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 literature, expression and structure databases 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. This 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? Which is the longest splice variant? 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?

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:

A. Visualization of cysteine 282 on the structure of the hemochromatosis protein

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 entry 1A6Z chain C provides the structure of 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 to the area of the disulphide bridges (colored in tan) by pressing "z" on the keyboard. Select the cysteine residues forming the disulphide bridges by double clicking on them. Mouse over the corresponding cysteine residues on the query line in the Alignment Viewer and read 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 that is mutated to tyrosine causing the hemochromatosis phenotype.

B. Visualization of hemochromatosis protein and beta-2-microglobulin complex

Return to the sequence alignment (Related Structures) page and select the link to MMDB (the Molecular Modeling Database). The graphic representation of the structure lists four chains. The PDB record, which can be accessed through the “1A6Z” link on the MMDB page, indicates that chains A and C represent the human hemochromatosis protein, while chains B and D represent human beta-2-microglobulin. Download the structure of the complex by clicking on the “View 3D
Structure” button on the MMDB page. For easier viewing, remove the helix and strand objects using Style→Edit Global Style -- unclick the boxes next to the Helix objects and Strand objects. To distinguish between the individual chains, select “Molecule” as the Color Scheme for the protein backbone. Click on the “Apply”, then “Done” buttons.

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.
The interaction of hemochromatosis protein with beta-2-microglobulin allows cell surface presentation of the complex. Once on cell surface, the hemochromatosis protein regulates iron absorption by regulating the interaction of the transferrin receptor with transferrin.
General protein information

Names
hemochromatosis protein
HFE class I-like protein HFE
hereditary hemochromatosis protein HLA-H

NCBI Reference Sequences (RefSeq)

Genomic

1. NM_000410.3 - NP_000601.1 hemochromatosis protein isoform 1 precursor

mRNA and Protein(s)

1. NM_000410.3 - NP_000601.1 hemochromatosis protein isoform 1 precursor

Source sequence(s)
AP019598.1; XM.13130
CCDS4878.1

Conserved Domains (2)
summary

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>cd40584</td>
<td>223-298</td>
</tr>
<tr>
<td>pfam06129</td>
<td>7-62</td>
</tr>
</tbody>
</table>

Descriptive
hemochromatosis protein isoform 1 precursor

Descriptions
Transcript Vrants: This variant (2) lacks a large 3’ region including the 3’ CDS and UTR but has an alternate 3’ exon, as compared to variant 1. The resulting protein (isoform 2) has a unique carboxy terminal.
124. determined rate-specific frequencies of the HFE mutations, C282Y and H63D
125. the homozygous Cys282Tyrsense mutation and high levels of serum ferritin. It is important to recognize symptoms of iron overload at an early stage because hereditary hemochromatosis needs to be treated immediately.
126. The effect of perturbing anion transport on cardiac autonomic function was studied in subjets with at least 1 copy of an HFE variant compared with wild-type subjects.
127. Our data suggest that the HFE gene is not a major disease gene for migraine.
128. analysis of the localization and functional effects of the HR and its chaperone protein beta2M
129. Prevalence of apolipoprotein B is significantly higher in symptoms of alcoholic fatty liver and fibrosis patients but not in healthy controls. The prevalence in alcohol fibrosis was as high as in the liver from Wilson’s disease and hemochromatosis patients.
130. The Ala51Val mutation may have a possible role on the cause of hemochromatosis in a Japanese case
131. REVHW (C282Y mutant gene product failed to associate with 2microglobulin and significantly reduced cell surface expression of the HFE-2m complex, thereby affecting the interaction with TFR and its interaction with transferrin.
132. 871 healthy unrelated subjects in Poland were collected to assess the relevant frequencies. Each subject was genotyped for the C282Y and H63D

Interactions

<table>
<thead>
<tr>
<th>Description</th>
<th>Interactant</th>
<th>Other Gene</th>
<th>Complex</th>
<th>Source</th>
<th>Pubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_000401.1</td>
<td>Beta 2 microglobulin</td>
<td>B2M</td>
<td></td>
<td>HPGR</td>
<td>PubMed</td>
</tr>
<tr>
<td>NP_000401.1</td>
<td>Transferrin receptor 2</td>
<td>TFRC</td>
<td></td>
<td>HPGR</td>
<td>PubMed</td>
</tr>
<tr>
<td>NP_000401.1</td>
<td>NP_003223.1</td>
<td>TFRC</td>
<td></td>
<td>HPGR</td>
<td>PubMed</td>
</tr>
<tr>
<td>Region</td>
<td>Contig position</td>
<td>mRNA pos</td>
<td>dbsNP rs#</td>
<td>cluster id</td>
<td>Heterozygosity</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>exon_1</td>
<td>16949420</td>
<td>161</td>
<td>rs2242955</td>
<td></td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exon_2</td>
<td>16949347</td>
<td>264</td>
<td>rs2389145</td>
<td></td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exon_3</td>
<td>16949528</td>
<td>437</td>
<td>rs2389145</td>
<td></td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exon_4</td>
<td>16951392</td>
<td>1005</td>
<td>rs800562</td>
<td></td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exon_5</td>
<td>16952084</td>
<td>1186</td>
<td>rs3991066</td>
<td></td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HEMOCHROMATOSIS; HFE

DESCRIPTION

The clinical features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatic cellular carcinoma (HCC, 114500), complicating cirrhosis, is responsible for about one-third of deaths in affected homozygotes. Since hemochromatosis is a relatively easily treated disorder if diagnosed, this is a form of preventable cancer.

