Correlating Disease Genes and Phenotypes
An NCBI Mini-Course

This mini-course focuses on the correlation of a disease gene to the phenotype. It demonstrates how NCBI resources such as the literature, expression and structure information can provide potential functional information for disease genes.

Mutations in the HFE gene are associated with the hemochromatosis disease. A laboratory working on the hemochromatosis disease wants to elucidate the biochemical and structural basis for the function of the mutant protein.

Outline:

In this exercise, we have the following goals:
1. Determine what is known about the HFE gene and protein (using Entrez Gene).
2. Determine identified SNPs and their locations in the HFE gene (using dbSNP).
3. Learn more about hemochromatosis and its genetic testing (using OMIM and Gene Tests)
4. Elucidate the biochemical and structural basis for the function of the wild type and mutant proteins, if possible.

During the first hour, an overview will be given using one disease gene, followed by an hour of hands-on session to practice using another disease gene. The following handout contains the screenshots of the overview.


Course Developed by Medha Bhagwat ([bhagwat@ncbi.nlm.nih.gov](mailto:bhagwat@ncbi.nlm.nih.gov))
Problem 1

Mutations in the HFE gene are associated with the hemochromatosis disease. A laboratory working on the hemochromatosis disease wants to elucidate the biochemical and structural basis for the function of the mutant protein.

Outline:

In this exercise, we have the following goals:
1. Determining what is known about the HFE gene and protein (using Entrez Gene).
2. Determining identified SNPs and their locations in the HFE gene (using dbSNP).
3. Learning more about the hemochromatosis disease and its genetic testing (using OMIM and Gene Tests)
4. Elucidating the biochemical and structural basis for the function of the wild type and the mutant protein, if possible (using CDD).

Step 1. Determining what is known about the HFE gene and protein (using Entrez Gene):

Search for 'HFE" in Entrez Gene. One entry is for the human HFE gene. Retrieve the entry by clicking on the HFE link.

What is the location and orientation of the HFE gene on the human genome? List the genes adjacent to it. How many alternatively spliced products have been annotated for the HFE gene when the RefSeq mRNA entries were reviewed? What are the differences in the spliced products? List some of the HFE gene aliases. What are the phenotypes associated with the mutations in the HFE gene? What is the conserved domain in the protein? To which cellular component(s) is the protein localized? Obtain the locations of exons and introns for each transcript by choosing "Gene Table" from the Display pull down menu.

Step 2. Determining identified SNPs and their locations in the HFE gene:

From the Links menu on the top right hand side of the page, click on the "SNP: GeneView" to access a list of the known SNPs (reported in dbSNP). By default, the SNPs in the coding region of a gene are reported. Additional SNPs such as in the upstream region or the introns can be viewed by clicking on the "in gene region" button. Currently, how many non-synonymous SNPs are placed on the longest hemochromatosis transcript variant, NM_000410? How many of these have links to OMIM? We will concentrate on the cys282tyr mutant in the following analysis.
**Step 3. Learning more about the hemochromatosis disease and its genetic testing:**

Click on the OMIM link next to the one of the SNPs in the SNP report. What are the clinical features of hemochromatosis? List the 5 types of iron-overload disorders labeled hemochromatosis. Which of these is associated with mutations in the HFE gene? How many allelic variants of the HFE gene have been reported? What is the phenotype associated with the Cys282Tyr mutant?

Click on the Gene Tests link at top of the page. Identify some of the laboratories performing the clinical testing for hemochromatosis. Now refer to the Reviews section. Mutation analysis is available for which of the HFE alleles? List one explanation for the hemochromatosis phenotype caused by the Cys282Tyr mutant.

