CHAPTER 3

The Normal Structure and Regulation of Human Globin Gene Clusters

Bernard G. Forget and Ross C. Hardison

The genes encoding the different globin chains of hemoglobin are members of an ancient gene family. In this chapter we will review the structural features of the globin genes, with particular attention to the sequences needed for proper regulation of gene expression. Some of these have been well- conserved during mammalian evolution and therefore are likely to provide a common function in many mammals. Others are only found in higher primates, and may play roles in lineage-specific regulation. We will first describe the structural characteristics of the human globin genes and then provide a comparative analysis of the genomic contexts, regulatory regions and evolutionary conservation of features present in the globin gene clusters.

NUMBER AND CHROMOSOMAL LOCALIZATION OF HUMAN GLOBIN GENES

Hemoglobin is a heterotetramer that contains two polypeptide subunits related to the α-globin gene subfamily (referred to here as α-like globins) and two polypeptide subunits related to the β-globin gene subfamily (β-like globins). Globin polypeptides bind heme, which in turn allows the hemoglobin in erythrocytes to bind oxygen reversibly and transport it from the lungs to respiring tissues. In humans, as in all vertebrate species studied, different α-like and β-like globin chains are synthesized at

progressive stages of development to produce hemoglobins characteristic of primitive (embryonic) and definitive (fetal and adult) erythroid cells (Figure 3.1).

Before precise knowledge of globin gene organization was gained by gene mapping and molecular cloning, a general picture of the number and arrangement of the human globin genes emerged from the genetic analysis of normal and abnormal hemoglobins and their pattern of inheritance. The number and subunit composition of the different normal human hemoglobins (Figure 3.1) suggested that there must exist at least one globin gene for each of the different globin chains: α, β, γ, δ, ε, and ζ. Evidence from the study of hemoglobin variants and the biochemical heterogeneity of the chains in fetal hemoglobin (HbF) showed that the α- and γ-globin genes were duplicated. Persons were identified whose red cells contained more than two structurally different α-globin chains that could be best explained by duplication of the α-globin gene locus, and the characterization of the structurally different $^{G}\gamma$ and $^{A}\gamma$ globin chains of HbF imposed a requirement for duplication of the γ-globin gene locus.

Studies of the pattern of inheritance of hemoglobin variants from persons carrying both an α chain and a β chain variant revealed that the α- and β-globin genes are on different chromosomes (or very widely separated if on the same chromosome). Variants of α-globin and β-globin chains were always observed to segregate independently in offspring of doubly affected parents (reviewed in 1). Linkage of the various β-like globin genes to one another was established from the study of interesting hemoglobin variants that contained fused globin chains, presumably resulting from nonhomologous crossover between different β-like globin genes. Characterization of Hb Lepore (2), with its δβ fusion chain, established that the δ-globin gene was linked to and located on the 5' (or N-terminal) side of the β-globin gene. Analysis of Hb Kenya (3), with its $^{A}\gamma\beta$ fusion chain, provided evidence for linkage of the $^{A}\gamma$ gene, and presumably the $^{G}\gamma$ gene as well, to the 5'-side of the δ- and β-globin genes.

Thus, the general arrangement of the globin genes that emerged from these various genetic analyses can be represented as illustrated in Figure 3.1. It was also assumed, but unsupported by genetic evidence, that the embryonic α-like (ζ) and β-like (ε) globin genes were likely to be linked to the loci encoding their adult counterparts.

By using rodent-human somatic hybrid cells containing only one or a few human chromosomes, Deisseroth and colleagues (4, 5) clearly established that the human α- and β-globin genes resided on different chromosomes. The α-like globin genes are located on chromosome 16, whereas the β-like globin genes are on chromosome 11. The latter results were obtained by hybridizing a solution of total cellular DNA from the various somatic hybrid cells to radioactive cDNAs, synthesized from α- and β-globin mRNAs by reverse transcriptase. These results were later confirmed and extended by various groups using the gene mapping procedure of Southern blot analysis with DNA from various hybrid cell lines containing different translocations or deletions of the involved chromosomes.

These studies also localized the globin gene loci to specific regions on their respective chromosomes: the β-globin gene cluster to the short arm of chromosome 11, and the α-globin gene cluster to the short arm of chromosome 16 (Figure 3.1). These chromosomal assignments were further confirmed and refined by *in situ* hybridization of radioactive cloned globin gene probes to metaphase chromosomes and by fluorescence-based *in situ* hybridization (FISH). Thus, the β-globin gene cluster was assigned to 11p15.5 and the α-globin gene cluster to 16p13.3. Subsequent DNA sequencing of entire human chromosomes and alignment with maps of chromosome bands places the β-globin gene cluster in 11p15.4. The α-globin gene cluster is only about 150 kilobase pairs (kb) from the telomere of the short arm of chromosome 16.

GLOBIN GENE STRUCTURE: INTRONS AND THEIR REMOVAL

The coding region of each globin gene in humans and other vertebrates is interrupted at two positions by stretches of noncoding DNA called intervening sequences (IVSs) or introns (6). In the β-like globin genes, the introns interrupt the sequence between codons 30 and 31 and between codons 104 and 105; in the α-globin gene family, the intervening sequences interrupt the coding sequence between codons 31 and 32 and between codons 99 and 100 (Figure 3.2.A). Although the precise codon position numbers at which the interruption occurs differ between the α- and β-like globin genes, the introns occur at precisely the same position in the aligned primary sequence of the α-

and β-globin chains. Thus, given the likely possibility that the α- and β-globin gene families originally evolved from a single ancestral globin gene (7), these gene sequences are homologous, and we infer that the presence of the introns at these positions predates the separation of α-globin and β-globin genes about 500 million years ago (in an ancestral jawed vertebrate). The first intervening sequence (IVS-1) is shorter than the second intervening sequence (IVS-2) in both α- and β-globin genes, but IVS-2 of the human β-globin gene is much larger than that of the α-globin gene (Figure 3.2.A).

The pattern of intron sizes of the ζ-like globin genes differs from that of the other α-like globin genes. Whereas the introns in the α and ψα genes are small, e.g. fewer than 150 base pairs (bp), those of the ζ and ψζ genes are larger (8). Furthermore, the first introns of the ζ and ψζ genes are much larger than their second introns; in fact they are eight to ten times larger than the first introns of any other globin gene.

The presence of intervening sequences that interrupt the coding sequences of structural genes imposes a requirement for some cellular process to remove these sequences in the mature mRNA. As illustrated in Figure 3.2.B, intervening sequences are transcribed into globin (and other) precursor mRNA molecules (9), but they are subsequently excised and the proper ends of the coding sequences joined to yield the mature mRNA (10). This posttranscriptional processing of mRNA precursors to remove introns has been termed splicing. A crucial prerequisite for the proper splicing of globin (and other) precursor mRNA molecules is the presence of specific nucleotide sequences at the junctions between coding sequences (exons) and intervening sequences (introns). Comparison of these sequences in many different genes has permitted the derivation of two different consensus sequences, which are almost universally found at the 5' (donor) and 3' (acceptor) splice sites of introns (11, 12). The consensus sequences thus derived are shown in Figure 3.2A, along with the consensus surrounding the branch point A involved in the initiation of splicing. The dinucleotides GT and AG shown in boldface, at the 5' and 3' ends, respectively, of the intron, are essentially invariant and are thought to be absolutely required for proper splicing. This is the so-called GT-AG rule. Rare examples have been described in which GC instead of GT is found at the donor splice site junction.

The importance of these consensus sequences is underscored by the fact that mutations that either alter them or create similar consensus sequences at new sites in a globin gene can lead to abnormal processing of globin mRNA precursors; these constitute the molecular basis for many types of thalassemia (Chapters 13 and 16). Throughout this chapter we will refer to human mutations that affect some aspect of the pathway for gene expression. Readers desiring more information may want to use databases such as HbVar (http://www.bx.psu.edu) (13) or the Locus Variants track on the UCSC Genome Browser (http://genome.ucsc.edu) (14) to find positions, genotypes and phenotypes for the greater than 1000 known globin gene variants.

DETAILED CHROMOSOMAL ORGANIZATION OF THE HUMAN GLOBIN GENES

A precise picture of the chromosomal organization of the α- and β-like human globin gene clusters, with respect to the number of structural loci and intergenic distances, was obtained by a number of different techniques: (1) restriction endonuclease mapping of genomic DNA (e.g. 15, 16) using the gel blotting procedure of Southern (17) and, (2) gene isolation and sequencing using recombinant DNA technology (e.g., 18). Sets of overlapping genomic DNA fragments spanning the entire α- and β-globin gene clusters were obtained by gene cloning, initially in bacteriophage λ and larger fragments in cosmid vectors. Detailed analysis of these recombinant DNA clones and complete DNA sequencing led to the determination of the gene organization illustrated in Figure 3.3. Some results were expected, such as the finding of single δ- and β-globin gene loci and duplication of the α- and γ-globin gene loci. In addition, single loci for the embryonic ζ and ε globin chains were found linked to the α- and β-globin gene clusters, respectively. It is noteworthy that the genes in each cluster are in the same transcriptional orientation and are arranged, in a 5' to 3' direction, in the same order as their expression during development.

An unexpected finding was the presence in the globin gene clusters of additional gene-like structures with sequence homology and an exon-intron structure similar to the actively expressed globin genes. These DNA segments have been called pseudogenes (19). One, called ψβ1, is in the β-like globin gene cluster between the γ- and δ-globin genes. At least two (and possibly four) are in the α-like globin gene cluster. The two clear examples are ψζ1 and ψα1, located between the active ζ-globin and α-globin genes (Fig. 3.3). All three (ψβ1, ψζ1, and ψα1) are characterized by the presence of one or more mutations that render them incapable of encoding a functional globin chain. This inability to encode a functional globin polypeptide does not necessarily render the pseudogenes inactive for transcription. The pseudogene ψβ1 is transcribed and spliced, as shown by several spliced ESTs (expressed sequence tags), whereas no evidence has been provided that ψα1 is transcribed. These pseudogenes appear to have arisen by gene duplication events within the globin gene clusters followed by mutation and inactivation of the duplicated gene and subsequent accumulation of additional mutations through loss of selective pressure.

Two other α-like globin genes have been identified and characterized in the α-globin gene cluster, but their roles, if any, in encoding globin polypeptides are still uncertain. The θ-globin gene located to the 3' or C-terminal side of the duplicated α-globin genes (20). It is more closely related to the α-globin genes than to the ζ-globin genes and is expressed at low levels in erythroid cells (21, 22). Clear homologs to the θ-globin gene are found in the homologous position in other mammalian α-like globin gene clusters. The μ-globin gene is located just 3' of the ψζ1-globin pseudogene (23, 24); it was initially called ψα2 (25) but with more accurate sequencing it is clear that this gene does not contain mutations that would render it inactive. It is a distant relative, being equally divergent from both α-globin and ζ-globin genes. Its closest relatives are the α^D-globin genes, which are actively expressed in red cells of reptiles and birds (24, 26). DNA sequences similar to that of the human μ-globin gene are found in other mammals, but in some species, such as mouse, the sequence has diverged so much that no obvious gene structure is found. Thus the presence of the θ-globin gene is conserved in all mammals examined but the μ-globin gene has been lost in some but not all lineages. Both

the θ-globin gene and the μ-globin gene are transcribed and spliced in erythroid cells, albeit at much lower levels than the α-globin gene. Curiously, no hemoglobin containing the θ-globin chain or the μ-globin chain has been identified, even by sensitive mass spectrometry (23). Furthermore, the predicted structure (translated amino acid sequence) of the θ-globin chain suggests that it would be unlikely to function normally as a hemoglobin subunit(27). Thus these genes remain a puzzle. They tend to be retained over mammalian evolution, hence indicating constraint for some function. They are expressed at the RNA level but do not appear to be translated into a polypeptide. Perhaps they or their RNA transcripts play some role that has yet to be discovered.

GENOMIC CONTEXT OF THE α-GLOBIN AND β-GLOBIN GENE CLUSTERS

The separation of α- and β-globin gene clusters to different chromosomes has allowed them to diverge into strikingly different genomic contexts, with paradoxical consequences for our understanding of their regulation. Given that all contemporary vertebrates have developmentally regulated hemoglobin genes encoding proteins used for oxygen transport in erythrocytes, it would have been reasonable to expect that the molecular mechanisms of globin gene regulation would be conserved in vertebrates. Certainly, the coordinated and balanced expression of α- and β-globin genes to produce the heterotypic tetramer $\alpha_2\beta_2$ in erythrocytes should be a particularly easy aspect of regulation to explain. Because the two genes would have been identical after the initial duplication in the ancestral vertebrate, with identical regulatory elements, it is parsimonious to expect selection to keep the regulatory elements very similar.

However, much has changed between the α- and β-like globin gene clusters since their duplication. Not only are they now on separate chromosomes in birds and mammals, but in mammals they are in radically different genomic contexts (28). A major determinant of the genomic environment is the G+C content. A G+C rich DNA segment has a high mole fraction of the nucleotides guanidylic acid (G) and cytidylic acid (C), whereas an A+T rich DNA segment has a high mole fraction of the nucleotides adenylic acid (A) and thymidylic acid (T). The G+C content for the human genome on average is

low (about 41%) but some segments can be much lower or higher, ranging from 30% to 65% in 20kb windows (29). Regions that are G+C rich tend to be enriched in genes, and those genes tend to be expressed in a broad range of tissues. They also tend to have islands with an abundance of the dinucleotide CpG (30). This is in stark contrast to the bulk of the genome, which has very few CpGs because these are the sites for DNA methylation, and substitution of CpG to TpG or CpA is very rapid on an evolutionary timescale (as much as 10 times faster than the rates of other substitutions). The CpG islands are thus short regions (a few hundred bp) in which the CpG dinucleotides are not methylated; these have been associated with important functions such as promoters for transcription.

The β-globin gene clusters in humans and other mammals are A+T rich, with no CpG islands (31), whereas the α-like globin gene clusters are highly G+C rich, with multiple CpG islands (32). This correlates with several important differences in the structure and regulation of the two gene clusters. Tissue-specific gene expression of the β-like globin genes is correlated with an increased accessibility of the chromatin only in expressing cells (33), and hence “opening” of a chromatin domain is a key step in activation of these genes. In contrast, the α-like globin genes, which are in constitutively open chromatin (28). The β-globin gene cluster is subject to tissue-specific DNA methylation(34), but, in keeping with the presence of CpG islands, the α-globin gene cluster is not methylated in any cell types (35). The β-globin gene clusters are replicated early in S phase only in cells expressing them, whereas the human α-globin genes are replicated early in all cells (36-38). Thus the mammalian α-globin genes have several characteristics associated with constitutively expressed "housekeeping" genes. The strikingly different genomic contexts of the two gene clusters affect several aspects of DNA and chromatin metabolism, including timing of replication, extent of methylation, and the type of chromatin into which the loci are packaged. Rather than selecting for similarities to insure coordinate and balanced expression, the processes of evolution at these two loci have made them quite different. The full implications of these differences may not yet be known. For instance, the two “healthy” genes with no known function in

the α-like globin gene cluster, θ and μ, are themselves CpG islands. Could this be a clue to a role for these genes outside the conventional one of coding for proteins?

The types of genes that surround the α-like and β-like globin gene clusters are quite different (Figure 3.3). The β-like globin gene cluster is surrounded by olfactory receptor (*OR*) genes, which encode G-protein coupled receptors expressed in olfactory epithelium (39). Several *OR* gene clusters containing about a thousand genes and pseudogenes are found in the human genome. The *OR* gene cluster surrounding the β-like globin genes is a particularly large one, with about a hundred genes extending almost 1 million bp (Mb) past *HBB* (the β-globin gene) and over 3 Mb toward the centromere from *HBE1* (the ε-globin gene). This arrangement is found in homologous regions in mammals and in chickens. Thus the erythroid-specific regulation of the β-like globin gene cluster is exerted in a chromosomal environment that is largely devoted to olfactory-specific expression. Perhaps this has had an impact on selection for a particularly powerful enhancer, to override the olfactory-specific regulation. As shown in Figure 3.3.A, several deletions causing β-thalassemia or Hereditary Persistence of Fetal Hemoglobin (HPFH) not only remove β-like globin genes, but they also fuse the remaining genes with sequences close to an *OR* gene. The phenotype of patients carrying such deletions may be explained in part by bringing positive or negative regulatory elements normally associated with *OR* genes into proximity of the β-like globin genes (40, 41).

In contrast, the α-like globin genes are surrounded by a variety of genes (Figure 3.3.B), many of which are widely expressed and carry out fundamental roles in cellular metabolism and physiology, such as *MPG* (encoding the DNA repair enzyme methyl purine glycosylase) and *POLR3K* (encoding a subunit of RNA polymerase III) (42). Although the α-like globin gene cluster and surrounding DNA is in constitutively open chromatin, histones are hyperacetylated (another mark of active loci) in erythroid cells in a more restricted region encompassing the globin genes and their regulatory sequences (43). Although the regions homologous to that surrounding the α-like globin gene cluster have undergone inter- and intra-chromosomal rearrangements in various vertebrates lineages, the genes from *POLR3K* through *HBQ1* have remained together in all species

examined from fish to mammals (44). This suggests that this region encompasses all the sequences needed in *cis* for appropriate regulation of the α-like globin genes.

Despite these many differences between α-like and β-like globin gene clusters in mammals, the appropriate genes are still expressed coordinately between the two loci, resulting in balanced production of α-like and β-like globins needed for the synthesis of normal hemoglobins. The mechanisms that accomplish this task still elude our understanding.

One important aspect that is common to the genomic contexts of both gene clusters is the presence of distal strong enhancers. The discovery of these was aided by mapping of deletions that result in β-thalassemia or α-thalassemia, which are inherited deficiencies in the amount of β-globin or α-globin, respectively (see Chapters 13 and 16). Some of these deletions removed distal sequences but retained all the globin genes, e.g. the deletions associated with Hispanic $(\varepsilon\gamma\delta\beta)^0$ thalassemia and the Ti~ α^0 thalassemia (Figure 3.3). Within the deleted intervals are critical long-range enhancers needed for high level expression of any gene in the linked globin gene clusters. These are the locus control region (LCR) for the β-globin gene cluster and HS-40 or major regulatory element (MRE) for the α-globin gene cluster. Thus regulation of expression of globin genes involves DNA sequences both close to the genes (proximal) and as much as 70 kilobase pairs (kb) away from the genes (distal). These will be examined in more detail in the next section.

EVOLUTIONARY INSIGHTS INTO REGULATION OF GLOBIN GENE CLUSTERS

Motivation

One avenue for improving the conditions of patients with hemoglobinopathies could involve regulation of expression of the globin genes. This hope is based on the normal human variation in phenotypes presented for a given mutant genotype. For

example, patients with naturally higher concentrations of HbF ($\alpha_2\gamma_2$) in their erythrocytes tend to have milder symptoms of either sickle cell disease or thalassemia (Chapters 17 & 19). The α-globin gene status can affect the severity of β-thalassemia, with more balanced production of α-globin and β-globin associated with milder disease. Thus considerable effort has gone into studying the stage-specific expression of the globin genes, with a long-term goal of enhancing or restoring production of embryonic or fetal hemoglobins in adult life or reducing expression of deleterious alleles. Although no current treatment by gene therapy is in practice as of this writing, much effort continues in this area. The use of hydroxyurea in treatment of sickle cell disease is an outgrowth of studies on mechanisms of regulation of globin genes. Current studies aim to discover more sophisticated and directed pharmacological methods for enhancing production of embryonic and fetal hemoglobins.

Studies over the past three decades have revealed much about the regulation of the human globin genes. In this section, we will summarize some of the information about DNA sequences needed in *cis* (i.e. on the same chromosome) for regulation of the globin genes. Chapter 4 will cover the proteins interacting with these regulatory DNA sequences.

Common versus lineage-specific regulation

Comparison of noncoding genomic DNA sequences among related species is a powerful approach to identifying and better understanding *cis*-regulatory modules (CRMs). However, it is important to distinguish what is similar and what is distinctive about the patterns of regulated expression of the genes in the species being compared. If one is searching for CRMs that carry out a function common to most or all mammals, then conservation across all mammals and evidence of strong constraint in noncoding DNA will provide good candidates for further experimental tests (e.g. 45, 46, 47). Such constrained noncoding sequences can have within them short, almost invariant regions that frequently correspond to transcription factor binding sites. These have been called phylogenetic footprints (48). However, if one is studying a type of regulation that only

occurs in higher primates, then searching for sequences conserved in other mammalian orders will be futile. Instead, the search should focus on sequences conserved in the species with a common mode of regulation but which differ from the homologous regions in species with a different regulation. These have been called differential phylogenetic footprints (49).

Regulatory features of globin genes common to many vertebrate species include tissue specificity and some aspects of developmental specificity. Expression of the α-like and β-like globin genes in all vertebrate species examined is restricted to the erythroid lineage. Thus some determinants of tissue specificity should be common to all these genes. One example is binding by the transcription factor GATA-1. As will be detailed in the following sections, either the promoter, enhancers or both for all globin genes have binding sites for GATA-1. Another feature common to all mammals is the expression of the ε-globin and ζ-globin genes exclusively in primitive erythroid cells, which are produced during embryonic life. Thus one might expect determinants of embryonic expression to be conserved in many species. Indeed, conservation of the upstream promoter regions of these genes in eutherian mammals is more extensive than is seen for other promoters in their globin gene clusters (50).

An example of lineage-specific regulation is the recruitment of the γ-globin genes for expression in fetal erythroid cells. In most eutherian mammals, the γ-globin genes are expressed in primitive erythroid cells, similar to the ε-globin gene, and the β-globin gene is expressed in definitive erythroid cells both during fetal and adult life. However, simian primates, including humans, express the γ-globin genes during fetal erythropoiesis, and the expression of the β-globin gene is delayed. The extent of delay varies in different primate clades, but in humans it is largely delayed until just before birth. Thus when examining interspecies alignments of the regulatory regions of the β-globin gene (*HBB*) and the γ-globin genes (*HBG1* and *HBG2*), one will be seeing a combination of CRMs used in common (e.g. for adult erythroid expression of *HBB*) and in a lineage-specific manner (e.g. fetal expression of *HBG1*).

Quantitative analysis of sequence alignments

Alignments of genomic DNA sequences reveal the segments that are similar between species, and often these reflect homology (descent from a common ancestor). These sequence matches tend to have highest similarity in the protein-coding exons, but significant stretches of noncoding sequences also align between mammalian species (for globin gene complexes, see (51-53). Further analysis is required to discern which sequence matches simply reflect common ancestry (aligned neutral DNA) versus those in sequences that are under constraint (sequences with a common function) (54, 55).

