

The Gene Gateway Workbook

A collection of activities introducing new users to the web resources that scientists access to learn about genetic disorders, genes, and proteins.

Human Genome Landmarks
Selected Genes, Traits, and Disorders

Multiple myeloma oncogene
Orofacial cleft
Leukemia, acute nonlymphocytic
Fanconi anemia, complementation group E
Ankylosing spondylitis
Stickler syndrome, type II
OSMED syndrome
Weissenbacher-Zweymuller syndrome
Deafness, nonsyndromic sensorineural
Dyslexia
Hemochromatosis ←
Porphyria variegata
Pemphigoid, susceptibility to
Immune suppression to streptococcal antigen
Sialidosis, types I and II
Panbronchiolitis, diffuse
Psoriasis susceptibility

To view the chromosomes of the Human Genome Landmarks poster online, order your free copy of the poster, or download additional copies of this workbook, go to the Gene Gateway website:

genomics.energy.gov/genegateway/

Using hereditary hemochromatosis as a model, access a variety of websites and databases to

- Learn about a genetic disorder and its associated gene.
- Identify mutations that cause the disorder.
- Find the gene on a chromosome map.
- Examine the gene's sequence and structure.
- Access the amino acid sequence of a gene's protein product.
- Explore the 3-D structure of the gene's protein product.

Acknowledgements

This workbook was last updated February 2011. With much appreciation, we would like to acknowledge Carrie Morjan, Aurora University, who facilitated the creation of this latest version by sharing her updates to The Gene Gateway Workbook for use in her genetics classes.

This workbook was first produced by the Biological and Environmental Research Information System at Oak Ridge National Laboratory, Oak Ridge, Tennessee, July 2003, with support from the U.S. Department of Energy Office of Biological and Environmental Research.

Office of Biological and Environmental Research
Office of Science
U.S. Department of Energy (DOE)



U.S. DEPARTMENT OF
ENERGY

Office of
Science

For More Information

This workbook is freely downloadable from the Gene Gateway website (see link below). For questions or comments concerning this document, contact Jennifer Bownas by email at bownasjl@ornl.gov.

Gene Gateway
genomics.energy.gov/genegateway/

Human Genome Project Information
www.ornl.gov/hgmis/home.shtml

DOE Genomic Science Program
genomicscience.energy.gov

DOE Office of Biological and Environmental Research
science.energy.gov/ber/

Table of Contents

Introduction.....	4
Why Use Hereditary Hemochromatosis as a Model?	5
Some Basic Concepts to Understand Before Starting	5
Activity 1.....	6
Online Mendelian Inheritance in Man (OMIM).....	6
GeneTests	11
Activity 2.....	14
NCBI Map Viewer	14
Activity 3.....	22
NCBI Entrez Gene and GenBank	22
Activity 4.....	32
UniProt Protein Knowledgebase and BLAST Searching	32
Activity 5.....	38
Protein Data Bank	38
Table of Standard Genetic Code for Translating DNA Sequence Records.....	50
Hereditary Hemochromatosis Worksheet	51

Introduction

The Gene Gateway Workbook is a collection of activities with screenshots and step-by-step instructions designed to introduce new users to genetic-disorder and bioinformatics resources freely available on the Web. It should take about 3 hours to complete all five activities.

The workbook activities were derived from more detailed guides and tutorials available at the Gene Gateway website (genomics.energy.gov/genegateway/). This website was created as a resource for learning more about the genes, traits, and disorders listed on the Human Genome Landmarks (HGL) poster, but it can be used to investigate any gene or genetic disorder of interest.

Many guides to using bioinformatic resources are designed for bioscience researchers and are too technical for nonexperts. This workbook and other Gene Gateway resources target a more general audience: teachers, high school and college students, patients with disorders and their families, and anyone else who wants to learn more about how life works at a molecular level.

This workbook shows you how to get started using bioinformatics resources that often intimidate and overwhelm new users. It also demonstrates how information from one resource, such as annotated protein sequence data from the UniProt Protein Knowledgebase, can be used to reinforce and clarify information available from another resource, such as three-dimensional (3-D) structures from Protein Data Bank (PDB). Gene Gateway provides users with a systematic approach to using multiple bioinformatics databases to gain a better understanding of how genes and proteins can contribute to the development of a particular genetic condition.

Using the genetic disorder hereditary hemochromatosis as a model, this workbook shows you how to access:

- Online Mendelian Inheritance in Man (OMIM) and GeneReviews to learn about a genetic disorder, its associated gene or genes, and common disease-causing mutations.
- NCBI Map Viewer to find a gene locus on a chromosome map.
- NCBI Entrez Gene and GenBank to examine the sequence and structure of a gene.
- UniProt Protein Knowledgebase to find the annotated amino acid sequence of a gene's protein product.
- Protein Data Bank to view and modify the 3-D structure of the gene's protein product.

Skills gained by working through the activities in this workbook can be applied to learning about other genetic disorders, genes, and proteins.

This workbook and other genome science resources are available from the website for the genome programs of the Office of Biological and Environmental Research, U.S. Department of Energy Office of Science (genomics.energy.gov/).

Why Use Hereditary Hemochromatosis as a Model?

