress has been made in understanding the clinical diversity of some other monogenic diseases through this route and, finally, try to assess what this information has told us about the problems we may encounter when we attempt to relate genotype to phenotype in the case of common multigenic diseases, that is disorders which probably reflect the variable activity of a number of different genes together with the interplay of numerous environmental factors.

2. PHENOTYPE-GENOTYPE RELATIONSHIPS FOR THE INHERITED DISORDERS OF HAEMOGLOBIN

The inherited disorders of haemoglobin are the commonest monogenic diseases; it is estimated that some 6–7% of the world's population are carriers (Weatherall *et al.* 1999). They fall naturally into two groups, the structural haemoglobin variants and the thalassaemias, that is disorders of the rate of production of the globin chains of haemoglobin.

The structure of human haemoglobin (Hb) changes during embryonic, foetal and adult life. All the normal haemoglobins are tetramers of two pairs of unlike globin chains. Adult and foetal haemoglobins have α -chains combined with β - (Hb A, $\alpha_2\beta_2$), δ - (Hb A₂, $\alpha_2\delta_2$) or γ -chains (Hb F, $\alpha_2\gamma_2$), 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, $\zeta_2\epsilon_2$), and α - and ϵ -chains form Hb Gower 2 ($\alpha_2\epsilon_2$). Embryonic haemoglobin production is confined to the yolk sac stage of development and thereafter is replaced by foetal haemoglobin synthesis up to shortly before term. After birth, Hb F is replaced by Hbs A and A₂ over the first year of life, though in normal adults small amounts of Hb F, constituting approximately 1% of the total haemoglobin, continue to be produced.

Although over 400 structural haemoglobin variants have been identified, only three, Hbs S, C and E reach polymorphic frequencies. Haemoglobin S, which in the homozygous state results in sickle-cell anaemia, is a major cause of disability throughout sub-Saharan Africa, the Mediterranean, the Middle East and parts of India. Haemoglobin C is less important because it is responsible for only mild forms of anaemia. Haemoglobin E, which occurs at very high frequencies in parts of India, Myanmar and throughout South-East Asia, because it behaves like a mild form of thalassaemia, is of major public health importance. The thalassaemias are classified into the α , β , $\delta\beta$ and $\gamma\delta\beta$ thalassaemias, based on the particular globin chains that are ineffectively synthesized. The discussions that follow are largely confined to the β and α thalassaemias since, clinically, they are the most important and by far the most common varieties.

Table 1. *The phenotypic heterogeneity of β thalassaemia*

homozygous or compound heterozygous states
β thalassaemia major; profound anaemia requiring life-long blood transfusion
β thalassaemia intermedia; moderate to mild anaemia; not transfusion-dependent
heterozygous states
β thalassaemia trait; mild anaemia
'silent' β thalassaemia; phenotypically normal
dominant β thalassaemia; moderate to severe anaemia

(a) *Mechanisms for the phenotypic diversity of the β thalassaemias*

(i) *Pathophysiology and clinical heterogeneity*

The hallmark of all the β thalassaemias is defective β -globin chain synthesis, which leads to imbalanced globin chain production and an excess of α -globin chains (Weatherall *et al.* 1965). The latter aggregate in red cell precursors and result in their abnormal maturation and premature destruction in the bone marrow. Red cells that survive to reach the peripheral blood have a markedly shortened survival. There is abundant evidence that the severity of the β thalassaemias is related to the degree of globin chain imbalance (Weatherall & Clegg 1999); the mechanisms whereby α -globin chain precipitation damages the red cell precursors and red cells has been reviewed recently (Weatherall 1998a). The overall result of the consequences of imbalanced globin chain synthesis is a profound degree of anaemia, which stimulates erythropoietin production which, in turn, leads to intense proliferation and expansion of the bone marrow. In normal adults some red cell precursors continue to synthesize the γ -chains of foetal haemoglobin, albeit at a very low level. In the face of imbalanced globin chain synthesis these cells come under intense selection because part of the excess of α -chains are bound to γ -chains to produce Hb F, thus reducing the magnitude of α -chain precipitation. Hence in all the severe forms of β thalassaemia there is a relatively high level of Hb F in the red cells.

Despite these well-defined consequences of defective β -globin chain production the clinical consequences are remarkably diverse (table 1). At their worst, the homozygous and compound heterozygous states for β thalassaemia are characterized by profound anaemia from the second or third month of life which, if not treated with blood transfusion, leads to death within the first two years, a disorder known as β thalassaemia major. On the other hand, some patients with apparently the same genotype have a milder illness. This ranges from a condition that is only a little less severe than the major form, through a spectrum of increasing haemoglobin levels, to one in which there are no symptoms and which is often ascertained on routine examination of the blood. This rather diverse collection of β thalassaemias of varying severity are called the β thalassaemia intermedias. Remarkably, the heterozygous states for β thalassaemia show equal phenotypic diversity. Typically, the β -thalassaemia trait, that is the inheritance of a single β -thalassaemia allele, is associated with extremely mild anaemia and morphological changes of the red cells.

However, in some cases it may be completely silent; that is, there is no anaemia or haematological abnormality. On the other hand, it may be almost as severe as the major forms of the illness. In other words there is also a dominantly inherited form of severe β thalassaemia.

Over the past 15 years, by a combination of analysis of the molecular pathology of different forms of thalassaemia, extensive family data collection and studies of large numbers of patients with intermediate forms of the disease, it has been possible to determine at least some of the mechanisms that underlie this remarkable phenotypic heterogeneity. But as we shall see, although a great deal of progress has been made many questions still remain unanswered. Perhaps the most surprising outcome of this work is that, although heterogeneity of the mutations of the β -globin gene locus that underlie the β thalassaemias can account for part of its phenotypic variability, a great deal cannot be explained in this way and it is clear that a number of other gene loci are involved.

Variable severity of β thalassaemia alleles

Close on 180 different mutations have been identified in the β -globin genes of patients with β thalassaemia (reviewed by Huisman *et al.* (1997) and Clegg & Weatherall (1999)). With the exception of a few deletions, the bulk are made up of point mutations or loss of one or two bases, which interfere with gene action either at the transcriptional, translational or post-translational levels. The resulting phenotypes are divided into the β^0 thalassaemias, in which there is no β -globin gene product, and the β^+ thalassaemias, in which there is a variable reduction in the output of β chains.

Some of the clinical heterogeneity of the β thalassaemias can be explained by the differing severity of particular alleles (figure 1). In the β^0 thalassaemias, which are mainly due to deletions, point mutations at the invariant intron-exon junction boundaries and nonsense or frameshift mutations, there is no production of β -globin chains. On the other hand, the β^+ thalassaemias, which in the main result from mutations of the promoters or which activate cryptic splice sites within exons or introns, show a wide spectrum of defective β -chain production. While in the case of the more severe alleles the level of β -globin chain production is little more than in β^0 thalassaemia, there are milder forms, usually due to the promoter mutations, in which there is only a moderate decrease in β -globin chain output. Thus some of the clinical heterogeneity of this disease in different populations can be explained by homozygosity or compound heterozygosity for these milder alleles. However, as illustrated in figure 2, in all the high-frequency populations of the world there are usually only one or at the most two mild alleles. Thus many of the milder forms of β thalassaemia can not be explained on this basis.

Furthermore, it has become apparent that there is marked phenotypic heterogeneity of the β thalassaemias, even in cases of homozygosity or compound heterozygosity for what are usually mild alleles. For example, in the Mediterranean region there is only one common mild allele, β IVS1-6-T \rightarrow C, that is a single base substitution at the sixth position of the first intervening sequence of the β -globin gene, which generates an alternate splice site. In the

homozygous state this usually produces a mild form of β thalassaemia which is not transfusion dependent and which is compatible with survival into adult life (Tamagnini *et al.* 1983; Efremov *et al.* 1994). Even within this fairly uniform phenotype many exceptions are observed; some patients with this genotype in Israel are much more severely affected and are transfusion dependent (Rund *et al.* 1997).

In southern and South-East Asian populations there is also only one common mild β -thalassaemia allele. In this case a GAG \rightarrow AAG mutation at codon 26 results in the production of an abnormal haemoglobin, Hb E, the synthesis of which is reduced because the output of β mRNA is decreased due to the generation of a cryptic alternative splice site in exon 1 of the β -globin gene. The fact that the output of β^E -globin is only slightly reduced is evidenced by the finding that homozygotes have an extremely mild form of anaemia. However, when this variant is inherited with a β^0 - or severe β^+ -thalassaemia allele, a condition called Hb E β thalassaemia, there may be profound anaemia with lifelong transfusion requirements. However, the clinical spectrum of this disorder is remarkably diverse, ranging from lethal anaemia in the first year or two of life to a condition which is so mild as to be encountered only by chance haematological examination. Recent studies have shown that this remarkable degree of phenotypic heterogeneity is encountered in cases with identical β -thalassaemia mutations, both within the same population and in different Asian populations (Rees *et al.* 1998a).

But perhaps the most striking example of the phenotypic heterogeneity of the β thalassaemias is the observation that not every homozygote or compound heterozygote for β^0 thalassaemia, in which there is no output of β chains at all, is severely affected (Knox-Macaulay *et al.* 1973; Weatherall *et al.* 1980). There are now many well-characterized examples of patients who are affected in this way who have mild, intermediate forms of β thalassaemia. Nor is this phenomenon restricted to those with particular β^0 -thalassaemia mutations, or to any racial group.

It is clear, therefore, that there is remarkable clinical diversity even among individuals who have inherited the same β -thalassaemia alleles. As well as being observed in individual cases, it has been well-documented among sibships, who have inherited not only identical β -thalassaemia mutations but, of course, the same β -globin gene clusters, an observation which suggests that whatever is responsible for this heterogeneity does not segregate within or near to the β -globin genes (Ho *et al.* 1998a).

Thus although some of the phenotypic heterogeneity of the β thalassaemias can be ascribed to interactions of β -thalassaemia alleles of different severity, it is clear that there are other genetic modifiers involved which are not related directly to the β -globin genes themselves. A number of loci have been identified that have this effect. First, there are the α -globin genes and, second, there appear to be several loci, which may or may not be linked to the β -globin gene cluster, which have the effect of augmenting Hb F production in patients with β thalassaemia.