Several bioinformatic tools have been developed to help interpret the alignments of multiple sequences. Results from two of these, each analyzing alignments of several mammals (human, chimpanzee, rhesus macaque, mouse, rat, dog, cow, and sometimes additional ones), are shown in Figure 3.3. The Conservation track plots the phastCons score at each position of the human sequence. This score is an estimate of the posterior probability that a given nucleotide is in the most strongly constrained (i.e. most slowly changing) portion of the genome (56). Higher scores are associated with a greater likelihood that a position or region is under strong purifying selection. Sequences that are needed for a feature that is common to these several placental mammals would be expected to have a high Conservation score.

A discriminatory analysis of the multiple alignments was used to generate a Regulatory Potential (RP) score (57). This machine-learning approach estimates the likelihood that a given aligning segment is a CRM, given the frequency of patterns in the alignments that are distinctive for CRMs as opposed to neutral DNA. The patterns are strings of alignment columns, and their discriminatory power is determined by the frequency of the patterns in training sets of alignments in CRMs versus alignments in neutral DNA. Although the RP score is influenced by features in addition to constraint, it is designed for finding CRMs that are common among species.

Basal promoters

Promoters are DNA sequences needed for accurate initiation of transcription. For some promoters including the globin gene promoters, one DNA segment interacts with RNA polymerase II and its accessory factors (such as TFIID and TFIIB) to determine the start site of transcription; this is the basal promoter (58). Five motifs have been associated with basal promoters, and these are found in the promoters of human globin genes (Figure 3.4.A). They include the familiar TATA box to which TBP binds, along with the BRE to which TFIIB binds and the Inr and DPE motifs to which components of TFIID binds (58).

Early studies revealed the presence of the ATAAA motif about 25-30 bp 5' to the start site of transcription of the globin genes (59), and this is by far the most restricted in its consensus, i.e. this motif appears to be under evolutionary constraint in globin genes. Recent studies on other promoters are revealing the roles of additional motifs close to the start site of transcription, but on both sides. Matches to these motifs can be found readily at the appropriate positions in the human globin genes (Figure 3.4.A). The motifs other than TATA do not have well-defined consensus sequences, either for genes in general or for the human globin genes, and thus their presence alone may not signify function. Also, only the TATA box, Inr and DPE show evidence of constraint in homologs in other mammalian species (Figure 3.5.A, conservation track). However, each of the motifs except BRE has been implicated in function by finding a mutation in at least one case of β-thalassemia. Every base in the TATA box has been altered in one or another thalassemia, and mutations in Inr, MTE and DPE also are associated with thalassemia (Figure 3.5.A, Compilation of Human Disease Variants and Other Mutations). The BRE overlaps with the β-direct repeat element (βDRE), which is a *cis*-regulatory element bound by βDRF and demonstrated to function in regulation of the β-globin gene by mutagenesis and expression in transfected cells (60). Thus, the mutagenesis data (natural and directed) indicate that all five motifs are important for appropriate expression of the β-globin gene. The presence of similar motifs in the basal promoters for other human globin genes suggests that they are active in these genes as well.

Although it is common to describe promoters recognized by RNA polymerase II by the motifs shown in Figure 3.4.A, it is important to realize that this is true for only a minority of human genes. Globin gene promoters fall into the category of promoters with well-defined TATA boxes at a restricted location and one major start site for transcription. Recent studies show that these comprise a small minority of promoters, perhaps only 10 to 20%. Most promoters are CpG islands with no obvious TATA box, and in some cases they have a broad distribution of start sites (61).

Upstream regulatory sequences

Adjacent to the basal promoter is the upstream regulatory region (58) , which in globin genes runs from about positions -40 to -250 (Figure 3.4.B). Only one motif in this region is found in all the highly expressed globin genes: the CCAAT box. Proteins such as NF-Y and CP1 bind to this motif (62, 63), and it has been implicated in promoter function because of its presence in many promoters and the results of mutagenesis and binding studies (59). It is missing from the δ-globin gene (*HBD*) promoter, but this gene is expressed at a low level (about 1-2% of *HBB*).

Two motifs are found in many but not all promoters. One is the CACC box, which is bound by transcription factors in the Krüppel-like zinc finger class (KLF). The first erythroid KLF discovered was EKLF, which binds to the CACC box in the *HBB* promoter and is needed for erythropoiesis (64, 65). The CACC boxes in globin promoters tend to be highly conserved in other mammals, albeit not as constrained as the CCAAT box (Figure 5.5.B). Mutations in almost every position in the proximal CACC box have been associated with β-thalassemia (Figure 5.5.B). Thus many lines of evidence point to the importance of this motif. Other KLFs may bind to the CACC boxes in other globin gene promoters, such as FKLF or KLF13 (66) for the *HBG1* and *HBG2* promoters.

The other motif occurring frequently in upstream regulatory regions is WGATAR, the binding site for GATA-1 and related proteins (Figure 3.4.B). GATA-1 plays a critical role in erythroid-specific gene activation and repression (67-69), and the binding sites in these upstream regions have been implicated in positive regulation of the

respective genes (70, 71). The GATA-1 binding sites upstream of *HBE1*, *HBG1*, *HBG2* and *HBZ2* are conserved in most mammals, but the ones upstream of *HBB* are not. GATA-1 binds to the promoter regions of β-globin genes in both human (63) and mouse (72), but the binding site motif occurs in different places in the two promoters (73). This is an example of alterations in the binding site being associated with changes in the pattern of regulation, e.g. the delay in onset of expression in humans.

A different set of binding sites is distinctive to each type of gene. For instance, βDRF (60) and BB1-binding protein (72, 74) have been implicated in the regulation of the β-globin gene but not other globin genes (Figure 3.4.B). Both binding sites are conserved in many placental mammals (Figure 3.5.B and 73). Likewise, binding of OCT1 and γPE has been shown for the upstream regions of g-globin genes but not others (75).

The *cis*-elements close to the γ-globin genes are key determinants of fetal versus embryonic expression. One of the clearest demonstrations of this is from transgenic mouse experiments using a construct containing an LCR to enhance expression of globin genes. The γ-globin gene of prosimians, e.g. the bush-baby galago, is expressed embryonically, and when it is included in the test construct in transgenic mice, the transgene is also expressed embryonically. In contrast, a human γ-globin gene, normally expressed during fetal life in humans, is expressed fetally when transferred into transgenic mice in an otherwise identical construct (76). Thus one would expect to find alterations in the regulatory regions of anthropoid (monkey, ape and humans γ-globin genes that are associated with this change in stage-specificity (i.e., sequences that are conserved in anthropoid primates but are different in prosimians and nonprimate mammals). Examination of aligned sequences for differential phylogenetic footprints (49) led to the identification of a stage selector element (SSE) in the human γ-globin gene promoter (Fig. 3.4.B). The SSE is a binding site for a factor called the stage-selector protein, or SSP, which has been implicated in the differential expression of γ- and β-globin genes (77). Additional DNA sequence that binds several proteins implicated in fetal silencing of the γ-globin gene (49). Parallel protein-binding and mutagenesis studies led to the discovery of a novel protein that binds to an element called the γPE the

upstream regulatory region of the γ-globin genes, which has also been implicated in regulation of this gene (75).

The most distinctive globin gene promoters are those of the α-globin genes (*HBA1* and *HBA2*). These promoters are CpG islands, and among the hemoglobin genes, only those encoding α-globin have this feature. (The θ-globin and μ-globin genes also have promoters in CpG islands, but as discussed above, it is not clear that they encode components of hemoglobin.) While the majority of mammalian promoters are CpG islands (61), most of the associated genes are expressed in multiple tissues and few if any are expressed at such a high level as the α-globin gene. Thus the presence of a CpG island in the promoter for a globin gene is curious, and it leads to several unanswered questions about the α-globin gene promoters. What prevents their expression in nonerythroid tissues? What sequences in addition to the CpG island lead to very high level expression in erythroid cells? No GATA-1 binding site is found in the α-globin gene promoters of most placental mammals (the mouse α-globin genes is a notable exception), so sequence-directed binding of this protein to the proximal sequences is not the answer. Several studies have shown that the CpG island is a key component of the *cis*-regulatory elements for the α-globin gene of humans and rabbits, possibly through its effects on chromatin structure (78, 79).

The differences in the arrays of proteins functioning at ε-, γ-, β- and α-globin genes indicate that a distinct battery of proteins functions in the promoter for each type of gene. Indeed, this is consistent with the observation that *cis*-acting sequences needed for stage-specific regulation of expression map close to the genes (80).

Proximal Enhancers

Enhancers are DNA sequences that increase the activity of promoters; they can be located on either side of a gene or internal to it, and they can act at considerable distances from genes (81). Two enhancers have been found close to genes in the β-globin gene cluster, one that is 3' to *HBB* and one that is 3' to *HBG1* (Figure 3.3.A). In both cases the enhancers are less than 1 kb downstream of the polyA additional signal for the respective genes. The *HBB* enhancer was discovered by its effect on developmental timing of

expression of globin transgenes when introduced into mice. High level expression of human γ- or β-globin transgene constructs in fetal erythroid cells (the normal onset of expression of mouse β-globin genes) is dependent on the presence of the enhancer (74, 82-84). The *HBG1* enhancer was discovered as the only DNA segment in a 22kb region surrounding the γ-globin genes for DNA segments that boosted expression of a reporter gene driven by a γ-globin gene promoter in transfected erythroid cells (85). Deletion of this enhancer from a large construct containing the human LCR and β-like globin genes had no effect on expression levels in transgenic mice (86), which could mean that it actually has no function, or that other sequences compensate for its loss, or that its function is not apparent in mice.

Indeed, comparative sequence analysis of these proximal enhancers strongly supports the conclusion that both play roles in higher primates but not in other species. As illustrated in Figure 3.4.C, both enhancers contain binding sites for GATA-1 (87, 88), and the *HBG1* enhancer also binds to the γPE protein (75). However, the DNA homologous to the *HBB* enhancer in other mammals is not strongly conserved, even in the GATA motifs. Furthermore, two of the GATA1-binding sites in the *HBG1* enhancer were introduced via an LTR-type transposable element that is present only in higher primates (Figure 3.6.A). Thus the presence of the *HBG1* proximal enhancer correlates with the fetal recruitment of γ-globin gene expression in anthropoids, and its function may not be observed in transgenic mice. Likewise, the presence of GATA-1 bindings sites only in higher primates suggests that the function of the *HBB* proximal enhancer may also be lineage-specific, perhaps related to the delay in expression of *HBB* in higher primates. In this case, an effect on developmental timing is readily demonstrable in transgenic mice, but because of the differences in timing of *HBB* expression in humans (the source of the transgene) and mouse (the host species), it is difficult to fully understand this function.

Distal Enhancers

In addition to the proximal promoters and enhancers, both the α-like and β-like globin gene clusters are regulated by distal control regions. The β-like globin cluster is

regulated by the distal LCR (reviewed in 89, 90), and the α-like globin gene cluster is regulated by HS-40 (91). In both cases, deletion of the distal control region is associated with thalassemia (Figure 3.3). Addition of the distal control regions has profound effects on expression of linked genes in transgenic mice. Without the LCR, erythroid expression of a β-globin transgene is not seen in all mouse lines (92), presumably because of integration in a repressive region of a chromosome (a position effect). With the LCR, the β-globin transgene is expressed at a high level in erythroid cells in almost all mouse lines, indicating strong enhancement and a reduction in position effects (93). HS-40 of the α-globin gene complex is a strong enhancer of globin gene expression, both in transgenic mice (91, 94) and in transfected cells (95).

The β-globin LCR is a very large regulatory region, containing at least five DNase hypersensitive sites in humans spread over about 17 kb (96-98) between *HBE1* and an *OR* gene (Figure 3.3.A). This region is highly conserved in mammals, with highly similar sequences indicative of constraint found both in the hypersensitive sites and between them (50, 90). This can be seen in Figure 3.3.A as the string of peaks of conservation and RP in this region.

The distal enhancer for the α-globin gene, HS-40, is much smaller than the LCR. It is about 250 bp in length (99), located in a widely expressed gene called *C16orf35* (Figure 3.3.B). Additional erythroid DNase hypersensitive sites are present in this large gene, but none have been shown to play a role in regulation of globin genes (26). HS-40 is sufficient for strong enhancement and high activity in erythroid cells of transgenic mice, especially during embryonic and fetal development (91). It is very strongly conserved in mammals, with obvious matches to species as distant as opossum (Figures 3.3.B and 3.6.B). Functional tests have shown that the homologous regions of chicken and fish also have enhancer activity, despite considerable divergence outside the protein-binding sites (44).

Regulatory activities in addition to tissue-specific enhancement have been attributed to the β-globin LCR, but they are not seen consistently in multiple lines of investigation (100). Examination of chromatin structure after deletion of the LCR led to the inference that the LCR is needed for tissue-specific chromosomal domain opening

(101). Chromosome 11 from a patient with the Hispanic $(\varepsilon\gamma\delta\beta)^0$ thalassemia (missing most of the LCR and some adjacent sequences) was transferred through multiple somatic cells to generate a hybrid murine erythroleukemia cell line containing the mutant human chromosome. The β-globin gene cluster in this hybrid cell line is inactive and is insensitive to DNase, indicating that the LCR is needed for opening a chromosomal domain (101). However, an engineered mouse line carrying a deletion of the mouse β-globin LCR and the sequences homologous to those lost in the Hispanic deletion retains an open chromatin conformation (accessible to DNase) in the mouse β-globin gene (102). Although expression of the mouse β-globin genes is reduced substantially, the locus is not silenced. Thus the repressive heterochromatin seen in the hybrid murine erythroleukemia cells carrying human chromosome 11 with the Hispanic deletion may have been produced during the chromosome transfers between cell lines. Currently, the DNA sequence determinants of chromatin opening have still not been discovered. The β-globin LCR has also been implicated in overcoming position effects in transgenic mice (103), in keeping with the inferred effect on opening a chromatin domain. However, transgene constructs containing the β-globin can still show position effect variegation (104). Both the β-globin LCR and the α-globin HS-40 are very strong, erythroid-specific enhancers needed for the expression of any of the linked globin genes. They also can overcome some but not all repressive effects after integration at a variety of chromosomal locations. This could be a consequence of the strong enhancement.

Three transcription factor-binding motifs are present in almost all DNase hypersensitive sites that have a strong function in the distal enhancers (Figure 3.4.D). All contain Maf-response elements, or MAREs, to which transcriptional activator proteins of the basic leucine zipper class can bind (105). A subfamily of proteins related to AP1, such as NF-E2, LCRF1/Nrf1, and Bach1, bind to this element (reviewed in 106, 107). All are heterodimers containing a Maf protein as one subunit, which is the basis for the name of the response element. All the hypersensitive sites have GATA motifs, to which GATA-1 and its relatives bind (108). The third common motif is CACC, to which a family of Zn-finger proteins including erythroid Krüppel-like factor (EKLF) can bind (64). At HS3 in the β-globin LCR, there is evidence that motifs related to CACC are

bound by additional Krüppel-like factors, such as Sp1 (109). HS2 of the β-globin LCR also has three E-boxes , which are the binding sites for TAL-1 and its heterodimeric partners (47). This protein has been implicated in regulation of hematopoiesis, and it appears to also play a role in enhancement by HS2.

Initial studies of protein binding at these and other CRMs used various *in vitro* methods and *in vivo* footprinting (99, 110-112). Recent experiments using chromatin immunoprecipitation have demonstrated occupancy of the CRMs by several of these proteins in erythroid cells (e.g., 113, 114-116). Many of the sites have implicated directly in activity by mutagenesis and gene transfer (e.g., 47, e.g., 117, 118, 119).

The protein binding sites in the distal positive regulators show some common patterns (Figure 3.4.D). A MARE plus two GATA motifs is present in most of the CRMs, and this arrangement has been shown to be needed for formation of a hypersensitive site at HS4 (120). The strongest enhancers (as assayed by gene transfer in somatic cells) are HS2 and HS-40. Both of these have two MAREs, and mutation of those MAREs removes much of the enhancing activity (117, 119). Thus the MAREs and proteins binding to them are critical for high-level enhancement, but the other binding sites contribute to function as well.

The CRMs marked by these hypersensitive sites in the distal positive regulators are conserved across almost all mammals (26, 90). The portion of the alignments for HS-40 shown in Figure 3.6.B indicates the very strong constraint seen in the known binding sites and additional short segments both for this enhancer and for HS2. Most of the binding sites in HS3 are also highly conserved, but some are not, likely reflecting both common and lineage-specific functions. HS4, with the MARE and two GATA motifs, is conserved across a wide span of placental mammals, but this DNA sequence is part of an LTR-type repeat, a member of the ERV1 repeat family. This appears to be an old transposable element (predating most of the mammalian radiation), but one that continues to provide a regulatory function.

CONCLUDING REMARKS

Molecular clones containing mammalian globin gene clusters were isolated about 30 years ago. Intense study since then has revealed much about their structure, evolution and regulation. However, understanding sufficient to lead to clinical applications continues to elude us. The myriad levels of regulation and function that operate within these gene clusters certainly confound attempts to find simplifying conclusions. Despite these challenges, studies of the globin gene clusters have consistently provided new insights into function, regulation and evolution. The lessons being learned as we try to integrate information from classical molecular biology and genetics, new high through-put biochemical assays, and extensive interspecies sequence comparisons are paving the way for applying these approaches genome-wide. The globin gene clusters illustrate the need to distinguish common from lineage-specific regulation. Although simple generalizations are rare, the extensive information that one needs for interpreting data in the context of comparative genomics is readily accessible. Throughout this chapter, we have illustrated points using output from the UCSC Genome Browser (http://genome.ucsc.edu), with special emphasis on the tracks showing Conservation, Regulatory Potential, and Locus Variants. Deeper information on the variants associated with disorders of the hemoglobins can be obtained from HbVar (http://www.bx.psu.edu). We hope that the examples presented here will be helpful in guiding interpretation of the multitude of data available to the readers now and in the future.

ACKNOWLEDGEMENTS

RH was supported by NIH grant R01 DK065806 and BGF was supported by NIH grants R01 DK19482 and P01 HL63357.

REFERENCES

1. Weatherall DJ and Clegg JB. Thalassemia Syndromes. 3rd. ed. 1981, Oxford: Blackwell Scientific Publications.
2. Baglioni C. The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. Proc Natl Acad Sci U S A 1962;48:1880-6.
3. Kendall AG, Ojwang PJ, Schroeder WA, and Huisman TH. Hemoglobin Kenya, the product of a gamma-beta fusion gene: studies of the family. Am J Hum Genet 1973;25:548-63.
4. Deisseroth A, Nienhuis A, Turner P, et al. Localization of the human alpha globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. Cell 1977;12:205-18.
5. Deisseroth A, Nienhuis AW, Lawrence J, Giles RE, Turner P, and Ruddle FH. Chromosomal localization of the human beta globin gene to human chromosome 11 in somatic cell hybrids. Proc Nat Acad Sci, USA 1978;75:1456-60.
6. Tilghman SM, Tiemeier DC, Seidman JG, et al. Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. Proc Natl Acad Sci USA 1978;75:725-9.
7. Goodman M, Czelusniak J, Koop B, Tagle D, and Slightom J. Globins: A case study in molecular phylogeny. Cold Spring Harbor Symp Quant Biol 1987;52:875-90.
8. Proudfoot NJ, Gil A, and Maniatis T. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 1982;31:553-63.
9. Tilghman SM, Curtis PJ, Tiemeier DC, Leder P, and Weissmann C. The intervening sequence of a mouse beta-globin gene is transcribed within the 15S beta-globin mRNA precursor. Proc Natl Acad Sci USA 1978;75:1309-13.
10. Krainer AR, Maniatis T, Ruskin B, and Green MR. Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell 1984;36:993-1005.
11. Mount SM. A catalogue of splice junction sequences. Nucleic Acids Res 1982;10:459-72.

12. Padgett RA, Grabowski PJ, Konarska MM, Seiler S, and Sharp PA. Splicing of messenger RNA precursors. Annu Rev Biochem 1986;55:1119-50.
13. Patrinos GP, Giardine B, Riemer C, et al. Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res 2004;32 Database issue:D537-D41.
14. Giardine B, Riemer C, Hefferon T, et al. PhenCode: connecting ENCODE data with mutations and phenotype. Hum Mutat 2007;28:554-62.
15. Jeffreys AJ and Flavell RA. The rabbit beta-globin gene contains a large large insert in the coding sequence. Cell 1977;12:1097-108.
16. Tuan D, Biro PA, deRiel JK, Lazarus H, and Forget BG. Restriction endonuclease mapping of the human gamma globin gene loci. Nucleic Acids Res 1979;6:2519-44.
17. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975;98:503-17.
18. Fritsch E, Lawn R, and Maniatis T. Molecular cloning and characterization of the human beta-like globin gene cluster. Cell 1980;19:959-72.
19. Zhang Z and Gerstein M. Large-scale analysis of pseudogenes in the human genome. Curr Opin Genet Dev 2004;14:328-35.
20. Hsu S, Marks J, Shaw J, et al. Structure and expression of the human theta 1 globin gene. Nature 1988;331:94-6.
21. Ley TJ, Maloney KA, Gordon JI, and Schwartz AL. Globin gene expression in erythroid human fetal liver cells. J Clin Invest 1989;83:1032-8.
22. Albitar M, Peschle C, and Liebhaber SA. Theta, zeta and epsilon globin messenger RNA are expressed in adults. Blood 1989;74:629-37.
23. Goh SH, Lee YT, Bhanu NV, et al. A newly discovered human alpha-globin gene. Blood 2005;106:1466-72.
24. Cooper SJ, Wheeler D, De Leo A, et al. The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation. Mol Phylogenet Evol 2006;38:439-48.

