

## DATABASES

# *HbVar*: A Relational Database of Human Hemoglobin Variants and Thalassemia Mutations at the Globin Gene Server

Ross C. Hardison,<sup>1\*</sup> David H.K. Chui,<sup>3</sup> Belinda Giardine,<sup>2</sup> Cathy Riemer,<sup>2</sup> George P. Patrinos,<sup>4</sup> Nicholas Anagnou,<sup>5</sup> Webb Miller,<sup>2</sup> and Henri Wajcman<sup>6</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania

<sup>2</sup>Department of Computer Science and Engineering, Pennsylvania State University, University Park, Pennsylvania

<sup>3</sup>Department of Pathology, McMaster University, Hamilton, Ontario, Canada

<sup>4</sup>Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

<sup>5</sup>Department of Basic Sciences, University of Crete, Heraklion, Greece

<sup>6</sup>INSERM-U468, Génétique Moléculaire et Physiopathologie, Hôpital Henri Mondor, Créteil, France

Communicated by Sergio Ottolenghi

We have constructed a relational database of hemoglobin variants and thalassemia mutations, called *HbVar*, which can be accessed on the web at <http://globin.cse.psu.edu>. Extensive information is recorded for each variant and mutation, including a description of the variant and associated pathology, hematology, electrophoretic mobility, methods of isolation, stability information, ethnic occurrence, structure studies, functional studies, and references. The initial information was derived from books by Dr. Titus Huisman and colleagues [Huisman et al., 1996, 1997, 1998]. The current database is updated regularly with the addition of new data and corrections to previous data. Queries can be formulated based on fields in the database. Tables of common categories of variants, such as all those involving the alpha1-globin gene (*HBA1*) or all those that result in high oxygen affinity, are maintained by automated queries on the database. Users can formulate more precise queries, such as identifying “all beta-globin variants associated with instability and found in Scottish populations.” This new database should be useful for clinical diagnosis as well as in fundamental studies of hemoglobin biochemistry, globin gene regulation, and human sequence variation at these loci. *Hum Mutat* 19:225–233, 2002. © 2002 Wiley-Liss, Inc.

**KEY WORDS:** hemoglobin variants; thalassemia; relational database; gene regulation; Globin Gene Server; HBE1; HBG2; HBG1; HBD; HBB; HBZ2; HBZ1; HBA2; HBA1

### DATABASES:

<http://globin.cse.psu.edu> (Globin Gene Server)

<http://globin.cse.psu.edu/globin/hbvar/menu.html> (*HbVar* Database)

## INTRODUCTION

The most common inherited diseases in humans result from mutations in the beta-globin gene complex and the alpha-globin gene complex. We refer to the beta-globin gene complex as *HBBC*, composed of the genes *HBE1* (MIM# 142100), *HBG2* (MIM# 142250), *HBG1* (MIM# 142200), *HBD* (MIM# 142000), and *HBB* (MIM# 141900), which encode the epsilon-, <sup>G</sup>gamma-, <sup>A</sup>gamma-, delta-, and beta-globin polypeptides, respectively. We refer to the alpha-globin gene complex as *HBAC*, composed of the genes *HBZ2* (MIM#

142310), *HBA2* (MIM# 141850), *HBA1* (MIM# 141800), and *HBQ1* (MIM# 142240), which encode the zeta-, alpha2-, alpha1-, and possibly theta-globin polypeptides, respectively.

Received 4 June 2001; accepted revised manuscript 12 November 2001.

\*Correspondence to: Ross Hardison, Department of Biochemistry and Molecular Biology, Pennsylvania State University, 206 Althouse Laboratory, University Park, PA 16802. E-mail: [rch8@psu.edu](mailto:rch8@psu.edu)

Contract grant sponsor: Public Health Service; Contract grant numbers: HG02238; DK27635.