- Hereditary hemochromatosis, a disorder in which too much iron accumulates in certain tissues and organs, is caused by changes in the DNA sequence of a single gene, so the genetic basis of this condition is easier to understand than more complex disorders caused by alterations in multiple genes.
- The gene and its protein product are relatively well studied. Three-dimensional structures of the protein product are available in PDB, the international repository for macromolecular structure data.
- Hereditary hemochromatosis is the most common autosomal recessive disorder affecting individuals of Northern European descent (about 1 in 200 Caucasians develop hereditary hemochromatosis).
- Effective methods for treatment are available with early diagnosis.

Some Basic Concepts to Understand Before Starting

- Genes are the basic physical and functional units of heredity. Each gene is located on a particular region of a chromosome and has a specific ordered sequence of nucleotides (the building blocks of DNA).
- Central dogma of molecular biology: DNA → RNA → Protein
 - Genetic information is stored in DNA.
 - Segments of DNA that encode proteins or other functional products are called genes.
 - Gene sequences are transcribed into messenger RNA intermediates (mRNA).
 - mRNA intermediates are translated into proteins that perform most life functions.
- Eukaryotic genes have introns and exons. Exons contain nucleotides that are translated into amino acids of proteins. Exons are separated from each other by intervening segments of DNA called introns. Introns do not code for protein, and they are removed when eukaryotic mRNA is processed. Exons are spliced back together to form the intron-free mRNA strand that is used as a template to make proteins.
- Special cellular components (ribosomes) use the triplet genetic code to translate the nucleotides of an mRNA sequence into the amino acid sequence of a protein. A Table of Standard Genetic Code is provided on page 50 of this workbook.
- There are 20 different amino acids. Proteins are created by linking amino acids together in a linear fashion to form polypeptide chains. See the Table of Standard Genetic Code on page 50 for single-letter and three-letter abbreviations for the 20 different amino acids.
- Polypeptide chains fold into 3-D structures that can associate with other molecular structures to perform specific functions.

Activity 1

Online Resources: OMIM and GeneTests

- Learn about the genetic disorder and its associated gene.
- Identify mutations that cause the disorder.

Online Mendelian Inheritance in Man (OMIM)

OMIM is a comprehensive database of human genes, genetic traits, and disorders created by researchers at Johns Hopkins University. The OMIM database, which is updated daily, is accessible through the National Center for Biotechnology Information (NCBI) suite of online resources. Each record in OMIM summarizes the body of research relevant to a particular gene, trait, or disorder.

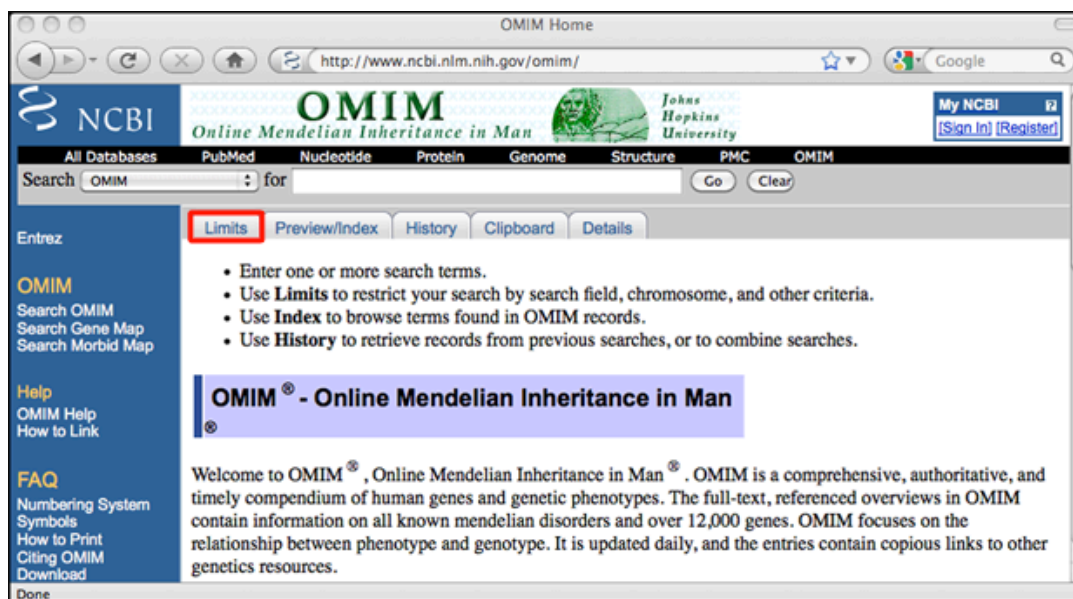
To access OMIM, let's go to the NCBI home page (www.ncbi.nlm.nih.gov) shown below, and then click on **OMIM** in the box on the upper right.



A screenshot of the OMIM home page is shown on the following page. The easiest way to begin a search is to simply type a disorder name in the search box at the top of the OMIM page and submit your search. However, NCBI also supports a variety of features for narrowing a search and browsing disorders alphabetically (using OMIM Morbid Map) or by chromosomal location (using OMIM Gene Map).

To narrow a search, NCBI has options for typing search field qualifiers into the search box [see OMIM Help (www.ncbi.nlm.nih.gov/Omim/omimhelp.html) for more information] or selecting search fields using the **Limits** tab. This exercise will demonstrate searches using the **Limits** tab.