25. Hardison RC, Sawada I, Cheng J-F, Shen C-KJ, and Schmid CW. A previously undetected pseudogene in the human alpha globin gene cluster. Nucleics Acids Research 1986;14:1903-11.
26. Hughes JR, Cheng JF, Ventress N, et al. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci USA 2005;102:9830-5.
27. Clegg JB. Can the product of the theta gene be a real globin? Nature 1987;329:465-6.
28. Craddock CF, Vyas P, Sharpe JA, Ayyub H, Wood WG, and Higgs DR. Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. EMBO J 1995;14:1718-26.
29. Lander ES,Linton LM,Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
30. Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986;321:209-13.
31. Collins FS and Weissman SM. The molecular genetics of human hemoglobin. Prog Nucl Acids Res & Mol Biol 1984;31:315-462.
32. Fischel-Ghodsian N, Nicholls RD, and Higgs DR. Unusual features of CpG-rich (HTF) islands in the human α-globin complex: association with nonfunctional pseudogenes and presence within the 3' portion of the ζ genes. Nucl Acids Res 1987;15:9215-25.
33. Groudine M, Kohwi-Shigematsu T, Gelinas R, Stamatoyannopoylos G, and Papyannopoulou T. Human fetal to adult hemoglobin switching: Changes in chromatin structure of the β-globin gene locus. Proc Natl Acad Sci, USA 1983;80:7551-5.
34. van der Ploeg LHT and Flavell RA. DNA methylation in the human γ-δ-β globin locus in erythroid and nonerythroid tissues. Cell 1980;19:947-58.

35. Bird A, Taggart M, Nicholls R, and Higgs D. Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene. EMBO J 1987;6:999-1004.
36. Epner E, Rifkind RA, and Marks PA. Replication of alpha and beta globin DNA sequences occurs during early S phase in murine erythroleukemia cells. Proc Natl Acad Sci USA 1981;78:3058-62.
37. Goldman MA, Holmquist GP, Gray MC, Caston LA, and Nag A. Replication timing of genes and middle repetitive sequences. Science 1984;224:686-92.
38. Dhar V, Mager D, Iqbal A, and Schildkraut CL. The co-ordinate replication of the human β-globin gene domain reflects its transcriptional activity and nuclease hypersensitivity. Mol Cell Biol 1988;8:4958-65.
39. Bulger M, Bender MA, von Doorninck JH, et al. Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse β-globin gene clusters. Proc Natl Acad Sci, USA 2000;97:14560-5.
40. Feingold EA and Forget BG. The breakpoint of a large deletion causing hereditary persistence of fetal hemoglobin occurs within an erythroid DNA domain remote from the beta-globin gene cluster. Blood 1989;74:2178-86.
41. Anagnou NP, Perez-Stable C, Gelinas R, et al. Sequences located 3' to the breakpoint of the hereditary persistence of fetal hemoglobin-3 deletion exhibit enhancer activity and can modify the developmental expression of the human fetal A gamma-globin gene in transgenic mice. J Biol Chem 1995;270:10256-63.
42. Flint J, Thomas K, Micklem G, et al. The relationship between chromosome structure and function at a human telomeric region. Nature Genetics 1997;15:252-7.
43. Anguita E, Johnson CA, Wood WG, Turner BM, and Higgs DR. Identification of a conserved erythroid specific domain of histone acetylation across the alpha-globin gene cluster. Proc Natl Acad Sci USA 2001;98:12114-9.
44. Flint J, Tufarelli C, Peden J, et al. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum Mol Genet 2001;10:371-82.

45. Gumucio DL, Heilstedt-Williamson H, Gray TA, et al. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human γ and ε globin genes. Mol Cell Biol 1992;12:4919-29.
46. Gumucio D, Shelton D, Zhu W, et al. Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the beta-like globin genes. Mol Phylog and Evol 1996;5:18-32.
47. Elnitski L. Conserved E boxes in the locus control region contribute to enhanced expression of beta-globin genes via TAL1 and other basic helix-loop-helix proteins. 1998, The Pennsylvania State University.
48. Tagle DA, Koop BF, Goodman M, Slightom J, Hess DL, and Jones RT. Embryonic ε and γ globin genes of a prosimian primate (*Galago crassicaudatus*): Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol 1988;203:7469-80.
49. Gumucio DL, Shelton DA, Blanchard-McQuate K, et al. Differential phylogenetic footprinting as a means to identify base changes responsible for recruitment of the anthropoid γ gene to a fetal expression pattern. J Biol Chem 1994;269:15371-80.
50. Hardison R and Miller W. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol Biol Evol 1993;10:73-102.
51. Margot JB, Demers GW, and Hardison RC. Complete nucleotide sequence of the rabbit beta-like globin gene cluster: Analysis of intergenic sequences and comparison with the human beta-like globin gene cluster. J Mol Biol 1989;205:15-40.
52. Shehee R, Loeb DD, Adey NB, et al. Nucleotide sequence of the BALB/c mouse β-globin complex. J Mol Biol 1989;205:41-62.
53. Hardison R, Krane D, Vandenbergh D, et al. Sequence and comparative analysis of the rabbit alpha-like globin gene cluster reveals a rapid mode of evolution in a G + C-rich region of mammalian genomes. J Mol Biol 1991;222:233-49.
54. Hardison RC. The nucleotide sequence of the rabbit embryonic globin gene β4. J Biol Chem 1983;258:8739-44.

55. Cooper GM, Brudno M, Stone EA, Dubchak I, Batzoglou S, and Sidow A. Characterization of evolutionary rates and constraints in three Mammalian genomes. Genome Res 2004;14:539-48.
56. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034-50.
57. Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, and Chiaromonte F. ESPERR: Learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 2006;16:1596-604.
58. Maston GA, Evans SK, and Green MR. Transcriptional Regulatory Elements in the Human Genome. Annu Rev Genomics Hum Genet 2006;7:29-59.
59. Efstratiadis A, Posakony JW, Maniatis T, et al. The structure and evolution of the human β-globin gene family. Cell 1980;21:653-68.
60. Stuve LL and Myers RM. A directly repeated sequence in the β-globin promoter regulates transcription in murine erythroleukemia cells. Mol Cell Biol 1990;10:972-81.
61. Carninci P, Sandelin A, Lenhard B, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 2006;38:626-35.
62. Cohen RB, Sheffery M, and Kim CG. Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α1-globin gene. Mol Cell Biol 1986;6:821-32.
63. deBoer E, Antoniou M, Mignotte V, Wall L, and Grosveld F. The human β-globin promoter; nuclear protein factors and erythroid specific induction of transcription. EMBO J 1988;7:4203-12.
64. Miller IJ and Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the *Kruppel* family of nuclear factors. Mol Cell Biol 1993;13:2776-86.
65. Perkins AC, Sharpe AH, and Orkin SH. Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature 1995;375:318-22.

66. Asano H, Li XS, and Stamatoyannopoulos G. FKLF, a novel Kruppel-like factor that activates human embryonic and fetal beta-like globin genes. Mol Cell Biol 1999;19:3571-9.
67. Pevny L, Simon MC, Robertson E, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature 1991;349:257-60.
68. Simon MC, Pevny L, Wiles MV, Keller G, Costantini F, and Orkin SH. Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. Nat Genet 1992;1:92-8.
69. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. Blood 2004;104:3136-47.
70. Martin D and Orkin S. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes & Dev 1990;4:1886-98.
71. Gong Q-H and Dean A. Enhancer-dependent transcripion of the ε-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NF-E2. Mol Cell Biol 1993;13:911-7.
72. Macleod K and Plumb M. Derepression of mouse β-major-globin gene transcription during erythroid differentiation. Mol Cell Biol 1991;11:4324-32.
73. Hardison R, Chao K-M, Schwartz S, Stojanovic N, Ganetsky M, and Miller W. Globin gene server: A prototype E-mail database server featuring extensive multiple alignments and data compilation. Genomics 1994;21:344-53.
74. Antoniou M, deBoer E, Habets G, and Grosveld F. The human β-globin gene contains multiple regulatory regions: Identification of one promoter and two downstream enhancers. EMBO J 1988;7:377-84.
75. Lloyd JA, Case SS, Ponce E, and Lingrel JB. Positive transcriptional regulation of the human γ-globin gene: γPE is a novel nuclear factor with multiple binding sites near the gene. J Biol Chem 1994;269:26-34.
76. TomHon C, Zhu W, Millinoff D, et al. Evolution of a fetal expression pattern via *cis*-changes near the γ-globin gene. J Biol Chem 1997;272:14062-6.

77. Jane SM, Ney PA, Vanin EF, Gumucio DL, and Nienhuis AW. Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the β-promoter. EMBO J 1992;11:2961-9.
78. Pondel M, Murphy S, Pearson L, Craddock C, and Proudfoot N. Sp1 functions in a chromatin-dependent manner to augment human alpha-globin promoter activity. Proc Natl Acad Sci USA 1995;92:7237-41.
79. Shewchuk BM and Hardison RC. CpG islands from the α-globin gene cluster increase gene expression in an integration-dependent manner. Mol Cell Biol 1997;17:5856-66.
80. Trudel M, Magram J, Bruckner L, and Costantini F. Upstream G gamma-globin and downstream beta-globin sequences required for stage-specific expression in transgenic mice. Mol Cell Biol 1987;7:4024-9.
81. Tjian R and Maniatis T. Transcriptional activation: A complex puzzle with few easy pieces. Cell 1994;77:5-8.
82. Trudel M and Costantini F. A 3' enhancer contributes to the stage-specific expression of the human β-globin gene. Genes & Devel 1987;1:954-61.
83. Behringer RR, Hammer RE, Brinster RL, Palmiter RD, and Townes TM. Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. Proc Natl Acad Sci USA 1987;84:7056-60.
84. Liu Q, Bungert J, and Engel JD. Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. Proc Natl Acad Sci USA 1997;94:169-74.
85. Bodine D and Ley T. An enhancer element lies 3' to the human A gamma globin gene. EMBO J 1987;6:2997-3004.
86. Liu Q, Tanimoto K, Bungert J, and Engel JD. The A gamma-globin 3' element provides no unique function(s) for human beta-globin locus gene regulation. Proc Natl Acad Sci USA 1998;95:9944-9.

87. Wall L, deBoer E, and Grosveld F. The human β-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes & Devel 1988;2:1089-100.

88. Puruker M, Bodine D, Lin H, McDonagh K, and Nienhuis AW. Structure and function of the enhancer 3' to the human Aγ-globin gene. Nucleic Acids Res 1990;18:407-7415.

89. Grosveld F, Antoniou M, Berry M, et al. The regulation of human globin gene switching. Philos Trans R Soc Lond 1993;339:183-91.

90. Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, and Miller W. Locus control regions of mammalian β-globin gene clusters: Combining phylogenetic analyses and experimental results to gain functional insights. Gene 1997;205:73-94.

91. Higgs D, Wood W, Jarman A, et al. A major positive regulatory region located far upstream of the human α-globin gene locus. Genes & Devel 1990;4:1588-601.

92. Chada K, Magram J, and Costantini F. Tissue- and stage-specific expression of a cloned adult beta globin gene in transgenic mice. Prog Clin Biol Res 1985;191:305-19.

93. Grosveld F, van Assendelft GB, Greaves D, and Kollias G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell 1987;51:975-85.

94. Sharpe JA, Chan-Thomas PS, Lida J, Ayyub H, Wood WG, and Higgs DR. Analysis of the human α-globin upstream regulatory element (HS-40) in transgenic mice. EMBO J 1992;11:4565-72.

95. Ren S, Luo X-n, and Atweh G. The major regulatory element upstream of the α-globin gene has classical and inducible enhancer activity. Blood 1993;81:1058-66.

96. Tuan D, Abelovich A, Lee-Oldham M, and Lee D. Identification of regulatory elements of human β-like globin genes, In: Stamatoyannopoulos G and Nienhuis AW, Editors Developmental Control of Globin Gene Expression. 1987, A. R. Liss, Inc.: New York. 211-20.

97. Forrester W, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, and Groudine M. Evidence for a locus activating region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucl Acids Res 1987;15:10159-77.
98. Dhar V, Nandi A, Schildkraut CL, and Skoultchi AI. Erythroid-specific nuclease-hypersensitive sites flanking the human β-globin gene cluster. Mol Cell Biol 1990;10:4324-33.
99. Jarman A, Wood W, Sharpe J, Gourdon G, Ayyub H, and Higgs D. Characterization of the major regulatory element upstream of the human α-globin gene cluster. Mol Cell Biol 1991;11:4679-89.
100. Higgs DR. Do LCRs open chromatin domains? Cell 1998;95:299-302.
101. Forrester WC, Epner E, Driscoll MC, et al. A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus. Genes & Devel 1990;4:1637-49.
102. Bender MA, Byron R, Ragoczy T, Telling A, Bulger M, and Groudine M. Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for beta-globin transcription. Blood 2006;108:1395-401.
103. Fraser P, Hurst J, Collis P, and Grosveld F. DNase I hypersensitive sites 1, 2 and 3 of the human β-globin dominant control region direct position-independent expression. Nucleic Acids Res 1990;18:3503-8.
104. Alami R, Greally JM, Tanimoto K, et al. beta-globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. Hum Mol Genet 2000;9:631-6.
105. Motohashi H, Shavit JA, Igarashi K, Yamamoto M, and Engel JD. The world according to Maf. Nucleic Acids Res 1997;25:2953-9.
106. Orkin S. Regulation of globin gene expression in erythroid cells. Eur J Biochem 1995;231:271-81.
107. Baron MH. Transcriptional control of globin gene switching during vertebrate development. Biochim Biophys Acta 1997;1351:51-72.

108. Evans T, Felsenfeld G, and Reitman M. Control of globin gene transcription. Annu Rev Cell Biol 1990;6:95-124.
109. Shelton DA, Stegman L, Hardison R, et al. Phylogenetic footprinting of hypersensitive site 3 of the β-globin locus control region. Blood 1997;89:3457-69.
110. Talbot D, Philipsen S, Fraser P, and Grosveld F. Detailed analysis of the site 3 region of the human β-globin dominant control region. EMBO J 1990;9:2169-78.
111. Strauss EC, Andrews NC, Higgs DR, and Orkin SH. In vivo footprinting of the human α-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. Mol Cell Biol 1992;12:2135-42.
112. Reddy PMS, Stamatoyannopoulos G, Papayannopoulou T, and Shen C-KJ. Genomic footprinting and sequencing of human β-globin locus: Tissue specificity and cell line artifact. J Biol Chem 1994;269:8287-95.
113. Forsberg EC, Downs KM, and Bresnick EH. Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells. Blood 2000;96:334-9.
114. Sawado T, Igarashi K, and Groudine M. Activation of beta-major globin gene transcription is associated with recruitment of NF-E2 to the beta-globin LCR and gene promoter. Proc Natl Acad Sci USA 2001;98:10226-31.
115. Letting DL, Rakowski C, Weiss MJ, and Blobel GA. Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. Mol Cell Biol 2003;23:1334-40.
116. Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WG, and Higgs DR. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J 2004;23:2841-52.
117. Ney P, Sorrentino B, McDonagh K, and Nienhuis A. Tandem AP-1-binding sites within the human β-globin dominant control region function as an inducible enhancer in erythroid cells. Genes & Devel 1990;4:993-1006.
118. Caterina JJ, Ciavatta DJ, Donze D, Behringer RR, and Townes TM. Multiple elements in human β-globin locus control region 5' HS2 are involved in enhancer

activity and position-independent transgene expression. Nucl Acids Res 1994;22:1006-11.

119. Gong Q, McDowell JC, and Dean A. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the ε-globin gene in vivo by 5' hypersensitive site 2 of the β-globin locus control region. Mol Cell Biol 1996;16:6055-64.

120. Stamatoyannopoulos JA, Goodwin A, Joyce T, and Lowrey CH. NFE2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human β-globin locus control region. EMBO J 1995;14:106-16.

121. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996-1006.

122. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, and Haussler D. The UCSC Known Genes. Bioinformatics 2006;22:1036-46.

123. Montgomery SB, Griffith OL, Sleumer MC, et al. ORegAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. Bioinformatics 2006;22:637-40.

FIGURE LEGENDS

Figure 3.1. Basic organization of human globin gene complexes. The locations of the alpha-globin gene complex very close to the telomere of the short arm of chromosome 16 and the beta-globin gene complex on the short arm of chromosome 11 are shown at the top. The genes are shown as boxes on the second line, named according to the globin polypeptide that is encoded. In both diagrams, the 5'-3' transcriptional orientation is from left to right. Note that the orientations with respect to the centromere (CEN) and telomere (TEL) are opposite; the alpha-like globin genes are transcribed toward CEN whereas the beta-like globin genes are transcribed toward TEL. The composition of hemoglobins produced at progressive developmental stages is given at the bottom.

Figure 3.2. Structure and expression pathway of globin genes.

(**A**) General structure of globin genes. The coding sequences of all globin genes in humans and other animals are separated by two introns (white boxes) into three exons. The first exon has a short 5' untranslated region (gray box) followed by a coding region (black box). All of the central exon codes for protein, while the third exon begins with coding sequences and ends with a 3' untranslated region. The relative sizes of the portions of the genes are indicated by the sizes of the boxes, and codon numbers are given above the boxes. The consensus sequence for critical sequences used in splicing are shown under the second intron of the beta-globin gene, and similar sequences are present in all introns. The vertical arrows show the splice site junctions within the consensus sequences where cleavage occurs during the process of joining the exons.

(**B**) The pathway for expression of globin genes. The RNA transcript is shown with short boxes corresponding to the untranslated regions (gray), coding regions (black) and introns (white) as in A, with processing and splicing steps occurring in the nucleus to form the mature mRNA. The mRNA is translated in the cytoplasm to generate a globin polypeptide to which the heme (gray disk) will bind. The diagram of the folded globin structure was provided by Dr. John Blamire at the Brooklyn College of the City University of New York.

Figure 3.3. Detailed maps of the human globin gene complexes, including genomic features and representative deletions.

(**A**) Detailed map of the β-like globin gene complex and surrounding olfactory receptor genes. The globin genes are named both by the encoded globin polypeptide and the official gene name. Pseudogenes are shown on a line below the genes. The known *cis*-regulatory modules are separated into distal elements such as the locus control region (shown as five DNase hypersensitive sites or HSs), promoters and enhancers close to the 3' ends of *HBG1* and *HBB*. The next two tracks show two features derived from multiple alignments of the human genomic sequence with sequences from six other placental mammals (chimp, rhesus macaque, mouse, rat, dog, and cow). The regulatory potential measures the similarity of patterns in the alignments to those that are distinctive for known regulatory regions versus neutral DNA (57). The conservation score estimates the likelihood that an alignment is in the most constrained portion of the genome, likely reflecting purifying selection (phastCons, 56). Positions of deletions that cause beta-thalassemia or Hereditary Persistence of Fetal Hemoglobin (HPFH) are shown in the lower portion.

(**B**) Detailed map of the α-like globin gene complex and surrounding genes. The conventions and tracks are similar to those in panel A. Positions of the distal erythroid HSs are from Hughes et al. (26). The deletions are grouped by those with deletion of a single alpha-globin gene (alpha-thalassemia-2), deletion of both alpha-globin genes (alpha-thalassemia-1), and a representative deletion (Ti~) that removes the distal enhancer (HS-40) but no structural genes. Coordinates of the deletions were provided by Dr. Jim Hughes.

These figures were generated starting with output from the UCSC Genome Browser (121), using the following tracks in addition to ones already mentioned: UCSC Known Genes (122), ORegAnno for *cis*-regulatory modules (123), and Locus Variants for the deletions (14). For panel A, the Genome Browser output was rotated 180^{o} so that the 5'-3' transcriptional orientation is left to right (note that the genome coordinates are decreasing from left to right). Both figures were edited for clarity. Information on

deletions and other variants is available both on the Locus Variants track as well as HbVar (13).

Figure 3.4. Motifs and binding sites in *cis*-regulatory modules of globin genes.

(**A**) Motifs in the basal promoter, based on those defined in the review by Maston et al. (58). Numbers along the top are relative to the transcription start site as +1, and ATG denotes the translation start site. The top consensus sequence is from Maston et al. (58). Corresponding positions in the globin genes are given for each motif, followed by the consensus derived for the globin genes. Symbols for ambiguous nucleotides are S = C or G, W = A or T, R = A or G, Y = C or T, D = A or G or T, H = A or C or T, V = A or C or G, and N = A or C or G or T.

(**B**) Motifs in the regulatory regions immediately upstream of the basal promoters. Motifs are indicated by sequence (CCAAT, CACC, and GATA), the name of the element (βDRE, αIRE, γPE, OCT) or the protein name followed by bs for "binding site" (BP2bs, NF1bs, BB1bs). Boxes for each motifs found in several upstream regions are shaded. The boxes were placed in the correct order but spacing is not indicated. The thick line for the *HBA* upstream regions (both *HBA1* and *HBA2*) denotes that it is a CpG island.

(**C**) Motifs in the proximal enhancers.

(**D**) Motifs in distal positive regulators, including three hypersensitive sites of the β-globin LCR and HS-40 for the α-globin gene cluster.

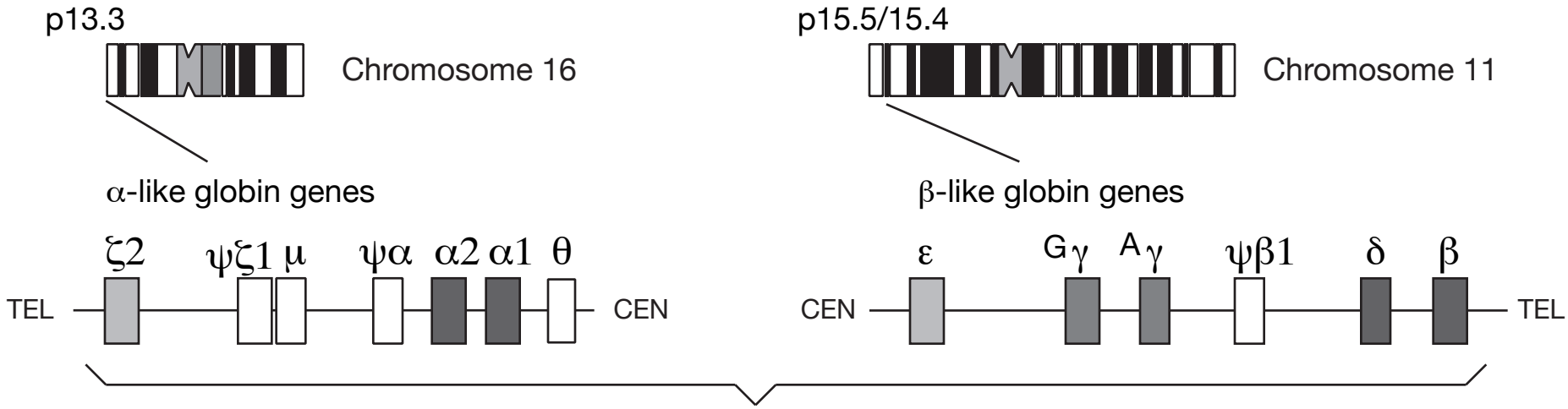
Figure 3.5. Conservation and mutations in globin gene promoters.