(ii) *The co-inheritance of α thalassaemia*

Although at first sight it might not seem obvious that it would be better to inherit two varieties of thalassaemia

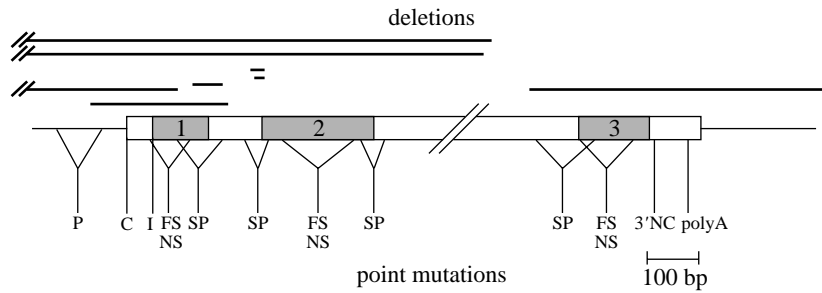


Figure 1. Summary of the main classes of mutation that underlie β thalassaemia. P, promoter; C, CAP site; I, initiation codon; FS, frameshift; NS, nonsense; SP, splice site. The unshaded boxes show the intervening sequences (IVS).

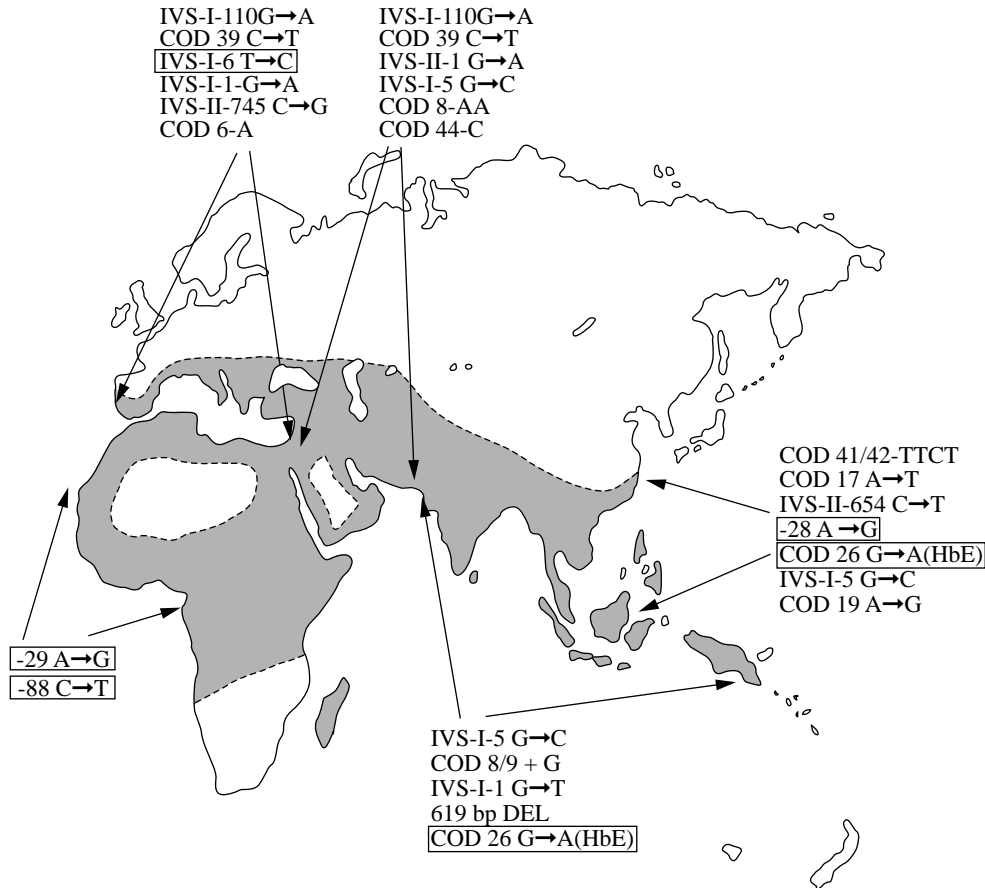


Figure 2. The global distribution of the different mutations that underlie β thalassaemia. Those shown in boxes are mild mutations.

instead of one, this is often the case. Indeed there is very strong evidence from a variety of ‘experiments of nature’ of this kind that the co-inheritance of α thalassaemia may have an important ameliorating effect on the phenotype of otherwise severe β thalassaemia.

Normal individuals have two α -globin genes per haploid genome and their genotype can be written $\alpha\alpha/\alpha\alpha$. There are two major varieties of α thalassaemia, α^+ and α^0 thalassaemia (Higgs 1993). In the α^+ thalassaemias one of the linked α -globin genes is lost by deletion, $-\alpha/\alpha\alpha$, or inactivated by a point mutation of a similar kind that results in β thalassaemia, $\alpha^T\alpha/\alpha\alpha$. In the α^0 thalassaemias both of the linked α -globin genes are lost, most commonly by deletions which involve part or all of the α -globin gene cluster; the heterozygous genotype is expressed as $-/\alpha\alpha$. Since α thalassaemia coexists with β thalassaemia at a high frequency in many populations,

it is not uncommon for individuals to inherit both types of the disease. Thus a homozygote or compound heterozygote for severe β -thalassaemia alleles may also be a heterozygote or a homozygote for α^+ thalassaemia, or a heterozygote for α^0 thalassaemia.

Because imbalanced globin chain synthesis is the major pathophysiological mechanism in β thalassaemia, resulting in the deleterious effects of excess α -chain production on red cell maturation and survival, it might be expected that the co-inheritance of α thalassaemia would have some beneficial effect on the more severe forms of β thalassaemia (figure 3). This is true to some extent; the co-inheritance of different α thalassaemia alleles may modify the homozygous or compound heterozygous states for β^0 thalassaemia, at least to some extent, and may convert those for the severe form of β^+ thalassaemia into milder, non-transfusion-dependent disorders. The many

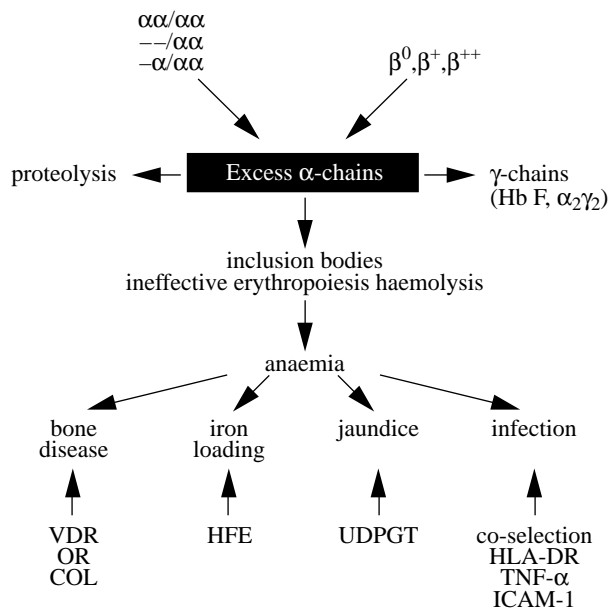


Figure 3. The factors that modify the phenotype of β thalassaemia. VDR, vitamin D receptor; OR, oestrogen receptor; COL, genes involved in collagen structure; HFE, hereditary haemochromatosis locus; UDPGT, bilirubin UDP-glucuronyl transferase; TNF, tumour necrosis factor; ICAM-1, intercellular adhesion molecule.

different interactions of this type that have been well characterized are reviewed by Weatherall & Clegg (1999). They are a particularly common cause of mild β -thalassaemia phenotypes in populations in which α thalassaemia is common, notably in the Mediterranean and South-East Asia. Since several hundred interactions of different α -thalassaemia alleles are possible (Higgs 1993), it follows that this mechanism alone offers the possibility of producing a broad range of β -thalassaemia phenotypes.

(iii) Genetic variability in the postnatal production of foetal haemoglobin

As soon as it was clear that some patients who were β^0 -thalassaemia homozygotes had a mild clinical phenotype, yet were able to maintain relatively high haemoglobin levels, and since all this haemoglobin was Hb F, it was clear that an unusual propensity for the production of foetal haemoglobin must be a factor in modifying the clinical course of β thalassaemia (Weatherall *et al.* 1980). Although it is now clear that this may be genetically determined, and that a number of loci are probably involved, our knowledge of the genetic basis for this important interaction is still rather limited.

There is good evidence that several genetic loci can modify γ -chain production during postnatal life, both in otherwise normal individuals and in β thalassaemics. While the clearest case has been made for the action of those that are not linked to the β -globin gene cluster, there is also evidence that loci that are encoded within the cluster may have this effect.

Determinants not linked to the β gene cluster

It is well established that some normal individuals produce slightly increased levels of foetal haemoglobin and that this is genetically determined. This condition, some-

times called the heterocellular form of hereditary persistence of foetal haemoglobin (HPFH) because the foetal haemoglobin, like that in normal people, is unevenly distributed among the red cells, is undoubtedly heterogeneous, but, to date, the genes that are involved have not been characterized. However, if families with patients with milder forms of β thalassaemia due to increased Hb F production are studied in detail, it is sometimes possible to observe unusually high levels of Hb F in one of the parents, or to find one or more unaffected relatives with slightly increased levels of Hb F. Work over many years in the author's laboratory, studying a very large kindred in which this condition segregates independently from the β -globin gene cluster, has assigned the locus involved to chromosome 6 (Thein & Weatherall 1989; Craig *et al.* 1996). However, studies of similar families indicate quite clearly that there are other forms of heterocellular HPFH which are not linked to chromosome 6, though they also interact to ameliorate β thalassaemia (Craig *et al.* 1997). There is a locus on the X chromosome which appears to have some effect on the level of Hb F in normal adults, although its role in determining the level in thalassaemia is not yet clear (Dover *et al.* 1992). But what these studies have told us is that there are a number of gene loci that are not linked to the β -globin gene complex that can fine-tune the level of Hb F, both in normal adults and in those with β thalassaemia. Presumably, at least some of these genes encode transcription factors that are involved with the activation or repression of γ -chain synthesis.

Determinants within the β -globin gene cluster

There are other forms of HPFH, which result from deletions involving the β -globin gene cluster or point mutations in the promoter regions of one or other of the duplicated γ -globin genes. However, these conditions are rare and, numerically, play a very small part in modifying the genotype of β thalassaemia. There is, however, a relatively common polymorphism at position -158 to the $\epsilon\gamma$ gene, which involves a C \rightarrow T change (Gilman & Huisman 1985). Although this seems to have little effect in normal individuals, there is reasonably good evidence that, at least in homozygotes, there may be an increased ability to produce Hb F under conditions of haemopoietic stress, and that this may have the effect of raising foetal haemoglobin levels in different forms of β thalassaemia (Labie *et al.* 1985; Thein *et al.* 1987; Galanello *et al.* 1989). While there is circumstantial evidence that there may be other factors encoded by the β -globin gene cluster that can modify the level of Hb F in response to haemopoietic stress, so far, and despite a great deal of work, they have not been identified.