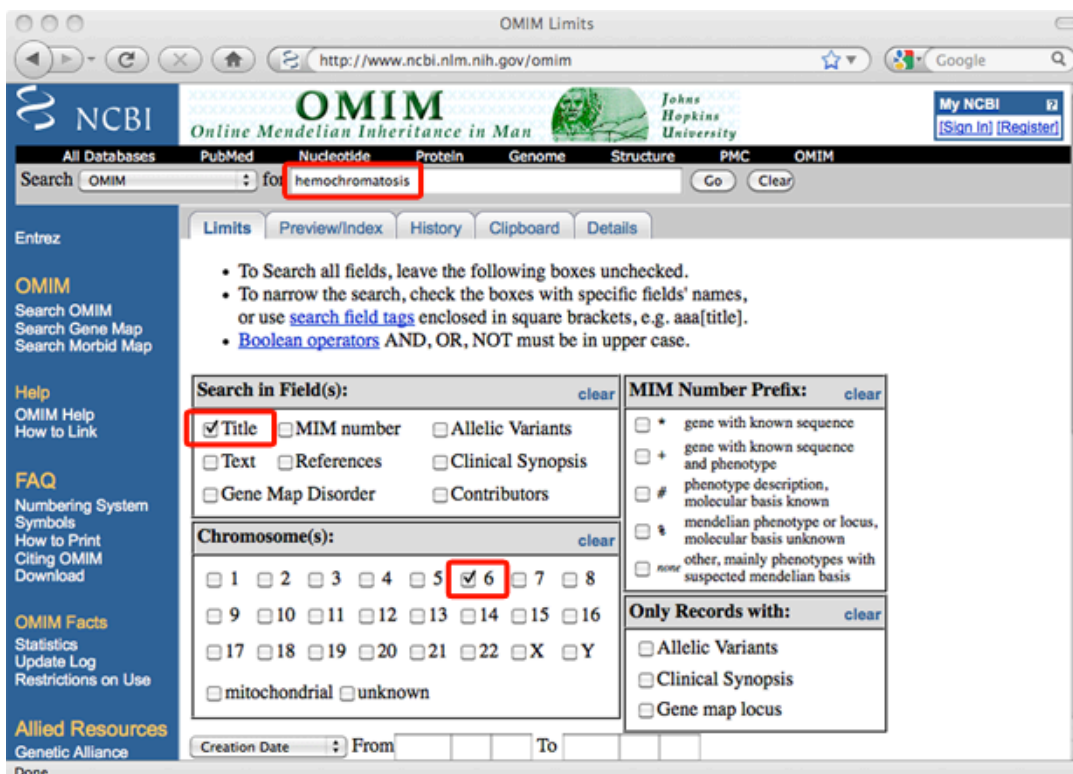
1. Select the **Limits** tab at the top of the OMIM page shown in the screenshot below.



URL for OMIM home page: www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Most genes, disorders and traits listed on the Human Genome Landmarks (HGL) poster were taken from the title fields of OMIM records, so we can narrow our search to look only for those records that have “hemochromatosis” in the title field. By selecting “hemochromatosis” from the HGL poster, we also know that the gene for this disorder is found on chromosome 6.

2. From the Limits page, enter **hemochromatosis** into the search box and select the **Title** box and chromosome **6** as shown in the screenshot below. Click **Go** to submit your search.



NOTE: Searching for OMIM records associated with multi-gene disorders, such as breast cancer or diabetes, which are caused by alterations in genes on different chromosomes, may provide multiple OMIM records in the search results. Limiting your search to just one chromosome for a multi-gene disorder may only retrieve a subset of all the records associated with that disorder.

3. The search should return one result: **MIM ID #235200**. A screenshot of the full OMIM record for the hemochromatosis disorder is shown below.

The screenshot shows a web browser window displaying the OMIM record for Hemochromatosis (HFE) with MIM ID #235200. The browser address bar shows the URL <http://www.ncbi.nlm.nih.gov/omim>. The page header includes the NCBI logo and the OMIM logo (Online Mendelian Inheritance in Man) with the Johns Hopkins University logo. The search bar contains the text "OMIM" and "for hemochromatosis". The search results show "Limits: Title, chromosome 6" and "Display: Detailed". The main content area displays the MIM ID #235200, the disorder name "HEMOCHROMATOSIS; HFE", and a description of the disorder. The description states: "Hereditary hemochromatosis is an autosomal recessive disorder of iron metabolism wherein the body accumulates excess iron (summary by Feder et al., 1996). Excess iron is deposited in a variety of organs leading to their failure, and resulting in serious illnesses including cirrhosis, hepatomas, diabetes, cardiomyopathy, arthritis, and hypogonadotropic hypogonadism. Severe effects of the disease usually do not appear until after decades of progressive iron loading. Removal of excess iron by therapeutic phlebotomy decreases morbidity and mortality if instituted early in the course of the disease. Classic hemochromatosis (HFE) is most often caused by mutation in a gene designated HFE on chromosome 6p21.3." The table of contents on the right side of the page lists various sections, with "Molecular Genetics" highlighted in red.

