Correlating Disease Genes to Phenotypes
An NCBI Mini-Course

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


Instructor:
Dr. Medha Bhagwat, NCBI
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 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 name and function of the protein encoded by 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 "Geneview in dbSNP" 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 first protein, NP_000401. 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, 1DE4G, represents the G chain of the hemochromatosis protein (complexed with transferrin receptor). Click on the blue dot next to 1DE4G to get the sequence alignment of the query protein to the G chain of 1DE4. Click on the "View 3D Structure" button. This downloads the structure of G chain of 1DE4 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 third 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.

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 cysteinie 282 in the immunoglobulin constant region domain in the HFE protein is involved in formation of a disulphide bridge. Its mutation to tyrosine will alter the folding of the protein.
1. **HFE hemochromatosis**

   **GeneID:** 3977  **Location:** 6p21.3

**Summary**

Official Symbol: **HFE** and Name: **hemochromatosis**

Gene type: protein-coding

Gene name: **HFE**

**Gene description:** hemochromatosis

**RefSeq status:** Reviewed

**Organism:** *Homo sapiens*

**Lineage:** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Scarnotheca; Primates; Catarrhini; Hominidae; Homo

**Gene aliases:** HFE1, HLA-H, M3C183790, A221C16.18.1

**Summary:** The protein encoded by this gene is a hemochromatosis protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (B2M). It is thought that this 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.

### Genomic context

- **Chromosome:** 6  **Location:** 6p21.3

### Bibliography

**PubMed links**

**GeneRef**

1. **HFE C282Y mutation significantly increases the risk of venous leg ulceration in primary cardiovascular diseases by almost 7 times.**

2. **multiple alleles in patients carrying the mutant C282Y allele exhibited earlier onset of disease symptom relative to other genotypes, but was warrants further study in a large cohort of MS patients.**

3. **HFE gene mutations considered in patients with chronic viral hepatitis in taiwan.**

4. **Additional risk of hereditary hemochromatosis given by class I HLA antigens may be secondary to the HFE gene linkage disequilibrium with certain class I alleles or to the existence of other non-hemochromatosis factors in our population.**
The hemochromatosis protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin.
### General gene information

**Gene Ontology**

<table>
<thead>
<tr>
<th>Function</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC class I receptor activity</td>
<td>IEA</td>
</tr>
<tr>
<td>transport</td>
<td>IEA</td>
</tr>
<tr>
<td>antigen presentation, endogenous antigen transport</td>
<td>IEA</td>
</tr>
<tr>
<td>antigen processing, endogenous antigen via MHC class I transport complex assembly</td>
<td>IEA</td>
</tr>
<tr>
<td>transport</td>
<td>TAS</td>
</tr>
<tr>
<td>transport complex assembly</td>
<td>TAS</td>
</tr>
<tr>
<td>receptor-mediated endocytosis</td>
<td>TAS</td>
</tr>
<tr>
<td>transport</td>
<td>ADS</td>
</tr>
</tbody>
</table>

**Component**

- MHC class I protein complex
- cytoplasm
- general to plasma membrane
- plasma membrane

**Homology**

- Mouse, Rat
  - Mac Viewer

**Phenotypes**

- Hemochromatosis DMH 235C0D
- Porphyria variegata DMH 13604C

**Markers (Sequence Tagged Sites/STS)**

- STS-U60319 (c-PCR)
- Alternate name BM75999
- Alternate name su-U60319
- FMC195311 (c-PCR)
- FMC19649.1E2 (c-PCR)
- Alternate name FMC23476.P1
- MRB 3982 (c-PCR)

### General protein information

**Names**: hemochromatosis protein MHC class I-like protein HFE, hereditary hemochromatosis protein HLA-II

### NCBI Reference Sequences (RefSeq)

- Reference NG_001325
- mRNA Sequence NM_004410

  **Transcriptional Variant**
  - Transcript Variant: This variant (1) encodes the longest isoform.
  - Source Sequence W82939
  - Product NE_000401 hemochromatosis protein isoform 1 precursor

  **Conserved Domains**: (2) summary
  - cl00402 MHC 1 Class I Heterocytotoxic antigen, domains alpha 1 and 2
  - cl00098 FC; Immunoglobulin domain constant region subfamily
    - Location: 223 - 291, Blast Score: 169