Finally, there is increasing evidence that some β -thalassaemia alleles may themselves favour a higher output of Hb F. This is certainly true in the case of promoter mutations or deletions which involve this region of the β -globin gene. This interesting phenomenon may reflect the competition between the γ - and β -promoters for rate-limiting regulatory proteins, particularly in adult life.

(iv) The phenotypic heterogeneity of heterozygous β thalassaemia

It is now clear that the 'silent' carrier states for β thalassaemia are due to mild alleles, usually involving the

promoter regions of the β -globin genes, which cause only a modest down-regulation of β -globin chain synthesis (reviewed by Weatherall & Clegg 1999).

The dominant forms of β thalassaemia are characterized by moderate anaemia, enlargement of the spleen and a thalassaemic blood picture, together with large inclusion bodies in the red cell precursors, similar to those seen in homozygotes (Weatherall *et al.* 1973). These disorders, too, can be related in part to particular mutations of the β -globin gene. Many of the recessive forms of β thalassaemia result from frameshift or nonsense mutations involving exons 1 or 2, or splice junction mutations in the same regions. In these cases, very little β -globin mRNA reaches the cell cytoplasm, a phenomenon known as nonsense-mediated decay. However, in the case of mutations which involve exon 3, full-length mRNA is usually produced; it turns out that the majority of the dominant forms of β thalassaemia are caused by mutations in this region, and result in the synthesis of truncated or elongated and highly unstable β -globin chains (Thein *et al.* 1990). These co-precipitate with excess α -chains to produce the inclusion bodies characteristic of this disorder. In short, the change from recessiveness to dominance is, to a large part, a reflection of the relative position of mutations along the β -globin gene; those that involve exons 1 and 2 are associated with recessive inheritance, while those that affect predominantly exon 3 result in a dominant form of inheritance.

To complicate matters still further, severe forms of heterozygous β thalassaemia are not always dominantly inherited. In some families there is a mixture of heterozygous phenotypes, some severe and some of the common, mild variety. The molecular basis for this surprising observation is now clear. It turns out that triplicated or even quadruplicated α -globin gene arrangements are found in most populations (reviewed by Weatherall & Clegg 1999). If they are inherited with a single β -thalassaemia allele, because each of the α -genes is expressed, they are able to generate sufficient globin chain imbalance to cause a moderately severe β -thalassaemia phenotype. And because the α - and β -globin genes are encoded on different chromosomes, it follows that the triplicated α -globin gene arrangements segregate independently from the β -thalassaemia alleles; hence in any particular family some β -thalassaemia heterozygotes will receive the triplicated α -globin gene arrangement and some will not.

Recently it has been found that the heterozygous state for the splice mutation, β IVS-2 654 C \rightarrow T, may result in either a dominant or recessive form of inheritance (Ho *et al.* 1998b). The mechanism has not yet been elucidated.

(v) *Phenotypic modification of other common genetic disorders of haemoglobin*

Much less progress has been made towards an understanding of the factors that can modify the phenotype of other common haemoglobin disorders. In the case of sickle-cell anaemia, the co-inheritance of α thalassaemia may modify the severity to some degree, while there is abundant evidence that unusually high levels of foetal haemoglobin, because Hb F interferes with the sickling process, may have a quite remarkable ameliorating effect. The genetic factors that are involved in the latter response

seem to be very similar to those that occur in β thalassaemia (reviewed by Weatherall & Clegg 1999).

Much of the phenotypic heterogeneity of the α thalassaemias results from variability of the primary mutations (Higgs 1993). As mentioned earlier, there are four α -globin genes in normal individuals, $\alpha\alpha/\alpha\alpha$; a variety of conditions of varying severity result from deletions which remove one, two, three or all of the α -genes. This is an excellent example of how a gene dosage effect can result in widely variable phenotypes. But perhaps the most striking examples of phenotypic variability in the α thalassaemias are seen on those forms associated with mental retardation (Weatherall *et al.* 1981). Unlike the common genetic forms of α thalassaemia, which are found in regions in which malaria was, or still is, endemic, these conditions occur in European populations and are not related to the tropics.

There are two distinct syndromes of this kind, one of which is encoded on chromosome 16, the other on the X chromosome (Wilkie *et al.* 1990a,b). In the first, ATR-16, there is usually mild mental retardation associated with large deletions or other chromosomal rearrangements that remove many genes, including the α -globin genes, from the short arm of chromosome 16. This, therefore, is an example of a contiguous gene syndrome and the variable phenotype is likely to result from the loss or gain of other genetic material on the affected chromosome. In the second, ATR-X, there is a complex phenotype characterized by severe mental retardation, a characteristic constellation of congenital anomalies, and a very mild form of α thalassaemia. It is due to different mutations that involve a gene on the X chromosome, *ATR-X*, which encodes for a transcription factor that is a member of the DNA helicase family (Gibbons *et al.* 1995). It seems likely that the complex genotype in this condition reflects the action of an altered transcription factor which is involved in early development in a number of tissues and organs, which also happens to be involved in transcription of the α -globin genes. This is the first example of heterogeneity of a globin disorder due to a mutation at a locus for a transcription factor. There are a number of disorders due to mutations at loci of this type, all of which are associated with complex phenotypes, reflecting the ubiquitous actions of these regulatory molecules during development.

(vi) *Secondary genetic modifiers*

So far we have only considered the action of genetic variability at two other globin gene loci which may modify the overall degree of globin chain imbalance in β thalassaemia and hence modify its phenotype. However, particularly now that patients with β thalassaemia are living longer, it is becoming apparent that genetic variability at loci which have nothing to do with globin chain production may have important phenotypic effects, related in particular to some of the complications of the disease. Although so far there is only limited information about these modifying genes, it seems likely that they will become of increasing importance, particularly as more thalassaemic patients survive to adult life.

Because of the rapid turnover of red cell precursors in thalassaemia, and the resulting breakdown of haem products, it is not uncommon for thalassaemic patients to

be mildly jaundiced. However, it has been known for a long time that there is a subset, which includes heterozygotes, who have unusually high bilirubin levels. Recently, evidence has been obtained that this results from a polymorphism in the promoter of the gene for bilirubin, UDP-glucuronosyltransferase (UGT-1). It appears that normal individuals may carry these mutations without being jaundiced. But in the presence of an increased rate of bilirubin turnover—as occurs in β thalassaemia—they may result in unusually high levels of bilirubin, which may lead, in some cases, to deep and persistent jaundice (Galanello *et al.* 1997; Sampietro *et al.* 1997).

Two other important complications of thalassaemia involve tissue damage from the effects of iron loading, not just from transfusion but from increased absorption, and bone disease, which has been ascribed to the results of both bone marrow expansion and secondary hypogonadism due to iron loading of the hypothalamic-pituitary axis. Particularly in the case of β -thalassaemic patients who are not transfusion dependent, there is a remarkable variability in the rate of iron loading. Although there have been few studies to date, preliminary data suggest that polymorphisms of the recently discovered gene for hereditary haemochromatosis may be involved in the variability of iron loading in some thalassaemic patients (Rees *et al.* 1998*b*). Similarly, studies of the polymorphisms that have been related to the development of idiopathic osteoporosis hint that at least two sets of alleles—those involving the loci that regulate the synthesis of collagen and the vitamin D receptor—may modify the degree of osteoporosis, even in transfusion-dependent β thalassaemics (Rees *et al.* 1998*c*; Wonke 1998). Other phenotypic modifiers of this type may be definable, notably the relative rate of proteolysis of excess α -globin chains, but as yet the loci involved have not been determined.

(vii) *Phenotypic variability by coevolution*

Both the α and β thalassaemias have reached extremely high gene frequencies in many parts of the tropical world. The observation that in every country in which the disease is common there is a completely different set of mutations (figure 2) suggests that these disorders have arisen independently and reached their high frequency by selection, probably augmented to some degree by drift and founder effect. Furthermore, studies of the patterns of restriction fragment length polymorphisms of the β -globin gene cluster, particularly those that carry β -thalassaemia mutations, suggest that, in evolutionary terms, the expansion of the thalassaemias must have been a fairly recent event (Flint *et al.* 1998). Although it was suggested by J. B. S. Haldane in 1949 that the selective agent might be malaria, and evidence has been obtained that this is the case for the sickle-cell gene, it is only in the last few years that it has been proved beyond reasonable doubt that α thalassaemia protects against the complications of *Plasmodium falciparum* malaria (Williams *et al.* 1996; Allen *et al.* 1997). Although formal proof is awaited, it seems very likely that this will also be the case for β thalassaemia and Hb E.

Recent studies have also shown that exposure to malaria has generated a great deal of variation of the

Table 2. *Malaria-related polymorphisms*

haemoglobin
Hbs S, C and E
α and β thalassaemia
red cell membrane
melanesian ovalocytosis
glycophorins
blood groups: Duffy; ABO(H); Le(a); Kidd
red cell metabolism
glucose-6-phosphate dehydrogenase; Na/K
HLA-DR
TNF α
ICAM-1

human genome, not all of which is confined to the products of the globin genes or, indeed, to red cell proteins (reviewed by Weatherall *et al.* 1997). For example, it is now clear that certain alleles of the HLA-DR system are protective and that susceptibility to malaria is also modified by genetic variability at loci involved in the regulation of cytokines and other mediators of response to infection. A list of some of the more important malaria-related polymorphisms is shown in table 2. Many of these different polymorphic systems, as well as modifying host response to malaria, may also be involved in variation of susceptibility to a wide range of other infectious pathogens.

These observations have important implications for further modification of the phenotype of the β thalassaemias. Just as the β -thalassaemia mutations are different in every part of the world, so the other malaria-related polymorphisms are equally unevenly distributed, presumably a reflection of the fairly recent exposure of human populations to malaria. It follows, therefore, that patients with thalassaemia from different parts of the world, as well as having different mutations, may have widely differing host-defence responses to infection (Weatherall 1998*b*, 1999). It can no longer be assumed, therefore, that variation in the response to infection in thalassaemic patients in different populations simply reflects different environmental conditions.