**Step 4. Elucidating the biochemical and structural basis for the function of the wild type and mutant proteins, if possible:**

Go back to the Entrez Gene report. Click on the protein accession number NP_000401 associated with the longest splice variant NM_000410. Select the Blink link. Click on the 3D structures button. The output contains a list of similar proteins with known 3D structures. The first entry, 1A6Z chain C, provides the structure of the part of human hemochromatosis protein. Click on the blue dot next to the accession number to get the sequence alignment of the query protein with 1A6Z chain C. Click on the "View 3D Structure" button. This downloads its 3D structure and its sequence alignment with the query protein. Zoom in the area of the disulphide bridge (colored in tan) by pressing "z" on the keyboard. Select the cysteine residues forming the disulphide bridge by double clicking on them. Mouse over the corresponding cysteine residues on the query line in the alignment and view the amino acid number at the bottom left of the window. One of them is the cysteine at position 282. It is the same cysteine which is mutated to tyrosine causing the hemochromatosis phenotype.

**You can now easily explain why the C282Y mutant has an altered function.**

**Summary:**

This mini-course describes how to obtain information about the HFE gene, known SNPs in it, and elucidate the biochemical and structural basis for the function of the wild type and Cys282Tyr mutant protein.

Summary: 1. The HFE gene is located on chromosome 6 and has at least 11 alternatively spliced products. 2. Currently, there are 8 non-synonymous SNPs annotated on the protein NP_000401. 3. The Cys282Tyr mutant is associated with the hemochromatosis disease and the site of mutation is used in hemochromatosis genetic testing.
4. The HFE protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin whereas the Cys282Tyr mutant fails to regulate this interaction leading to iron overload. The conserved cysteine 282 in the immunoglobulin constant region domain of the HFE protein is involved in formation of a disulphide bridge. Its mutation to tyrosine will alter the folding of the protein.
Transferrin receptor

hemochromatosis
1. In hemodialysis patients, hyperferritinemia reflects a relative increase in iron availability and a decrease in non-specific antioxidant activity, is favored by HFE mutations, and represents a risk factor for advanced cardiovascular damage.

2. Our data suggest that the HFE gene is not a major disease gene for migraine.

3. Iron overload has been associated with HFE mutations (C282Y and H63D).

4. The HFE IVS5+1 G/A splice site mutation is not the major explanation for unexpectedly high prevalence of high transferrin saturation and high serum ferritin in North American Asians.

5. H63D mutations in HFE play a role in the pathogenesis of ALS in various populations. This association might involve a later-onset subset of ALS.

6. The effect of particulate air pollution on cardiac autonomic function was shielded in subjects with at least 1 copy of an HFE variant compared with wild-type subjects.

7. This study confirms the high frequency of C282Y mutation in patients with potheria cutanea tarda and its relationship with iron overload.

8. Demonstrate a correlation between the presence of the H63D mutation and the occurrence of...
### General Protein Information

**Names**
- Hemochromatosis protein
- HFE (Hemochromatosis)
- Hereditary hemochromatosis protein HLA-H

### NCBI Reference Sequences (RefSeq)

#### Genomic

1. **NM_000419.3**
   - **Reference**: NM_000419.3
   - **Range**: 7163 - 8077
   - **GenBank**: TACX

#### mRNA and Protein(s)

1. **NM_000419.3**
   - **Transcript**: NM_000419.3
   - **Name**: Hemochromatosis protein isoform 1 precursor
   - **Description**: Transcript variant: This variant (1) encodes the longest isoform.
   - **Source sequence(s)**: GenBank: NM_000419.3
     - **NM_000419**
       - **Conserved Domain(s)**: CUB domain
       - **Gene**: HFE
       - **Organism**: Homo sapiens
       - **Location**: 7163 - 8077
       - **NCBI**: 5163
       - **RefSeq**: TACX

2. **NM_000419.3**
   - **Transcript**: NM_000419.3
   - **Name**: Hemochromatosis protein isoform 2 precursor
   - **Description**: Transcript variant: This variant (2) lacks a large 3' region including the 3' UTR and 5' UTR but has an alternate 3' exons, compared to variant 1. The resulting protein (isoform 2) has a unique carboxy terminus.