Inherited disorders of hemoglobin have been reviewed recently [Forget et al., 2001]. Single nucleotide substitutions can lead to amino acid replacements that cause hemolytic anemias, such as sickle-cell disease, or hemoglobins that are unstable or have altered oxygen affinity. Substitutions or deletions, which occur in any of several regions of the genes, can reduce expression of *HBB*, *HBA2*, or *HBA1* to cause the inherited anemia called thalassemia. Some other sequence changes have little or no effect on hemoglobin function, but are useful polymorphisms for genetic studies. Previous compilations [Bunn and Forget, 1986; Huisman et al., 1996, 1997, 1998] revealed a large number of hemoglobin variants, affecting almost every amino acid in the globin polypeptides. New hemoglobin variants and thalassemias continue to be discovered, and it is critical to maintain an up-to-date and accessible repository of this information.

The books that recorded information on hemoglobin variants and thalassemias [Huisman et al., 1996, 1997, 1998] are a rich source of information not only about the mutations, but also about the methods used in detection and analysis, their biochemical properties, associated clinical effects, ethnic distribution, and other data. This richness presents a great opportunity for users but it presents formidable challenges for database construction. Indeed, our efforts to move this information into a database format have occurred in three discrete stages. Our initial work converted *A Syllabus of Human Hemoglobin Variants (1996)* [Huisman et al., 1996] to an HTML format that could be freely accessed on the web [Chui et al., 1998; Hardison et al., 1998b]. Although this resource is valuable because of its ease of access and capacity to support simple text queries, it has many limitations. It is not a true database, it does not use a controlled vocabulary, and the query capacity is limited. Much of the information in the online *Syllabus* was then converted into an SRS database at the European Bioinformatics Institute [Hardison et al., 2001]. This is accessible at <http://srs.ebi.ac.uk>. However, full control of the vocabulary was not achieved and no mechanism was established for updating the information. Much of *A Syllabus of Thalassemia Mutations* [Huisman et al., 1997] was converted into HTML and is available online, but it was not converted into a database.

We now report on the construction of a true relational database capturing much of the information in *A Syllabus of Thalassemia Mutations* and *A Syllabus of Human Hemoglobin Variants (Second Edition)* [Huisman et al., 1997, 1998]. We will describe the construction and curation of the database and give examples of its use.

#### **Informatics: Sources, Design, Implementation, and Access to HbVar**

**Sources.** The primary sources of information in *HbVar* are the books *A Syllabus of Human Hemoglobin Variants (Second Edition)* [Huisman et al., 1998] and *A Syllabus of Thalassemia Mutations, 1997* [Huisman et al., 1997]. Additional entries and corrections are made continually by the curators. Dr. Henri Wajcman is the curator for the hemoglobin variants, and Drs. George Patrinos and Nicholas Anagnou are the curators for the thalassemia mutations.

**Design and implementation.** The two *Syllabi* provide several types of information about the hemoglobin variants and thalassemia mutations, and the database was built to accommodate this information. We elected to build *HbVar* as a relational database, using Oracle as the database management system. This choice was guided partly by the desire to support efficient interfaces with other databases and sequence analysis output. Our previous efforts in developing a database of experimental results on gene expression [Riemer et al., 1998] used a custom database management system tailored for extensive nesting of data. The structure of the hemoglobin gene mutation data is somewhat simpler, and indeed the use of multiple tables in the relational database accommodated the lesser degree of nesting in these data. The hemoglobin variants and the thalassemias were combined into a single database because most of the fields were the same for both classes of mutations, and some variants also cause thalassemia. One can easily restrict a search to one class or the other on the query page.

The various types of information available for each variant are illustrated for Hb Henri Mondor in Figure 1. Not only is the position and type of amino acid substitution given, but substantial additional material about the biochemistry, function, and clinical effects of the variant are presented. The principal components of the schema required to accommodate these fields are listed in Table 1.