(viii) *The role of the environment*

It has always seemed self-evident that if a child is born with a serious genetic anaemia their early environment will play a major role in determining the course of the illness. In the past, many children with β thalassaemia were born in poor countries in which standards of hygiene, nutrition and healthcare were grossly inadequate. Indeed, many of them died within the first year of life without a diagnosis being made. However, during the past 30 years there has been a major demographic transition in many of these countries, with a remarkable fall in childhood mortality rates. Hence children with β thalassaemia are now surviving and receiving treatment, which, though it may not be as effective as in richer, Western countries, is nevertheless keeping them alive. Although few comparisons have been made, from limited studies carried out in the author's laboratory over the last few years it is quite apparent that children with identical β -thalassaemic genotypes who live in rural, tropical

Table 3. *Some mechanisms for the phenotypic diversity of β thalassaemia*

heterogeneity of mutations at the β -globin gene locus
action of modifiers
primary: α -globin genes; γ -globin genes
secondary: multiple loci, involving bone, iron and bilirubin metabolism; others
variability due to co-selection
acquired factors
environment
geographical, social, cultural

environments have different patterns of disease to those who live in Western countries. This is particularly noticeable in disorders such as Hb E β thalassaemia, which are not always transfusion-dependent. Although the environmental factors that produce such different phenotypes in association with identical genotypes are not yet known, it appears that the pattern of exposure to infection and even the climate itself may be of considerable importance.

Social and religious backgrounds of patients with the thalassaemias are also important in determining their reactions to the disease and the way in which those with similar genotypes may have varying degrees of functional defect. Both patient and parental attitudes to disease differ widely among different cultures. For example, in some countries, disease is considered to be a serious stigma and patients are isolated from society, with all the psychological problems that follow. Differences in religious beliefs and their influence on decision making about the management of inherited diseases also are playing an increasingly important role in determining how individual patients respond to disease.

(ix) *To what extent is it possible to explain the overall phenotypic variability of β thalassaemia?*

The factors that have been clearly demonstrated to contribute towards the phenotypic heterogeneity of the β thalassaemias are summarized in table 3. It has been difficult to carry out formal studies of the degree of heritability which determines the clinical phenotype of this disease. There have been no twin studies comparing phenotypes of the thalassaemias, and comparisons between sibling pairs are fraught with problems unless it is possible to collect enough pairs that have been raised in different locations. Hence, while as mentioned in the preceding section, there is growing evidence that β -thalassaemic patients with identical genotypes who live in different environments have major differences in their clinical course, it is not possible to calculate the relative roles of the environment in determining the widely differing phenotypes.

It is possible, however, to obtain some indication of the relative importance of different genetic factors which modify the phenotype of β thalassaemia. A study that addressed this problem (Ho *et al.* 1998a) is summarized in figure 4. The question that was asked was what genetic factors are involved in determining the phenotypes of intermediate severity between a fully transfusion-dependent disorder and a symptomless trait. In this study,

75% of the patients had received two β -thalassaemia alleles. Out of these, 64 patients in all, it was possible to identify a genetic modifier in a high proportion of cases. About half of them reflected the interaction of mild β -thalassaemia alleles. Out of the other half, one or other genetic determinant for elevating foetal haemoglobin was demonstrated in 19 patients, while in seven no modulating factors could be determined. One interesting feature of this study was that there were eight pairs of siblings, seven of whom were discordant with respect to their clinical phenotype. In seven of these families, it was possible to demonstrate that the major difference between the siblings was in the steady-state level of Hb F, which ranged from 1 g dl⁻¹ to as much as 8–9 g dl⁻¹. These data underline the importance of variability at the γ -gene locus in determining the phenotype.

In short, while different environments undoubtedly play an important role in determining the phenotype of β thalassaemia, it is clear that variability at two primary modifying loci, the α - and γ -gene loci, plays a major part, an observation that is extremely encouraging for genetic counselling purposes, at least when the genes that are involved in the variable γ -globin gene response are identified.

3. PHENOTYPE–GENOTYPE RELATIONSHIPS FOR OTHER MONOGENIC DISEASES

Nearly every monogenic disease shows a wide range of phenotypic variability. As genes for these conditions have been isolated and their structure determined, it has become apparent that, as in the case of the thalassaemias, at least some of the clinical diversity can be related to the different mutations involved. And although, in most cases, this work is at a much earlier stage than is the case for thalassaemia, it is already apparent that other genetic modifiers must also exist.

(a) *Phenotypic heterogeneity due to the interaction of alleles at different loci*

There are a few examples of phenotypic variability of monogenic disease due to the interactions of alleles at separate loci. Obviously, for this to be of importance, at least one set of the alleles must be reasonably common in any particular population. For example, there is variety of inherited forms of muscular dystrophy. One of the commonest, Duchenne muscular dystrophy (DMD), results from mutations at the *DMD* gene, which maps to the short arm of the X chromosome and which encodes a cytoskeletal protein called dystrophin that belongs to the spectrin family. Boys with DMD develop severe muscle weakness early in life, while those with the related Becker form of muscular dystrophy have a milder illness with a later onset. This heterogeneity can be related to the underlying molecular defects at the DMD locus. However, there is a congenital variety of muscular dystrophy, the Fukuyama form, which is inherited in an autosomal recessive fashion and which involves the central nervous system. It turns out that in the latter condition at least about 10% of patients are dystrophin negative. To explain the high frequency of dystrophin deficiency in males with Fukuyama muscular dystrophy, it has been suggested that the Fukuyama muscular

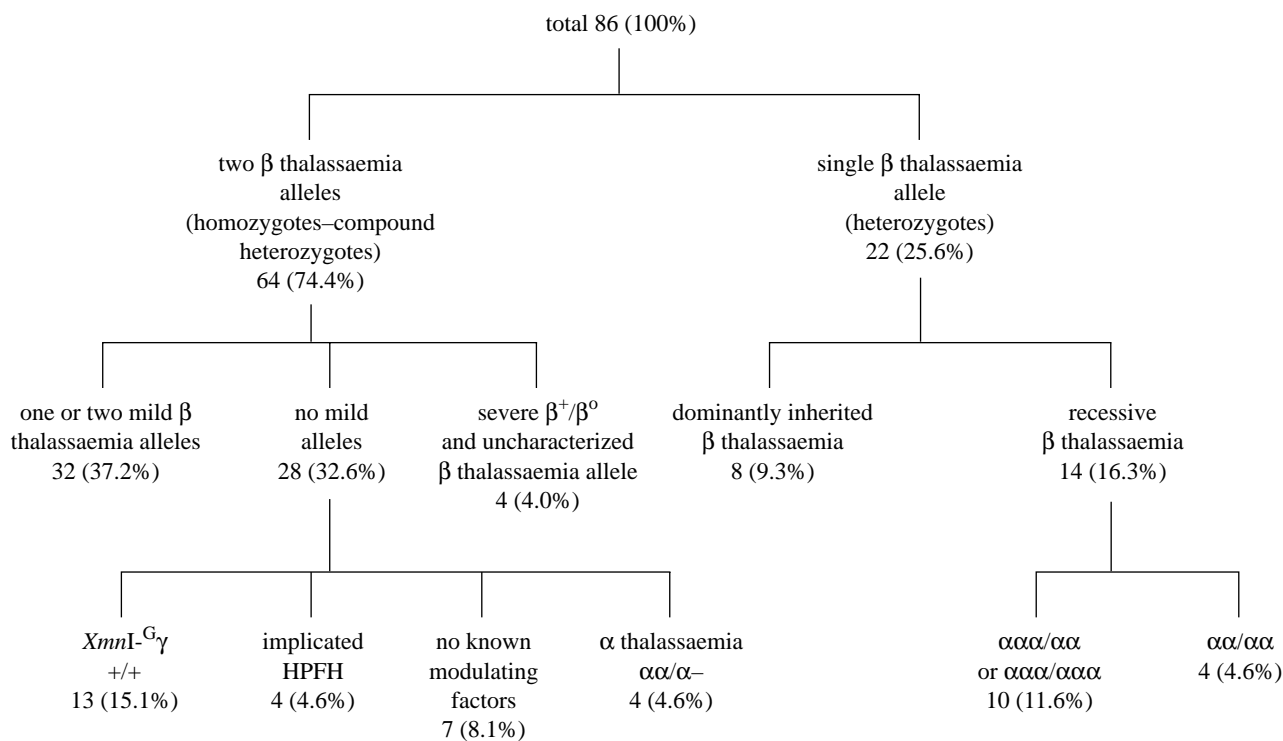


Figure 4. A flow chart showing the molecular pathology in patients in the UK with the milder, intermediate forms of β thalassaemia. HPFH, hereditary persistence of foetal haemoglobin. The *XmnI* polymorphism in the $G\gamma$ -gene promoter is described in the text (from Ho *et al.* 1998a)

dystrophy gene product normally interacts with dystrophin, and that the phenotype in these severely affected males reflects heterozygosity for the Fukuyama mutation coupled with a DMD gene mutation (Beggs *et al.* 1992).

Retinitis pigmentosa is the name given to a set of inherited degenerations of the retina. This condition is very heterogeneous, both clinically and with respect to its inheritance, and is X linked. Both autosomal dominant and recessive forms have been described. Kajiwara *et al.* (1994) reported three families in which a form of retinitis pigmentosa segregated; affected individuals were double heterozygotes for a specific peripherin-*RDS* gene mutation and a mutation in a second gene, *ROM1*. It is thought that the products of these two loci, which are on different chromosomes, interact non-covalently in the rim region of the photoreceptor outer segment disc membrane. Persons heterozygous for only one of these mutations have no symptoms.

Another condition that shows marked phenotypic heterogeneity is inherited porphyria cutanea tarda, in which heterozygotes are predisposed to photosensitive, cutaneous lesions. The phenotype seems to be exacerbated by a variety of environmental factors, including iron overload and alcohol abuse. The condition results from a deficiency of uroporphyrinogen decarboxylase (URO-D). Not all patients with mutations at the gene that encodes URO-D are symptomatic. Recently it has been found that there is an increased frequency of the alleles of the haemochromatosis gene, notably C282Y and H63D, in both Argentinian and Italian patients with porphyria cutanea tarda. Since the haemochromatosis alleles are associated with increased iron absorption and iron overload, and because iron loading exacerbates

porphyria cutanea tarda, it seems likely that this is another example of the deleterious interaction of two alleles, in this case with completely different functions (Mendez *et al.* 1998; Sampietro *et al.* 1998).