(**A**) Basal promoter and (**B**) Upstream promoter for *HBB*. In each panel, the sequence of an 80 bp segment is shown, along with positions of mutations associated with β-thalassemia, conservation scores, and alignments with many mammals, chicken and frog (*X. tropicalis*). The display is from the UCSC Genome Browser in genome coordinates (top line), and the direction of transcription is from right to left (opposite that used in previous figures). The start site of transcription is denoted by the vertical line leading to a leftward arrow. Boxes are drawn around motifs, which are labeled by name and proteins that bind to them (bottom line in each panel).

Figure 3.6. Wide range of conservation in globin gene enhancers.

(**A**) Proximal enhancer for *HBG1*, showing the sequence of part of the 3' enhancer, alignments with sequences of other anthropoid primates, the encompassing repetitive element, and binding motifs.

(**B**) Distal enhancer for the α-globin gene cluster, HS-40. The panel shows an 80 bp segment of the enhancer, along with the Ti~ α-thalassemia deletion that removes this DNA and more, the conservation track and alignments with several eutherian mammals and the marsupial opossum. Binding sites are boxed, and labeled by name and proteins binding to them.



Embryonic	Fetal	Adult
Hb Gower 1: $\zeta_2\varepsilon_2$	Hb F: $\alpha_2\gamma_2$	Hb A: $\alpha_2\beta_2$
Hb Gower 2: $\alpha_2\varepsilon_2$		Hb A_2: $\alpha_2\delta_2$
Hb Portland: $\zeta_2\gamma_2$		

Fig. 1

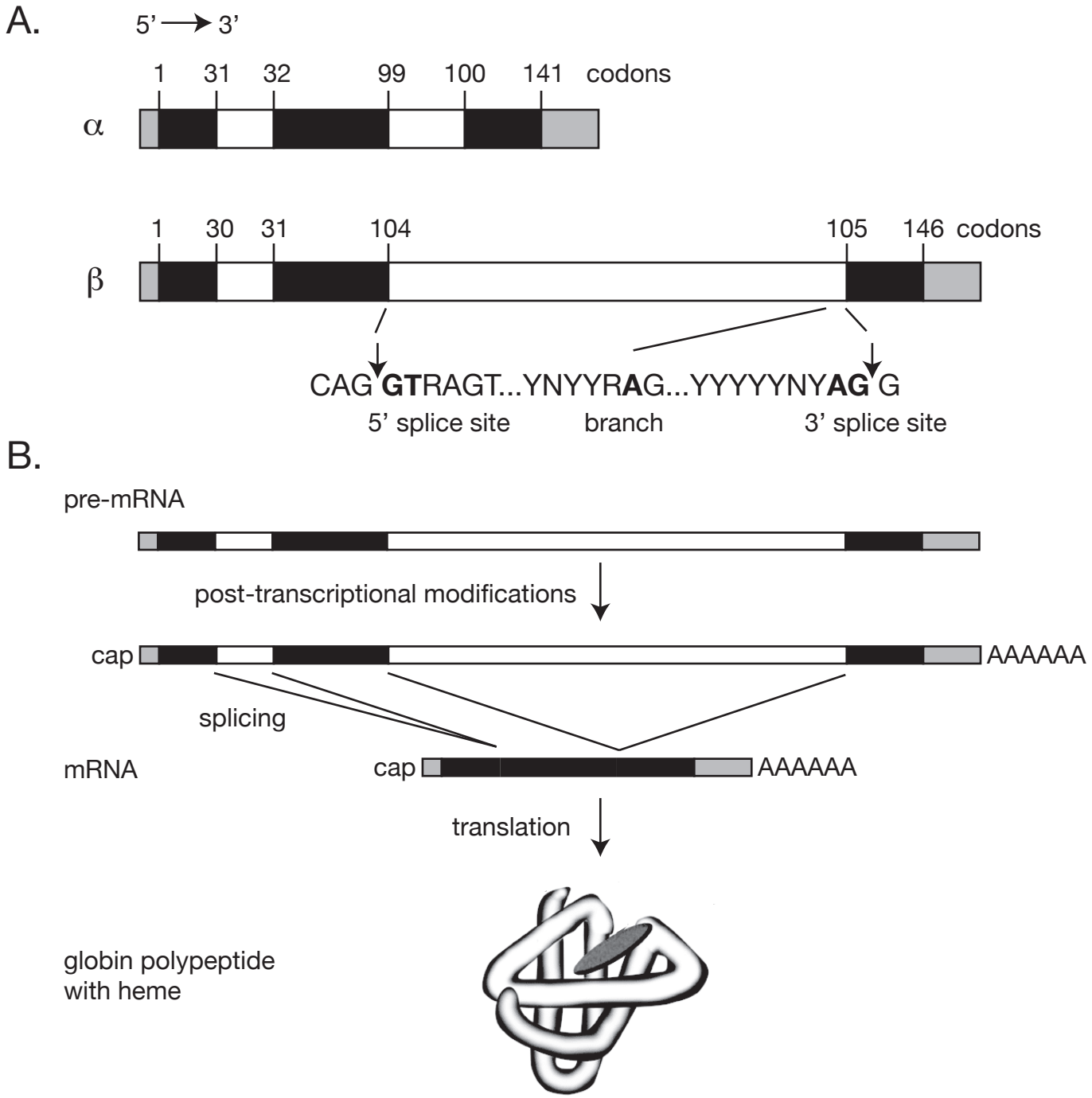

A.
5' → 3'
1 31 32 99 100 141 codons
α
1 30 31 104 105 146 codons
β
CAG GTRAGT...YNYYRAG...YYYYYNYAG G
5' splice site
branch
3' splice site
B.
pre-mRNA
post-transcriptional modifications
cap
AAAAAA
splicing
mRNA
cap
AAAAAA
translation
globin polypeptide
with heme

Fig. 2

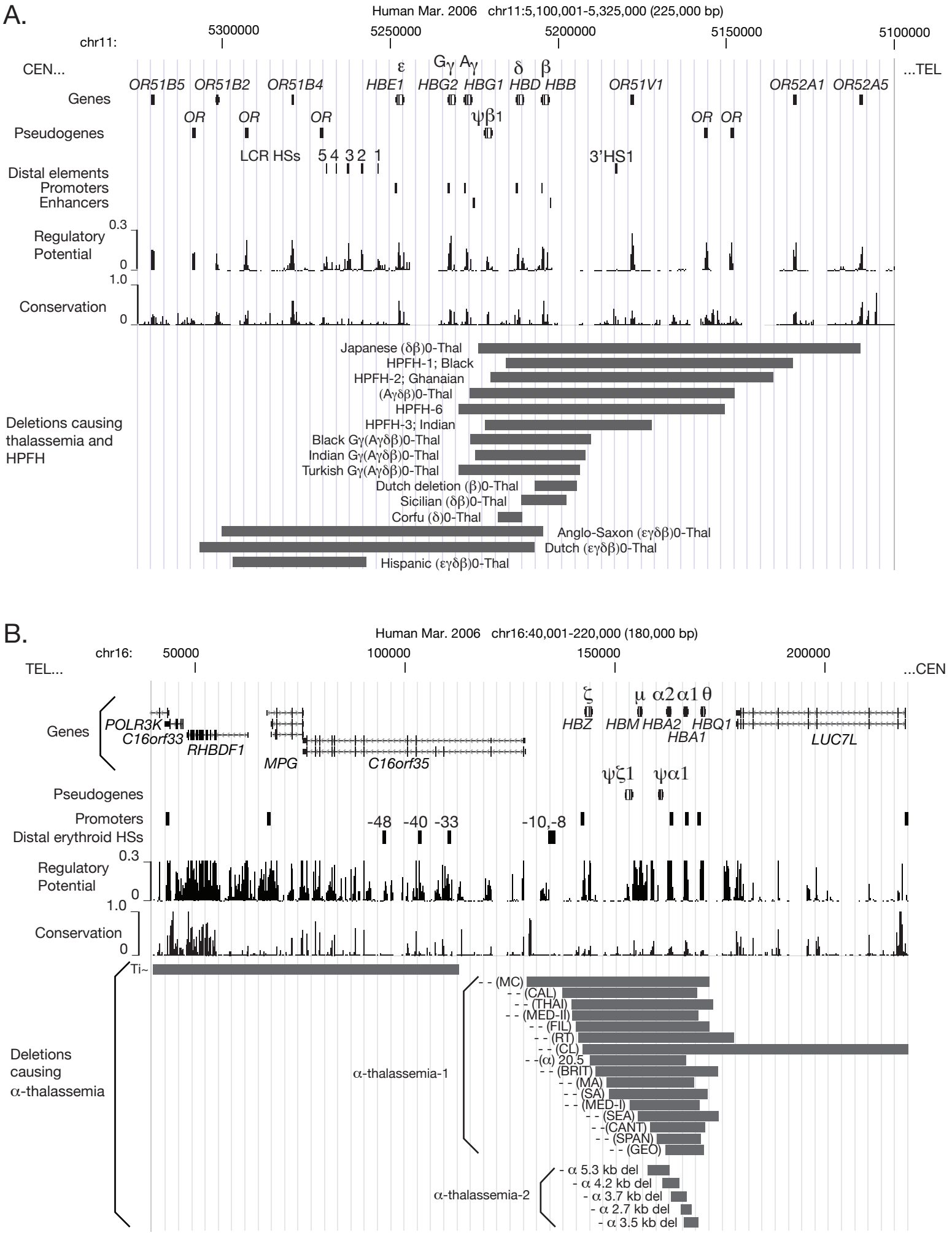
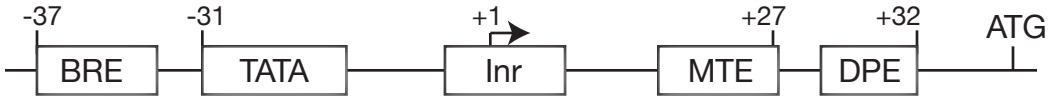

A.
Human Mar. 2006 chr11:5,100,001-5,325,000 (225,000 bp)
chr11:
5300000
5250000
5200000
5150000
5100000
CEN...
...TEL
ε Gγ Aγ δ β
OR51B5 OR51B2 OR51B4 HBE1 HBG2 HBG1 HBD HBB OR51V1 OR52A1 OR52A5
Genes
OR OR OR ψβ1 OR OR
Pseudogenes
LCR HSs 5 4 3 2 1
3'HS1
Distal elements
Promoters
Enhancers
Regulatory Potential
0.3
0
Conservation
1.0
0
Deletions causing thalassemia and HPFH
Japanese (δβ)0-Thal
HPFH-1; Black
HPFH-2; Ghanaian
(Aγδβ)0-Thal
HPFH-6
HPFH-3; Indian
Black Gγ(Aγδβ)0-Thal
Indian Gγ(Aγδβ)0-Thal
Turkish Gγ(Aγδβ)0-Thal
Dutch deletion (β)0-Thal
Sicilian (δβ)0-Thal
Corfu (δ)0-Thal
Anglo-Saxon (εγδβ)0-Thal
Dutch (εγδβ)0-Thal
Hispanic (εγδβ)0-Thal
B.
Human Mar. 2006 chr16:40,001-220,000 (180,000 bp)
chr16:
50000
100000
150000
200000
TEL...
...CEN
ζ μ α2 α1 θ
Genes
POLR3K C16orf33 RHBDF1 MPG C16orf35 HBZ HBM HBA2 HBQ1 HBA1 LUC7L
ψζ1 ψα1
Pseudogenes
Promoters
-48 -40 -33 -10,-8
Distal erythroid HSs
Regulatory Potential
0.3
0
Conservation
1.0
0
Deletions causing α-thalassemia
Ti~
- - (MC)
- - (CAL)
- - (THAI)
- - (MED-II)
- - (FIL)
- - (RT)
- - (CL)
- -(α) 20.5
- - (BRIT)
- - (MA)
- - (SA)
- - (MED-I)
- - (SEA)
- -(CANT)
- - (SPAN)
- - (GEO)
α-thalassemia-1
- α 5.3 kb del
- α 4.2 kb del
- α 3.7 kb del
- α 2.7 kb del
- α 3.5 kb del
α-thalassemia-2

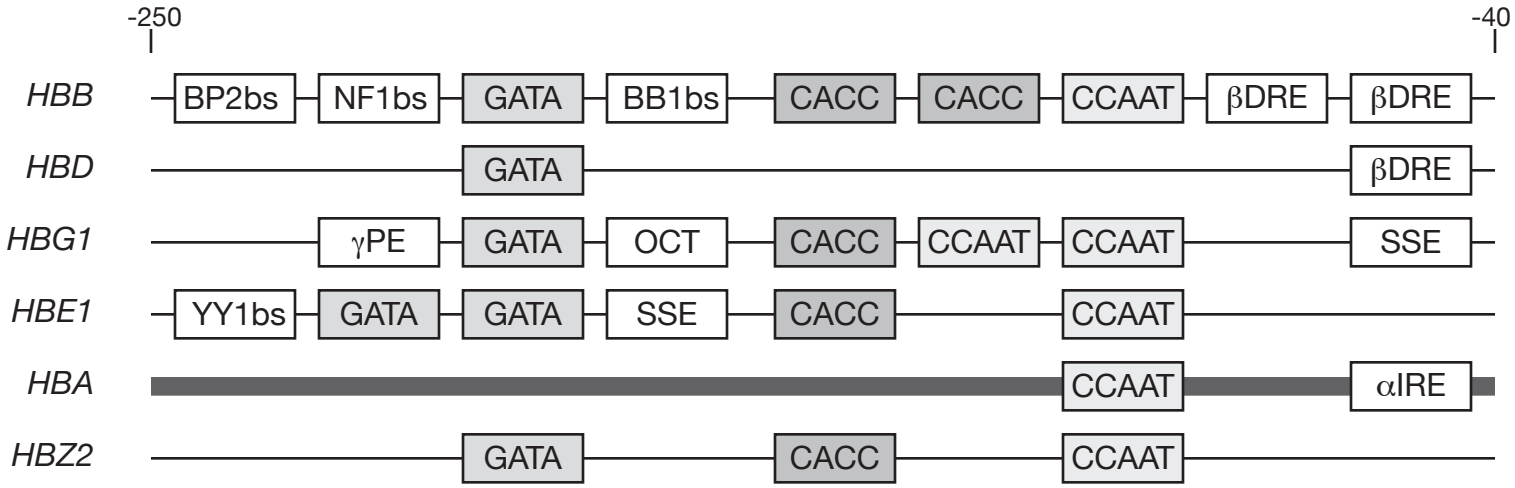
Fig. 3

A. Basal promoter

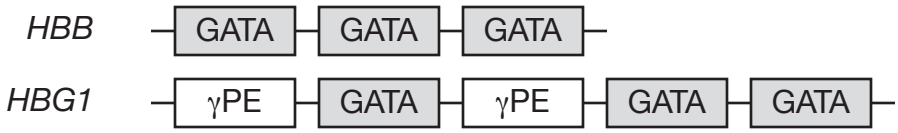


Consensus	SSRCGCC	TATAWAAR	YY**A**NWYY	SSAACGS	RGWYV
HBB	GGGCTGG	CATAAAAG	TTACATT	ACAACTG	AGCAA
HBD	GGACACA	CATAAAAG	TTACACT	ATAACAG	AGCAA
HBG1	CGGCGGC	AATAAAAG	ACACACT	GGAACGT	AGGTT
HBE1	CTTCGGC	AATAAAAG	GCACATA	GCAATCA	AGCAA
HBA	CCGCGCC	CATAAACC	GCACTCT	CAGACTC	AGAGA
HBZ2	CAGCTCC	TATATAAG	CTACCAA	CCAACTC	AGTGC
Consensus for globin genes	SSRCDSC	HATAAAAG	NYACAYW	VNAACNN	AGNTW

B. Upstream regulation



C. Proximal enhancers 3' to genes



D. Distal positive regulators 5' to genes

HBB locus control region

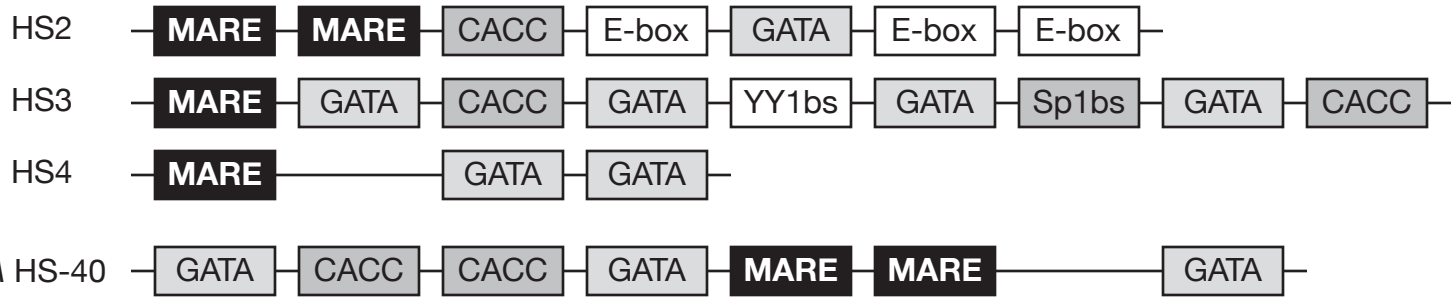
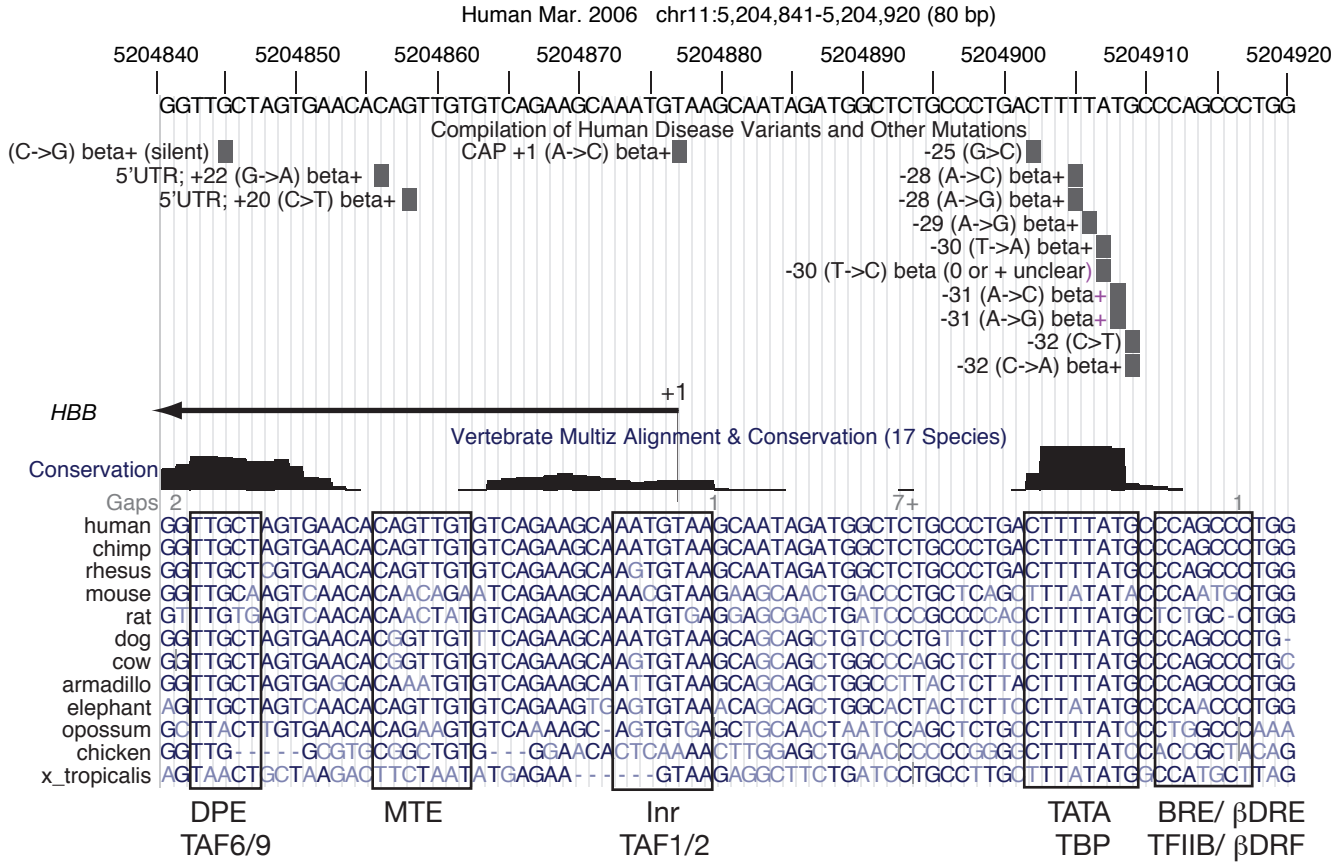