Human Hemoglobin Variants and Thalassemias

Hb Henri Mondor beta 26(B8) Glu>Val

Category  
Hbvar  
3D position  
External

Hematology and Clinical Presentation

	Clinical presentation	Laboratory findings	Comments
Heterozygote	<ul style="list-style-type: none"> <li>● mild Anemia</li> </ul>	<ul style="list-style-type: none"> <li>● Microcytosis</li> <li>● Hypochromia</li> <li>● Hb X 37.5 % of total Hb</li> </ul>	

Electrophoresis  
Electrophoresis comment: Hb X and Hb A<sub>2</sub> can be separated at alkaline pH; Hb X moves between Hb F and Hb S

Chromatography  
DEAE-Sephadex: Hb X was isolated

Structure Studies  
Methods for obtaining structural information

- Protein analysis
  - Cleavage method
    - Trypsin digestion of isolated globin chain
  - Separation method
    - Cation exchange chromatography
    - Fingerprinting (peptide separation)
  - Determination
    - Amino acid composition
      - Amino acid analysis
    - Amino acid sequence
      - Amino acid sequence determination

Mutation sequence analysis  
DNA sequence changes are Computed;  
Protein sequence changes are Experimental;  
GAG>GTG at codon 26 in beta

Stability  
Relative stability  
□ Mildly unstable

Occurrence  
Ethnic background  
African  
Occurrence Comment  
Found in a 9-year-old African female

References

Ref ID	Medline ID	Authors	Title	Journal	Year	Volume	Num	Pp
579	1001469	Blouquit Y Arous N Machado PF Garel MC	Hb Henri Mondor: beta26 (B8) Glu leads to Val: a variant with a substitution localized at the same position as that of HbE beta26 Glu leads to Lys.	FEBS Lett	1976	72	1	5-7

New Query  
<http://globin.cse.psu.edu/globin/hbvar/menu.html> May 24, 2001

FIGURE 1. Detailed entry for Hb Henri Mondor from the *HbVar* database. The information in the various fields is organized by the database management system and returned in a format similar to that in Huisman et al. [1998].

The description of the variant includes the name, the position and alteration in the amino acid sequence, the computed DNA sequence change, and the position of the affected amino acid in hemoglobin, both in the secondary structure (e.g., position 8 in helix B for Hb Henri Mondor, Fig. 1) and in the tertiary structure (e.g., external). Detailed information about a mutation is entered only once, and other information is deduced from that automatically. Many hemoglobin variants have been and continue to be discovered and characterized as altered proteins and polypeptides. Thus the experimental data are more often from protein structural information (amino acid sequence or mass spectrometry) than from DNA sequence. We incorporated both forward and reverse translation, along with the wild-type sequence, so the program can compute the predicted DNA sequence changes from the observed variant and wild-type sequences, or compute the predicted amino acid changes from observed DNA se-

quence changes. It is important to have both protein and DNA data, since post-translational modifications can differ as a result of some amino acid changes. For example, the substitution of glutamate for valine at position 1 of the alpha-globin chain in Hb Thionville causes it to retain the initiator methionine, which also is acetylated. The substitution of cysteine for serine at position 9 of the beta-globin chain in Hb Pôrto Alegre leads to polymerization of hemoglobin tetramers by formation of intermolecular disulfide bridges. Also, a mutation may encode a residue differing from that found in the mature protein. For instance, deamidation of asparagine to form aspartate in Hb Osler mimics Hb Nancy, in which the mutation directly codes for an aspartate at the same position [Préhu et al., 1998]. This information on post-translational modification is currently entered as free comments.

An additional complication is that five different numbering systems are used for globins and the genes that encode them. The amino acid

TABLE 1. A Synopsis of the Schema for *HbVar*—  
a Database of Human Hemoglobin Variants  
and Thalassemias

Name
Category
Type of Thalassemia
Description:
Chain
Residue number
Substitutions
Insertions
Deletions
Fusion gene Hbs
Contact
Haplotype
Hematology:
Genotype
Hematological findings
Modifier
Condition
Laboratory findings
Assay
Range
Units
Other factors
Electrophoresis
Method
Quantitative result
Chromatography
Method
Stability
Relative stability
Dissociation
Other stability information
Occurrence
Ethnic background
Frequency
Structure studies
Separation of hemoglobins
Separation of globin chains
Methods
Protein analysis
DNA analysis
Functional studies
Study
Result
What the study covered
Comments on the variant
References
Authors/editors
Journal articles
Other references