(b) *Unexplained phenotypic heterogeneity in common monogenic disease*

There are numerous examples of unexplained phenotypic variability in monogenic diseases that have been well defined at the molecular level (Summers 1996; Wolf 1997) (table 4). Here we will consider a few examples.

(i) *Cystic fibrosis*

Cystic fibrosis (CF), a particularly common monogenic disease in Europeans, shows remarkable phenotypic variability. It may involve pancreatic, pulmonary and reproductive function, all of which seem to result from a primary failure of chloride ion transport across cell membranes. The protein that is involved, the cystic fibrosis transmembrane regulator (CFTR), consists of two membrane-spanning domains, an R domain and two intracellular nucleotide-binding folds. More than 400 mutations of the CFTR gene have been reported, the commonest of which is a deletion of the phenylalanine codon at position 508 (ΔF 508), which accounts for about 70% of all Caucasian CF chromosomes.

There is a reasonably good correlation between phenotype and genotype for those who suffer from pancreatic insufficiency, most of whom are homozygous for ΔF 508, or, less frequently, for other subsets of severe alleles, or compound heterozygotes (Hamosh & Cutting 1993). On the other hand, there is marked variability in the severity and course of pulmonary disease for all genotypes, both

Table 4. *Examples of marked phenotypic diversity of monogenic disease in association with identical mutations*

(References will be found in Wolf (1997) and Weatherall & Clegg (1999).)

gene	localization	mutation	product	disease, with variable phenotype
HBB	11p.11.5	many	β -globin	β thalassaemia
HBB	11p.11.5	Glu 6 Val	β -globin	sickle-cell anaemia
HBA	16p.13.3	many	α -globin	α thalassaemia
CFTR	7q31-q32	DF508/R334W	CFTR	cystic fibrosis
ALD	Xq28	5.7 kb deletion	ABC transporter	adrenoleukodystrophy
APC	5q21-q22	Arg 302 stop	APC	familial adenomatous polyposis coli
FBN1	15q21.1	Cys 1409 Ser	fibrillin 1	Marfan syndrome
FMR1	Xq27.3	variable CGG expansion	FMR1 protein	fragile-X syndrome
HD	4p16.3	variable CAG expansion	huntingtin	Huntington's disease
OTC	Xp21.1	Arg 40 His	ornithine transcarbamylase	OTC deficiency
PAH	12q22-q24.2	identical haplotype	phenylalanine hydroxylase	hyperphenylalaninaemia
PAX3	2q35-q37	916 nt deletion	paired box protein 3	Waardenburg syndromes, 1 and 3
PKD1	16p13.3	Try 3818 stop	polycystin	polycystic kidney disease
PRNP	20p12-pter	Asp 178 Asn (+129 Met/Val)	prion protein	familial fatal insomnia Creutzfeldt–Jakob disease
RDS	6p21.2-cen	3 bp deletion	peripherin	retinitis pigmentosa
SRY	Yp11.3	several	testis-determining factor	XY gonadal dysgenesis or normal male
TTR	18q12.1	haploidenity	transerythrin	familial polyneuropathy
WT1	11q13	Arg 394 Trp	zinc finger TF	Denys–Drash syndrome
SCN4A	17q23	several	α -subunit Na ⁺ channel protein 4	periodic paralyses, myopathies, paramyotonia congenita
FGFR2	10q25.3-q26	many	fibroblast growth factor 2 receptor	several craniosynostoses syndromes; marked phenotypic variability with identical mutations

within and between families (Kerem *et al.* 1990; Welsh *et al.* 1995). The male reproductive phenotype, characterized by bilateral absence of the vas deferens, is also remarkably variable. At least 60% of males with this syndrome alone are simple or compound heterozygotes for mutations of CFTR. On the other hand, genetically identical brothers have been reported who are discordant for this complication. Interestingly, in a study of men with obstructive azoospermia in the absence of congenital absence of the vas deferens, it was found that at least half carried alleles, many of them mild, associated with abnormalities in the production of CFTR (Jarvi *et al.* 1995).

In short, extensive clinical, molecular and family data indicate that the pancreatic phenotype is primarily under the control of the CFTR locus. On the other hand, pulmonary disease is much less influenced by the molecular pathology at this locus and must be modified by other factors, both genetic and environmental.

(ii) *The hyperphenylalaninaemias*

Another common autosomal recessive disease, hyperphenylalaninaemia (HPA), also shows remarkable phenotypic variability (Scriver *et al.* 1995). Phenylalanine is converted to tyrosine by the action of phenylalanine hydroxylase (PAH) in the liver. Reduced activity of this enzyme results in the accumulation of phenylalanine in serum, which leads to brain damage and mental retardation unless treated by dietary restriction shortly after birth.

Over 70 mutations of the PAH gene have been identified in patients with HPA. At least some of the heterogeneity of this disease can, like the thalassaemias, be

explained by variable function of different alleles at the PAH gene locus. However, there are several factors that suggest that other genes must be involved. Families have been reported in which siblings with identical PAH genotypes have widely differing phenotypes (DiSilvestre *et al.* 1991). Furthermore, there appears to be considerable heterogeneity with respect to the neurological effects and degree of mental retardation in patients with similar genotypes and phenylalanine levels (reviewed by Summers 1996). Interestingly, a reduction in activity of PAH causes lowered levels of tyrosine, which is a precursor for several neurotransmitters. It is clear, therefore, that there is a potential for many different gene interactions in this disorder, although, to date, it has not been possible to identify any phenotypic modifiers.

(iii) *Gaucher's disease*

Gaucher's disease is a lysosomal glycolipid storage disorder characterized by the accumulation of glucosylceramide in tissues (Beutler & Grabowski 1995). It has an extremely varied phenotype and three main varieties have been defined. In type 1, multiple organs may be involved, but the nervous system is always spared; in type 2, there is an early onset with severe disease of the nervous system and death within the first few years of life; and patients with type 3 have a late onset of neurological disease, together with all the other features of the disorder seen in type 1 disease. Glucosylceramide is a normal intermediate in the catabolism of globoside and gangliosides. Its hydrolysis to ceramide and glucose is catalysed by a lysosomal acid β -glucosidase, the gene for which has been located on chromosome 1. The majority of the disease alleles that cause Gaucher's disease are

mis-sense mutations involving this locus. There is a pseudogene with a high degree of homology, approximately 16 kb downstream from the active gene. Apart from the more conventional types of mutation, some forms of Gaucher's disease have been associated with gene fusion events between the active locus and the pseudogene, and also by gene conversion (Latham *et al.* 1990).

As in the other monogenic diseases, some of the heterogeneity of this condition can be related to events of the glucocerebrosidase locus (reviewed by Beutler & Grabowski 1995). The severe neuropathic form is usually associated with homozygosity for a particular mutation, while in other populations, homozygosity for a different mutation usually produces type 3 Gaucher's disease. However, it is apparent that in addition to these clear differences in the clinical expression of different genotypes there is a great deal of intra-group variation. Ninety per cent of homozygotes for the N3705 mutation are asymptomatic or have a mild type 1 disease, while 10% present early with a much more severe form. Thus although a family of β -glycosidase activators, the saposins, have been defined, and indeed a genetically determined deficiency of saposin C produces a storage disorder not unlike Gaucher's disease, to date there is no clear evidence that genetic variability at the loci that control these activators is involved in the phenotypic heterogeneity of Gaucher's disease, nor have other loci with a similar effect been determined.

(iv) *Familial adenomatous polyposis (FAP)*

This is an autosomal dominant condition in which affected family members have hundreds of colonic polyps and in which colon cancer often develops by the age of 30 years. The gene responsible, *APC*, codes for a protein that probably interacts with cell adhesion molecules (Joslyn *et al.* 1991; Kinzler *et al.* 1991). Mutations of the *APC* gene segregate with the disease in FAP families and are found in sporadic tumours. The condition shows wide phenotypic variability; in some families there are only a few polyps in some members, while others have the full-blown phenotype (Leppert *et al.* 1990). In Gardner's syndrome there are also extra-colonic manifestations, including bone tumours, soft tissue tumours and congenital anomalies. Given the extraordinary phenotypic heterogeneity of this condition it seems likely that other genetic or environmental factors must be involved.

A mouse model has provided some insights into the genetic factors that may modify the phenotype of FAP (Balmain & Nagase 1998). The mouse equivalent of the *APC* locus is the *Min-1* (multiple intestinal neoplasia) locus. Breeding experiments with mice carrying one copy of an abnormal allele show that the number of polyps varies dramatically depending on the genetic background in which the *Min* mutation is segregating. A tumour-modifier gene, *Mom-1*, has been identified, and further investigations have led to its identification as the secreted-phospholipase 2a (*Pla-2g2*) gene. Although the precise mechanism whereby this tumour-modifier alters the phenotype of the mouse homologue of FAP is not yet clear, interestingly, *Mom-1* maps to chromosome 4 in the mouse, which shows conserved synteny with human chromosome 1p35-p36, a region which is frequently

deleted in a variety of human tumours. A candidate for a human homologue of this tumour suppressor, the phospholipase A2 gene, has been identified.

(c) ***Other mechanisms for phenotypic heterogeneity of monogenic disease***

There are a variety of other mechanisms that may be responsible for generating phenotypic diversity of diseases that are encoded at a single locus. Some of these, which have been loosely labelled epigenetic, are understood, at least in principle, although the molecular mechanisms involved have not been established and the prediction of the phenotype from the underlying genotype may be extremely difficult.

(i) *X inactivation*

Because X inactivation is usually random, female carriers for a particular X-linked trait are, in effect, mosaics, with each cell population functionally hemizygous for a particular trait. Hence carriers would be expected to produce approximately half of an abnormal gene product or express about the same amount of a product. However, because X inactivation occurs early during embryogenesis, there is wide variation in the expression of X-linked mutant genes in females; considerable skewing of the distribution of values is encountered. Interestingly, there is a high frequency of discordant phenotypes among twins with X-linked diseases (reviewed by Willard 1995). This may be because X inactivation precedes twinning, and non-randomness reflects asymmetrical splitting of the inner cell mass. Furthermore, it seems likely that some X-linked disorders may have a deleterious effect on cell function during early embryogenesis, a phenomenon that may lead to extreme skewing of the distribution of cell populations. The situation is further complicated by the fact that not all parts of the X chromosome are inactivated.