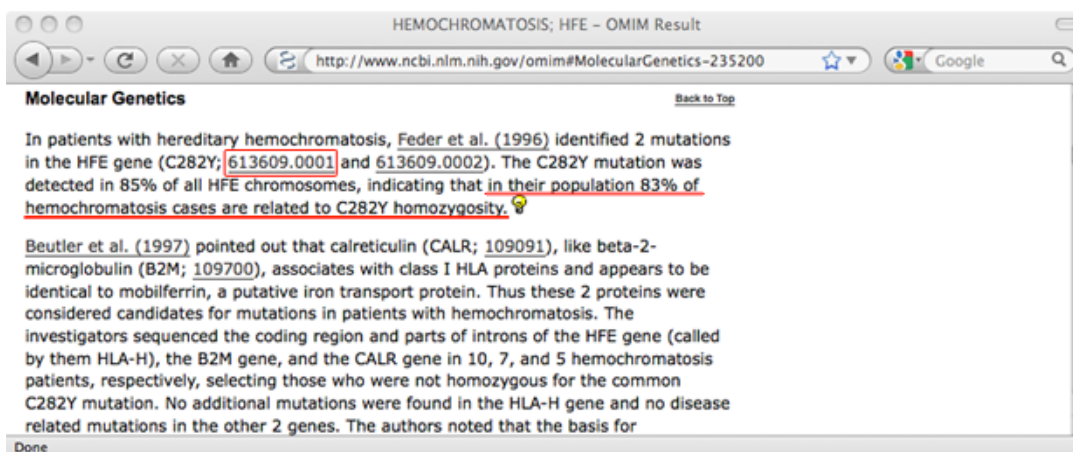
4. Let's examine some of the features of this record:
- Each OMIM record is assigned a unique six-digit **MIM ID** number located at the top of each entry. For hemochromatosis, the MIM ID is 235200. As a unique identifier for a disorder, the MIM ID can be used to search other databases for information about a particular disorder.
 - The number sign (#) prefix in front of the MIM ID means that this entry refers to the description of a phenotype, and the molecular basis for this phenotype is known. For more information about other MIM number prefixes, see OMIM Help (www.ncbi.nlm.nih.gov/Omim/omimhelp.html#MIMnumberPrefix).
 - Below the MIM ID, you will find the disorder name and the official gene symbol (shown in the image on the next page). The official gene symbol, which is **HFE** for hemochromatosis, serves as a unique identifier for a gene. To be "official," a gene symbol must have been approved by the HUGO Gene Nomenclature Committee (www.genenames.org). **The gene**

symbol is especially useful when searching other databases (such as sequence, genome-mapping, and structure databases) for gene-specific information.

MIM ID #235200
HEMOCHROMATOSIS; HFE
 — Disorder Name
 Official Gene Symbol

NOTE: For a disorder like hemochromatosis, which is primarily caused by mutations in a single gene, the official gene symbol may be included in the record title. For complex disorders like breast cancer, official symbols for associated genes will be described in the first paragraph of text.

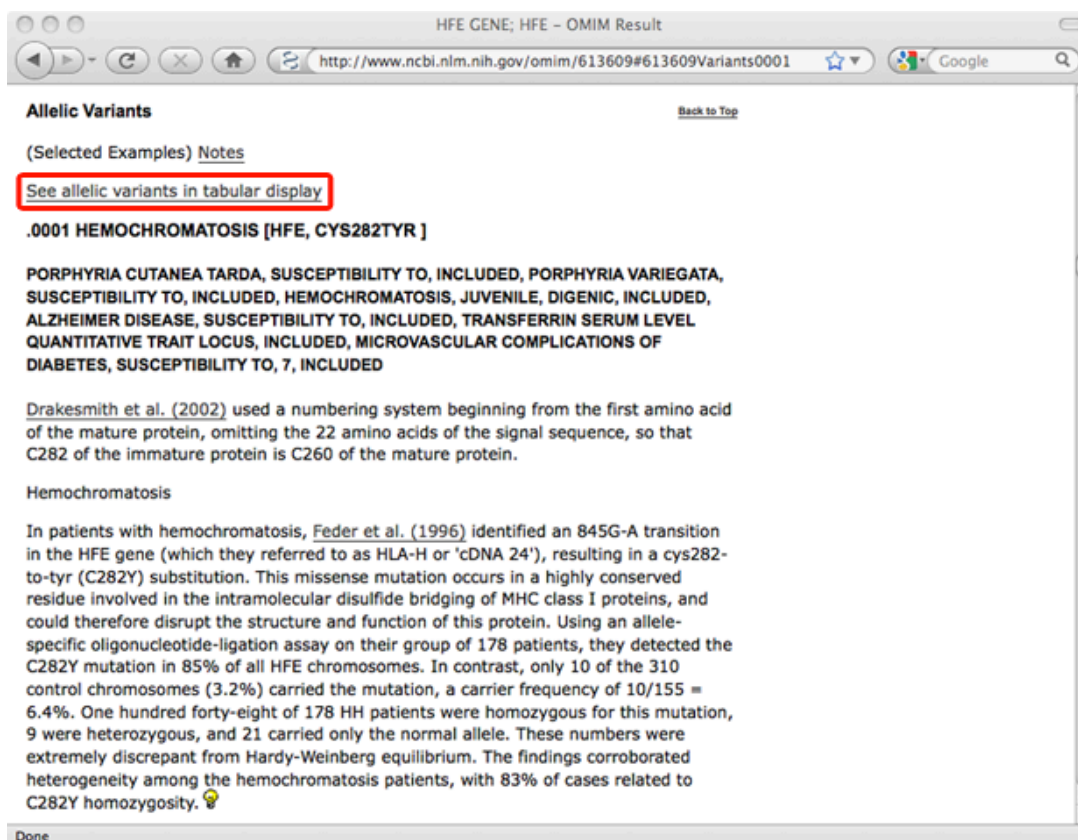
- The **Gene map locus** describes where a gene can be found on a chromosome. For the gene locus **6p21.3**, 6 is the chromosome number, p indicates the short arm of the chromosome, and 21.3 is a number assigned to a particular region of the chromosome. Clicking on a gene map locus opens the OMIM Gene Map, a table of genes organized by chromosomal location.
 - The amount of text within an OMIM record varies according to what is known about a particular gene, disorder, or trait. Since hemochromatosis is well studied, a lot of information is known about this disorder and its gene. Some different types of information that may be included in an OMIM record are disorder description, inheritance, molecular genetics, genotype and phenotype correlations, diagnosis, population genetics, and animal models.
 - Each record includes a **Table of Contents** box on the right with quick links to different sections within the record.
5. To learn more about the molecular basis of hemochromatosis, select the **Molecular Genetics** link in the Table of Contents box (see screenshot on previous page). The Molecular Genetics section of the OMIM record for hemochromatosis is shown below.