numbering system that is commonly used by biochemists studying hemoglobins starts with the amino acid after the initiator methionine as +1, whereas the HUGO recommended nomenclature assigns the initiator methionine itself as +1. The DNA numbering system for the globin genes that is commonly used in publications assigns +1 to the “capped” nucleotide, which is the one corresponding to the first nucleotide in the mRNA. The HUGO recommended nomenclature uses the A in the ATG initiation codon as +1. Finally, we have chosen GenBank sequence

files that extend through the entire gene clusters for use as reference sequences; currently these are U01317 for *HBBC* and Z84721 for *HBAC*. They have the start sites for the genes at particular known positions in the files. The database management system has been programmed to automatically convert among these five numbering systems. Users need to specify the desired numbering system for their queries, and of course curators must choose an appropriate system for entering data as well. The ability to automatically convert among the different systems allows considerable flexibility and frees the user from having to make the conversions manually, which can be an error-prone process. The correlations between the amino acid position, the helix position in the secondary structure, and the position in the tertiary structure are also made automatically, based on tables provided by the curators.

The section on hematology and clinical presentation is a summary of published results (Fig. 1, Table 1). In this and subsequent sections, considerable effort was made to enforce a uniform, controlled vocabulary. Descriptions in the *Syllabi* expressed the same concept in multiple ways; e.g., “anemia,” “hemolytic anemia,” and “hemolysis” are all highly related and equivalent for the purpose of this database. Thus, with advice from the curators, we went through many cycles of reducing multiple terms to a single synonym. Data entry and queries are now performed primarily via menus and lists to enforce this controlled vocabulary. The Boolean operators AND and OR can be used for some but not all fields, depending on what is being described. For example, an anemia cannot be both mild and severe, but a patient can have both anemia and hypochromia. Additional information on laboratory findings, such as hematocrit, hemoglobin levels, and other hematological measurements is recorded quantitatively for some published cases (often the initial observations).

Techniques used to identify, isolate, and establish the structure of each hemoglobin variant are recorded as various electrophoretic and chromatographic methods (Fig. 1, Table 1). The information from the *Syllabi* was greatly enhanced by the addition of detailed results for electrophoresis while developing the current relational database. A quantitative scale has been developed to describe mobility of hemoglobins

in standard electrophoretic techniques [Wajcman et al., 2001]. Values for electrophoretic mobility for each of the hemoglobin variants are being recorded as they are measured.

Additional fields record (and allow querying upon) stability of variant hemoglobins, functional studies (e.g., 2,3-diphosphoglycerate effect, Bohr effect, cooperativity), ethnic group and geographic location of the published occurrences of the mutations, and reference. As for other categories, a controlled vocabulary has been developed and enforced.

**Access.** The *HbVar* database can be accessed on the web at the Globin Gene Server [Hardison et al., 1994, 1995, 1998a, 2001; Chui et al., 1998], at <http://globin.cse.psu.edu>. Instructions and frequently asked questions are also available from the *HbVar* page.

**Examples of Use**

*HbVar* can be examined in two different ways: through categorized listings or user-generated queries.

An up-to-date tabulation of common groups of hemoglobin variants and thalassemia mutations is maintained automatically. The results of these automated queries are computed and returned when one clicks on “Summaries of mutation categories” at the *HbVar* homepage. The totals as of October 2001 are listed in Table 2. Note that the set of hemoglobin variants (with 832 entries) overlaps with the set of thalassemia mutations (336 entries), so the sum of these ex-

ceeds the total number of entries, which is 1,125. One can click on the button for each category to obtain a listing of all entries therein, and clicking on a specific entry will then return a page similar to the one shown for Hb Henri Mondor in Figure 1. The format of the page returned for each variant or mutation is similar to that used in the original *Syllabi*.