### Related Sequences

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genbank</td>
<td>Genbank</td>
</tr>
<tr>
<td>AF181284</td>
<td>AAF04222</td>
</tr>
<tr>
<td>AF331965</td>
<td>AAE16502</td>
</tr>
<tr>
<td>AF55359</td>
<td>AAM62068</td>
</tr>
<tr>
<td>AF524999</td>
<td>AAM11950</td>
</tr>
<tr>
<td>NM_002514</td>
<td>AAF04244</td>
</tr>
<tr>
<td>J93136</td>
<td>AAB22813</td>
</tr>
<tr>
<td>Y99581</td>
<td>CAQ0934</td>
</tr>
<tr>
<td>Z92910</td>
<td>CAB37492</td>
</tr>
<tr>
<td>mRNA</td>
<td>mRNA</td>
</tr>
<tr>
<td>AF074807</td>
<td>AAC26464</td>
</tr>
<tr>
<td>AF074808</td>
<td>AAC26454</td>
</tr>
<tr>
<td>AF074809</td>
<td>AAC26449</td>
</tr>
</tbody>
</table>
**Gene Information**

**Symbol**: Homo sapiens

**Lineage**: Eukaryota; Animalia; Chordata; Vertebrata; Euteleostomi; Mammalia; Primates; Homo sapiens

**Gene Name**: LS1, KIF3, KIF4A, MGC18790, G25C10.10.1

**Summary**: The protein encoded by this gene is a conserved protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (B2M). It is thought that this 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.

**Exon Information**

- **mRNA** length: 2717 bp, number of exons: 7
- **Protein** length: 349 aa, number of exons: 6

**Exon Coords**

- **NM_000410**: 297 bp - 297, length: 222 - 297, 76 bp
- **NM_139004**: 264 bp - 3685, length: 3622 - 3685, 264 bp
- **NM_139009**: 276 bp - 4370, length: 4095 - 4370, 276 bp
- **NM_139007**: 276 bp - 5741, length: 5466 - 5741, 276 bp
- **NM_139010**: 114 bp - 6013, length: 5900 - 6013, 114 bp
- **NM_139011**: 1056 bp - 7007, length: 6967 - 7007, 41 bp
- **NM_139002**: 434 bp - 9610, length: 9177 - 9610, 434 bp

**Coding Exon Coords**

- **NM_000410**: 2717 bp - 2776, length: 222 - 2776, 66 bp
- **NM_139004**: 264 bp - 3634, length: 3622 - 3634, 22 bp
- **NM_139009**: 276 bp - 4367, length: 4095 - 4367, 276 bp
- **NM_139007**: 276 bp - 5731, length: 5466 - 5731, 276 bp
- **NM_139010**: 114 bp - 6002, length: 5900 - 6002, 114 bp
- **NM_139011**: 1056 bp - 7015, length: 6967 - 7015, 48 bp
- **NM_139002**: 434 bp - 9608, length: 9177 - 9608, 434 bp

**Intron Coords**

- **NM_000410**: 2776 bp - 3324 bp, length: 298 - 3621 bp
- **NM_139004**: 3634 bp - 209 bp, length: 3886 - 4094 bp
- **NM_139009**: 4367 bp - 1095 bp, length: 4371 - 5465 bp
- **NM_139007**: 5731 bp - 158 bp, length: 5742 - 5899 bp
- **NM_139010**: 6002 bp - 953 bp, length: 6014 - 6966 bp
- **NM_139011**: 7015 bp - 1154 bp, length: 8023 - 9176 bp

**mRNA Accessions**

- **NM_000410**: 2717 bp, length: 7 exons
- **NM_139004**: 1922 bp, length: 5 exons
- **NM_139009**: 1280 bp, length: 6 exons
- **NM_139007**: 1085 bp, length: 5 exons
- **NM_139010**: 809 bp, length: 4 exons
- **NM_139011**: 533 bp, length: 3 exons
- **NM_139005**: 1140 bp, length: 5 exons
- **NM_139003**: 804 bp, length: 5 exons
- **NM_139006**: 1045 bp, length: 6 exons
- **NM_139008**: 781 bp, length: 5 exons
- **NM_139002**: 726 bp, length: 4 exons

**Protein Accessions**

- **NP_000401**: 349 aa, length: 6 exons
- **NP_620573**: 257 aa, length: 5 exons
- **NP_620572**: 326 aa, length: 6 exons
- **NP_620571**: 216 aa, length: 5 exons
- **NP_620570**: 169 aa, length: 4 exons
- **NP_620580**: 77 aa, length: 3 exons
- **NP_620574**: 277 aa, length: 5 exons
- **NP_620572**: 243 aa, length: 5 exons
- **NP_620575**: 335 aa, length: 6 exons
- **NP_620577**: 247 aa, length: 5 exons
### Gene Model (mRNA alignment) Information from genome sequence