ALLELIC VARIANTS

- 0001 HEMOCROMATOSIS [HFE, CYS282TYR] dbSNP: PORPHYRIA VARIEGATA, INCLUDED
- 0002 HEMOCROMATOSIS [HFE, H636ASP] dbSNP: PORPHYRIA VARIEGATA, INCLUDED
- 0003 HEMOCROMATOSIS [HFE, SER106CT] dbSNP
- 0004 HFE INTRONIC POLYMORPHISM (HFE, 40A>C) dbSNP
- 0005 HFE POLYMORPHISM (HFE, VAL31MET) dbSNP
- 0006 HFE POLYMORPHISM (HFE, VAL158MET) dbSNP
- 0007 PORPHYRIA VARIEGATA (HFE, G151D) dbSNP
- 0008 HEMOCROMATOSIS [HFE, A939G] dbSNP
- 0009 HEMOCROMATOSIS [HFE, E1103C] dbSNP
- 0010 HEMOCROMATOSIS [HFE, G199AR] dbSNP
- 0011 HEMOCROMATOSIS [HFE, G282PR] dbSNP
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

- **Laboratories offering clinical testing:**
  - Pediatric, Italy
  - Alberta Lions, BSc, PhD; Antonina D’Amico, BSc, PhD; Elisa Del Giudice, BSc, PhD
  - AMLD Laboratories
  - National Genetics Laboratory
  - Salt Lake City, UT
  - Elaine Lions, PhD; Pang Nei, MD; Edward A. Fleshner, MD; Maria Panza, MD
  - Emory University Hospitals
  - Emory University Clinic, Atlanta, Georgia
  - Ender Atick, MD, PhD
  - Alberta Children’s Hospital
  - Molecular Diagnostics Laboratory
  - Calgary, Alberta, Canada
  - Phelan Bridge, PhD, FCQMG, PACMG; Jillian Farb, MD, FCQMG
  - Baylor College of Medicine
  - Medical Genetics Laboratories
  - Houston, TX
  - Christine K. Eng, MD, FACMG; William T. O'Brien, PhD; Lee-lin Wong, PhD; Sue W. Cheung, PhD
  - Benelux, BV, NMA
  - Molecular Biology Laboratories
  - Kutten, Czech Republic
  - Fachschaft Medizin, MIV
  - Biochemistry Center of Wisconsin
  - Molecular Diagnostics Laboratory
  - Milwaukee, WI
  - Daniel B. Heidtman, PhD
  - Boston University School of Medicine
  - Center for Human Genetics
  - Boston, MA
  - Aubry Wilkinsky, MD, DSc
  - brunette Molecular Genetics Diagnostic and Research Laboratory
  - Istanbul, Turkey

- **Additional testing labs:**
  - **Testing:**
  - **Research:**
  - **Review:**
  - **Resources:**
### HFE-Associated Hereditary Hemochromatosis

**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 hypogonadism.

**Gene Symbol**: HFE

**Gene Entry**: 11680

**Ensembl**: ENSG00000115952

**OMIM Entries for HFE-Associated Hereditary Hemochromatosis**

<table>
<thead>
<tr>
<th>OMIM Entry</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>235200</td>
<td>HEMOCHROMATOSIS, HFE</td>
</tr>
</tbody>
</table>

**Genomic Databases for HFE-Associated Hereditary Hemochromatosis**

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Entry Gene</th>
<th>HGMD</th>
<th>GeneCards</th>
<th>GDB</th>
<th>Genetals</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFE</td>
<td>235200</td>
<td>HFE</td>
<td>HFE</td>
<td></td>
<td>HFE</td>
</tr>
</tbody>
</table>

**Normal allele variants**: The HFE gene is about 10 kb in size and contains seven exons [Feder et al 1995; Akyu 1998]. HFE gives rise to at least seven alternative transcripts encoding four to seven exons.

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

- **Cys282Tyr** (p.C82Y): nucleotide B455C (c.1364T>C).
  - 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.
- **His63Pro** (p.H63P): nucleotide C191G (c.57C>G).
  - 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 most predicted primary translation product is 348 amino acids, which gives rise to a mature protein of about 321 amino acids after cleavage of the signal sequence. The HFE protein is similar to the alpha-2-macroglobulin at the primary [Feder et al 1996] and tertiary structure [Budin et al 1998] levels. The mature protein is expressed on the cell surface as a heterodimer with beta-2-microglobulin, and this interaction is necessary for normal presentation on the cell surface. The normal HFE 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 as yet unknown [Keverne et al 2004].

**Abnormal gene product**: The p.C82Y 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 as yet unclear.

**Resources**

...
The interaction of hemochromatosis protein with beta-2-microglobulin allows cell surface presentation of the complex. Once on cell surface, the hemochromatosis protein regulates iron absorption by regulating the interaction of the transferrin receptor with transferrin.
Molecular components in the MMDB structure are listed below. The icons indicate macromolecular chains, 3D domains, protein classifications and ligands. Please hold the mouse over each icon for more information on the component.
The interaction of hemochromatosis protein with beta-2-microglobulin allows cell surface presentation of the complex. Once on cell surface, the hemochromatosis protein regulates iron absorption by regulating the interaction of the transferrin receptor with transferrin.
Beta-2-microglobulin

Transferrin

Transferrin receptor

Beta-2-microglobulin

Cys282

hemochromatosis
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?

What is the name and function of the protein encoded by the HBB gene? Beta globin is a subunit of which protein? Name other subunit(s) in that protein.

**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
(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.