Fig. 4

A. Basal promoter for *HBB*



B. Upstream promoter for *HBB*

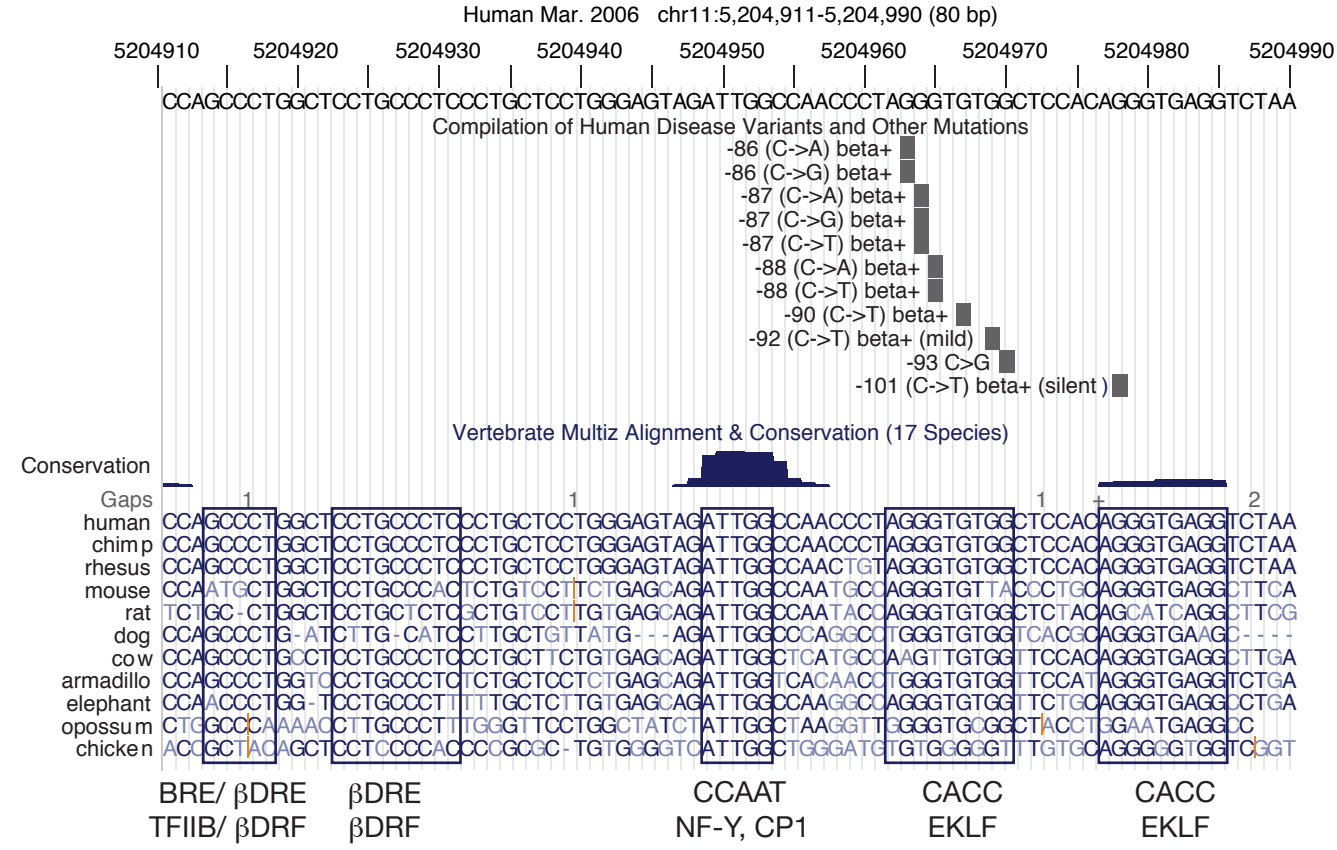
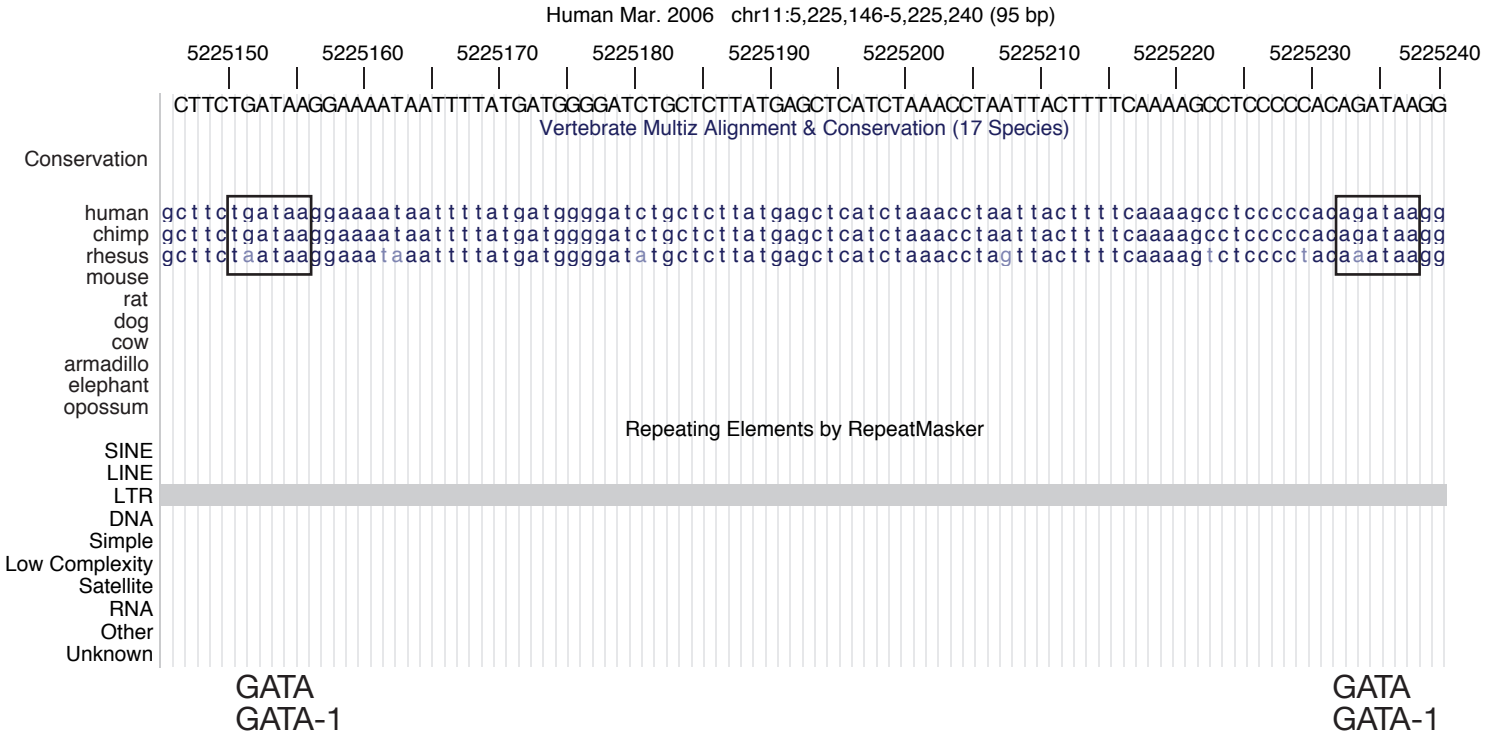


Fig. 5

A. 3’ enhancer for *HBG1*



B. Distal enhancer for *HBA*

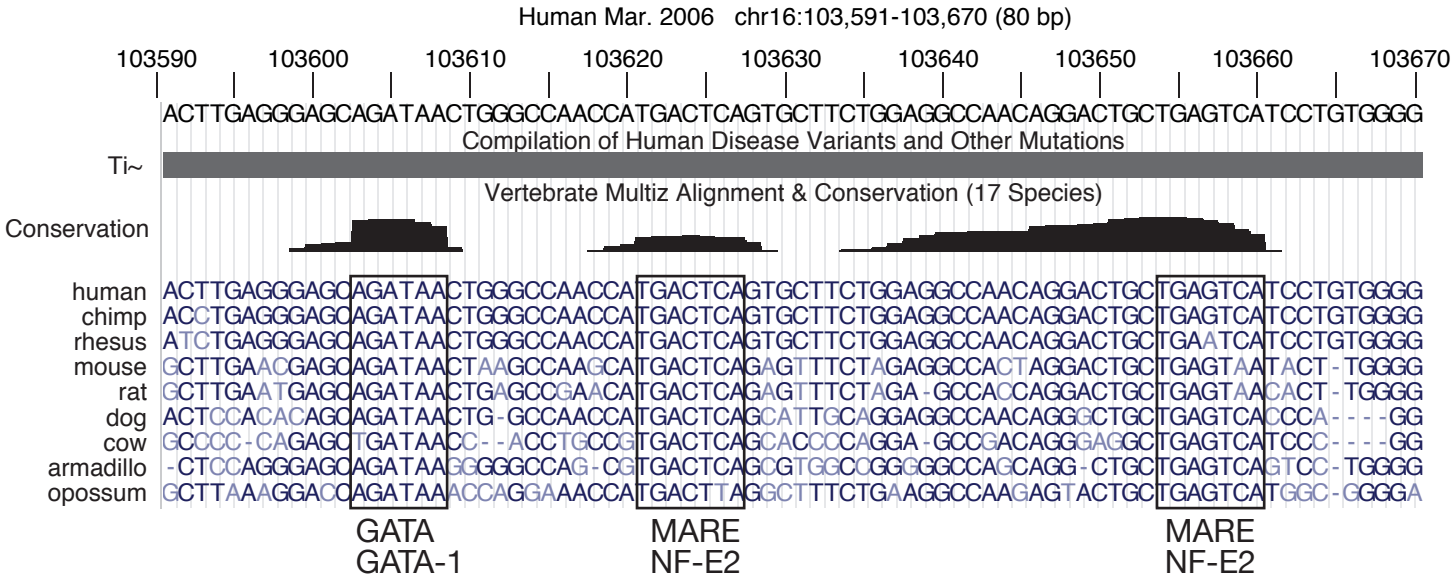


Fig. 6

nature medicine

Lysine-specific demethylase 1 is a therapeutic target for fetal hemoglobin induction

Lihong Shi[1], Shuaiying Cui[1], James D Engel[1] & Osamu Tanabe[1,2]

Enhanced fetal γ-globin synthesis alleviates symptoms of β-globinopathies such as sickle cell disease and β-thalassemia, but current γ-globin–inducing drugs offer limited beneficial effects. We show here that lysine-specific demethylase 1 (LSD1) inhibition by RNAi in human erythroid cells or by the monoamine oxidase inhibitor tranylcypromine in human erythroid cells or β-type globin–transgenic mice enhances γ-globin expression. LSD1 is thus a promising therapeutic target for γ-globin induction, and tranylcypromine may serve as a lead compound for the development of a new γ-globin inducer.

β-globinopathies such as sickle cell disease and β-thalassemia arise from genetic defects in the synthesis of the adult hemoglobin β chain (in HbA; $\alpha_2\beta_2$). Elevated synthesis of fetal hemoglobin (HbF; $\alpha_2\gamma_2$), which contains the γ chain instead of β, reduces morbidity of β-globinopathies[1]. γ-globin inducers such as hydroxyurea and decitabine have been used to treat β-globinopathies, but they show long-term benefit in only fewer than half of compliant patients[1]. These drugs are antineoplastic agents that can cause deleterious effects such as bone marrow suppression and reproductive toxicity. Therefore, finding more effective and safer γ-globin inducers is desirable.

Because the nuclear receptors TR2 and TR4 repress the γ-globin–encoding genes[2,3], the signaling pathways that these receptors regulate could be a therapeutic target for γ-globin induction. TR2 and TR4 associate with co-repressors, including DNA methyltransferase I (DNMT1) and LSD1, as components of a core heterotetrameric complex[4]. LSD1 removes methyl groups from mono- and dimethyl histone H3 lysine 4 (H3K4)[5], an activating epigenetic signature. To investigate a possible role for LSD1 in globin gene regulation, we examined the association of LSD1 with the gene promoters of β-type globins in human erythroid cells differentiated *ex vivo* from $CD34^+$ progenitors. Cells appeared to be proerythroblasts after 8 d in culture (**Fig. 1a**), and after 14 d 26% of the cells had undergone enucleation. We confirmed erythroid differentiation by flow cytometry (**Supplementary Fig. 1a**). LSD1 protein was expressed at comparable levels throughout differentiation (**Fig. 1b**). TR2 and TR4 bound the gene promoters of embryonic ε- and fetal γ-globins, but not to the promoter of the gene encoding adult β-globin (in accordance with the lack of their binding motifs therein; **Supplementary Fig. 2**)[2,3]. In contrast, LSD1 was detected at all three promoters in differentiating erythroid cells (**Fig. 1c**), which is not surprising given that LSD1 can bind multiple transcription factors[4,6,7]. As the cells differentiated, LSD1 binding to all three promoters diminished.

Tranylcypromine (TCP), a monoamine oxidase inhibitor that has been used clinically as an antidepressant since 1960, inhibits LSD1 with a half-maximum inhibitory concentration of ~2 μM (ref. 8). To investigate a possible role for LSD1 in γ-globin regulation, we examined the effects of TCP on human erythroid cells differentiated *ex vivo*. TCP administration at 0.5 or 1.5 μM did not alter cell proliferation or viability, but 5 μM TCP reduced cell proliferation and delayed differentiation without affecting cell viability (**Fig. 1d** and **Supplementary Fig. 1b**). HbF synthesis, detected by HPLC, was enhanced by TCP from 4.6% of total hemoglobin to 31% (**Fig. 1e**,**f**). Flow-cytometric analysis showed that HbF was induced in all of the cells in a dose-dependent manner (**Fig. 1g**), and γ-globin mRNA expression was induced up to 9.4-fold (**Fig. 1h**). These results suggest that LSD1 has a crucial role in γ-globin silencing in adult erythroid cells and that TCP might be used to treat β-globinopathies, as the effective concentration for HbF induction is within the therapeutic plasma concentration in treated depressed patients[9].

To quantify global H3K4 dimethylation (H3K4me2), an activating epigenetic signature, we subjected cell extracts to immunoblotting with antibodies to total histone H3 or H3K4me2. Global H3K4me2 increased during erythroid differentiation (**Supplementary Fig. 3a**), and TCP administration further enhanced global H3K4me2, confirming the inhibitory effect of TCP on LSD1 activity (**Supplementary Fig. 3b**).

To assess whether LSD1 inhibition might account for the observed HbF induction, we examined the effect of TCP on H3K4me2 abundance at each β-type globin promoter by chromatin immunoprecipitation (ChIP). In immature erythroid cells (day 8), H3K4me2 was undetectable, but in more mature cells (on days 11 and 14), H3K4me2 accumulated at all three promoters (**Supplementary Fig. 3c**), presumably owing to dissociation of LSD1 from these promoters over this time course (**Fig. 1c**). TCP treatment resulted in statistically significant enhancement of H3K4me2 accumulation at the γ-globin promoter (**Fig. 2a**), but not at the β-globin promoter, indicating that one possible mechanism for γ-globin induction by TCP is H3K4me2 accumulation at its promoter through LSD1 inhibition.

We also examined the β-type globin promoters for two repressive regulatory marks: DNMT1 association and H3K9 dimethylation (H3K9me2). DNMT1 is another co-repressor bound by TR2 and TR4, and is presumably also involved in γ-globin silencing[4]. DNMT1 may be involved in γ-globin induction by TCP, as LSD1 can demethylate and thereby stabilize DNMT1 (ref. 10). H3K9me2 is an epigenetic signature implicated in transcriptional silencing and heterochromatin formation. We detected both DNMT1 association and H3K9me2 accumulation at the γ-globin promoter, and both diminished upon

[1]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, Michigan, USA. [2]Department of Integrative Genomics, Tohoku Medical Megabank, Tohoku University, Sendai, Japan. Correspondence should be addressed to J.D.E. (engel@umich.edu) or O.T. (otanabe@umich.edu).

Received 20 November 2012; accepted 22 January 2013; published online 17 February 2013; doi:10.1038/nm.3101

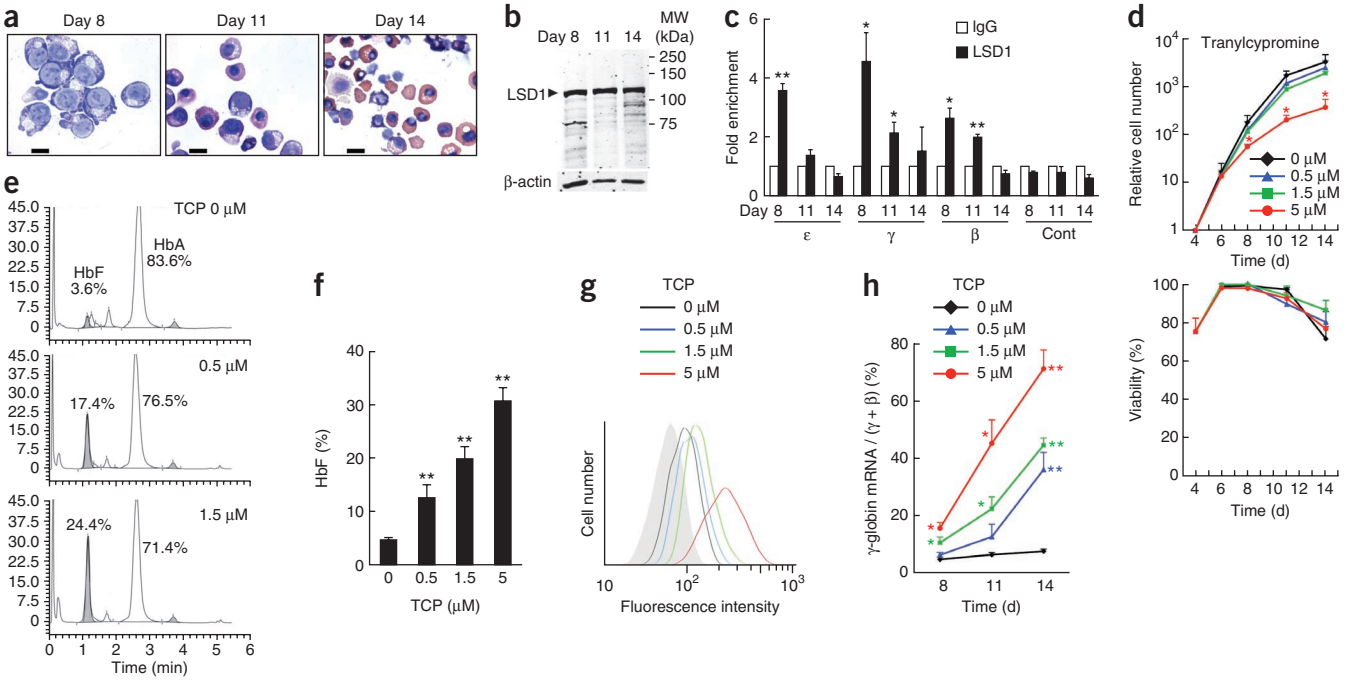


Figure 1 Induction of fetal hemoglobin by an LSD1 inhibitor, TCP, in adult erythroid cells. (**a**) Morphology of primary human erythroid cells differentiated *ex vivo* from CD34+ progenitor cells at the indicated times of cell culture. Scale bars, 10 μm. (**b**) Immunoblot showing LSD1 and β-actin (loading control) abundance in differentiating human erythroid cells. (**c**) ChIP assay examining LSD1 binding to the gene promoters of embryonic ε-, fetal γ- and adult β-globins after 8, 11 or 14 d of differentiation from CD34+ cells. The abundance of DNA precipitated with an LSD1-specific antibody was normalized to that precipitated with control IgG. As a negative control (Cont), an intergenic region between the ε- and ^{G}γ-globin–encoding genes was used. $n = 2$ independent immunoprecipitations. (**d**) Proliferation and viability of differentiating erythroid cells exposed to the indicated concentrations of TCP. $n = 3$ independent experiments. (**e**) Representative HPLC chromatograms showing HbF abundance (shaded area) in TCP-treated and untreated human erythroid cells 14 d after differentiation induction. Numbers indicate the percentages of HbF and HbA in total hemoglobin. The *y* axis represents the abundance of hemoglobin in the elution in an arbitrary unit. (**f**) Average HbF percentages in total hemoglobin, determined by HPLC, in TCP-treated and untreated human erythroid cells on day 14. $n = 3$–5 independent experiments. (**g**) Flow-cytometric analysis of HbF synthesis in TCP-treated and untreated cells on day 14 (colored lines). The shaded area indicates staining with control IgG. (**h**) Relative γ-globin mRNA abundance normalized to total β-type globin mRNAs (fetal γ- plus adult β-globin) in TCP-treated and untreated cells. $n = 3$ independent experiments. Statistically significant differences between the LSD1-specific antibody and control IgG (**c**) or between TCP-treated and untreated cells (**d**,**f**,**h**) are indicated (*$P < 0.05$; **$P < 0.01$). Error bars represent s.e.m.

TCP administration (**Supplementary Fig. 4**), suggesting additional indirect mechanisms for γ-globin induction by TCP. Specific activation of γ-globin transcription (unaccompanied by β-globin induction) by TCP thus seems to be a consequence of the combination of H3K4me2 gain, DNMT1 loss and H3K9me2 loss at the γ-globin promoter. Possible roles for TR2 and TR4 in γ-globin induction by TCP were not further investigated.

When tested alongside other HbF inducers, the effect of TCP was far superior to hydroxyurea, and roughly equivalent to decitabine at their respective optimal concentrations for primary human erythroid cells[11,12] (**Fig. 2b**). We next asked whether TCP can augment HbF induction by hydroxyurea or decitabine, anticipating that their combinatorial use might achieve a higher HbF level that would not be attainable with these single agents without cytotoxicity. The combination of TCP and decitabine had a greater than additive effect, resulting in almost 50% HbF accumulation in differentiating erythroid cells, whereas the combination of TCP and hydroxyurea was less effective (**Fig. 2b** and **Supplementary Table 1**).

To further explore the role of LSD1 in γ-globin silencing and address the mechanism for γ-globin induction by TCP, we examined the effect of LSD1 knockdown on γ-globin synthesis. We infected primary erythroid cells with lentivirus expressing either of two shRNAs targeting LSD1. Infection with either virus significantly reduced LSD1 mRNA and protein abundance as compared to a control virus (**Fig. 2c** and **Supplementary Fig. 5**). LSD1 knockdown resulted in enhanced HbF and γ-globin mRNA synthesis (**Fig. 2d**,**e** and **Supplementary Fig. 5**). These results validate the role of LSD1 in γ-globin silencing and are consistent with the concept that TCP induces γ-globin synthesis through LSD1 inhibition.

To examine possible global effects of TCP on erythroid gene expression, we conducted microarray analyses after 8 d of differentiation and detected a small number of transcripts whose level of expression was altered after TCP treatment, suggesting that only a few early erythroid genes are regulated by LSD1 (**Supplementary Figs. 6**, **7** and **Supplementary Table 2**). Gene ontology terms associated with the greatest changes in gene expression after TCP treatment (**Supplementary Table 3**) did not allow for a simple interpretation of how TCP treatment has effects specifically on γ-globin accumulation.