<table>
<thead>
<tr>
<th>Total gene model (contig mRNA transcript):</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>transcript</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>NM_004010</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_004010</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
<tr>
<td>NM_159501</td>
<td>plus strand</td>
</tr>
</tbody>
</table>
HEMCHROMATOSIS; HFE

ALTERNATIVE TITLES; SYMBOLS

HLA-H
HEMCHROMATOSIS, HEREDITARY; HH
HEMCHROMATOSIS GENE, INCLUDED; HFE, INCLUDED

Gene map locus 6p21.3

TEXT

DESCRIPTION
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*

---

### Laboratories offering clinical testing:

<table>
<thead>
<tr>
<th>Laboratory Name</th>
<th>Sequencing of entire coding region</th>
<th>Sequencing of select exons</th>
<th>Mutation detection</th>
<th>Prenatal diagnosis</th>
<th>Preimplantation diagnosis</th>
<th>Clinical confirmation of mutations identified in a research lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARUP Laboratories, Inc.</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>ARUP Laboratories</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Salt Lake City, UT</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Blaine Lyons, PhD; Hong Mao, MD; Edward R Ashwood, MD; Narzia Pastewka, PhD</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Adolab HealthCare Group</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Adolab Genetic Diagnostic Center</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Istanbul, Turkey</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Enders Atik, MD, PhD</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Alberta Children's Hospital</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Molecular Diagnostic Laboratory</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Calgary, Alberta, Canada</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Peter Bridge, PhD, FCOG, FSCMG; Jillian Parboosingh, PhD, FCOG</td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="X" /></td>
</tr>
</tbody>
</table>
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

Molecular Genetics

Information in the Molecular Genetics tables may differ from that in the text; tables may contain more recent information. —JLG.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Chromosomal Location</th>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFE</td>
<td>6p21.3</td>
<td>Hereditary Hemochromatosis protein</td>
</tr>
</tbody>
</table>

Data are compiled from the following standard references: Gene symbol from HGNC; chromosomal locus, gene name, clinical group, complementation group from OMIM; protein name from SWISS-prot.

OMIM Entries for HFE- Associated Hereditary Hemochromatosis
- 235200: HEMOCHROMATOSIS, HFE

Genomic Databases for HFE- Associated Hereditary Hemochromatosis

- Gene Symbol | Entrez Gene | HGMD | GeneCards | GEO | GenAtlas |
- HFE          | 235200      | 119939 | HFE     | 119939 |    HFE   |

For a description of the genomic database listed, click here.

Normal allelic variants: A snine at position 65 to cysteine (65C) has been identified. The effect of this mutation is unknown.

Pathologic allelic variants: Two missense mutations have been identified, a cysteine at position 282 to tyrosine (282Y); histidine at position 67 to aspartic acid (67D).
- Cys282Tyr (synonymous: C282Y; nucleotide 886G→A): This missense mutation removes a highly conserved cysteine residue that normally forms an intramolecular disulfide bond, and thereby prevents the protein from being expressed on the cell surface.
- His67Asp (synonymous: H67D; nucleotide 167C→T): This missense mutation may impair interaction of the HFE-encoded protein with the transferrin receptor on the cell surface.

Normal gene product: A cell-surface protein of 321 amino acids with sequence similarity to HLA Class I molecules. The normal protein forms a heterodimer with beta-2-microglobulin, and this interaction is necessary for normal presentation on the cell surface. The normal protein binds to the transferrin receptor, and may act by modulating its affinity for transferrin.

Abnormal gene product: an impaired cell-surface protein is apparently formed. This protein does not migrate to the cell surface and does not bind transferrin (bound to deficient iron). Therefore, lack of internalization of transferrin into the small bowel absorptive cell may lead to compensatory increase in iron absorption (Baum et al 1995).

Resources

GenetReviews provides information about selected national organizations and resources for the benefit of the reader. GenetReviews is not responsible for information provided by other organizations. —JLG.

