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 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.


Course Developed by Medha Bhagwat (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 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 "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 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.

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 where as the Cys282Tyr mutant fails to regulate this interaction leading to iron overload. The conserved cysteine 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.
Entrez Gene is a searchable database of genes, from *Homo sapiens*, and defined by sequence and/or located in the NCBI Map Viewer.

**1: HFE homochromatosis [Homo sapiens]**

**GeneID**: 3077

**Official Symbol**: HFE and Name: homochromatosis [Homo sapiens]

**Other Aliases**: HFE1, HH, HLA-H, MGC103500, d221c16.10.1

**Other Designations**: MHC class I-like protein HFE; homochromatosis protein; hereditary hemochromatosis protein HLA-H

**Chromosomes**: 6

**Location**: 6p21.3

**MIM**: 235200

**Genetic**: 3077

**Summary**

- **Official Symbol**: HFE
- **Homo sapiens**
- Provided by NCBI

- **Primary source**: NCBI-GenBank
- **See related**: HPPO-01950, MIM:216260
- **Gene type**: protein coding
- **Status**: Reviewed

- **Organism**: Homo sapiens
- **Lineage**: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
- **Also known as**: HH; HFE1; HLA-H; MGC103500; d221c16.10.1

**Summary**

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 iron.
<table>
<thead>
<tr>
<th>Region</th>
<th>Contig</th>
<th>mRNA pos</th>
<th>dbsNP rsID</th>
<th>cluster id</th>
<th>Heterozygosity</th>
<th>Validation</th>
<th>3D</th>
<th>OMM</th>
<th>Function</th>
<th>dbsNP allele</th>
<th>Protein residue</th>
<th>Codon pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon_1</td>
<td>16949347</td>
<td>825</td>
<td>rs24125855</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>C</td>
<td>Thr [T]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949347</td>
<td>825</td>
<td>rs1733949</td>
<td>0.139</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>G</td>
<td>Asp [D]</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949347</td>
<td>813</td>
<td>rs1801752</td>
<td>0.139</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>C</td>
<td>His [H]</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949356</td>
<td>813</td>
<td>rs1801752</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>T</td>
<td>Cys [C]</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949356</td>
<td>813</td>
<td>rs1039520</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>C</td>
<td>Thr [T]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949356</td>
<td>813</td>
<td>rs1039520</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>coding reference</td>
<td>T</td>
<td>Leu [I]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949356</td>
<td>802</td>
<td>rs29734725</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>C</td>
<td>His [H]</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16949356</td>
<td>802</td>
<td>rs29734725</td>
<td>N.D.</td>
<td>H</td>
<td>Yes</td>
<td>coding reference</td>
<td>T</td>
<td>Leu [I]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16951137</td>
<td>871</td>
<td>rs29734750</td>
<td>0.003</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>T</td>
<td>Leu [I]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16951137</td>
<td>871</td>
<td>rs1805026</td>
<td>0.043</td>
<td>H</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>A</td>
<td>Tyr [Y]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16951137</td>
<td>871</td>
<td>rs1805026</td>
<td>0.043</td>
<td>H</td>
<td>Yes</td>
<td>coding reference</td>
<td>C</td>
<td>Cys [C]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12

HOMOCYSTINURIA; MTHFR

ALLERIC VARIANTS

- 0001 HOMOCYSTINURIA [HFE, CYS282TYR] dbSNP
- 0002 HOMOCYSTINURIA [HFE, HIS63ASP] dbSNP
- 0003 HOMOCYSTINURIA [HFE, SER65CYS] dbSNP
- 0004 HFE ECTOPIC POLYMORPHISM [HFE, 5569G-A] dbSNP
- 0005 HFE POLYMORPHISM [HFE, VAL55MET] dbSNP
- 0006 HFE POLYMORPHISM [HFE, H240W] dbSNP
- 0007 PORPHYRIA VARIETAT [HFE, GLN127HES] dbSNP
- 0008 HOMOCYSTINURIA [HFE, ARG353MET] dbSNP
- 0009 HOMOCYSTINURIA [HFE, L121T] dbSNP
- 0010 HOMOCYSTINURIA [HFE, GLY382ARG] dbSNP
- 0011 HOMOCYSTINURIA [HFE, GLN238FEO] dbSNP