Finally, we tested the effects of TCP on γ-globin gene expression in transgenic mice harboring a yeast artificial chromosome containing the human β-type globin locus (β-YAC mice)[3]. In mice, the genes encoding human fetal γ-globin behave essentially as an embryonic gene and are strongly silenced in adults[3,13]. We administered TCP at two different doses (3 or 6 mg per kg body weight per day, 5 d a week) for 4 weeks. Expression of the human γ-globin–encoding genes in bone marrow cells was induced by TCP in a dose-dependent manner up to ninefold as compared to untreated mice (**Fig. 2f**). Furthermore, the effect of TCP was sustained for at least 18 d after cessation of drug

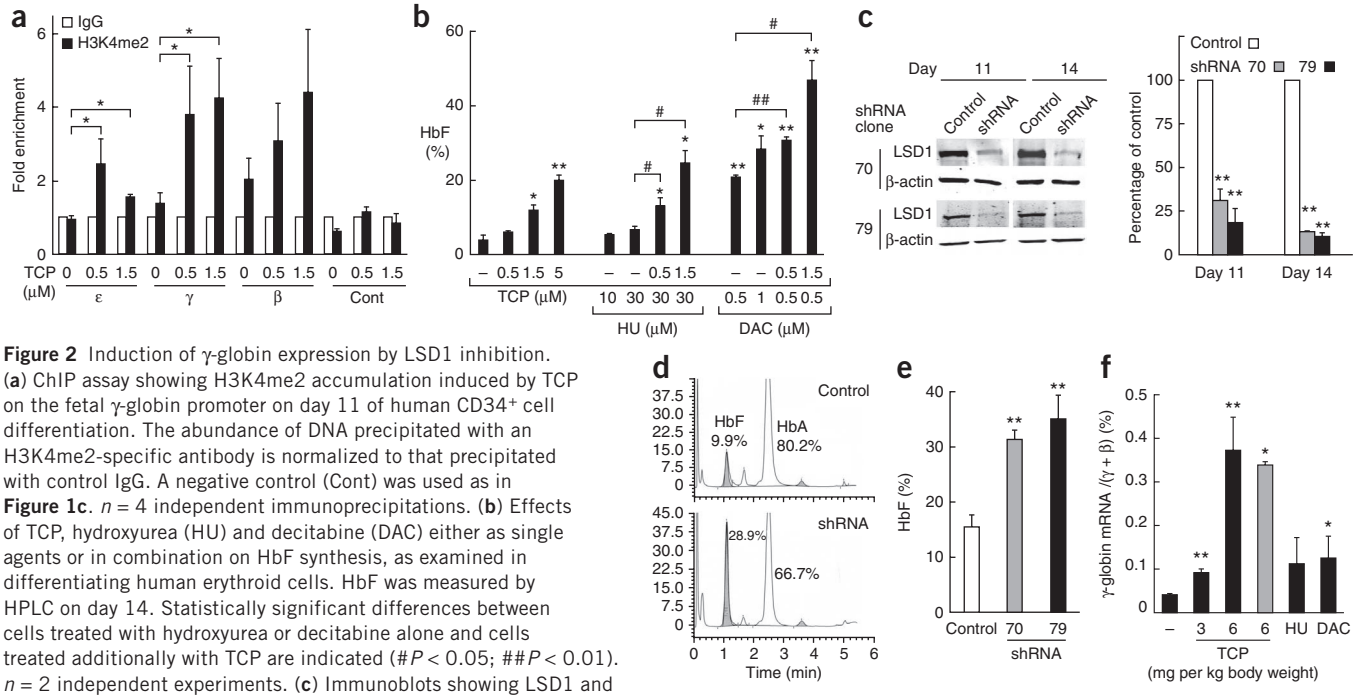


Figure 2 Induction of γ-globin expression by LSD1 inhibition. (**a**) ChIP assay showing H3K4me2 accumulation induced by TCP on the fetal γ-globin promoter on day 11 of human CD34$^+$ cell differentiation. The abundance of DNA precipitated with an H3K4me2-specific antibody is normalized to that precipitated with control IgG. A negative control (Cont) was used as in **Figure 1c**. $n = 4$ independent immunoprecipitations. (**b**) Effects of TCP, hydroxyurea (HU) and decitabine (DAC) either as single agents or in combination on HbF synthesis, as examined in differentiating human erythroid cells. HbF was measured by HPLC on day 14. Statistically significant differences between cells treated with hydroxyurea or decitabine alone and cells treated additionally with TCP are indicated (#$P < 0.05$; ##$P < 0.01$). $n = 2$ independent experiments. (**c**) Immunoblots showing LSD1 and β-actin (internal control) in shRNA virus– or control virus–infected cells. Two different shRNA clones (70 and 79) targeting different segments of LSD1 mRNA were used. The bar graph shows the relative abundance of LSD1 normalized to β-actin and to control cells. $n = 3$ independent experiments. (**d**) HPLC chromatograms showing HbF abundance in LSD1 shRNA (clone 70) virus– or control virus–infected cells. The *y* axis represents the abundance of hemoglobin in the elution in an arbitrary unit. (**e**) Average HbF abundance, determined by HPLC, in LSD1 shRNA virus– or control virus–infected cells. (**f**) γ-globin mRNA abundance, normalized to total β-type globin mRNAs, in bone marrow cells of β-YAC mice injected with TCP at 3 or 6 mg per kg body weight per day or saline 5 d a week for 4 weeks. A subset of mice injected with TCP (6 mg per kg body weight) for 4 weeks were analyzed 18 d after cessation of drug administration (gray bar). For comparison, mice were treated with hydroxyurea (200 mg per kg body weight per day, 5 d a week for 4 weeks) or decitabine (5 mg per kg body weight per day for 5 d) to examine their individual inductive effects[14,15]. $n = 6$ mice for the control group, $n = 3$ mice for the TCP 3 mg per kg body weight group, $n = 4$ mice for all other groups. Statistically significant differences between drug-treated and untreated cells (**a**,**b**), between shRNA virus– and control virus–infected cells (**c**,**e**), or between drug-injected and control mice (**f**) are indicated (*$P < 0.05$; **$P < 0.01$). Error bars represent s.e.m.

injection (**Fig. 2f**). γ-globin induction by TCP exceeded the efficacy achieved by hydroxyurea or decitabine when these agents were used at their published optimal doses in the β-YAC mice (**Fig. 2f**), or as compared to their efficacy in other mouse models[14,15]. Induced expression of the genes encoding human γ-globin as well as those encoding endogenous mouse embryonic β-type globins by TCP injection was observed in FACS-sorted, stage-matched erythroid progenitors from the bone marrow (**Supplementary Fig. 8**). After administration of TCP to the β-YAC mice, we did not observe any toxic effects or significant changes in blood cell counts (**Supplementary Table 4**) or hematopoietic precursor populations (**Supplementary Fig. 9**).

Several transcriptional repressors for the γ-globin–encoding genes have been identified recently[13,15–17], but the absence of pharmacological inhibitors for these repressors poses considerable challenges to the therapeutic application of those findings. In contrast, targets of current HbF-inducing medications have not been identified, greatly impeding efforts to improve their safety or efficacy. This study demonstrates that LSD1 plays a crucial role in fetal γ-globin silencing and that an LSD1 inhibitor, TCP, can enhance HbF synthesis. TCP has been widely used as an antidepressant with manageable side effects; it exerts its antidepressive effect by elevating serotonin levels[18] and does not seem to have cytotoxic or mutagenic effects. Although it was recently reported that LSD1 inhibition by RNAi in mice can cause hematopoietic defects[19], hematological toxicity has not been associated with TCP in humans except in rare cases of transient thrombocytopenia attributable to overdosing[20]. TCP may thus serve as a lead compound for developing a new HbF-inducing medication that is complementary to, or more favorable than, current medications.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Microarray data were deposited in the Gene Expression Omnibus (GEO) database with accession number GSE42078.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health grants DK086956 (to O.T. and J.D.E.) and HL24415 (to J.D.E.) and an American Heart Association postdoctoral fellowship to L.S. We thank D. Giacherio and D. Harro for hemoglobin HPLC analysis; C. Yu, A.D. Campbell, K.C. Lim, T. Hosoya, B. Godfrey and M. Sierant for discussions, comments and technical assistance; and D. Lavelle and J. DeSimone for discussions regarding the inductive properties of decitabine. Lentiviruses were produced by the University of Michigan Vector Core. Microarray analysis was performed at the University of Michigan Microarray Core.

AUTHOR CONTRIBUTIONS

O.T. and J.D.E. conceived of the study; L.S., S.C., J.D.E. and O.T. designed experiments. L.S. and S.C. performed experiments. L.S., S.C., J.D.E. and O.T. analyzed data and wrote the paper.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nm.3101.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Testa, U. *Ann. Hematol.* **88**, 505–528 (2009).
2. Tanabe, O. *et al. EMBO J.* **21**, 3434–3442 (2002).
3. Tanabe, O. *et al. EMBO J.* **26**, 2295–2306 (2007).
4. Cui, S. *et al. Mol. Cell Biol.* **31**, 3298–3311 (2011).
5. Shi, Y. *et al. Cell* **119**, 941–953 (2004).
6. Hu, X. *et al. Proc. Natl. Acad. Sci. USA* **106**, 10141–10146 (2009).
7. Saleque, S., Kim, J., Rooke, H.M. & Orkin, S.H. *Mol. Cell* **27**, 562–572 (2007).
8. Lee, M.G., Wynder, C., Schmidt, D.M., McCafferty, D.G. & Shiekhattar, R. *Chem. Biol.* **13**, 563–567 (2006).
9. Mallinger, A.G. & Smith, E. *Psychopharmacol. Bull.* **27**, 493–502 (1991).
10. Wang, J. *et al. Nat. Genet.* **41**, 125–129 (2009).
11. Moutouh-de Parseval, L.A. *et al. J. Clin. Invest.* **118**, 248–258 (2008).
12. Banzon, V. *et al. Exp. Hematol.* **39**, 26–36. e1 (2011).
13. Sankaran, V.G. *et al. Nature* **460**, 1093–1097 (2009).
14. Buller, A.M., Elford, H.L., DuBois, C.C., Meyer, J. & Lloyd, J.A. *Blood Cells Mol. Dis.* **25**, 255–269 (1999).
15. Xu, J. *et al. Science* **334**, 993–996 (2011).
16. Menzel, S. *et al. Nat. Genet.* **39**, 1197–1199 (2007).
17. Wilber, A., Nienhuis, A.W. & Persons, D.A. *Blood* **117**, 3945–3953 (2011).
18. Frieling, H. & Bleich, S. *Eur. Arch. Psychiatry Clin. Neurosci.* **256**, 268–273 (2006).
19. Sprüssel, A. *et al. Leukemia* **26**, 2039–2051 (2012).
20. Eyer, F., Jetzinger, E., Pfab, R. & Zilker, T. *Clin. Toxicol. (Phila.)* **46**, 261–263 (2008).

ONLINE METHODS

***Ex vivo* differentiation of purified human CD34$^+$ cells.** We purchased cryopreserved vials of purified human CD34$^+$ hematopoietic progenitor cells from the Fred Hutchinson Cancer Research Center, which collected the cells from healthy volunteers in full compliance with federal and institutional regulations on informed consent and confidentiality. The CD34$^+$ cells were isolated from the peripheral blood after mobilization by administration of granulocyte colony–stimulating factor (G-CSF). We grew and differentiated the cells *ex vivo* into the erythroid lineage in 14 d by a two-phase culture method described previously[4,21]. Cell number and viability were determined with a hemocytometer by trypan blue staining. Cell morphology was examined by Wright-Giemsa staining (Sigma-Aldrich) of cytospins. For cell proliferation and viability analysis, data from three biological replicates are presented.

Drug treatment of differentiating primary human erythroid cells. Tranylcypromine (TCP; Tocris Bioscience) dissolved in water at 50 mM was added to the culture medium at final concentrations of 0.5, 1.5 or 5 μM on days 4–14 of the differentiation culture. Hydroxyurea (HU; Sigma-Aldrich) was added to the culture medium at final concentrations of 10 or 30 μM on days 4–14 of the culture. Decitabine (DAC; Sigma-Aldrich) was added to the culture medium at final concentrations of 0.5 or 1 μM on days 7–14 of the culture. In some experiments, either hydroxyurea or decitabine was administered in combination with TCP. Every 2–3 d, culture media were replaced with media containing freshly added drugs.

Flow cytometry and cell sorting. For cell surface marker analysis, the indicated amounts of antibodies were used to stain 10^6 cells suspended in 100 μL PBS with 2% FBS. Human erythroid cells were stained with phycoerythrin (PE)-Cy7–conjugated CD34-specific (eBioscience, 25-0349, 0.125 μg), PE-conjugated CD71-specific (eBioscience, 12-0719, 0.03 μg), or PE-Cy5–conjugated glycophorin A–specific (BD Biosciences, 559944, 0.015 μg) antibody. Mouse bone marrow cells were stained with allophycocyanin (APC)-conjugated Ter-119–specific (BioLegend, 116212, 0.5 μg), PE-conjugated CD71-specific (BioLegend, 113808, 0.5 μg), PE-Cy5–conjugated Mac-1–specific (eBioscience, 15-0112-81, 0.5 μg), APC-conjugated Gr-1–specific (BioLegend, 108411, 0.5 μg) or PE-conjugated CD41-specific (eBioscience, 12-0411-81, 0.5 μg) antibody. For cytoplasmic HbF analysis, 10^5 cells were fixed in 0.05% glutaraldehyde for 10 min, permeabilized in 0.1% Triton X-100 for 5 min and then stained with 5 μL of an APC-conjugated HbF-specific antibody (Invitrogen, MHFH05) in 80 μL of PBS with 0.1% BSA. Stained cells were analyzed using a FACSCanto II (BD Biosciences) or sorted on a FACSAria I (BD Biosciences). Fractional percentages of cell populations represent the averages of two to four biological replicates.

Immunoblotting. Cells were lysed in Laemmli sample buffer and subjected to SDS-PAGE. Proteins were then transferred to a nitrocellulose membrane (Li-Cor) and probed with antibodies to LSD1 (Abcam, ab17721, 1:1,000 dilution), total histone H3 (Abcam, ab1791, 1:20,000 dilution) or dimethyl histone H3 Lysine 4 (H3K4me2) (Abcam, ab11946, 1:5,000 dilution), and fluorescence-conjugated secondary antibodies (Li-Cor). Proteins were visualized and quantified with the Odyssey Infrared Imaging System (Li-Cor). Quantification data represent the results of three or four biological replicates.

ChIP assays. Chromatin immunoprecipitation (ChIP) was performed as described previously using 10^6 cells for each assay[4]. For TR2, TR4 and LSD1 binding, ethylene glycol bis(succinimidyl succinate) was used as a cross-linker in addition to formaldehyde. Ten micrograms of antibodies to TR2 and TR4 (ref. 3), LSD1 (Abcam, ab17721), H3K4me2 (Millipore, 05-1338), dimethyl histone H3 Lysine 9 (H3K9me2) (Abcam, ab1220), or DNMT1 (Abcam, ab92453) were used for immune complex formation in 500 μL of immunoprecipitation dilution buffer[4]. Precipitated DNA was quantified by real-time quantitative PCR assay with primer pairs for human embryonic ε-, fetal γ- and adult β-globin promoter sequences[4], as well as a primer pair for an intergenic region between the ε- and ^{G}γ-globin genes used as a negative control (5′-TCC CACTCTGTGGGTTGTCTGTTT-3′ and 5′-CCCTTCTACACATTGGCTTA GGAAAGG-3′). Data representative of two to four independent immunoprecipitations are presented.

Hemoglobin analysis by HPLC. Cells were lysed and analyzed for hemoglobin composition using the Bio-Rad Variant II Hemoglobin Testing System equipped with an ion-exchange HPLC column (Hercules). Data of three to five biological replicates are presented.

RT-qPCR assay. mRNAs for human and mouse β-type globins, human LSD1, and 16 other human genes that were selected for validation of the microarray analysis, as well as 18S rRNA, were quantified by reverse transcription and quantitative real-time PCR (RT-qPCR) assay as described previously[4]. Relative abundance of human fetal γ- or mouse embryonic εy- and βH1-globin mRNAs was determined using the respective human or mouse adult β-globin mRNA as an internal control, based on threshold cycle (C_t) values and the experimentally determined amplification efficiency for each primer pair. Relative mRNA abundance for LSD1 and all other mRNAs was similarly determined, using 18S rRNA as an internal control. For gene expression analysis with primary human erythroid cells, results of three biological replicates are presented. All the primer pairs except for 18S rRNA were designed to span introns. Primer sequences are as follows: human γ-globin, 5′-GATGCCATAAAGCACCTGGATG-3′ and 5′-TTGCAGAATAAAGCCTATCCTTGA-3′; human β-globin, 5′- AACTG TGTTCACTAGCAACCTCAA-3′ and 5′- GAGTGGACAGATCCCCAAA GGA-3′; LSD1, 5′-TGGCCATTCTCAAAGGGATT-3′ and 5′-CAGCACGCC AACGAGACA-3′; and 18S rRNA, 5′-ACCGCAGCTAGGAATAATGGA-3′ and 5′-GCCTCAGTTCCGAAAACCA-3′. Primer sequences for mouse β-type globins were previously published[4]. Sequences for all other primers will be provided upon request.

Expression of LSD1 shRNAs by lentiviral vector. The pLKO.1-puro lentiviral vectors used to express LSD1 shRNAs (short hairpin RNAs) and the puromycin resistance gene were purchased from Sigma-Aldrich (TRCN0000046070 and TRCN0000382379). The control vector was generated by deleting a segment (between EcoRI and NdeI sites) with the hairpin sequence from the shRNA vector. Lentiviruses were generated by transient transfection of the vectors into a packaging cell line. For infection, cells were exposed to virus on day 4 of culture for 24 h and then to 1 μg ml^{-1} puromycin on days 6–14 to select for infected cells.

Drug administration to mice. All animal experiments were approved by the University Committee on Use and Care of Animals at the University of Michigan. TCP, hydroxyurea, or decitabine dissolved in saline was administered to 7- to 19-week-old transgenic mice harboring a yeast artificial chromosome containing the intact human β-type globin locus (β-YAC mice)[3]. TCP was administered by subcutaneous injection at a dose of 3, 6 or 10 mg per kg body weight per day, 5 d a week for 4 weeks. Hydroxyurea was administered by intraperitoneal injection at a dose of 200 mg per kg body weight per day, 5 d a week for 4 weeks. Decitabine was administered by subcutaneous injection at a dose of 5 mg per kg body weight per day for 5 d. Control mice were injected with saline only. Following the stated periods of injections, bone marrow cells were harvested for flow-cytometric analysis, cell sorting or RT-qPCR analysis for gene expression of human or mouse globins. For gene expression analysis of human γ-globin in unfractionated bone marrow cells, 11 male and 14 female mice were used, and data from three to six mice are presented for each experimental condition. For flow-cytometric analysis and sorting of bone marrow cells, three males and one female were used. For analysis of hematological parameters of peripheral blood, four males and three females were used.

Microarray analysis. Total RNA was extracted from primary human erythroid cells after 8 days of *ex vivo* differentiation from CD34$^+$ cells with or without TCP treatment (0.5 μM, 1.5 μM or 5 μM). The integrity of RNA samples was verified on the Bioanalyzer (Agilent). The microarray experiments were performed using the Human Genome U219 Array Strip (Affymetrix) with the GeneAtlas 3′ IVT Express Kit. Gene expression values were calculated with the Robust Multi-array Average algorithm[22].

Statistical analysis. Student's *t*-test was used for all the statistical analyses. Error bars represent s.e.m.

21. Giarratana, M.C. *et al. Nat. Biotechnol.* **23**, 69–74 (2005).
22. Irizarry, R.A. *et al. Biostatistics* **4**, 249–264 (2003).

nature publishing group

Emerging science of hydroxyurea therapy for pediatric sickle cell disease

Nancy S. Green[1] and Sandra Barral[2]

Hydroxyurea (HU) is the sole approved pharmacological therapy for sickle cell disease (SCD). Higher levels of fetal hemoglobin (HbF) diminish deoxygenated sickle globin polymerization *in vitro* and clinically reduce the incidence of disease morbidities. Clinical and laboratory effects of HU largely result from induction of HbF expression, though to a highly variable extent. Baseline and HU-induced HbF expression are both inherited complex traits. In children with SCD, baseline HbF remains the best predictor of drug-induced levels, but this accounts for only a portion of the induction. A limited number of validated genetic loci are strongly associated with higher baseline HbF levels in SCD. For induced HbF levels, genetic approaches using candidate single-nucleotide polymorphisms (SNPs) have identified some of these same loci as being also associated with induction. However, SNP associations with induced HbF are only partially independent of baseline levels. Additional approaches to understanding the impact of HU on HbF and its other therapeutic effects on SCD include pharmacokinetic, gene expression–based, and epigenetic analyses in patients and through studies in existing murine models for SCD. Understanding the genetic and other factors underlying the variability in therapeutic effects of HU for pediatric SCD is critical for prospectively predicting good responders and for designing other effective therapies.

Healthy People 2020, the federal public health agenda, has set a goal of "Increase(ing) the proportion of persons with hemoglobinopathies who receive disease-modifying therapies" (1). For the vast majority of people with sickle cell disease (SCD), the Healthy People goal will be reached through increased use of hydroxyurea (HU). Critical questions surrounding its use include how this drug works to ameliorate the clinical severity of SCD and what subpopulation of children with SCD benefit most from its use. This review addresses these questions from a translational science perspective.

SCD affects an estimated 90,000 people in the United States (2), with more than 1,900 newborns detected annually through universal newborn screening (2). Infant screening, early preventive therapy, and parental guidance have largely eliminated early child mortality from SCD (3–5). Moreover, specialized care and ongoing preventive services have prolonged average life expectancy (6). Despite these successes, multiorgan damage and mortality accumulate by early adulthood, resulting in shortened life span (6).

HU holds expanding promise for improved clinical outcomes. More than 2 decades ago, the seminal Multicenter Study of Hydroxyurea phase III trial for adults demonstrated the striking clinical impact of HU: 40% reduction in the incidence of acute pain episodes, acute chest syndrome, and hospitalization (7). These results led to approval in 1998 of HU for use in symptomatic SCD by the US Food Drug Administration (FDA). HU remains the only FDA-approved drug for SCD, but approval does not extend to pediatric use. The approval gap for children is partially attributed to the lack of a commercial pharmaceutical sponsor. Helping to span the gap is the FDA's recent commissioning of a pediatric study of the pharmacokinetics of HU and its relative bioavailability of a liquid formulation (http://clinicaltrials.gov/show/NCT01506544).

Clinical efficacy of HU treatment varies among individuals, although most patients with severe phenotypes benefit from its use (7,8). This review describes newly identified mechanisms for the effects of HU, including genetic regulation of fetal hemoglobin (HbF) as a disease modifier and the biologic effects of HU on blood vessels and gene regulation. These recent advances improve the prospects for prospectively assessing efficacy of HU therapy, are inspiring clinical trials for additional salutatory effects of HU, and may guide future drug development.

CLINICAL EFFECTS

The profound clinical effects of HU in children with SCD have been recently reviewed (9–11); these are summarized here in **Table 1**. Much of the work on HU in children with SCD has come from phase II and III trials trials led by Ware *et al.*, including pivotal studies such as phase I/II trial of HU in children with sickle cell anemia (HUG KIDS) (12–14), Hydroxyurea Safety and Organ Toxicity Trial (HUSOFT) (15), Pediatric Hydroxyurea Phase III Clinical Trial (BABY HUG) (16–18) and an early pediatric trial published in 1999 (12). French investigators have also contributed insights into the impact of HU (19,20). Randomized pediatric trials with HU have demonstrated decreased occurrences of pain episodes (18), acute chest syndrome, hospitalization (8,11,18), transfusion, and splenic autoinfarction (18), along with improved quality of life (21,22).