The variety and depth of information recorded in *HbVar* allows one to submit highly specific queries. For example, one can ask for all hemoglobin variants involving substitutions in the beta-globin chain that are unstable and found in Scottish populations. This requires selecting beta-globin as the chain, any substitution as the change, unstable or any form of this for all tests of hemoglobin stability, and Scottish as the ethnic group (illustrated in part in Fig. 2A). Three hemoglobins, listed in Figure 2B, are returned, and one can find more detailed information by following the hyperlinks to the individual entries, which are again displayed in a format similar to Figure 1. One can imagine a query similar to this one being useful in a clinical setting, not for a definitive diagnosis, but rather as a source of data to help reach a diagnosis.

Other queries are useful for biochemical studies. A recent study used *HbVar* to collect all the hemoglobin variants that cause erythrocytosis [Rai et al., 2001].

One can also ask questions relevant to gene expression studies, such as “Find all mutations between positions -92 and -86 of *HBB* that are associated with beta+ thalassemia.” The nine mutations returned (Table 3) are alterations in many of the positions of the binding site for EKLF, with some positions mutated to different nucleotides in different thalassemia patients.

These examples show only a fraction of the power of this database. One can query on any of the properties outlined in Table 1, thereby allowing combinations of criteria involving biochemical, genetic, and clinical aspects of globin gene mutations.

The *HbVar* database and associated resources, such as the online *Syllabi*, are currently in use worldwide. In April of 2001, we recorded 8,154 accesses to the online *Syllabi* from 1,418 unique IP addresses. *HbVar* was accessed 1,791 times from 400 unique IP addresses, despite the fact that is only now being announced.

TABLE 2. Summaries of Mutation Categories in *HbVar*—a Database of Human Hemoglobin Variants and Thalassemias

Query	Count of results
Total entries in database	1,125
Total hemoglobin variant entries	832
Total thalassemia entries	336
Total entries in both variant and thalassemia categories	43
Entries involving the alpha1 gene	212
Entries involving the alpha2 gene	250
Entries involving the beta gene	635
Entries involving the delta gene	57
Entries involving the <sup>A</sup> gamma gene	44
Entries involving the <sup>G</sup> gamma gene	52
Entries with a fusion gene mutation	8
Entries with a substitution mutation	899
Entries with an insertion mutation	46
Entries with a deletion mutation	116
Hemoglobins with high oxygen affinity	79
Unstable hemoglobins	121
Methemoglobins	9

**A.**

Location: [http://globin.cse.psu.edu/cgi-bin/hbvar/query\\_vare?](http://globin.cse.psu.edu/cgi-bin/hbvar/query_vare?)

**Stability**

**Relative stability**

Any test: Normal to mildly unstable, Unstable, Very unstable

Deduced from low abundance: Hyperunstable

Heat stability: Inconclusive, Mildly unstable

Isopropanol precipitation: Inconclusive, Mildly unstable

**Dissociation**

Decreased dissociation, Associates into dimers, Associates into dimers (oxyHb)

**Other stability information**

Abnormal polymerization, Fetus inclusion bodies, Increased autooxidation

Joined with:  OR  AND

**Occurrence**

Ethnic background: Saudi Arabian, Scandinavian, Scottish, Sephardic Jewish, Serbian

Joined with:  OR  AND

Frequency range from: to

Text description:

**B.**

### Human Hemoglobin Variants and Thalassemias -- Query Results

There are 3 matches to your query

Name	Mutation	Mutation, HUGO nomenclature
<a href="#">Hb Warwickshire</a>	beta 5(A2) Pro>Arg	HBB g.17C>G
<a href="#">Hb J-Auckland</a>	beta 25(B7) Gly>Asp	HBB g.77G>A
<a href="#">Hb Sabine</a>	beta 91(F7) Leu>Pro	HBB g.405T>C

[New Query](#)

<http://globin.cse.psu.edu/globin/hbvar/menu.html> May 25, 2001

FIGURE 2. Construction and results of the query “Find all hemoglobin variants involving substitutions in the beta-globin chain that are unstable hemoglobin and found in Scottish populations.” **A:** Part of the query page is shown, with items highlighted that are selected for the query. The various classes of unstable hemoglobins are all selected under “Relative stability: Any test,” and “Scottish” is chosen under “Occurrence: Ethnic background.” Other parts of the query page (not shown) have restricted the search to the beta-globin chain and substitutions. **B:** The initial output from the query. The names are hyperlinks to further information on each variant, in the format of Figure 1.