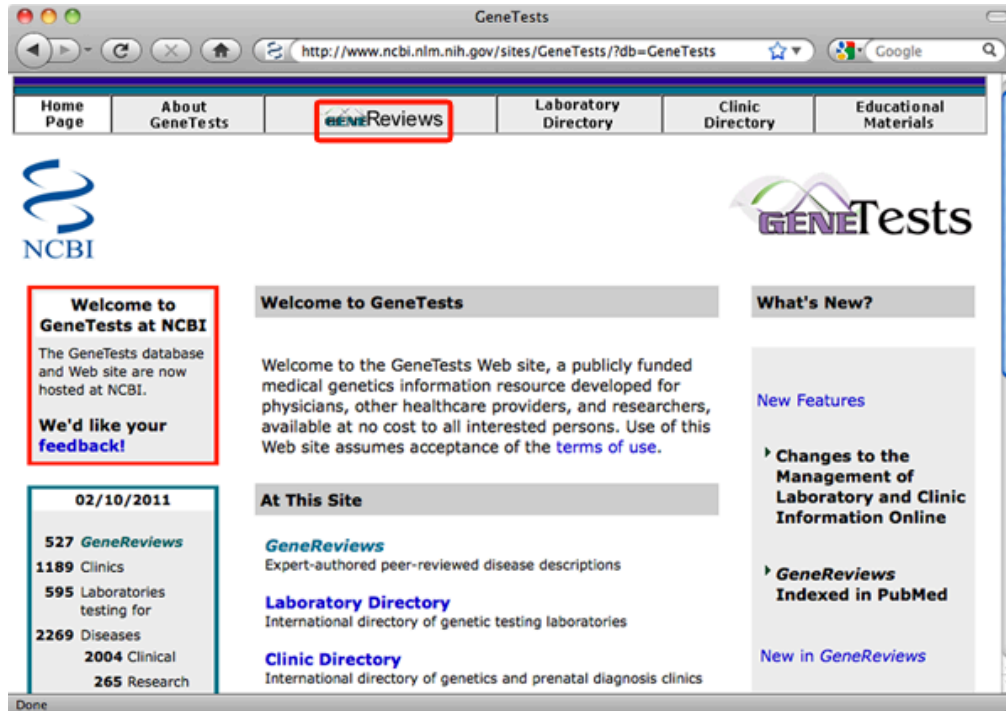
- One study showed that about 83% of hemochromatosis cases are related to the C282Y mutation. The "C282Y" notation means that a mutation occurs in the DNA sequence that changes the amino acid at position 282 in the protein product from a cysteine (C) to a tyrosine (T).
6. Click on the first link for the C282Y mutation, [613609.0001](#). This link will take you to the OMIM record for the HFE gene (MIM ID *613609; the asterisk prefix indicates the record represents a gene of known sequence). OMIM often maintains separate records for

phenotypes (such as the disorder hemochromatosis) and the genes associated with those phenotypes.