These different issues make for considerable heterogeneity in the expression of mutant genes in female carriers for X-linked disorders. There is increasing evidence that negative selection may be a major factor for the skewed distribution of cell populations in female carriers. For example, Coleman *et al.* (1993) described a female with the genes for both incontinentia pigmenti and haemophilia A. It appeared that the presence of the gene for incontinentia pigmenti on an X chromosome had unmasked the factor VIII gene mutation on the other chromosome, presumably by negative selection of the former. The latter phenomenon also occurs in the ATR-X syndrome, mentioned earlier.

(ii) *Genomic imprinting*

Genomic imprinting describes differential marking of maternally and paternally inherited alleles of genes or chromosome regions during gametogenesis, which leads to their differential expression during development. Although there is only limited knowledge of the role of imprinting in determining the phenotypic variability of genetic disease, it is quite clear that it plays an important part in determining the clinical picture of disorders such as the Prader-Willi (PWS) and Angelman (AS) syndromes, Beckwith-Weidemann syndrome and a variety of tumours. PWS and AS are both associated with major

developmental and behavioural disorders, but whereas patients with PWS have neonatal hypertonia, and later in infancy develop hyperphagia and severe obesity, hypogonadism and short stature, those with AS have a different clinical picture characterized by ataxia, seizures, sleep disorders, hyperactivity, speech abnormalities and severe mental retardation. These conditions are both caused by defects in imprinted-gene inheritance in a region of chromosome 15q11–q13. Although there seem to be several genetic mechanisms that can lead to these phenotypes, each involves parental imprinting, such that paternal gene expression is lost in PWS and maternal gene expression is lost in AS.

It appears that PWS involves loss of a function of several maternally expressed genes, while mutations in a single gene, which encodes a ubiquitin-protein ligase, which is also subjected to tissue-restricted imprinting, is the basis for at least some cases of AS. Even more remarkably, identification of mutations in the imprinting processes in these diseases has led to the definition of an imprinting centre, which regulates the initiation of imprint switching for all genes in a 2 Mb domain during gametogenesis. As further information about the molecular basis for PWS has been obtained, it appears that it can result from paternal interstitial deletions, maternal uniparental disomy, or mutations of the imprinting process. Furthermore, it is becoming clear that there are significant phenotypic differences associated with these different classes of mutation (Cassidy 1997).

Another interesting example of phenotypic modification by imprinting has come from studies of Turner's syndrome, a sporadic disorder of females in which all or part of one X chromosome is deleted. In this condition, intelligence is usually normal, but social adjustment problems are particularly common. In a study of 80 females with this disorder, and a single X chromosome, in 55, the X chromosome was maternally derived while in 25 it was of paternal origin. Members of the former group were significantly better adjusted, with superior verbal and executive function skills than those who had received the paternal X. These observations were interpreted as indicating that there is a gene locus for social cognition, which is imprinted and is not expressed from the maternally derived X chromosome. Further studies suggested that the locus escapes X inactivation, and probably lies on the long arm of the X chromosome or close to the centromere on the short arm (Skuse *et al.* 1997).

The molecular basis for imprinting, and other conditions in which this occurs, are reviewed by Nicholls *et al.* (1998).

(iii) *Mutations of mitochondrial DNA*

Mitochondrial DNA is found at high copy number, 10^3 to 10^4 copies, in nearly all the cells of the body. The mitochondrial genome is all inherited maternally and most of the copies are identical at birth, a condition termed homoplasmy. Occasionally, subpopulations of mitochondrial DNA carry mutations. When such heteroplasmic DNA is present during embryogenesis it can lead to a variety of disorders, particularly affecting the musculo-skeletal and nervous systems (Wallace 1999).

It seems likely that the wide phenotypic diversity of these disorders reflects the complexities of mitochondrial

DNA genetics, which is still poorly understood. Some of these issues have been discussed in detail recently (Lightowers *et al.* 1997). Among the many unanswered questions are how these mutations arise, whether they can be repaired, what influences the segregation and fixation of heteroplasmic mitochondrial DNA, whether levels of heteroplasmy fluctuate during life and, most importantly, whether it is possible to predict the segregation and transmission of a mutant genome.

(iv) *Expansion*

At least 12 neurological diseases are known to result from expansion of CTG, CGG, CAG or GAA repeats. The remarkable feature associated with triplet repeat disease is the tendency for these regions to undergo extensive expansion, a phenomenon which has been well documented for disorders such as fragile-X syndrome, myotonic dystrophy and Friedreich's ataxia (Ashley *et al.* 1995). What is known of the mechanisms of this remarkable phenomenon has been reviewed recently (Sinden 1999).

Fragile-X syndrome, which is a common cause of mental retardation in males, is a good example of how phenotypic heterogeneity can be generated by expansion of regions containing trinucleotide repeats. In normal individuals, a DNA segment at Xq27.3 contains between six and 60 copies of the repeat CGG. In some persons, the number is increased to between 60 and 200. In this case, the disorder is clinically silent; pre-mutations of this kind are characteristic of normal-transmitting males and some mentally normal female carriers. Full mutations, which arise in the offspring of pre-mutation carriers, consist of many hundreds or thousands of copies of the repeat and lead to the full expression of the clinical phenotype and the characteristic cytogenetic abnormality. Heterozygous carriers have an extremely variable phenotype; approximately 50% of females carrying the full mutation show some mental impairment although those who carry a pre-mutation are usually normal. Similarly, female carriers may or may not show some of the somatic changes associated with the disease in males. Although X inactivation may be partly responsible, this does not seem to be the whole story.

Huntington's disease, another neurodegenerative disorder, also shows considerable phenotypic heterogeneity, particularly with respect to age of onset and rate of progression. At the 5'-region of the locus involved there is a CAG repeat sequence, which ranges from ten to 30 copies in normal individuals and which is expanded to beyond 35 copies in patients with Huntington's disease. There appears to be some relationship between the length of repeats and age of onset. Remarkably, the juvenile onset of cases shows a preponderance of paternal transmission.

Overall, these conditions show considerable phenotypic heterogeneity. This is reflected at the molecular level, where there is wide variation in the extension of the length of the repeats required to produce an abnormal phenotype, not all diseases are associated with a pre-mutation length, and the relationship between the expansion of trinucleotide repeats and the causation of the disease is not clear. In some cases, those in which the repeats involve promoters, it is thought that methylation

of the promoter regions lead to gene silencing, while in those that involve CAG repeats, it is thought that polyglutamine expansion may play a role. But what is clear is that genotypic instability of this type is responsible for neurodegenerative disorders with widely differing phenotypes. Currently, the reasons for this predilection for neurological disease are not clear.

4. COMMON MULTIGENIC DISEASES

In recent years there has been a major effort directed at trying to define the genetic component of the common diseases of Western society: heart disease, hypertension, diabetes, the major psychoses, asthma and many others. Twin studies have given rather inconsistent results, but it is clear that there is a major difference between these conditions with respect to the importance of the genetic component. Concordance rates for type 1 (insulin-dependent) diabetes have ranged from 0.23 to 0.6, those for type 2 (insulin-resistant) diabetes from 0.27 to 0.83 (see Hawkes 1997). Much of this variability reflects different methods of ascertainment and analysis. Concordance rates also vary for other common diseases, but, overall, they tend to be quite low. These observations, and the fact that the inherited component probably reflects the action of a number of different genes, suggest that attempts to identify the genes in these multigenic diseases are going to be fraught with difficulty.

A wide variety of approaches is being used to dissect the genetic factors in multigenic disease. Many of the approaches involve linkage analyses using extended pedigrees, sibpairs, ancestral haplotype matching and the like. Many different classes of genetic markers have been used, including candidate genes, variable number tandem repeats (VNTRs), microsatellite DNA and, most recently, single nucleotide polymorphisms (SNPs). In the future, linkage disequilibrium mapping and ancestral haplotyping, combined with the knowledge gained from the human genome project, will undoubtedly speed up this process. The development of an SNP catalogue should allow the identification of regions of DNA in which variants can be defined that would affect protein sequence or expression (VAPSEs). It should also be possible to define homogeneous collections of families and large populations to determine subsets of variants that are associated with disease and hence to explain the association of linkage markers to particular regions.

So far many of these studies have given equivocal results and false leads. At least part of the problem, particularly in psychiatric genetics, is a reflection of difficulties in defining phenotypes. But even where this is possible, many difficulties have been encountered. The study of the genetics of type 1 diabetes, a disease characterized by immune damage to the pancreas and insulin-dependence early in life, exemplifies the possibilities and complications of this new field.

The study of type 1 diabetes offers a number of advantages. While there may be difficulties in defining the disease in its very earliest stages, the fully established phenotype is clear-cut. Furthermore, there is an excellent mouse model of autoimmune type 1 diabetes, the non-obese diabetic (NOD) strain. However, twin studies in humans have shown approximately a 50% concordance

rate in monozygotic twins and only 5% in dizygotic, indicating that there is a major environmental component to this disease. Formal linkage studies in the mouse indicate that there might be up to 14 loci involved, while in humans there could be over ten. More encouragingly, however, in the mouse it appears that three of the loci may account for nearly 80% of the variance (Todd 1999). In mice and in humans, the class 2 genes of the major histocompatibility complex were found to play an important role in susceptibility to the disease. In both cases, a region has also been implicated which is characterized by a variable number of tandem repeats (VNTR) about 600 bp 5' to the insulin gene. In mice, a third gene has been identified, most probably that for interleukin-2. Although a third locus has not yet been definitely defined in humans, the strongest linkage outside the MHC has been localized to chromosome 11p11, and the peak linkage in this region is near *ICF8*, which encodes for a negative regulator of IL2 expression.

Thus three loci have been established as playing a role in the variation of susceptibility to type 1 diabetes, both in mouse and man. It should be noted, in passing, that they were defined by the 'candidate gene' approach, and only retrospectively by linkage. It is assumed, therefore, that variability at these loci somehow invokes a state of increased susceptibility (or resistance) for an autoimmune response to various environmental triggers, presumably infection, although extensive studies have failed to define what these agents might be. Recently, Todd (1999) has reviewed the extremely complex interactions of these three loci and has discussed ways in which variability of their function may be directly involved in generating the autoimmune background required to develop type 1 diabetes in response to multiple environmental factors, acting in both positive and negative ways during different stages of development. The remarkable diversity of the immune system and high degeneracy of T-cell recognition provides a situation in which there is an extremely fine balance between response to infection and the self-destruction of tissues through immunological mimicry. It seems likely that the majority of individuals are resistant to autoimmune disease because they have a particular complement of alleles of their immunoregulatory genes that ensure that T- and B-cell lymphocyte responses to self-antigens are either avoided in the first place by clonal deletion in the thymus, or are tightly and quickly regulated. But while the identification of at least three of the loci involved in this process is encouraging, the very complexity of function at each of them, the numerous ways in which this may be modified and the lack of any hint as to the identity of the environmental triggers, offers some indication of how far away we are from a genuine understanding of the pathogenesis of this common condition, let alone using this information for predictive genetics. The other common multigenic diseases are unlikely to be much less complex.