### FUTURE PROSPECTS

Plans for further development include adding new variants and thalassemias as they are discovered, improving the presentation of current information, and building links to other resources. Curators are currently adding new variants and thalassemias. Users who would like to have new variants entered should contact Dr. Wajcman ([wajcman@im3.inserm.fr](mailto:wajcman@im3.inserm.fr)) and those who would like to have new thalassemias entered should contact Dr. Chui ([chuid@fhs.cmu.edu](mailto:chuid@fhs.cmu.edu)), Dr. Patrinos ([patrinos@ch1.fgg.eur.nl](mailto:patrinos@ch1.fgg.eur.nl)), or Dr. Anagnou ([anagnou@imbb.forth.gr](mailto:anagnou@imbb.forth.gr)). These same individuals or Dr. Hardison ([\[psu.edu\]\(mailto:rch8@psu.edu\)\) should be contacted to resolve difficult cases. It would be appropriate for journals that publish new mutations to establish an agreement that publication would also entail recording the mutation in the \*HbVar\* database, so the journals and database would be connected and mutually enhance the value of the data. We see \*HbVar\* as the primary repository for information on hemoglobin variants and thalassemia mutations, building on the legacy of Prof. Huisman’s remarkably comprehensive \*Syllabi\*. The online versions of the \*Syllabi\* will continue to be offered, but no plans are in place for updating them in their current format. However, it would be possible](mailto:rch8@</a></p>
</div>
<div data-bbox=)

to use *HbVar* as the source to construct a book equivalent to a *Syllabus*. Appropriate journals may be interested in publishing summaries of information from the database as well.

Some classes of mutations are more difficult to record precisely in a database. For example, several beta-thalassemia mutations are very large deletions of at least 100 kb, removing much or all of *HBBC* and the upstream locus control region [e.g., Kioussis et al., 1983; Taramelli et al., 1986; Feingold and Forget, 1989]. Some large deletions that remove *HBD* and *HBB* result in increased expression of *HBG1* and *HBG2*, resulting in a form of hereditary persistence of fetal hemoglobin [e.g., Kosteas et al., 1997]. Ideally, these should be recorded as deletions between two specified nucleotides in a reference sequence. However, almost all of these were initially characterized by the genomic restriction fragments that were lost as a result of the deletion, and only some of them have been sequenced across the breakpoint junction. Furthermore, despite the recent publications of working draft human genome sequences [International Human Genome Sequencing Consortium, 2001; Venter et al., 2001] and an almost-finished sequence of about 385 kb encompassing *HBBC* in human [Bulger et al., 2000], the region between *HBB* and some of the large deletion breakpoints has not been filled. This region has been resistant to multiple efforts at sequencing [Bulger et al., 2000], and it appears to rearrange frequently in large genomic clones [Bepler et al., 1999; Imam et al., 2000]. Thus it may fall into the “unsequenceable” class of genomic DNA, and a continuous reference sequence may not be available for some time (if ever). Additional efforts are needed to sequence all breakpoints, and strategic decisions are needed to record such information in the potential absence of a complete reference sequence.

The hemoglobin mutation data are most useful when analyzed in concert with other information, such as experimental data on gene regulation and protein function, and sequence conservation in other species. Interspecies sequence alignments for *HBBC* have been available at the Globin Gene Server for several years [Hardison et al., 1994, 1998a, 2001]. A prototype database of experimental results on globin gene expression has been built [Riemer et al., 1998], and progress has been made in moving this to a new relational database with an updated schema to facilitate broader coverage of the literature on globin gene regulation. One of our future goals is to build user interfaces and programs that will allow integration of information among these resources.