[1]Division of Pediatric Hematology/Oncology/Stem Cell Transplantation, Department of Pediatrics, Columbia University, New York, New York; [2]G.H.Sergievsky Center, Department of Neurology, Columbia University, New York, New York. Correspondence: Nancy S. Green (nsg11@columbia.edu)

Received 30 March 2013; accepted 25 October 2013; advance online publication 15 January 2014. doi:10.1038/pr.2013.227

Table 1. Clinical effects of hydroxurea on children with SCD[a]

Effects	Blood/circulation	Organ/whole body
Laboratory/physiological measurement	Elevates HbF levels (8,14,23) and stabilizes high infant levels (18)	Brain: improves TCD flow velocity (26,27)
	Increases hemoglobin (13–15)	Spleen: preserves blood flow (15,18)
	Increases MCV (8,14,15,22)	Lungs: decreases acute chest syndrome (15,18)[b]
	Decreases hemolysis (13–15)	Renal: decreases hyperfiltration (24) and hyposthenuria (25)
	Decreases WBC and platelet counts (8,13–15)	
Clinical/well-being	Fewer acute pain crises (8,13,18)[b]	Reduces mortality (30–32)[b,c]
	Reduces dactylitis (18)[b]	Improves growth (13,15)[b]
	Fewer transfusions (18)[b]	Improves quality of life (21,22)
	Fewer hospitalizations (8,13,18)[b]	
Not yet known	Stabilizes HbF as adults	Improves overall life span for children[b]
	Reduces alloimmunization (through reducing transfusion)	Improves cognitive development[b]
	Reduces transfusion-related iron toxicity (through reducing transfusion)	Protects from stroke/infarct
		Prevents long-term renal, lung, and cardiac effects
		Reduces cholelithiasis
		Reduces retinopathy
		Normalizes timing of physical maturation[b]
		Maintains fertility[b]

HbF, fetal hemoglobin; MCV, red cell mean corpuscular volume; SCD, sickle cell disease; TCD, transcranial Doppler; WBC, white blood cell.

[a]Some of the effects have not been demonstrated across all pediatric age ranges nor have they been tested in prospective randomized trials. Reports demonstrating effect by randomized trials are preferentially cited. [b]Patient-oriented outcomes. [c]Established long-term outcome for adults.

Prolonged use sustains the laboratory effects of decreased anemia, markers of hemolysis, and counts of white blood cells and platelets, in addition to increased red cell mean corpuscular volume (23). Early HU use stabilizes renal hyperfiltration (24), hyposthenuria (25), and age-dependent decrease in HbF (18). Induction of HbF is described below.

Of note, although the laboratory effects of HU apply across the pediatric ages tested, many of the various clinical improvements noted for one age range have not necessarily been assessed for other ranges. For example, reduced dactylitis, hyposthenuria, and transfusions were noted in the BABY HUG trial of children enrolled at age 9–13 mo (17,18). Improved transcranial Doppler blood flow through large cerebral arteries has been demonstrated in school-aged children (26,27). Despite positive findings, some of these trials had mixed results. For example, the primary end points of the BABY HUG study were not met (18). In the Stroke with transfusions changing to hydroxyurea (SWiTCH) study for secondary stroke prevention, continued chronic transfusion was advantageous relative to HU with phlebotomy (28). Moreover, HU reduces but does not eliminate the symptoms of and morbidity in SCD. For example, the SWiTCH trial demonstrated that chronic transfusions more effectively prevented pain episodes than HU with phlebotomy (29).

Two long-term studies demonstrated substantially improved life spans from prolonged use of HU in adults, including a study based on the Multicenter Study of Hydroxyurea in Sickle Cell Anemia trial (30,31). Prospective life span data for children taking HU are not yet available due to later uptake into pediatric trials. Nonetheless, a recent retrospective study from Brazil reported improved childhood mortality for those taking HU for up to 6 y (32). Collectively, these data are increasingly persuasive about the enduring impact of HU on SCD.

The pharmacokinetics of HU appears to follow a biphenotypic metabolism in children (33). Multiple single-nucleotide polymorphisms (SNPs) are associated with two apparent pharmacokinetic profiles of HU uptake and excretion. However, these genotypes do not correlate with response by the biomarker HbF.

FETAL HEMOGLOBIN

The clinical severity of SCD is highly variable. Children experience multiple different clinical complications of differing severity levels and frequencies. HbF is of critical importance in the major sickle subtype homozygous sickle cell disease (HbSS; and sickle (HbS)-β-zero thalassemia, herein collectively referred to as HbSS). Lower HbF levels correlate with overall severer disease manifestations (34). Unlike "adult" hemoglobin A (HbA), HbF actively inhibits the polymerization of HbS, the underlying pathophysiology of SCD. In solution, HbF concentration higher than 15% prevents sickle globin polymerization (35). The cutoff for defining lower risk of severe complications has been estimated at 20% (36).

Sharp declines of HbF during infancy occur as HbF-producing γ-globin is replaced by β-globin. This switch leads to the predominant expression of either HbA or HbS. The F-to-S switch in children affected by HbSS (37) occurs more gradually than the F-to-A switch in nonanemic children. HbF levels in toddlers with SCD stabilize by age 3 or 4 y and are generally constant throughout childhood. Despite bearing the same β-globin sickle variant, affected populations with African

ancestry exhibit wide variations in HbF levels (37–40). In the United States, pediatric levels vary from 3 to 20% of total hemoglobin, compared with only 0.5–2% for nonanemic individuals. The average HbF level in the US SCD pediatric population is ~10% (36).

USE OF HU

Mechanism of Action

The physiology of the HU effect is complex and can generally be generally categorized into two overlapping pathways: effects on HbF production and improved blood flow through reduced intercellular adhesion (**Figure 1**). HU is a short-acting cytotoxic drug that induces a state of "stress erythropoiesis." Enhanced HbF production from intermittent mild marrow toxicity is believed to stem from the steady shifting of marrow physiology to the stressed state. The marrow responds to the repetitive pharmacological injury of daily use by enhanced erythropoiesis and increased HbF production (34,41). Paradoxically, the net effect of marrow toxicity is induced HbF and stabilization of cellular hemoglobin solubility. These effects lead to decreased levels of red blood cell (RBC) membrane damage and hemolysis (34,41).

HbF induction usually occurs within the first few months after initiating HU and is reversible on cessation or diminution of dosing (**Figure 2**). Relevance of HU induction of HbF was demonstrated through a proof-of-principle murine model for SCD. Lack of expression of human HbF precluded HU induction in those mice. In the murine model, HU itself had no effect on improving anemia or protecting organs from SCD damage. In contrast, HbF gene therapy markedly improved the blood smear, microscopic, and organ-level pathological effects of SCD (42).

HU appears to influence RBC–endothelial interactions. Decreased expression of RBCs, white blood cells, and endothelial integrins and other adhesion molecules probably improves microvascular blood flow and reduces proinflammatory cell–cell interactions (43,44). Microvascular effects of SCD and HU appeared to be replicated using an interesting microfluidic model of blood flow and endothelialized microfluidic channels (44). Whole-blood samples from SCD lead to microvascular occlusion and thrombosis. Blood samples from patients with SCD had diminished velocity and greater tendency to obstruct in the microchannels. These effects nearly normalized using blood from patients on HU (44). HU may be associated with reduced generation of microparticles, suggesting a reduction in markers of inflammation and thrombosis (45).

HU may reduce cellular adhesion in general and/or adhesion provoked by infection or inflammation. Integrins and other cell surface glycoproteins regulate neutrophil migration and RBC flow through endothelial interactions. In a murine model for SCD and pneumococcal pneumonia and sepsis, HU provided some protection by decreasing the recruitment of neutrophils into infected lungs. Mice genetically engineered to lack E-selectin were not protected by HU (46). This finding strengthens the view that HbF-independent effects of HU include decreasing leukocyte–endothelial adhesion.

HU may also stimulate nitric oxide (NO) production as an NO donor or through stimulation of intermediates (discussed below). As a potent vasodilator, NO repletion contributes to improved vascular health in SCD (**Figure 1**) (47). Along with decreased "sticky" interaction between blood cells and the endothelium, enhanced NO-induced local vasodilation may also benefit blood flow (**Figure 1**) (48,49). However, questions have arisen regarding these effects of NO (50). In all, decreased pathology from damaged RBCs and pathological interactions between RBCs and endothelial cells appear to synergistically reduce clinical signs, symptoms, and morbidities of the disease (**Table 1**). The ameliorative effects of HU appear to persist for as long as it is taken and the pharmacokinetics are maintained.

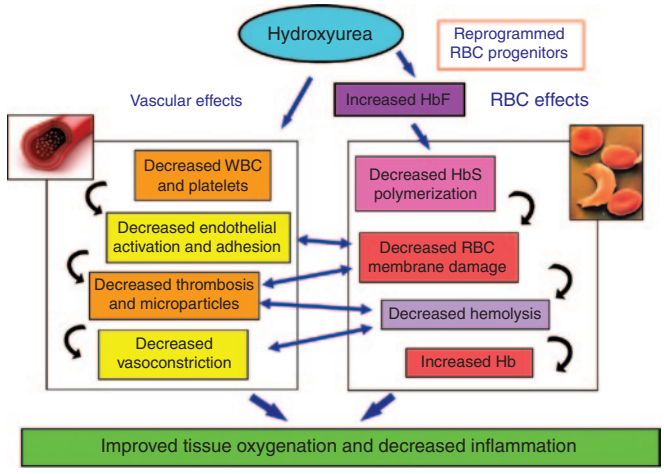


Figure 1. Physiological effects of hydroxyurea on sickle cell disease (SCD). Hydroxyurea has pleiotropic effects in ameliorating SCD, with complex and interacting effects of vascular and red blood cell (RBC) components. Hb, hemoglobin; HbF, fetal Hb; HbS, sickle hemoglobin; WBC, white blood cells.

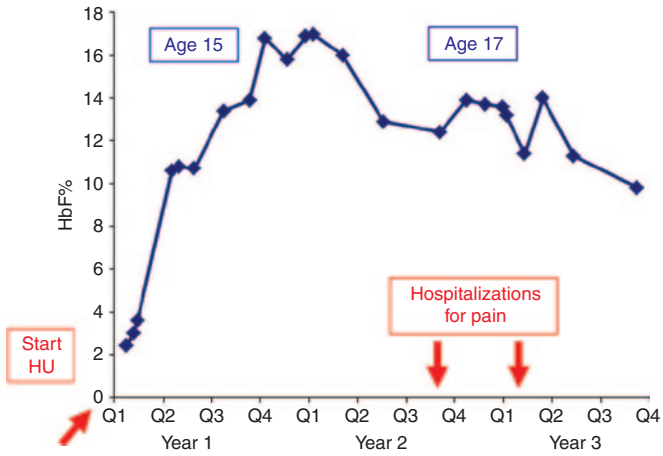


Figure 2. Fetal hemoglobin (HbF) levels of a teenager with homozygous sickle hemoglobin (HbSS) on hydroxyurea (HU). Before HU use, this teenager had two to three hospitalizations for pain each year. She had no admissions for 1.7 y after beginning HU. Her baseline HbF was 2.4%, and maximum recorded HbF level was 16.9%. She acknowledged intermittent adherence in the years 2 and 3, during which time she had two admissions for acute pain episodes. Blue diamonds refer to HbF data points.

HbF Response to HU

The individual extent of HU-induced HbF is highly variable. Standard pediatric dosing of HU adjusts for dose-dependent myelotoxicity (14,33). Under these conditions, HU generally induces HbF by an additional 8–18% relative to the baseline levels (14,33,51,52). In contrast with the biomarker glycosylated Hb (HbA1c) for diabetes, no absolute HbF target exists. Nonetheless, peak attained HbF levels remain fairly constant in childhood (23). No absolute limit to the therapeutic amount of HbF induction has been described. For example, people of Southeast Asia or Saudi Arabia with SCD have baseline HbF levels averaging 16–20%. HU induction raises their levels 1.5- to 2-fold, associated with further diminution of their already tempered clinical symptoms (53).

Children with SCD generally have higher baseline HbF levels than adults and more pronounced HbF response to HU (14,54). Factors responsible for differences may include the need for highly regenerative marrow RBC precursors and, for HU, normal renal function for prompt excretion. Adults normally experience age-dependent decreased marrow cellularity. In SCD, disease-related marrow infarcts and other age-related physiological effects could exacerbate normal marrow regression. Age-related diminution of response to HU increases the likelihood that genetic studies using pediatric populations may reveal more precise basic biologic insights.

Genetic Analysis of HU-Induced HbF

Analyses of HbF regulation are crucial to understanding the spectrum of SCD severity, variability of HU response, and design of novel therapies. In addition to the established observations of ethnic variability of HbF levels, several key observations drive the rationale for identifying genetic components of HU induction of HbF in US populations of SCD:

1. Baseline HbF levels in SCD have high heritability (55,56);
2. HbF induction from HU therapy is also a heritable trait (55);
3. Genome-wide SNP studies in normal nonanemic adults identified a few major loci associated with variation of low HbF levels. These regions are both cis and trans of the β-globin gene locus (56,57);
4. These same loci are associated with baseline HbF in people with SCD in the United States (56,58–62). Additional loci have been identified but are not yet replicated;
5. A modest correlation in children exists between levels of HbF at baseline and on HU (14,33,51).

Taken together, these findings lead to the prediction that genetic regulation of HbF expression at baseline overlaps with the control of HU-induced HbF The three major loci related to HbF expression in normal and SCD populations are the following: a SNP upstream of the γ-globin gene within the globin locus on chromosome 11, previously identified by restriction enzyme analysis as the XmnI site (37,60,61); *BCL11A*, the gene encoding a transcription factor now recognized as a major silencer of HbF expression (51,58,59,61,63); and the intergenic interval between *HBS1L* and *MYB* (56,58,61). Additional loci have been identified and await replication (59,64,65), in addition to evaluation of known and probable epigenetic effects (**Figure 3**) (66).

Only a few published studies report on the genetics of HbF response to HU in SCD (33,51,54,64). Compared with genomic studies of more common disorders, sample sizes of studies on HU effects in SCD are inevitably modest. Using the retrospective cohort from the Multicenter Study of Hydroxyurea adult trial and assessing more than two dozen candidate genes, Ma *et al.* (54) reported significant associations between SNPs and HbF response to HU in loci of genes involved in the metabolism of arginine to NO and in a transcription factor that induces DNA bending. This report predated the identification of BCL11A as a central regulator of HbF expression. Most of the Multicenter Study of Hydroxyurea patients exhibited a small HbF response to HU (54), with less than 5% change in HbF from baseline. This blunted response is not universal in US adults with SCD and may be influenced by both patient characteristics and adherence to HU regimen.

Whether HbF induction by HU occurs through the direct influence of BCL11A is a concept awaiting direct testing. Effects of BCL11A on HbF are probably mediated through its protein partners, upstream or downstream effectors, chromatin structure, and/or telomerase function (recently reviewed in ref. (67)). Other reports include associations between HU response in SCD and polymorphisms in the guanosine triphosphate–binding protein gene *sar1a* (64), underscoring the complexity of the genetic pathways regulating the HbF response to HU (**Figure 3**).

Two pediatric pharmacogenetic analyses using candidate SNP markers suggested that just a few genes are associated with baseline HbF, including several SNPs within *BCL11A* (**Table 2**). SNP associations with induced HbF are generally not independent of baseline HbF levels (33,51). In contrast

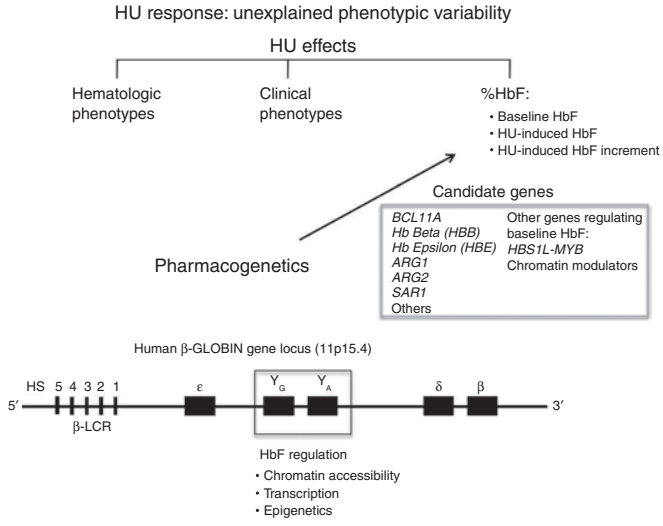


Figure 3. Phenotypic variability in hydroxyurea (HU) response. A diagram synthesizes the varying clinical and genetic effects of HU in sickle cell disease. The β-globin locus is shown below. ARG1 and 2, arginase 1 and 2; Hb, hemoglobin; HbF, fetal Hb; LCR, locus control region; MYB, myeloblastosis oncogene.

Table 2. SNPs associated with HbF

Gene	Phenotype	Clinic	Population	Ethnicity	*N*	SNP	β	*P*	Reference
ARG1	HbF induced by HU	Pediatric	USA	AA	174	rs17599586	NA	4×10^{-3}	(32)
ARG2	HbF induced by HU	Pediatric	USA	AA	174	rs2295644	NA	3×10^{-3}	(32)
ARG2	HbF induced by HU	Adult	USA	AA	137	rs10483801	NA	1×10^{-3}	(53)
BCL11A	HbF baseline	Pediatric	USA	AA, H	108	rs4671393	2.88	5×10^{-5}	(50)
BCL11A	HbF baseline	Pediatric	USA	AA	174	rs4671393	NA	3×10^{-4}	(32)
BCL11A	HbF baseline	Adult	USA	AA	255	rs766432	NA	2×10^{-10}	(62)
BCL11A	HbF baseline	Pediatric + adult	USA	AA	1032	rs4671393	0.60	4×10^{-37}	(59)
BCL11A	HbF baseline	Pediatric + adult	USA	AA	1275	rs4671393	0.60	2×10^{-42}	(60)
BCL11A	HbF baseline	Pediatric + adult	Brazil	NA	350	rs4671393	0.50	3×10^{-8}	(60)
BCL11A	HbF induced by HU	Pediatric	USA	AA, H	47	rs1186868	3.37	0.019	(50)
FTL1	HbF induced by HU	Adult	USA	AA	137	rs2182008	NA	0.003	(53)
GLP2R	F cells baseline	Pediatric + adult	USA	AA	440	rs12103880	−1.36	3×10^{-8}	(58)
HAO2	HbF induced by HU	Adult	USA	AA	137	rs10494225	NA	2×10^{-3}	(53)
HBB	HbF baseline	Pediatric	USA	AA, H	108	rs7482144	3.88	2×10^{-4}	(50)
HBB	HbF baseline	Pediatric + adult	USA	AA	1032	rs10128556	0.42	2×10^{-9}	(59)
HBB	HbF baseline	Pediatric + adult	USA	AA	1275	rs7482144	0.41	4×10^{-7}	(60)
HBE	HbF baseline	Pediatric	USA	AA, H	108	rs7130110	2.86	6×10^{-5}	(50)
HBE	HbF baseline	Pediatric	USA	AA	174	rs7130110	NA	3×10^{-5}	(32)
HBE	HbF induced by HU	Pediatric	USA	AA, H	38	rs7130110	6.04	0.004	(50)
HBS1L-MYB	HbF baseline	Pediatric + adult	USA	AA	1032	rs9402686	0.65	2×10^{-13}	(59)
HBS1L-MYB	HbF baseline	Pediatric + adult	USA	AA	1275	rs9399137	0.60	5×10^{-11}	(60)
NOS1	HbF induced by HU	Adult	USA	AA	137	rs7977109	NA	0.023	(53)
OR51B5/B6	HbF induced by HU	Pediatric	USA	AA	1153	rs5006884	0.20	3×10^{-8}	(64)
OR51B6	HbF baseline	Pediatric	USA	AA, H	108	rs5024042	1.70	0.031	(50)
SAR1A	HbF induced by HU	Adult	USA	AA	32	rs4282891	NA	<0.05	(63)

AA, African American; H, Hispanic; HbF, fetal hemoglobin; HU, hydroxyurea; *N*, sample size; NA, not applicable; SNP, single-nucleotide polymorphism.

with the induced HbF level, the treatment-associated increment appears to be a less relevant marker. Both of these observations probably reflect the association between baseline and induced levels.

In our own smaller multisite analysis, baseline levels of the candidate genes were significantly associated with SNPs within the *BCL11A* and the β- and ε- globin loci (*HBB* and *HBE*, respectively), with an additive attributable variance from these loci of 23% (Table 2) (51). Consistent with studies by Ware *et al.*(14,33), we reported that baseline HbF levels explained 33% of the variance in induced levels. The variant in *HBE* accounted for an additional 13% of the variance in induced levels, whereas variants in the *HBB* and *BCL11A* loci did not contribute beyond baseline levels. Thus, our data suggest that the combined effects of baseline HbF and one SNP marker contributed an estimated 46% of the variance in HbF (51).

By trend analysis, children with an allele associated with higher HbF ("favorable" allele) in one of the *BCL11A* and/or either globin marker had significantly higher average values of baseline HbF than those who lacked a favorable allele (51). Effects on baseline HbF from a SNP in each these two genes were additive and were associated with two-fold higher HbF for patients with favorable alleles in both loci. Similarly, having at least one favorable allele in either globin locus and in *BCL11A* was associated with a higher level of induced HbF. Statistical significance did not withstand adjustment for baseline HbF, probably reflecting the interrelatedness of HbF regulation under both physiological conditions. Genetic studies examining larger pediatric populations on HU, unusual responders, and the influence of specific sequence variants are needed to evaluate the contribution of these and other genetic loci responsible for HbF response.

OTHER PHYSIOLOGICAL EFFECTS OF HU

Cellular Biology

The effects of HU largely depend on its effects on nucleic acid synthesis in dividing RBC progenitors. HU affects the S-phase by inhibiting ribonucleotide reductase, an enzyme important for DNA synthesis. Depletion of DNA precursors by HU causes arrest of the replication fork, leading to cell death. A cell-based *ex vivo* assay for HbF induction, burst-forming unit erythroid colonies grown in methylcellulose from blood of children with SCD, demonstrated that HU decreases the number of burst-forming unit erythroid colonies. HU and other ribonucleotide reductase inhibitors increase HbF production in that system. Interestingly, other cytotoxic agents that are not of that drug category, such as cytarabine and alkylating agents, decreased burst-forming unit erythroid counts but did not induce HbF (41,68).