5. IMPLICATIONS FOR THE FUTURE

Given the complexities of genotype-phenotype relations, how will the information that will undoubtedly arise from the human genome project modify clinical practice in the future?

In the case of monogenic diseases, it seems likely that we will be able to define what is really important in determining major phenotypic differences and what amounts to no more than fine-tuning of the phenotype. In the case of the thalassaemias, it is quite clear that really important phenotypic differences can be related to the action of many different mutations of the same locus and the interaction of two major modifying loci. We already have the tools with which to analyse the primary mutations of the β -locus and those at the α -globin locus, which can significantly modify the phenotype. We are also starting to learn a little about the different factors that vary the output of γ -globin genes and hence fix the level of Hb F, another factor of major importance in determining the clinical course. Thus, although there are many other layers of complexity that will have some effect on the phenotype, we are approaching the stage when we can offer a reasonable idea of the prognosis in any individual patient.

Although less progress has been made so far, it seems very likely that we will achieve this level of diagnostic capability in the case of many common monogenic diseases. It is possible to envisage a time when it will be possible to provide more accurate genetic counselling, prenatal diagnosis for the more severe monogenic disorders and a more rational basis on which decisions can be made about the application of experimental forms of treatment, gene therapy for example. But the clear lesson that has been learnt from the haemoglobin field is that we will only reach this position when we have a genuine understanding of how mutations result in abnormal function, that is of the pathophysiology of particular diseases.

As we have learnt from studies of diabetes, and some other common multigenic diseases, it should be possible to obtain information about some of the major genes involved in these complex multigenic systems. The insights that are being gained into the way in which our genetic make-up may make us prone to autoimmune disease, though still at a very early stage, offer some indication of the value of this approach for the study of common diseases of completely unknown aetiology, the major psychoses for example. However, it is already clear that we are dealing with a level of complexity that is infinitely greater than that which underlies the phenotypic diversity of the monogenic diseases, and it may take a very long time before we are in a position to use this information in clinical practice. Indeed, it seems very likely that we will have to develop sophisticated mathematical approaches to try to estimate the interaction of multiple alleles at many different loci, models which may also have to take into account the wide variation in the environmental triggers that may be involved in generating disease in genetically susceptible individuals. However, even if this goal is only likely to be achieved in the very distant future, if ever, along the way we should obtain valuable information about disease mechanisms, which should result in more rational forms of therapy.

The genetic approach to polygenic disease, the development of methods for comparing the genetic make-up of large populations and relating it to the risk of developing these conditions over long periods of time, and the use to microchip technology for screening the expression of

genes should provide the pharmaceutical industry with a wide range of potential targets for drug design. Currently, one of the problems is that there are already too many candidates of this kind for any one company to cope with. It is too early to say whether it will be possible to define the really important genes in the genesis of multigenic diseases, though the studies of diabetes, outlined earlier, suggest that this may be possible. As evidenced by the study of the rare, monogenic forms of Alzheimer's disease, a detailed analysis of these 'experiments of nature' may, in the long run, prove even more informative.

The remarkable complexity of the genotype-phenotype relationship has undoubtedly been underestimated during the early period of the revolution in the biomedical sciences that followed the DNA era. It has led to many statements being made about the imminence of accurate predictive genetics that are simply not true. Indeed, we have only reached this stage, albeit incompletely, in the case of one or two monogenic diseases. We are still a long way off understanding the phenotypic diversity of most single-gene disorders and it is far from certain that we will ever reach a stage in which we can accurately predict the occurrence of some of the common disorders of Western society at any particular stage in an individual's life.

In arriving at the rather primitive level of understanding of phenotype-genotype relationships outlined in this review, it is clear, particularly from work in the haemoglobin field, that almost everything that has been learnt has resulted from careful analyses of the unusual patient or family. As genetics develops in the new millennium it will be important to remember that computer modelling or other sophisticated approaches to unravelling biological complexity can never replace the study of these remarkable 'experiments of nature'. It is vital, therefore, that the field evolves as a closely integrated partnership between the clinical and basic biological sciences; they will continue to have much to offer each other.

I am grateful to Dr Andrew Wilkie for a critical review of this manuscript and for many useful suggestions. The section on the heterogeneity of the thalassaemias is the fruit of discussions over many years with my colleagues in the MRC Molecular Haematology Unit in Oxford, UK, notably John Clegg, Douglas Higgs, Bill Wood and SweeLay Thein. I am grateful to Mrs Janet Watt for preparing this manuscript.

REFERENCES

- Allen, S. J., O'Donnell, A., Alexander, N. D. E., Alpers, M. P., Peto, T. E. A., Clegg, J. B. & Weatherall, D. J. 1997 α^+ -thalassaemia protects children against disease due to malaria and other infections. *Proc. Natl Acad. Sci. USA* **94**, 14 736-14 741.
- Ashley, C. T. & Warren, S. T. 1995 Trinucleotide repeat expansion and human disease. *A. Rev. Genet.* **29**, 703-728.
- Balmain, A. & Nagase, H. 1998 Cancer resistance genes in mice: models for the study of tumour modifiers. *Trends Genet.* **14**, 139-144.
- Beggs, A. H., Neumann, P. E., Arahata, K., Arikawa, E., Nonaka, I., Anderson, M. S. & Kunkel, L. M. 1992 Possible influences on the expression of X-chromosome-linked dystrophin abnormalities by heterozygosity for autosomal recessive Fukuyama congenital muscular dystrophy. *Proc. Natl Acad. Sci. USA* **89**, 623-627.

- Beutler, E. & Grabowski, G. A. 1995 Gaucher disease. In *The metabolic and molecular bases of inherited disease*, 7th edn (ed. C. R. Scriver, A. L. Beaudet, W. S. Sly & D. Valle), pp. 2641–2670. New York: McGraw-Hill.
- Cassidy, S. B. 1997 Prader-Willi syndrome. *J. Med. Genet.* **34**, 917–923.
- Coleman, R., Genet, S. A., Harper, J. I. & Wilkie, A. O. M. 1993 Interaction of incontinentia pigmenti and factor VIII mutations in a female with biased X inactivation, resulting in haemophilia. *J. Med. Genet.* **30**, 497–500.
- Craig, J. E., Rochette, J., Fisher, C. A., Weatherall, D. J., Marc, S., Lathrop, G. M., Demenais, F. & Thein, S. L. 1996 Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nature Genet.* **12**, 58–64.
- Craig, J. E., Rochette, J., Sampietro, M., Wilkie, A. O. M., Barnetson, R., Hatton, C. S. R., Demenais, F. & Thein, S. L. 1997 Genetic heterogeneity in heterocellular hereditary persistence of fetal haemoglobin. *Blood* **90**, 428–434.
- DiSilvestre, D., Koch, R. & Groffen, J. 1991 Different clinical manifestations of hyperphenylalaninemia in three siblings with identical phenylalanine hydroxylase genes. *Am. J. Hum. Genet.* **48**, 1014–1016.
- Dover, G. J., Smith, K. D., Chang, Y. C., Purvis, S., Mays, A., Meyers, D. A., Sheils, C. & Serjeant, G. 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.
- Efremov, D. (and 10 others) 1994 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.
- Flint, J., Harding, R. M., Boyce, A. J. & Clegg, J. B. 1998 The population genetics of the haemoglobinopathies. *Clin. Haematol.* **11**, 1–51.
- Galanello, R. (and 11 others) 1989 Molecular analysis of β^0 -thalassaemia intermedia in Sardinia. *Blood* **74**, 823–827.
- Galanello, R., Perseu, L., Melis, M. A., Cipollina, L., Barella, S., Giagu, N., Turco, M. P., Maccioni, O. & Cao, A. 1997 Hyperbilirubinaemia in heterozygous β -thalassaemia is related to co-inherited Gilbert's syndrome. *Br. J. Haematol.* **99**, 433–436.
- Gibbons, R. J., Picketts, D. J., Villard, L. & Higgs, D. R. 1995 Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassaemia (ATR-X syndrome). *Cell* **80**, 837–845.
- Gilman, J. G. & Huisman, T. H. J. 1985 DNA sequence variation associated with elevated fetal $\text{G}\gamma$ globin production. *Blood* **66**, 783–787.
- Haldane, J. B. S. 1949 Disease and evolution. *Ricera Scient.* **19**, 2–12.
- Hamosh, A. & Cutting, G. R. 1993 Genotype/phenotype correlations in cystic fibrosis. In *Cystic fibrosis—current topics* (ed. J. A. Dodge, D. H. Brock & J. H. Widdicombe), pp. 69–89. Chichester, UK: Wiley.
- Hawkes, C. H. 1997 Twin studies in medicine—what do they tell us. *Q. J. Med.* **90**, 311–321.
- Higgs, D. R. 1993 α -thalassaemia. *Clin. Haematol.* **6**, 117–150.
- Ho, P. J., Hall, G. W., Luo, L. Y., Weatherall, D. J. & Thein, S. L. 1998a Beta thalassaemia intermedia: is it possible to predict phenotype from genotype? *Br. J. Haematol.* **100**, 70–78.
- Ho, P. J., Hall, G. W., Watt, S., West, N. C., Wimperis, J. W., Wood, W. G. & Thein, S. L. 1998b Unusually severe heterozygous β -thalassaemia: evidence for an interacting gene affecting globin translation. *Blood* **92**, 3428–3435.
- Huisman, T. H. J., Carver, M. F. H. & Erol Baysal, E. 1997 *A syllabus of thalassaemia mutations*. Augusta, ME: The Sickle Cell Anemia Foundation.
- Jarvi, K., Zielenski, J., Wilschanski, M., Durie, P., Buckspan, M., Tullis, E., Markiewicz, D. & Tsui, L.-C. 1995 Cystic fibrosis transmembrane conductance regulator and obstructive azoospermia. *Lancet* **345**, 1578.
- Joslyn, G. (and 22 others) 1991 Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell* **66**, 601–613.
- Kajiwara, K., Berson, E. L. & Dryja, T. P. 1994 Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/*RDS* and *ROM1* loci. *Science* **264**, 1604–1608.
- Kerem, E., Corey, M., Kerem, B., Rommens, J., Markiewicz, D., Levison, H., Tsui, L.-C. & Durie, P. 1990 The relationship between genotype and phenotype in cystic fibrosis—analysis of the most common mutations. *N. Engl. J. Med.* **323**, 1517–1522.
- Kinzler, K. W. (and 20 others) 1991 Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661–665.
- Knox-Macaulay, H. H. M., Weatherall, D. J., Clegg, J. B. & Pembrey, M. E. 1973 Thalassaemia in the British. *Br. Med. J.* **iii**, 150–155.
- Labie, D., Pagnier, J., Lapoumeroulie, C., Rouabhi, F., Dunda-Belkhdja, O., Chardin, P., Beldjord, C., Wajcman, H., Fabry, M. E. & Nagel, R. L. 1985 Common haplotype dependency of high $\text{G}\gamma$ -globin gene expression and high Hb F levels in β -thalassaemia and sickle cell anemia patients. *Proc. Natl Acad. Sci. USA* **82**, 2111–2114.
- Latham, T., Grabowski, G. A., Theophilus, B. D. M. & Smith, F. I. 1990 Complex alleles of the acid β -glucosidase gene in Gaucher disease. *Am. J. Hum. Genet.* **47**, 79–86.
- Leppert, M., Burt, R., Hughes, J. P., Samowitz, W., Nakamura, Y., Woodward, S., Gardner, E., Lalouel, J.-M. & White, R. 1990 Genetic analysis of an inherited predisposition to colon cancer in a family with a variable number of adenomatous polyps. *N. Engl. J. Med.* **322**, 904–908.
- Lightowers, R. N., Chinnery, P. F., Turnbull, D. M. & Howell, N. 1997 Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet.* **13**, 450–455.
- Mendez, M., Sorkin, L., Rossetti, M. V., Astrin, K. H., del C. Batlle, A. M., Parera, V. E., Aizencang, G. & Desnick, R. J. 1998 Familial porphyria cutanea tarda: characterization of seven novel uroporphyrinogen decarboxylase mutations and frequency of common hemochromatosis alleles. *Am. J. Hum. Genet.* **63**, 1363–1375.
- Nicholls, R. D., Saitoh, S. & Hirsthemke, B. 1998 Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.* **14**, 194–200.
- Rees, D. C., Styles, J., Vichinsky, E. P., Clegg, J. B. & Weatherall, D. J. 1998a The hemoglobin E syndromes. *Ann. NY Acad. Sci.* **850**, 334–343.
- Rees, D. C., Singh, B. M., Luo, L. Y., Wickramasinghe, S. & Thein, S. L. 1998b Nontransfusional iron overload in thalassaemia: association with hereditary hemochromatosis. *Ann. NY Acad. Sci.* **850**, 490–494.
- Rees, D. C., Basran, R. K., Hum, B., Peto, T. E. A., Weatherall, D. J. & Olivieri, N. F. 1998c Genetic influences on bone disease in thalassaemia. *Blood* **532**, a.
- Rund, D., Oron-Karni, V., Filon, D., Goldfarb, A., Rachmilewitz, E. & Oppenheim, A. 1997 Genetic analysis of β -thalassaemia intermedia in Israel: diversity of mechanisms and unpredictability of phenotype. *Am. J. Hematol.* **54**, 16–22.
- Sampietro, M., Lupica, L., Perrero, L., Comino, A., di Montemuros, F. M., Cappellini, M. D. & Fiorelli, G. 1997 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.
- Sampietro, M. (and 11 others) 1998 High prevalence of the His63Asp HFE mutation in Italian patients with porphyria cutanea tarda. *Hepatology* **27**, 181–184.