### Related Sequences

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic</td>
<td>Protein</td>
</tr>
<tr>
<td>AF185934.1</td>
<td>AAF5122.1</td>
</tr>
<tr>
<td>AF204969.1</td>
<td>None</td>
</tr>
<tr>
<td>AF331065.1</td>
<td>AAK1501.1</td>
</tr>
<tr>
<td>AF523559.1</td>
<td>AAM9000.1</td>
</tr>
<tr>
<td>AF523599.1</td>
<td>AAM9100.1</td>
</tr>
<tr>
<td>CH471087.1</td>
<td>EWSS516.1</td>
</tr>
<tr>
<td>EWSS517.1</td>
<td>EWSS518.1</td>
</tr>
<tr>
<td>EWSS519.1</td>
<td>EWSS520.1</td>
</tr>
<tr>
<td>EWSS521.1</td>
<td>EWSS522.1</td>
</tr>
<tr>
<td>EWSS523.1</td>
<td>EWSS524.1</td>
</tr>
<tr>
<td>EWSS525.1</td>
<td>EWSS526.1</td>
</tr>
<tr>
<td>EWSS527.1</td>
<td>EWSS528.1</td>
</tr>
<tr>
<td>CS182780.1</td>
<td>CA12968.2.1</td>
</tr>
<tr>
<td>CS18283.1</td>
<td>AAS0440.1</td>
</tr>
<tr>
<td>CS18284.1</td>
<td>AAS0883.1</td>
</tr>
<tr>
<td>CS18285.1</td>
<td>CA70443.1</td>
</tr>
<tr>
<td>Z92910.1</td>
<td>CAB07742.1</td>
</tr>
<tr>
<td>mRNAs</td>
<td></td>
</tr>
<tr>
<td>AF079587.1</td>
<td>AAC62746.1</td>
</tr>
<tr>
<td>AF079608.1</td>
<td>AAC62747.1</td>
</tr>
<tr>
<td>AF079629.1</td>
<td>AAC62748.1</td>
</tr>
<tr>
<td>AF090385.1</td>
<td>AAS52104.1</td>
</tr>
<tr>
<td>mRNAs</td>
<td></td>
</tr>
<tr>
<td>AF115264.1</td>
<td>AAG29571.1</td>
</tr>
<tr>
<td>AF115265.1</td>
<td>AAG29572.1</td>
</tr>
<tr>
<td>mRNAs</td>
<td></td>
</tr>
<tr>
<td>AF143239.1</td>
<td>AAG29573.1</td>
</tr>
<tr>
<td>mRNAs</td>
<td></td>
</tr>
<tr>
<td>AF143240.1</td>
<td>AAG29574.1</td>
</tr>
<tr>
<td>mRNAs</td>
<td></td>
</tr>
<tr>
<td>AF143240.1</td>
<td>AAG29575.1</td>
</tr>
</tbody>
</table>
The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (B2M). It is thought that the protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin. The iron storage disorder, hereditary hemochromatosis, is a recessive genetic disorder that results from defects in this gene. At least eleven alternatively spliced variants have been described for this gene. Additional variants have been found but their full-length nature has not been determined.
<table>
<thead>
<tr>
<th>Region</th>
<th>Contig position</th>
<th>mRNA position</th>
<th>dbsNP rs#</th>
<th>Cluster id</th>
<th>Heterozygosity</th>
<th>Validation</th>
<th>OMIM</th>
<th>Function</th>
<th>dbsNP allele</th>
<th>Protein residue</th>
<th>Codon pos</th>
<th>Amino acid pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon_1</td>
<td>16945920</td>
<td>161</td>
<td>G-</td>
<td>rs12429855</td>
<td>Yes</td>
<td>start codon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>exon_2</td>
<td>16949347</td>
<td>264</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>T</td>
<td>nonsynonymous</td>
<td>C</td>
<td>Thr</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>16949430</td>
<td>547</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>T</td>
<td>nonsynonymous</td>
<td>A</td>
<td>Asp</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>16949436</td>
<td>553</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>C</td>
<td>nonsynonymous</td>
<td>G</td>
<td>His</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>16949523</td>
<td>437</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>A</td>
<td>nonsynonymous</td>
<td>C</td>
<td>Thr</td>
<td>2</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>16950555</td>
<td>474</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>T</td>
<td>nonsynonymous</td>
<td>T</td>
<td>Thr</td>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>exon_3</td>
<td>16949833</td>
<td>541</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>A</td>
<td>nonsynonymous</td>
<td>C</td>
<td>His</td>
<td>3</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>16951197</td>
<td>810</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>A</td>
<td>nonsynonymous</td>
<td>T</td>
<td>Thr</td>
<td>2</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>16951392</td>
<td>1005</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>C</td>
<td>nonsynonymous</td>
<td>A</td>
<td>Tyr</td>
<td>2</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>16952063</td>
<td>1186</td>
<td>N.D.</td>
<td></td>
<td>Yes</td>
<td>contig reference</td>
<td>C</td>
<td>nonsynonymous</td>
<td>T</td>
<td>Tyr</td>
<td>3</td>
<td>342</td>
</tr>
</tbody>
</table>
In patients with hemochromatosis, Falk et al. (1996) identified an HFE-A mutation in the HFE gene (which they referred to as HLA-A or HLA-3, M4), resulting in a cysteine-to-proline (C282Y) substitution. This residue mutation occurs in a highly conserved residue involved in the intracellular trafficking and function of MHC class II proteins, and could therefore disrupt the structural and functional properties of this protein. Using allele-specific oligonucleotide (ASO) analysis on this group of patients, they detected the C282Y mutation in 60% of HFE hemochromatosis. In contrast, only 19 of the 38 control chromosomes (4.3%) carried the mutation, a carrier frequency of 0.015 ± 0.04. One hundred thirty-eight of 184 HFE patients were homozygous for the mutation, 9 were heterozygous, and 21 were normal. These numbers were essentially unchanged from Hardy-Weinberg equilibrium. The findings confirmed heterozygosity among the hemochromatosis patients, with 89% of those tested in C282Y homozygosity.