- National Digestive Diseases Information Clearinghouse (NDDIC) Hemochromatosis
- National Human Genome Research Institute Learning About Hereditary Hemochromatosis
- National Library of Medicine Genetics Home Reference Hemochromatosis
The hemochromatosis protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin.
HEMOCHROMATOSIS; HFE

ALLELIC VARIANTS
(selected examples)

- **0001 HEMOCHROMATOSIS** [HFE, CYS282TYR] (dbSNP)
- **0002 HEMOCHROMATOSIS** [HFE, H63GASP] (dbSNP)
- **0003 HEMOCHROMATOSIS** [HFE, SER65CYS] (dbSNP)
- **0004 HFE INTRONIC POLYMORPHISM** [HFE, 5565G>A] (dbSNP)
- **0005 HFE POLYMORPHISM** [HFE, VAL53MET] (dbSNP)
- **0006 HFE POLYMORPHISM** [HFE, VAL59MET] (dbSNP)
- **0007 Porphryia Varaigna** [HFE, GLN127His] (dbSNP)
- **0008 HEMOCHROMATOSIS** [HFE, AR1003MET] (dbSNP)
- **0009 HEMOCHROMATOSIS** [HFE, LEU811FHR] (dbSNP)
- **0010 HEMOCHROMATOSIS** [HFE, GLY93ARG] (dbSNP)
- **0011 HEMOCHROMATOSIS** [HFE, GLN283PRO] (dbSNP)
Query: gi|4504377| homochromatosis protein isoform 1 precursor [Homo sapiens]
Matching gi: 57114069, 38502807, 29709343, 22854810, 20250786, 15115850, 14100030, 11094315, 2497915, 2370111, 2088551460970

200 BLAST hits to 23 unique species Sort by taxonomy proximity

<table>
<thead>
<tr>
<th>Domain</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>0</td>
</tr>
<tr>
<td>Bacteria</td>
<td>199</td>
</tr>
<tr>
<td>Metazoa</td>
<td>0</td>
</tr>
<tr>
<td>Fungi</td>
<td>0</td>
</tr>
<tr>
<td>Plants</td>
<td>0</td>
</tr>
<tr>
<td>Viruses</td>
<td>0</td>
</tr>
<tr>
<td>Other Eukaryotes</td>
<td>0</td>
</tr>
</tbody>
</table>

Keep only | Cut-Off | Select | Reset |

348 aa

<table>
<thead>
<tr>
<th>SCORE</th>
<th>P</th>
<th>ACCESSION</th>
<th>GI</th>
<th>PROTEIN DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1517</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain G, Homochromatosis Protein Hfe</td>
</tr>
<tr>
<td>1517</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain D, Homochromatosis Protein Hfe</td>
</tr>
<tr>
<td>1517</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain A, Homochromatosis Protein Hfe</td>
</tr>
<tr>
<td>1517</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain C, Hfe (Human) Homochromatosis</td>
</tr>
<tr>
<td>525</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain A, The Crystal Structure Of Hfe</td>
</tr>
<tr>
<td>525</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain D, Crystal Structures Of The</td>
</tr>
<tr>
<td>525</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain A, Crystal Structures Of The</td>
</tr>
<tr>
<td>525</td>
<td>1</td>
<td>T0346</td>
<td>4680560</td>
<td>Chain D, Crystal Structures Of The</td>
</tr>
</tbody>
</table>

Query: gi|4504377| homochromatosis protein isoform 1 precursor [Homo sapiens]
Matching gi: 57114069, 38502807, 29709343, 22854810, 20250786, 15115850, 14100030, 11094315, 2497915, 2370111, 2088551460970

Get CasID Now!

200 BLAST hits to 4 unique species Sort by taxonomy proximity

<table>
<thead>
<tr>
<th>Domain</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>0</td>
</tr>
<tr>
<td>Bacteria</td>
<td>260</td>
</tr>
<tr>
<td>Metazoa</td>
<td>0</td>
</tr>
<tr>
<td>Fungi</td>
<td>0</td>
</tr>
<tr>
<td>Plants</td>
<td>0</td>
</tr>
<tr>
<td>Viruses</td>
<td>0</td>
</tr>
<tr>
<td>Other Eukaryotes</td>
<td>0</td>
</tr>
</tbody>
</table>

Keep only | Cut-Off | Select | Reset |
Bacon et al. Gastroenterology, 116:193-207, Figure 4
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 "Geneview in dbSNP" 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 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 first on the NP_000509 protein link and then on Blink. Click on the 3D structures button. The output contains a list of similar proteins with 3D structures known. Find the entry, for example 1DXTD, representing the structure of deoxyhemoglobin. Click on the blue dot next to 1DXTD to get the sequence alignment of the query protein to the D chain of 1DXT. 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 Primary Citation). 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 unclick 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 7 non-synonymous SNPs and 3 synonymous 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.