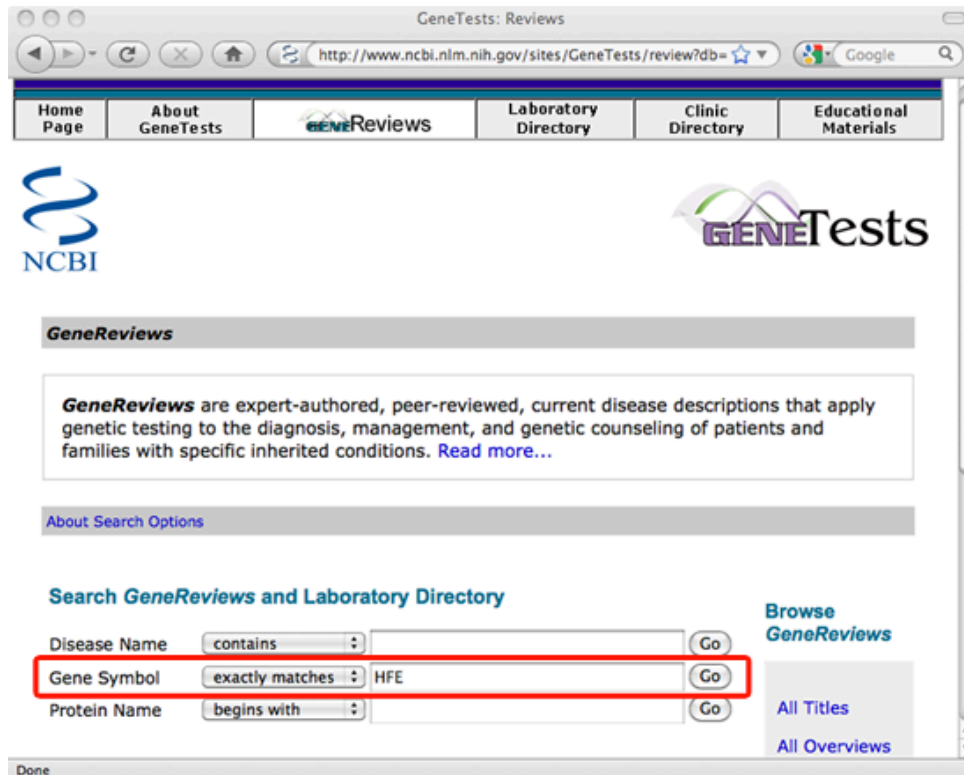
- The **Allelic Variants** section of the OMIM record for the HFE gene is shown in the screenshot below. This section typically describes some of the most notable gene mutations (also called allelic variants) that produce disease phenotypes. Note that the C282Y mutation is also known as the CYS282TYR mutation, and it is the first of several mutations that have been identified for the HFE gene. To see a listing of the different mutations for the HFE gene, click on the “**See allelic variants in tabular display**” link.



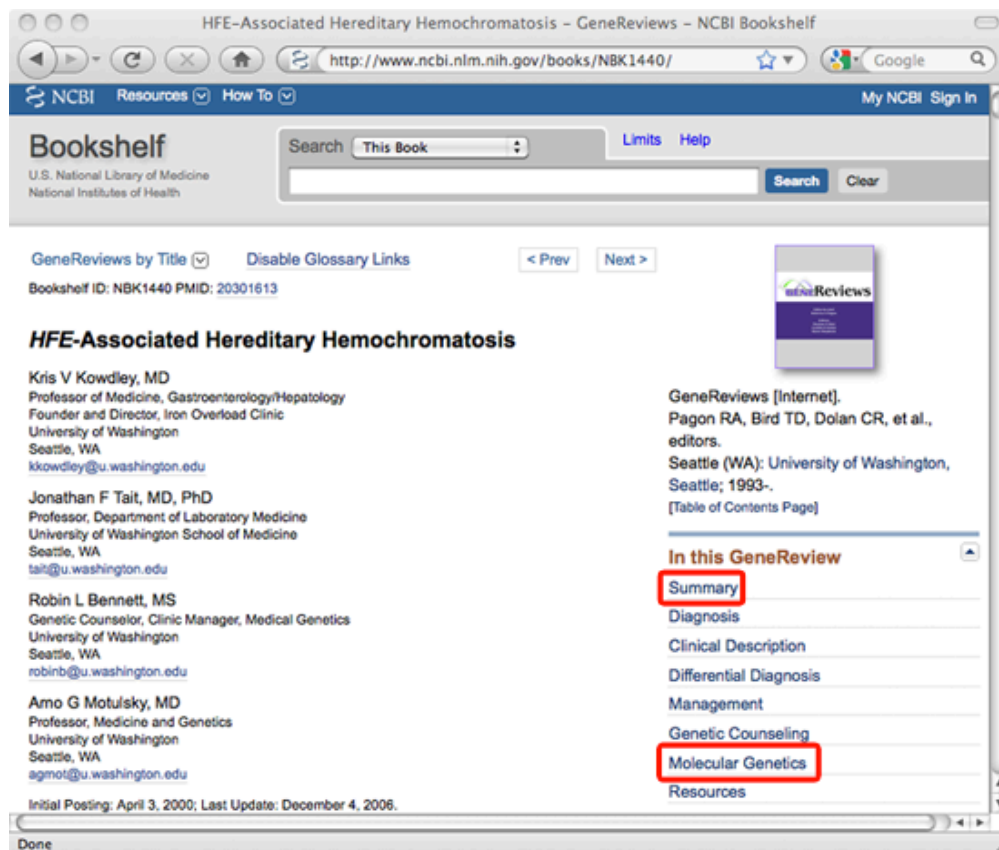
- Now you are ready to answer Questions 1–2 for Activity 1 in the worksheet on page 51.
- Scroll to the top of this OMIM record, and click on the **Limits** tab. Let's use options on the Limits page to determine how many genes in the human genome have been described in OMIM.
 - Uncheck the boxes for **Title** and chromosome **6**.
 - Check the boxes beside the MIM Number Prefix options for *** gene with known sequence** and **+ gene with known sequence and phenotype** as shown in the screenshot on the next page.
 - Then click the **Go** button beside the search box at the top of the page.