HU's lethal effect of on ribonucleotide reductase and cell survival are also seen in laboratory bacteria such as *Escherichia coli* (69). Whether the bactericidal effects influence investigation using animal models or even in patients has not previously been studied. Direct bacterial effects on the HU-dampened expression of adhesion molecules should be addressed in a murine model of bacterial infection.

HbF response to temporary marrow toxicity is probably attributable to transcriptional and epigenetic effects on the progenitor developmental program (66,70). HU signaling appears to involve cGMP (cyclic guanine monophosphate), cAMP (cyclic adenosine monophosphate), p38MAPK (mitogen-activated protein kinase), and other pathways. Activation of cGMP may induce HbF via enhancing production of NO (47,68,71,72). NO may also support HbF production (47). HU induces a small guanosine triphosphate–binding protein, the secretion-associated and Ras-related protein (SAR) (73). SAR may be involved in the activation of transcription factors and signal transduction pathways in erythroleukemia K562 cells and in human bone marrow–derived progenitor cells. HU may also function through kinase and signal transduction pathways, such as *globin transcription factor-1*, to enhance γ- and β-globin synthesis in erythroid cells (47).

Gene Expression

Comparing whole mRNA at the pre- and post-initiation stages of therapy revealed that HU affects expression of a number of genes involved in transcription, translation, ribosome assembly, and chromosomal organization (66,70). Results may vary with age, dosing, or other clinical conditions. Variation in cell source, whether from bone marrow or purified early reticulocytes, would be expected to affect detection of expressed genes. HU may also affect expression of genes that link HU and HbF to BCL11A (66,70). Epigenetic analysis of the γ-globin promoter did not reveal much impact from HU (74). Interestingly, HU appears to upregulate specific microRNAs (74). These results require further investigation but underscore the view that HU is involved with complex pathways of gene regulation.

Testing for Oncogenicity

The primary effect is damaging DNA replication by inhibiting ribonucleotide reductase. This effect raises concerns about an oncogenic potential, especially after prolonged use. These fears have been amplified by its original use as chemotherapeutic agent for chronic myeloid leukemia, the latent phase of acute leukemia. Although links with acute leukemia outside of chronic myeloid leukemia have been disproven (75), concerns for the safety of long-term use in children persist. Several studies have tested DNA and cellular toxicity from pediatric HU users. No genotoxicity was detected using several different assays *in vitro*, including karyotype, illegitimate VDJ recombination of the variable regions of rearranged T-cell receptor genes, and chromatid breaks (9). Increased reticulocyte micronuclei were observed, but this effect was highly variable among patients and did not increase with time (76). In all, oncogenicity of HU is probably quite low or nonexistent. A few cases of acute leukemia were reported in patients after many years of HU treatment, but these do not appear to be more frequent than in the untreated population (75).

POTENTIAL PHARMACOLOGICAL ALTERNATIVES TO HU

Other HbF inducers have been assessed during the past few decades, including nucleoside analogs such as 5-azacytidine and decitabine. However, they are often poorly tolerated, potentially oncogenic, and lack proof of effectiveness comparable with HU (recently reviewed in ref. (41)). Additional HbF-inducing drugs are histone deacetylase inhibitors, erythropoietin (already high in SCD and shown not to induce HbF in SCD), valproate, thalidomide derivatives (e.g., pomalidimide), and kit ligand. In all, a variety of cellular stresses and stimuli can promote coordinated stress responses, including activation of the γ-globin gene (66,70). Based on results from SCD mouse models, inhibitors of phosphodiesterase 9 (71) or hypoxia-inducible factor-1α (HIF-1α) (77), alone or in combination with HU, may be clinically useful to stimulate cGMP and NO for HbF production and/or to enhance its antisickling impact (71).

BARRIERS TO USE OF HU

Outside of clinical trials with HU, ample documentation exists of incomplete clinical effectiveness of HU. Uneven drug

adherence has been well documented (78,79). Provider non- and underutilization is well documented (75,80). Our recent multisite survey of parents of children with SCD revealed several family barriers to use of HU, such as lack of FDA approval, near-universal safety concerns, and highly varied knowledge about its benefits, including many for whom its basic property of decreasing episodes of pain was unknown (81). Use of HU was positively correlated with fundamental knowledge of parents regarding the basic positive effects of HU on disease, independent of parental demographics such as education level, language spoken, or ethnicity. Barriers in effective communication between providers and families may be exacerbated by issues arising from medical delivery systems.

The mixed uptake of HU by families may also reflect family perspectives on the long-term effects of SCD. A single-site survey of parents revealed that the majority believed that the disease effects were going to diminish over time and would not affect life goals or life span (82). These poignant perceptions will need to be addressed if families are to embrace the long-term benefits of HU against its inconveniences and largely theoretical risks.

CONCLUSION

HU is a remarkably effective drug for a large proportion of children with SCD. Efforts to achieve expanded understanding of the scientific underpinnings of its effects on SCD, predict individual response, and perfect the clinical applications for modifying disease effects are ongoing. Clinical trials will continue to test the uncertain benefits of HU (**Table 1**), such as primary prevention of brain infarcts (clinicaltrials.gov/show/NCT01389024). Murine models will facilitate insight into the benefits provided by induced HbF, altered expression of adhesion molecules, reduced BCL11A levels, and other mechanisms. Genetic epidemiology will be used to identify specific variants in regulatory genes and gene pathways.

The accumulating science of HU is anticipated to lead to three direct effects for children with HbSS: (i) use at earlier ages; (ii) wider clinical indications; and (iii) delineation of children who are less likely to enjoy substantive benefit from HU. For this last group, more aggressive consideration of chronic transfusion, hematopoietic stem cell transplantation, or trial of emerging alternative agents may be warranted. To date, early clinical trials of other experimental HbF-inducing drugs have demonstrated considerable short- and long-term toxicity compared with HU. Therefore, HU is predicted to remain the mainstay of pharmacological therapy for SCD in the foreseeable future.

Progress toward the Healthy People 2020 goals will occur through increased use of HU. Nonetheless, the entire SCD population may not benefit from HU alone. Dampened impact on clinical complications and HbF induction occurs for those with certain genotypes, many patients with HbSC (compound heterozygous for HbS and HbC), some adult patients, and those with renal compromise. New, effective, and safe therapies, alone or in combination with HU, are still needed to maximize pharmacological benefit for everyone living with SCD. Constructive engagement must be made to assist families in undertaking long-term HU use to help them to balance the optimism of HU treatment and its potential toxicities with the risk of accumulating disease consequences. Although several important crisis modulators are currently under investigation, disease modifiers that prevent crises and other morbidities are arguably the primary therapeutic targets.

STATEMENT OF FINANCIAL SUPPORT

The authors acknowledge support from the Clinical Translational Science Award at Columbia University, 5UL1RR024156 (H.N. Ginsberg, principal investigator).

Disclosure: No commercial or other conflict of interest is reported by the authors.

REFERENCES

1. Healthy People 2020. (http://healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=4.) Accessed 6 September 2012.
2. Hassell KL. Population estimates of sickle cell disease in the U.S. Am J Prev Med 2010;38:Suppl:S512–21.
3. Hamideh D, Alvarez O. Sickle cell disease related mortality in the United States (1999-2009). Pediatr Blood Cancer 2013;60:1482–6.
4. Quinn CT, Rogers ZR, McCavit TL, Buchanan GR. Improved survival of children and adolescents with sickle cell disease. Blood 2010;115:3447–52.
5. Yanni E, Grosse SD, Yang Q, Olney RS. Trends in pediatric sickle cell disease-related mortality in the United States, 1983-2002. J Pediatr 2009;154:541–5.
6. Lanzkron S, Carroll CP, Haywood C Jr. Mortality rates and age at death from sickle cell disease: U.S., 1979-2005. Public Health Rep 2013;128:110–6.
7. Charache S, Terrin ML, Moore RD, et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. N Engl J Med 1995;332:1317–22.
8. Ferster A, Vermylen C, Cornu G, et al. Hydroxyurea for treatment of severe sickle cell anemia: a pediatric clinical trial. Blood 1996;88:1960–4.
9. McGann PT, Ware RE. Hydroxyurea for sickle cell anemia: what have we learned and what questions still remain? Curr Opin Hematol 2011;18:158–65.
10. Strouse JJ, Heeney MM. Hydroxyurea for the treatment of sickle cell disease: efficacy, barriers, toxicity, and management in children. Pediatr Blood Cancer 2012;59:365–71.
11. Strouse JJ, Lanzkron S, Beach MC, et al. Hydroxyurea for sickle cell disease: a systematic review for efficacy and toxicity in children. Pediatrics 2008;122:1332–42.
12. Kinney TR, Helms RW, O'Branski EE, et al. Safety of hydroxyurea in children with sickle cell anemia: results of the HUG-KIDS study, a phase I/II trial. Pediatric Hydroxyurea Group. Blood 1999;94:1550–4.
13. Wang WC, Helms RW, Lynn HS, et al. Effect of hydroxyurea on growth in children with sickle cell anemia: results of the HUG-KIDS Study. J Pediatr 2002;140:225–9.
14. Ware RE, Eggleston B, Redding-Lallinger R, et al. Predictors of fetal hemoglobin response in children with sickle cell anemia receiving hydroxyurea therapy. Blood 2002;99:10–4.
15. Hankins JS, Ware RE, Rogers ZR, et al. Long-term hydroxyurea therapy for infants with sickle cell anemia: the HUSOFT extension study. Blood 2005;106:2269–75.
16. Thornburg CD, Dixon N, Burgett S, et al. A pilot study of hydroxyurea to prevent chronic organ damage in young children with sickle cell anemia. Pediatr Blood Cancer 2009;52:609–15.
17. Thornburg CD, Files BA, Luo Z, et al.; BABY HUG Investigators. Impact of hydroxyurea on clinical events in the BABY HUG trial. Blood 2012;120:4304–10; quiz 4448.
18. Wang WC, Ware RE, Miller ST, et al.; BABY HUG investigators. Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). Lancet 2011;377:1663–72.

19. Bernaudin F, Verlhac S, Arnaud C, et al. Impact of early transcranial Doppler screening and intensive therapy on cerebral vasculopathy outcome in a newborn sickle cell anemia cohort. Blood 2011;117:1130–40; quiz 1436.
20. de Montalembert M, Brousse V, Elie C, Bernaudin F, Shi J, Landais P; French Study Group on Sickle Cell Disease. Long-term hydroxyurea treatment in children with sickle cell disease: tolerance and clinical outcomes. Haematologica 2006;91:125–8.
21. Thornburg CD, Calatroni A, Panepinto JA. Differences in health-related quality of life in children with sickle cell disease receiving hydroxyurea. J Pediatr Hematol Oncol 2011;33:251–4.
22. Dampier C, Lieff S, LeBeau P, et al.; Comprehensive Sickle Cell Centers (CSCC) Clinical Trial Consortium (CTC). Health-related quality of life in children with sickle cell disease: a report from the Comprehensive Sickle Cell Centers Clinical Trial Consortium. Pediatr Blood Cancer 2010;55:485–94.
23. Zimmerman SA, Schultz WH, Davis JS, et al. Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. Blood 2004;103:2039–45.
24. Aygun B, Mortier NA, Smeltzer MP, Shulkin BL, Hankins JS, Ware RE. Hydroxyurea treatment decreases glomerular hyperfiltration in children with sickle cell anemia. Am J Hematol 2013;88:116–9.
25. Alvarez O, Miller ST, Wang WC, et al.; BABY HUG Investigators. Effect of hydroxyurea treatment on renal function parameters: results from the multi-center placebo-controlled BABY HUG clinical trial for infants with sickle cell anemia. Pediatr Blood Cancer 2012;59:668–74.
26. Kratovil T, Bulas D, Driscoll MC, Speller-Brown B, McCarter R, Minniti CP. Hydroxyurea therapy lowers TCD velocities in children with sickle cell disease. Pediatr Blood Cancer 2006;47:894–900.
27. Zimmerman SA, Schultz WH, Burgett S, Mortier NA, Ware RE. Hydroxyurea therapy lowers transcranial Doppler flow velocities in children with sickle cell anemia. Blood 2007;110:1043–7.
28. Ware RE, Helms RW; SWiTCH Investigators. Stroke With Transfusions Changing to Hydroxyurea (SWiTCH). Blood 2012;119:3925–32.
29. Alvarez O, Yovetich NA, Scott JP, et al.; Investigators of the Stroke With Transfusions Changing to Hydroxyurea Clinical Trial (SWiTCH). Pain and other non-neurological adverse events in children with sickle cell anemia and previous stroke who received hydroxyurea and phlebotomy or chronic transfusions and chelation: results from the SWiTCH clinical trial. Am J Hematol 2013;88:932–8.
30. Steinberg MH, McCarthy WF, Castro O, et al.; Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia and MSH Patients' Follow-Up. The risks and benefits of long-term use of hydroxyurea in sickle cell anemia: a 17.5 year follow-up. Am J Hematol 2010;85:403–8.
31. Voskaridou E, Christoulas D, Bilalis A, et al. The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: results of a 17-year, single-center trial (LaSHS). Blood 2010;115:2354–63.
32. Lobo CL, Pinto JF, Nascimento EM, Moura PG, Cardoso GP, Hankins JS. The effect of hydroxcarbamide therapy on survival of children with sickle cell disease. Br J Haematol 2013;161:852–60.
33. Ware RE, Despotovic JM, Mortier NA, et al. Pharmacokinetics, pharmacodynamics, and pharmacogenetics of hydroxyurea treatment for children with sickle cell anemia. Blood 2011;118:4985–91.
34. Platt OS. Hydroxyurea for the treatment of sickle cell anemia. N Engl J Med 2008;358:1362–9.
35. Bunn HF. Subunit assembly of hemoglobin: an important determinant of hematologic phenotype. Blood 1987;69:1–6.
36. Powars DR, Weiss JN, Chan LS, Schroeder WA. Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? Blood 1984;63:921–6.
37. Green NS, Fabry ME, Kaptue-Noche L, Nagel RL. Senegal haplotype is associated with higher HbF than Benin and Cameroon haplotypes in African children with sickle cell anemia. Am J Hematol 1993;44:145–6.
38. Makani J, Menzel S, Nkya S, et al. Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. Blood 2011;117:1390–2.
39. Nagel RL, Erlingsson S, Fabry ME, et al. The Senegal DNA haplotype is associated with the amelioration of anemia in African-American sickle cell anemia patients. Blood 1991;77:1371–5.
40. Solovieff N, Hartley SW, Baldwin CT, et al. Ancestry of African Americans with sickle cell disease. Blood Cells Mol Dis 2011;47:41–5.
41. Fathallah H, Atweh GF. Induction of fetal hemoglobin in the treatment of sickle cell disease. Hematology Am Soc Hematol Educ Program 2006;1:58–62.
42. Lebensburger JD, Pestina TI, Ware RE, Boyd KL, Persons DA. Hydroxyurea therapy requires HbF induction for clinical benefit in a sickle cell mouse model. Haematologica 2010;95:1599–603.
43. Gambero S, Canalli AA, Traina F, et al. Therapy with hydroxyurea is associated with reduced adhesion molecule gene and protein expression in sickle red cells with a concomitant reduction in adhesive properties. Eur J Haematol 2007;78:144–51.
44. Tsai M, Kita A, Leach J, et al. *In vitro* modeling of the microvascular occlusion and thrombosis that occur in hematologic diseases using microfluidic technology. J Clin Invest 2012;122:408–18.
45. Nébor D, Romana M, Santiago R, et al. Fetal hemoglobin and hydroxycarbamide moduate both plasma concentration and cellular origin of circulating microparticles in sickle cell anemia children. Haematologica 2013;98:862–7.
46. Lebensburger JD, Howard T, Hu Y, et al. Hydroxyurea therapy of a murine model of sickle cell anemia inhibits the progression of pneumococcal disease by down-modulating E-selectin. Blood 2012;119:1915–21.
47. Lou TF, Singh M, Mackie A, Li W, Pace BS. Hydroxyurea generates nitric oxide in human erythroid cells: mechanisms for gamma-globin gene activation. Exp Biol Med (Maywood) 2009;234:1374–82.
48. Kato GJ, Hebbel RP, Steinberg MH, Gladwin MT. Vasculopathy in sickle cell disease: biology, pathophysiology, genetics, translational medicine, and new research directions. Am J Hematol 2009;84:618–25.
49. King SB. Nitric oxide production from hydroxyurea. Free Radic Biol Med 2004;37:737–44.
50. Bunn HF, Nathan DG, Dover GJ, et al. Pulmonary hypertension and nitric oxide depletion in sickle cell disease. Blood 2010;116:687–92.
51. Green NS, Ender KL, Pashankar F, et al. Candidate sequence variants and fetal hemoglobin in children with sickle cell disease treated with hydroxyurea. PLoS ONE 2013;8:e55709.
52. Meier ER, Byrnes C, Weissman M, Noel P, Luban NL, Miller JL. Expression patterns of fetal hemoglobin in sickle cell erythrocytes are both patient- and treatment-specific during childhood. Pediatr Blood Cancer 2011;56:103–9.
53. Italia K, Jain D, Gattani S, et al. Hydroxyurea in sickle cell disease–a study of clinico-pharmacological efficacy in the Indian haplotype. Blood Cells Mol Dis 2009;42:25–31.
54. Ma Q, Wyszynski DF, Farrell JJ, et al. Fetal hemoglobin in sickle cell anemia: genetic determinants of response to hydroxyurea. Pharmacogenomics J 2007;7:386–94.
55. Steinberg MH, Voskaridou E, Kutlar A, et al. Concordant fetal hemoglobin response to hydroxyurea in siblings with sickle cell disease. Am J Hematol 2003;72:121–6.
56. Thein SL, Menzel S. Discovering the genetics underlying foetal haemoglobin production in adults. Br J Haematol 2009;145:455–67.
57. Garner C, Tatu T, Reittie JE, et al. Genetic influences on F cells and other hematologic variables: a twin heritability study. Blood 2000;95:342–6.
58. Bae HT, Baldwin CT, Sebastiani P, et al. Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. Blood 2012;120:1961–2.
59. Bhatnagar P, Purvis S, Barron-Casella E, et al. Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. J Hum Genet 2011;56:316–23.
60. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet 2010;42:1049–51.
61. Lettre G, Sankaran VG, Bezerra MA, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci USA 2008;105:11869–74.
62. Sebastiani P, Solovieff N, Hartley SW, et al. Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. Am J Hematol 2010;85:29–35.

63. Sedgewick AE, Timofeev N, Sebastiani P, et al. BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. Blood Cells Mol Dis 2008;41:255–8.
64. Kumkhaek C, Taylor JG 6th, Zhu J, Hoppe C, Kato GJ, Rodgers GP. Fetal haemoglobin response to hydroxycarbamide treatment and sar1a promoter polymorphisms in sickle cell anaemia. Br J Haematol 2008;141:254–9.
65. Solovieff N, Milton JN, Hartley SW, et al. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. Blood 2010;115:1815–22.
66. Flanagan JM, Steward S, Howard TA, et al. Hydroxycarbamide alters erythroid gene expression in children with sickle cell anaemia. Br J Haematol 2012;157:240–8.
67. Bauer DE, Orkin SH. Update on fetal hemoglobin gene regulation in hemoglobinopathies. Curr Opin Pediatr 2011;23:1–8.
68. Yang YM, Pace B. Pharmacologic induction of fetal hemoglobin synthesis: cellular and molecular mechanisms. Pediatr Pathol Mol Med 2001;20:87–106.
69. Bollenbach T, Quan S, Chait R, Kishony R. Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. Cell 2009;139:707–18.
70. Costa FC, da Cunha AF, Fattori A, et al. Gene expression profiles of erythroid precursors characterise several mechanisms of the action of hydroxycarbamide in sickle cell anaemia. Br J Haematol 2007;136:333–42.
71. Almeida CB, Scheiermann C, Jang JE, et al. Hydroxyurea and a cGMP-amplifying agent have immediate benefits on acute vaso-occlusive events in sickle cell disease mice. Blood 2012;120:2879–88.
72. Cokic VP, Andric SA, Stojilkovic SS, Noguchi CT, Schechter AN. Hydroxyurea nitrosylates and activates soluble guanylyl cyclase in human erythroid cells. Blood 2008;111:1117–23.
73. Tang DC, Zhu J, Liu W, et al. The hydroxyurea-induced small GTP-binding protein SAR modulates gamma-globin gene expression in human erythroid cells. Blood 2005;106:3256–63.
74. Walker AL, Steward S, Howard TA, et al. Epigenetic and molecular profiles of erythroid cells after hydroxyurea treatment in sickle cell anemia. Blood 2011;118:5664–70.
75. Lanzkron S, Strouse JJ, Wilson R, et al. Systematic review: hydroxyurea for the treatment of adults with sickle cell disease. Ann Intern Med 2008;148:939–55.
76. Flanagan JM, Howard TA, Mortier N, et al. Assessment of genotoxicity associated with hydroxyurea therapy in children with sickle cell anemia. Mutat Res 2010;698:38–42.
77. Kaul DK, Fabry ME, Suzuka SM, Zhang X. Antisickling fetal hemoglobin reduces hypoxia-inducible factor-1a expression in normoxic sickle mice: microvascular implications. Am J Physiol Heart Circ Physiol 2013;304:H42–50.
78. Candrilli SD, O'Brien SH, Ware RE, Nahata MC, Seiber EE, Balkrishnan R. Hydroxyurea adherence and associated outcomes among Medicaid enrollees with sickle cell disease. Am J Hematol 2011;86:273–7.
79. Thornburg CD, Rogers ZR, Jeng MR, et al.; BABY HUG Investigators. Adherence to study medication and visits: data from the BABY HUG trial. Pediatr Blood Cancer 2010;54:260–4.
80. Haywood C Jr, Beach MC, Lanzkron S, et al. A systematic review of barriers and interventions to improve appropriate use of therapies for sickle cell disease. J Natl Med Assoc 2009;101:1022–33.
81. Oyeku SO, Driscoll MC, Cohen HW, et al. Parental and other factors associated with hydroxyurea use for pediatric sickle cell disease. Pediatr Blood Cancer 2013;60:653–8.
82. Roth M, Krystal J, Manwani D, Driscoll C, Ricafort R. Stem cell transplant for children with sickle cell anemia: parent and patient interest. Biol Blood Marrow Transplant 2012;18:1709–15.