A real-life example illustrates the need for such integration. We received a request by e-mail from a user in Denmark for help in finding mutations in the binding site for EKLF that are associated with beta-thalassemia. We instructed this user to fill in the query form of *HbVar* so as to specify the region including -92 through -86 in the *HBB* promoter, since we knew that was the binding site for EKLF (Table 3). Ideally, this information should be readily accessible at the same time one is searching the database. It currently is available from many sources, such as a figure in Huisman et al. [1997], or other databases such as TRRD [Kolchanov et al., 1999], or the literature [Feng et al., 1994]. However, more complete recording of information in databases and tight integration among them will facilitate access to these data by a wide community of users.

Information on human polymorphisms that do not cause a change in amino acid sequence or affect expression of the globin genes is not currently recorded in *HbVar* in an organized way, although some polymorphisms that are linked

TABLE 3. Matches to the Query “Find all Mutations Between Positions -92 and -86 of *HBB* That are Associated With beta+ Thalassemia”

Name	Mutation	Mutation, HUGO nomenclature
-92 (C->T) beta+ (mild)	beta nt -92 C>T	HBB g.-142C>T
-90 (C->T) beta+	beta nt -90 C>T	HBB g.-140C>T
-88 (C->A) beta+	beta nt -88 C>A	HBB g.-138C>A
-88 (C->T) beta+	beta nt -88 C>T	HBB g.-138C>T
-87 (C->A) beta+	beta nt -87 C>A	HBB g.-137C>A
-87 (C->T) beta+	beta nt -87 C>T	HBB g.-137C>T
-87 (C->G) beta+	beta nt -87 C>G	HBB g.-137C>G
-86 (C->A) beta+	beta nt -86 C>A	HBB g.-136C>A
-86 (C->G) beta+	beta nt -86 C>G	HBB g.-136C>G

to another hemoglobin abnormality are noted in free-text comments. However, considerable polymorphism information is already available for some regions [Fullerton et al., 1994; Harding et al., 1997], and much more is being collected in the human genome projects. Addition of this polymorphism information, organized so it could be queried upon, could be useful both for genetic mapping of novel or complex traits and for evaluation of the role of selection in conserved sequences [Hudson et al., 1987; McDonald and Kreitman, 1991].

### ACKNOWLEDGMENTS

This work was supported by grants from Public Health Service (HG02238 to W.M and DK27635 to R.C.H.).