1. Click on **GeneReviews** in the navigation bar at the top.
2. At the GeneReviews search page (shown below), use the **Gene Symbol** search option, select **exactly matches** from the drop-down menu, and enter **HFE** into the search box. Click **Go** to submit your search.



3. Beside the search result “*HFE*-Associated Hereditary Hemochromatosis,” select the [Reviews](#) link to access the hereditary hemochromatosis review shown below.



The screenshot shows a web browser window displaying the NCBI Bookshelf page for "HFE-Associated Hereditary Hemochromatosis". The page includes the following information:

- Bookshelf ID:** NBK1440 PMID: 20301613
- Authors:**
 - Kris V Kowdley, MD: Professor of Medicine, Gastroenterology/Hepatology, Founder and Director, Iron Overload Clinic, University of Washington, Seattle, WA. Email: kkwowdley@u.washington.edu
 - Jonathan F Tait, MD, PhD: Professor, Department of Laboratory Medicine, University of Washington School of Medicine, Seattle, WA. Email: jtait@u.washington.edu
 - Robin L Bennett, MS: Genetic Counselor, Clinic Manager, Medical Genetics, University of Washington, Seattle, WA. Email: robinb@u.washington.edu
 - Arno G Motulsky, MD: Professor, Medicine and Genetics, University of Washington, Seattle, WA. Email: agmot@u.washington.edu
- Editors:** Pagon RA, Bird TD, Dolan CR, et al. (Seattle (WA): University of Washington, Seattle; 1993-). [Table of Contents Page]
- Initial Posting:** April 3, 2000; Last Update: December 4, 2006.
- Navigation Column (Right Side):**
 - Summary (highlighted with a red box)
 - Diagnosis
 - Clinical Description
 - Differential Diagnosis
 - Management
 - Genetic Counseling
 - Molecular Genetics (highlighted with a red box)
 - Resources

4. On the right side of the screen is a navigation column with links to different sections of the *HFE*-Associated Hereditary Hemochromatosis GeneReview.
5. Access the **Summary** section to learn about disease characteristics and treatment for hemochromatosis. This section can help answer Question 3 for Activity 1 in the worksheet on page 51.
6. Access the **Molecular Genetics** section for a brief overview of this disorder’s molecular basis. Within this section you can find information about:
- official symbol for the gene associated with this disorder.
 - chromosomal locus of the gene.
 - gene size and the number of exons in the gene.
 - name of the gene’s protein product.
 - description of the protein’s function.
 - mutations in nucleotide and amino acid sequences that cause abnormal protein products and disease phenotypes.
 - links to scientific literature and other databases for more information.

Activity 2

Online Resource: NCBI Map Viewer

- Find the hereditary hemochromatosis gene on a chromosome map.

NCBI Map Viewer

NCBI Map Viewer is a Web-based tool for viewing and searching an organism's complete genome. Users also can view maps of individual chromosomes and zoom in to specific regions within chromosomes to explore the genome at the sequence level.

Map Viewer provides access to several different types of maps for different organisms. Many of these maps are meaningful only to scientific researchers. A discussion of all the different types of maps and genomic data is beyond the scope of this activity, which will focus only on how to locate a specific gene locus on a chromosome map.

- Go to the NCBI Map Viewer website (www.ncbi.nlm.nih.gov/mapview/). In the list of **Primates**, click on the **Build 37.2** link for *Homo sapiens* (human).

The screenshot shows the NCBI Map Viewer interface. The main content area displays a hierarchical list of organisms and their genome builds. The 'Primates' section is expanded, showing a table of organisms with their scientific names, common names, and available genome builds. The 'Build 37.2' link for *Homo sapiens* (human) is highlighted with a red box.

Scientific name	Common name	Build	Tools
<i>Callithrix jacchus</i>	white-tufted-ear marmoset	Build 1.1	Q B R
<i>Homo sapiens</i>	human	Build 37.2	Q B R Cf G
		Build 36.3	Q B R Cf
<i>Macaca mulatta</i>	rhesus macaque	Build 1.2	Q B R G
<i>Pan troglodytes</i>	chimpanzee	Build 2.1	Q B R G
<i>Pongo abelii</i>	Sumatran orangutan	Build 1.2	Q B R

2. The Map Viewer page for the entire human genome is shown in the screenshot below.

The screenshot shows the NCBI Map Viewer interface. At the top, there's a navigation bar with links for PubMed, Nucleotide, Protein, Genome, Gene, Structure, PopSet, Taxonomy, and Help. Below this is a search bar with the text "Search for" and a dropdown menu set to "on chromosome(s)". The main content area displays "Homo sapiens (human) genome view" with a karyotype of human chromosomes. A sidebar on the left contains "Map Viewer" and "NCBI Resources" links. A news banner at the bottom of the main content area mentions a November 2010 update to the human genome reference assembly.

Homo sapiens genome view: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606

3. In Activity 1, we learned that the official symbol for the hereditary hemochromatosis gene is HFE, and its locus is 6p21.3. Let's find the HFE gene on chromosome 6.

What is a locus?

The locus for a particular gene describes the region of a chromosome where that gene can be found. For the **6p21.3** locus: **6** is the chromosome number, **p** indicates the short arm of the chromosome, and **21.3** is the number assigned to a particular band or region on a chromosome. When chromosomes are stained in the lab, light and dark bands appear, and each band is numbered. The higher the number, the farther away the band is from the centromere. A locus containing **q** is found on the long arm of a chromosome.