Joung et al. (1996) provided convincing evidence that the C282Y mutation in hemochromatosis in the HFE gene is the cause of hemochromatosis. In studies in Australia, patients were classified at the genotype and phenotype level. All showed homozygosity for the C282Y allele and substitution. The expression of the C282Y allele in all patients correlated with the disease, and the increased risk of liver disease was present in all populations and ranging at least 4.5-fold, with a probability of 0.0005.

Joung et al. (1996) commented on the significance of the C282Y mutation in a study of a group of 1,477 individuals who had been under study in France for 40 years and identified by stringent criteria. Homozygosity for the C282Y mutation was found in 36 of 1,477 individuals (2.4%) of the patients were compound heterozygotes for the C282Y mutation and the H63D mutation (C282Y/D63H); 1 was homozygous for the H63D mutation, and 2 were heterozygous for the mutation. These results corresponded to an allele frequency of 2.4% for the C282Y allele and 2.8% for the H63D allele. In the study of 1,477 individuals, the C282Y mutation was well observed in the Hardy-Weinberg expected frequencies. The Hardy-Weinberg model was also used to control groups. The C282Y mutation was not observed in the HFE genotype, and the Hardy-Weinberg model was not observed. The C282Y mutation was not observed in the HFE genotype, and the Hardy-Weinberg model was not observed.

Joung et al. (1996) reported on the incidence of the C282Y mutation in 147 patients with hereditary hemochromatosis and 156 controls. Of the 147 patients, 137 (93.5%) were homozygous for the C282Y mutation, while 10 (6.5%) were heterozygous. Of the 156 controls, 146 (93.8%) were homozygous for the C282Y mutation, while 10 (6.2%) were heterozygous. Overall, the C282Y mutation was also found in 94% of the patients for the wild-type allele (92.2% for the HFE allele and 97.8% for the HFE allele).

**OMIM** (Online Mendelian Inheritance in Man) shows additional information on hemochromatosis. The OMIM number for hemochromatosis is 235200. The database includes the following information:

- **Allelic Variants**
  - **Hemochromatosis**
    - **Variant ID**: HFE, C282Y
    - **Variant Description**: Porphyria varietae, included
    - **Alleles**: Heme, Deferoxamine, included
    - **Contributors**: Hemo, Deferoxamine, included
    - **Gene Symbol**: HFE
    - **Uniprot**: P04703
    - **GenBank**: NM_000461
    - **NCBI**: NM_000461
    - **ORCID**: 0000-0001-7777-7777
    - **Gene Ontology**: C0000057, C0000060, C0000061
    - ** OMIM**: 235200

This information is based on the literature and databases available at the time of the study.
The result of your search (below) includes a group of related disorders with your search term in bold or an alphabetical listing of the individual entries that match your search term. For more information about search results, see Interpreting Your Search Results.