- Scriber, C. R., Kaufman, S., Eisensmith, R. C. & Woo, S. L. C. 1995 The hyperphenylalaninemia. In *The metabolic and molecular bases of inherited disease*, 7th edn (ed. C. R. Scriber, A. L. Beaudet, W. S. Sly & D. Valle), pp. 1015–1076. New York: McGraw-Hill.
- Skuse, D. H., James, R. S., Bishop, D. V. M., Coppin, B., Dalton, P., Aamodt-Leeper, G., Bacarese-Hamilton, M., Creswell, C., McGurk, R. & Jacobs, P. A. 1997 Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* **387**, 705–708.
- Sinden, R. R. 1999 Human genetics '99: trinucleotide repeats. Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet.* **64**, 346–353.
- Summers, K. M. 1996 Relationship between genotype and phenotype in monogenic diseases: relevance to polygenic diseases. *Hum. Mutat.* **7**, 283–293.
- Tamagnini, G. P., Lopes, M. C., Castanheira, M. E., Wainscoat, J. S. & Wood, W. G. 1983 β^+ thalassaemia—Portuguese type: clinical, haematological and molecular studies of a newly defined form of β thalassaemia. *Br. J. Haematol.* **54**, 189–200.
- Thein, S. L., Hesketh, C., Taylor, P., Temperley, P., Hutchison, R. M., Old, J. M., Wood, W. G., Clegg, J. B. & Weatherall, D. J. 1990 Molecular basis for dominantly inherited inclusion body β thalassaemia. *Proc. Natl Acad. Sci. USA* **87**, 3924–3928.
- Thein, S. L., Sampietro, M., Old, J. M., Cappellini, M. D., Fiorelli, G., Modell, B. & Weatherall, D. J. 1987 Association of thalassaemia intermedia with a beta-globin gene haplotype. *Br. J. Haematol.* **65**, 370–373.
- Thein, S. L. & Weatherall D. J. 1989 A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the β -globin gene complex. In *Hemoglobin switching. B. Cellular and molecular mechanisms* (ed. G. Stamatoyannopoulos, A. W. Nienhuis & R. Alan), pp. 97–111. New York: Liss.
- Todd, J. A. 1999 From genome to aetiology in a multifactorial disease, type 1 diabetes. *BioEssays* **21**, 164–174.
- Wallace, D. C. 1999 Mitochondrial diseases in man and mouse. *Science* **283**, 1482–1488.
- Weatherall, D. J. 1998a Pathophysiology of thalassaemia. *Clin. Haematol.* **11**, 127–146.
- Weatherall, D. J. 1998b Thalassaemia in the next millennium. *Ann. NY Acad. Sci.* **850**, 1–9.
- Weatherall, D. J. 1999 Harvey Lectures. The phenotypic diversity of monogenic disease: lessons from the thalassaemias. (In the press.)
- Weatherall, D. J. & Clegg, J. B. 1999 *The thalassaemia syndromes*, 4th edn. Oxford, UK: Blackwell Science. (In the press.)
- Weatherall, D. J., Clegg, J. B. & Naughton, M. A. 1965 Globin synthesis in thalassaemia: an *in vitro* study. *Nature* **208**, 1061–1065.
- Weatherall, D. J., Clegg, J. B., Knox-Macaulay, H. H. M., Bunch, C., Hopkins, C. R. & Temperley, I. J. 1973 A genetically determined disorder with features both of thalassaemia and congenital dyserythropoietic anaemia. *Br. J. Haematol.* **24**, 681–702.
- Weatherall, D. J., Clegg, J. B., Wood, W. G., Old, J. M., Higgs, D. R., Pressley, L. & Darbre, P. D. 1980 The clinical and molecular heterogeneity of the thalassaemia syndromes. *Ann. NY Acad. Sci.* **344**, 83.
- Weatherall, D. J., Higgs, D. R., Bunch, C., Old, J. M., Hunt, D. M., Pressley, L., Clegg, J. B., Bethlenfalvay, N. C., Sjolin, S., Koler, R. D., Magenis, E., Francis, J. L. & Bebbington, D. 1981 Hemoglobin H disease and mental retardation. A new syndrome or a remarkable coincidence? *N. Engl. J. Med.* **305**, 607–612.
- Weatherall, D. J., Clegg, J. B. & Kwiatkowski, D. 1997 The role of genomics in studying genetic susceptibility to infectious disease. *Genome Res.* **7**, 967–973.
- Weatherall, D. J., Clegg, J. B., Higgs, D. R. & Wood, W. G. 1999 The hemoglobinopathies. In *The metabolic basis of inherited disease*, 8th edn (ed. C. R. Scriber, A. L. Beaudet, W. S. Sly & D. Valle). New York: McGraw-Hill. (In the press.)
- Welsh, M. J., Tsui, L.-C., Boat, T. F. & Beaudet, A. L. 1995 Cystic fibrosis. In *The metabolic and molecular bases of inherited disease*, 7th edn (ed. C. R. Scriber, A. L. Beaudet, W. S. Sly & D. Valle), pp. 3799–3878. New York: McGraw-Hill.
- Wilkie, A. O. M., Buckle, V. J., Harris, P. C., Lamb, J., Barton, N. J., Reeders, S. T., Lindenbaum, R. H., Nicholls, R. D., Barrow, M., Bethlenfalvay, N. C., Hutz, M. H., Tölmie, J. L., Weatherall, D. J. & Higgs, D. R. 1990a Clinical features and molecular analysis of the α thalassaemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. *Am. J. Hum. Genet.* **46**, 1112–1126.
- Wilkie, A. O. M., Zeitlin, H. C., Lindenbaum, R. H., Buckle, V. J., Fischel-Ghodsian, N., Chui, D. H. K., Gardner-Medwin, D., MacGillivray, M. H., Weatherall, D. J. & Higgs, D. R. 1990b Clinical features and molecular analysis of the α -thalassaemia/mental retardation syndromes. II. Cases without detectable abnormality of the α globin complex. *Am. J. Hum. Genet.* **46**, 1127–1140.
- Willard, H. F. 1995 The sex chromosomes and X chromosome inactivation. In *The metabolic and molecular bases of inherited disease*, 7th edn (ed. C. R. Scriber, A. L. Beaudet, W. S. Sly, D. Valle, B. Childs & B. Vogelstein), pp. 719–738. New York: McGraw-Hill.
- Williams, T. N., Maitland, K., Bennett, S., Ganczakowski, M., Peto, T. E. A., Newbold, C. I., Bowden, D. K., Weatherall, D. J. & Clegg, J. B. 1996 High incidence of malaria in α -thalassaemic children. *Nature* **383**, 522–525.
- Wolf, U. 1997 Identical mutations and phenotypic variation. *Hum. Genet.* **100**, 305–321.
- Wonke, B. 1998 Bone disease in thalassaemia major. *Br. J. Haematol.* **103**, 897–901.