Copyright © 1996-2006 Johns Hopkins University
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:                      | Sequencing of entire coding region | Sequencing of exonic events | Analysis of variant regions | Clinical diagnosis | Laboratory identification
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ARUP Laboratories, Inc.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ARUP Laboratories</td>
</tr>
<tr>
<td>ARUP Laboratories</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ARUP Laboratories</td>
</tr>
<tr>
<td>Salt Lake City, UT</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Salt Lake City</td>
</tr>
<tr>
<td>Elaine Lyce, PhD, Rong Mao, MD, Edward R. Ashwood, MD, Marzia Pasqual, PhD</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Elaine Lyce, MD</td>
</tr>
<tr>
<td>Askadem Healthcare Group</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Askadem Healthcare</td>
</tr>
<tr>
<td>Askadem Genetic Diagnostic Center</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Askadem Genetic</td>
</tr>
<tr>
<td>Istanbul, Turkey</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Istanbul, Turkey</td>
</tr>
<tr>
<td>Erdal Altun, MD, PhD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Erdal Altun, MD</td>
</tr>
<tr>
<td>Alberta Children's Hospital</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Alberta Children's</td>
</tr>
<tr>
<td>Molecular Diagnostic Laboratory</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Molecular Diagnostic</td>
</tr>
<tr>
<td>Calgary, Alberta, Canada</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Calgary, Alberta</td>
</tr>
<tr>
<td>Piotr Biedrzycki, PhD, PCCMG, FACC, Jillian Parboosingh, PhD, FACC</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Piotr Biedrzycki, PhD</td>
</tr>
</tbody>
</table>
**HFE-Associated Hereditary Hemochromatosis**

**Authors:**
- Klys W Krawadl, MD
- Jonathan F. Tall, MD, PhD
- Robin L. Bessette, MS
- Anna G. Mihaylik, MD

**About the Authors**
- Initial Posting: 3 April 2000
- Last Update: 4 December 2008

**Summary**

**Disease characteristics:** HFE-associated hereditary hemochromatosis (HFE-HHC) is characterized by inappropriately 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. Hepatitis/hepatomegaly 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 hypogonadism.

This description applies to individuals with clinical expression of HFE-HHC. A large, but yet undefined, fraction of heterozygotes for HFE-HHC do not develop clinical symptoms (i.e., penetrance is low).

**Diagnosis/Testing:** The diagnosis of HFE-HHC in individuals with clinical symptoms consistent with HFE-HHC and/or biochemical
Normal allelic variants: The HFE gene is about 13 kb in size and contains seven exons [Feder et al 1996, Abaz 1998]. HFE gives rise to at least eleven alternative transcripts encoding four to seven exons.

Pathologic allelic variants: At least 28 distinct mutations have been reported, most being missense or nonsense mutations. Two observed mutations account for the vast majority of disease-causing alleles in the population:

- Cys282Try (p.Cys282Try, nucleotide 846G>T). This missense mutation removes a highly conserved cysteine residue that normally forms an intermolecular disulfide bond with beta-2-microglobulin, and thereby prevents the protein from being expressed on the cell surface.
- He1K46R (p.His46Arg; nucleotide 147C>G). This missense mutation may alter a pI-dependent intramolecular salt bridge, possibly affecting interaction of the HFE protein with the transferrin receptor.

Abnormal gene products: The p.Cys282Try 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 mechanistic basis 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.
Query: hemochromatosis protein isoform 1 precursor [Homo sapiens]
Reference: [NMDE] [PubMed]

Structure: IDE4 Chain G, Hemochromatosis Protein Hfe Complexed With Transferrin Receptor

E-value = 7e-168, Bit score = 588, Aligned length = 275, Sequence Identity = 100%
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 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 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 7 non-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.