Search Result for OMIM# 235200
HFE-Associated Hereditary Hemochromatosis

Select all clinical laboratories:

[Table with information on laboratories offering clinical testing]
HFE-Associated Hereditary Hemochromatosis

Authors: Kris V. Knobley, MD  Jonathan F. Fab, MD, PhD  Robin L. Bennett, MS  Arne G. Mollison, MD

About the Authors
Initial Posting: 3 April 2006  Last Update: 4 December 2006

Summary

Disease characteristics: HFE-associated hereditary hemochromatosis (HFE-HHC) is characterized by inappropriate high absorption of iron by the gastrointestinal mucosa, resulting in excessive storage of iron particularly in the liver, skin, pancreas, heart, joints, and testes. Abdominal pain, weakness, lethargy, and weight loss are early symptoms. Without therapy, males may develop symptoms between age 40 and 60 years and females after menopause. Hepatic fibrosis or cirrhosis may occur in untreated individuals after age 40 years. Other findings in untreated individuals may include progressive increase in skin pigmentation, diabetes mellitus, congestive heart failure and/or arrhythmias, arthritis, and hypertension.

This disease is characterized by hepatic iron overload. About 1/10 of patients develop cirrhosis and about 1/20 develop hepatocellular carcinoma. HFE-HHC is a progressive disease with a median disease-free survival age of 65 years. The incidence of HFE-HHC is about 1/100 in the general population and 1/1000 in blacks.

HFE-HHC can be transmitted to all offspring. In males, the disease is transmitted to the female offspring and the male offspring are carriers. In females, the disease is transmitted to the male offspring and the female offspring are carriers.

Diagnosis/Testing: The diagnosis of HFE-HHC is based on clinical symptoms consistent with HFE-HHC and/or biochemical data. HFE-HHC can be diagnosed by measuring the level of serum ferritin and/or transferrin saturation.

Resources

Pathologic allelic variants: At least 20 distinct mutations have been reported, most being missense or nonsense mutations. Missense mutations account for the vast majority of disease-causing alleles in the population.

- Cys282Tyr (p.C282Y; nucleotide 843T>G): This missense mutation removes a highly conserved cysteine residue that normally forms an intramolecular disulfide bond with beta-2 microglobulin, and thereby prevents the protein from being transported to the cell surface.
- H63D (p.H63D; nucleotide 187C>T): This missense mutation may alter a pH-dependent intramolecular salt bridge, possibly affecting interaction of the HFE protein with the transferrin receptor.

Normal gene product: The largest predicted primary translation product is 540 amino acids, which gives rise to a mature protein of about 221 amino acids after cleavage of the signal sequence. The HFE protein is similar to HLA Class I molecules at the primary [Fidler et al 1996] and tertiary structural [Leonard et al 1998] levels. The mature protein is expressed on the cell surface as a homodimer with beta-2 microglobulin, and this interaction is necessary for normal presentation on the cell surface. The normal protein binds to transferrin receptor-1 on the cell surface and may reduce cellular iron uptake; however, the exact means by which the HFE protein regulates iron uptake is not clear at present.

Abnormal gene product: The p.C282Y mutation destroys a key cysteine residue that is required for disulfide bonding with beta-2 microglobulin. As a result, the HFE protein does not mature properly and becomes trapped in the endoplasmic reticulum and Golgi apparatus, leading to decreased cell-surface expression. The mechanism for the phenotypic effect of other HFE mutations is not clear at present.
The hemochromatosis protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin.
Related Structures

Query: hemochromatosis protein isofrom 1 precursor (Homo sapiens)
[gi: 4504377]
Structure: 1A82 Chain C, Hfe (Human) Hemochromatosis Protein
Reference: [NMDB] [PubMed]