### REFERENCES

- Bepko G, O'Brian KC, Kim Y-C, Schreiber G, Pitterle DM. 1999. A 1.4-Mb high-resolution physical map and contig of chromosome segment 11p15.5 and genes in the LOH11A metastasis suppressor region. *Genomics* 55: 164–175.
- Bulger M, Bender MA, von Doorninck JH, Wertman B, Farrell C, Felsenfeld G, Groudine M, Hardison R. 2000. Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse  $\beta$ -globin gene clusters. *Proc Natl Acad Sci USA* 14560–14565.
- Bunn HF, Forget BG. 1986. Hemoglobin: molecular, genetic and clinical aspects. Philadelphia: W.B. Saunders Co.
- Chui DHK, Hardison R, Riemer C, Miller W, Carver MFH, Molchanova TP, Efremov GD, Huisman THJ. 1998. An electronic database of human hemoglobin variants on the World Wide Web. *Blood* 91:2643–2644.
- Feingold EA, Forget BG. 1989. 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* 74:2178–2186.
- Feng WC, Southwood CM, Bieker JJ. 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* 269:1493–1500.
- Forget BG, Higgs DR, Steinberg M, Nagel RL. 2001. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press.
- Fullerton SM, Harding RM, Boyce AJ, Clegg JB. 1994. Molecular and population genetic analysis of allelic sequence diversity at the human beta-globin locus. *Proc Natl Acad Sci USA* 91:1805–1809.
- Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* 60:772–789.
- Hardison R, Chao K-M, Schwartz S, Stojanovic N, Ganetsky M, Miller W. 1994. Globin gene server: a prototype e-mail database server featuring extensive multiple alignments and data compilation. *Genomics* 21:344–353.
- Hardison R, Elnitski L, ElSherbini A, Goldstrohm A, Jackson J, Peck J, Riemer C, Schwartz S, Stojanovic N, Miller W. 1995. The Globin Gene Server: an aid to studying the regulation of mammalian globin gene clusters. In: Stamatoyanopoulos G, editor. *Molecular biology of hemoglobin switching*. Andover, Hampshire, UK: Intercept Ltd. p 405–426.
- Hardison R, Riemer C, Chui DHK, Huisman THJ, Miller W. 1998a. Electronic access to sequence alignments, experimental results, and human mutations as an aid to studying globin gene regulation. *Genomics* 47:429–437.
- Hardison RC, Chui DHK, Riemer C, Miller W, Carver MFH, Molchanova TP, Efremov GD, Huisman THJ. 1998b. Access to a syllabus of human hemoglobin variants (1996) via the World Wide Web. *Hemoglobin* 22:113–127.
- Hardison RC, Chui DHK, Riemer C, Giardine B, Lehväsliho H, Wajcman H, Miller W. 2001. Databases of human hemoglobin variants and other resources at the Globin Gene Server. *Hemoglobin* 25:183–193.
- Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
- Huisman THJ, Carver MFH, Efremov GD. 1996. A syllabus of human hemoglobin variants (1996). Augusta, GA: The Sickle Cell Anemia Foundation.
- Huisman THJ, Carver MFH, Baysal E. 1997. A syllabus of thalassemia mutations (1997). Augusta, GA: The Sickle Cell Anemia Foundation.
- Huisman THJ, Carver MFH, Efremov GD. 1998. A syllabus of human hemoglobin variants. 2nd edition. Augusta, GA: The Sickle Cell Anemia Foundation.
- Imam AM, Patrinos GP, de Krom M, Bottardi S, Janssens RJ, Katsantoni E, Wai AW, Sherratt DJ, Grosveld FG. 2000. Modification of human beta-globin locus PAC clones by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 28:E65.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Kioussis D, Vanin E, deLange T, Flavell R, Grosveld F. 1983.  $\beta$ -globin gene inactivation by DNA translocation in  $\gamma\beta$ -thalassemia. *Nature* 38:662–664.
- Kolchanov NA, Ananko EA, Podkolodnaya OA, Ignatieva EV, Stepanenko IL, Kel-Margoulis OV, Kel AE, Merkulova TI, Goryachkovskaya TN, Busygina TV, Kolpakov FA, Podkolodny NL, Naumochkin AN, Romashchenko AG. 1999. Transcription regulatory regions database (TRRD): its status in 1999. *Nucleic Acids Res* 27:303–306.
- Kosteas T, Palena A, Anagnou NP. 1997. Molecular cloning of the breakpoints of the hereditary persistence of fetal

- hemoglobin type-6 (HPFH-6) deletion and sequence analysis of the novel juxtaposed region from the 3' end of the beta-globin gene cluster. *Hum Genet* 100:441–445.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 354:114–116.
- Préhu C, Godart C, Vigneron C, Wajcman H. 1998. Hb Nancy and Hb Osler: two distinct genetic variants with identical clinical and hemoglobin phenotype. *Comptes Rendus Acad Sci Paris, Sciences de la vie* 321:373–376.
- Rai DK, Griffiths WJ, Alvelius G, Landin B. 2001. Electrospray mass spectrometry: an efficient method to detect silent hemoglobin variants causing erythrocytosis. *Clin Chem* 47:1308–1311.
- Riemer C, ElSherbini A, Stojanovic N, Schwartz S, Kwitkin PB, Miller W, Hardison R. 1998. A database of experimental results on globin gene expression. *Genomics* 53:324–337.
- Taramelli R, Kioussis D, Vanin E. 1986. Gamma delta beta-thalassemias 1 and 2 are the result of a 100k bp deletion in the human beta-globin cluster. *Nucl Acids Res* 14:7017–7029.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C. 2001. The sequence of the human genome. *Science* 291:1304–1351.
- Wajcman H, Préhu C, Bardakdjian-Michau J, Promé D, Riou J, Godart C, Mathis M, Hurtrel D, Galactéros F. 2001. Abnormal hemoglobins: the laboratory methods. *Hemoglobin* 25:169–181.