<table>
<thead>
<tr>
<th>E-value</th>
<th>7e-168</th>
<th>Bit score</th>
<th>588</th>
<th>Aligned length</th>
<th>275</th>
<th>Sequence Identity</th>
<th>100%</th>
</tr>
</thead>
</table>

```
| gi:6904377 | 20 | HIKSmLHmKLHAEASQGEGMFFLAGVYSGQYVYRDIRYERFPEWYRSQIISSQWLQGLQLSLRQENFTVDF | 102 |
| 1A82_C     | 1 | HIKSmLHmKLHAEASQGEGMFFLAGVYSGQYVYRDIRYERFPEWYRSQIISSQWLQGLQLSLRQENFTVDF | 60 |

| gi:4504377 | 103 | W1MEM9K5552H7QV1LC3TMQ1U652W0QPYVQDQGMLFLFCFTDQLDRAAAWRQYVTQLEHERW1IKIDRQHAY | 160 |
| 1A82_C     | 81 | W1MEM9K5552H7QV1LC3TMQ1U652W0QPYVQDQGMLFLFCFTDQLDRAAAWRQYVTQLEHERW1IKIDRQHAY | 160 |

| gi:4504377 | 183 | L1S0CFAQ6QGQLCGRGVEDQGQVPHYKQVKTQVASSVTLVCCSAHVFPQHNTNLMRLKQGDCREQFECFVLDQGQDG | 262 |
| 1A82_C     | 161 | L1S0CFAQ6QGQLCGRGVEDQGQVPHYKQVKTQVASSVTLVCCSAHVFPQHNTNLMRLKQGDCREQFECFVLDQGQDG | 260 |

| gi:4504377 | 263 | YQW6I16LAVF0EEQRYTQOHG0LDQMLVYIN | 257 |
| 1A82_C     | 241 | YQW6I16LAVF0EEQRYTQOHG0LDQMLVYIN | 276 |
```
Problem 2:

Mutations in the HBB gene are associated with sickle cell anemia. A laboratory working on sickle cell anemia wants to elucidate the biochemical and structural basis for the function of the mutant HBB protein.

**Step 1. Determining what is known about the HBB gene and protein (using Entrez Gene):**

Search for 'HBB' in Entrez Gene. One entry is for the human HBB gene. Retrieve the entry by clicking on the HBB link.

What is the location and orientation of the HBB gene on the human genome? List the genes adjacent to it. How many alternatively spliced products have been annotated for the HBB gene when the RefSeq mRNA entries were reviewed? List some of the HBB gene aliases. What are the phenotypes associated with the mutations in the HBB gene? Where are the mouse and rat HBB genes located?

What is the name and function of the protein encoded by the HBB gene? What is the conserved domain in the protein? To which cellular component(s) is the protein localized? Beta hemoglobin is a subunit of which protein? Name other subunit(s) in that protein.

Obtain the locations of exons and introns for each transcript by choosing "Gene Table" from the Display pull down menu. Go back to the description page.

**Step 2. Determining other identified SNPs and their locations in the HBB gene:**

From the Links menu on the top right hand side of the page, click on the "SNP: GeneView" to access a list of the known SNPs (reported in dbSNP). By default, the SNPs in the coding region of a gene are reported. Additional SNPs such as in the upstream region or the introns can be viewed by clicking on the "in gene region" button. Currently, how many coding SNPs are placed on the beta hemoglobin transcript NM_000518? How many of these have links to OMIM? We will concentrate on the Glu7Val mutant in the following analysis.

**Step 3. Learning more about sickle cell anemia disease and its genetic testing:**

Go back to the Entrez Gene report. Click on the OMIM link and then HBB link. What are the phenotypes caused by mutations in HBB, the absence of HBB and reduced amounts of HBB? What is the clinical synopsis of sickle cell anemia? What is its prominent feature? What is its mode of inheritance? How many allelic variants of the HBB gene have been reported? As mentioned in the OMIM report, the allelic variants are listed for the mature beta hemoglobin protein which lacks
an initiator methionine. Hence, the allelic variants in the OMIM report are off by one amino acid compared to the precursor protein in NP_000509. Click on the Allelic Variant “View list” to get information about the mutant proteins from patients. Is the Glu6Val variant mentioned in the list? (It is the variant number 0243). Which phenotype does it cause? What is the name of the mutant hemoglobin (hemoglobin S).

Click on the Gene Tests link at top of the page. Identify some of the laboratories performing the clinical testing for sickle cell anemia. Now refer to the Reviews section for Sickle Cell Disease, Mutation analysis is available for which of the HBB alleles? List one explanation for the sickle cell anemia phenotype caused by the Glu7Val mutant beta hemoglobin.

**Step 4. Elucidating the biochemical and structural basis for the function of the wild type and mutant proteins, if possible:**

**A. Information about the wild type protein**

Go back to the OMIM report by clicking the back button on the web browser. Go to the Gene report through the Links menu. Based on the RefSeq summary and the PubMed articles, describe the biochemical functions of beta hemoglobin and hemoglobin S. PubMed articles in the Entrez Gene report indicate that the 3-D structure of hemoglobin S is available.

Let us first take a look at the structure of the wild type protein. Click on the NP_000509 protein link and select Blink. Click on the “Show identical” button and then on the “3D structures” button. The output contains a list of similar proteins with 3D structures known. The entry, 1DXTD, represents the structure of deoxyhemoglobin chain D. Click on the blue dot next to 1DXTD to get the sequence alignment of the query protein to the D chain of 1DXTD. To view the 3D structure of deoxyhemoglobin (all chains, 2 alpha and 2 beta), click on the MMDB link. That takes us to the MMDB structure summary page for 1DXT. Access the PDB entry, by clicking on 1DXT. Note that the chains A and C in the structure represent alpha chains, and B and D represent beta chains. Go back to the MMDB summary page. View the deoxyhemoglobin tetramer by clicking on the "View 3D Structure button".

Search for the structure of the mutant (deoxyhemoglobin S) in the structure database. Two entries, 1HBS and 2HBS, are retrieved. Click on the 2HBS link. Then click on the PubMed link from the MMDB and PDB entries (under Reference). The abstracts indicate that the mutated valine residue of the beta chain contacts with another hemoglobin tetramer molecule to form hemoglobin polymers which are building blocks for the sickle cell fiber.
B. To show the side chains of the mutant residue and view its interaction with another hemoglobin molecule: Download the structure 2HBS by clicking on View 3D Structure. For easier viewing, remove the helix and strand objects using Style--Edit global style, and unclick the boxes next to the Helix objects and Strand objects. Highlight valine 6 from the H chain (one of the beta chains). To show the side chains of the residue, use the Structure window--Style--Annotate--new. Give a name to this annotation such as "valine" and then click on Edit Style. Change the protein backbone "Rendering" to "Space Fill", Color Scheme to "charge" or "hydrophobicity". Repeat these steps for the Protein Sidechains row and click the Protein Sidechains on. To show the amino acid number, choose the Labels panel, and change the Protein Backbone spacing to 1. Click on the “Done”, “OK” then “Done” buttons. The valine interacts with a pocket between the two helices on another tetramer. Identify the residues from other molecules within 4 angstroms of the valine, use Show/Hide--Select by distance--other molecules. To unselect the highlighted residues, click on the white portion of the sequence window.

You can now easily explain why the Glu7Val mutant has an altered function.

Summary:

This mini-course describes how to obtain information about the HBB gene, known SNPs in it, and elucidate the biochemical and structural basis for the function of the wild type and Glu7Val mutant protein.

Summary: 1. The HBB gene is located on chromosome 11 and has no alternatively spliced products annotated. 2. Currently, there are 301 coding SNPs annotated on the protein NP_000509. 3. The Glu7Val mutant is associated with the sickle cell anemia disease and the site of mutation is used in sickle cell anemia genetic testing. 4. The HBB gene encodes beta hemoglobin which is a part of hemoglobin along with alpha hemoglobin. Hemoglobin is a tetramer consisting of 2 beta and 2 alpha chains. Mutation of the 7th negatively charged amino acid, glutamic acid, to hydrophobic valine leads to polymerization of hemoglobin forming a sickle fiber that changes the shape of red blood cells leading to sickle cell anemia.