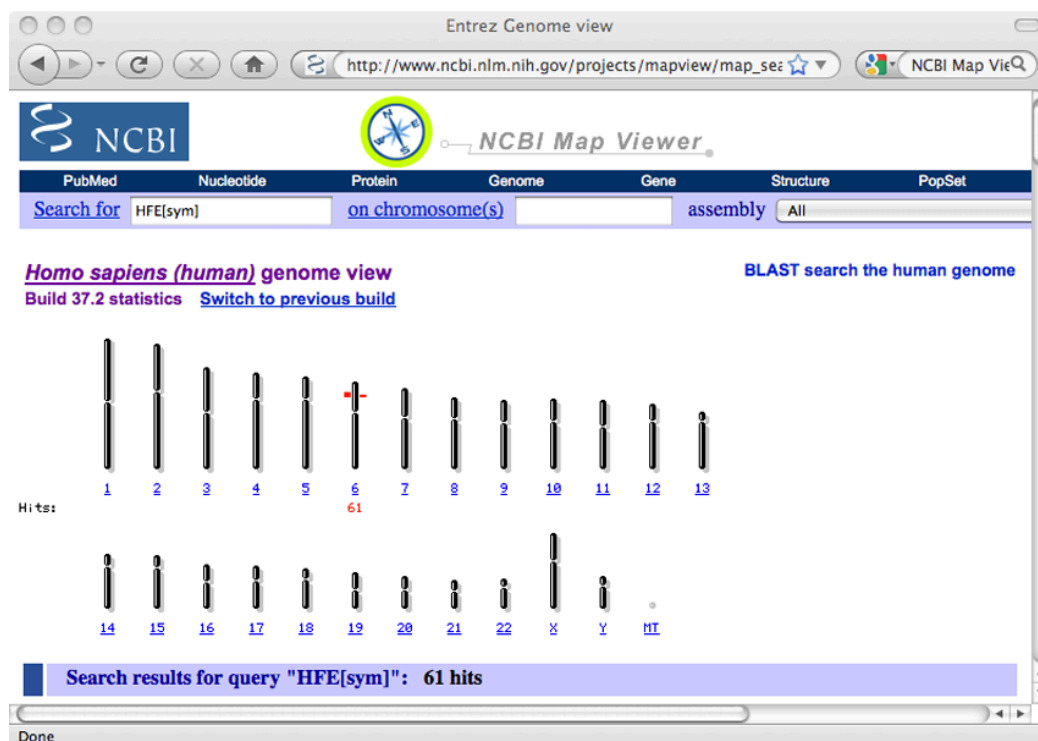
Short and Long Arms of a Chromosome

The diagram shows a vertical chromosome with a central centromere. The upper portion is labeled 'p short arm' and the lower portion is labeled 'q long arm'. A red band is shown on the short arm, and a blue band is shown on the long arm.

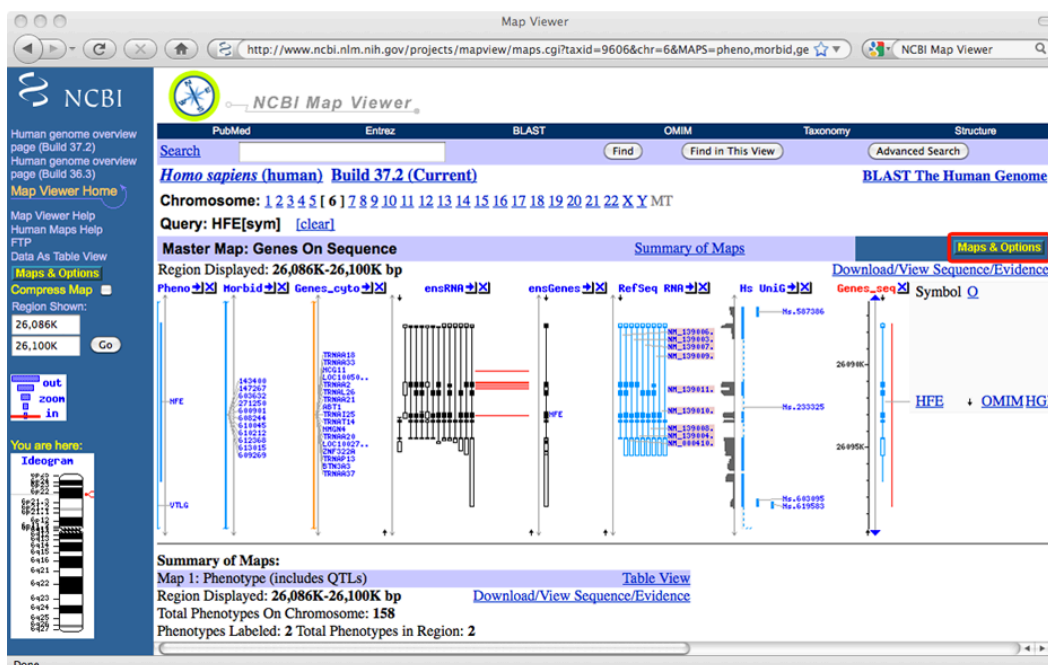
4. In the search box at the top of the Map Viewer page, enter **HFE[sym]** and then click the **Find** button to submit your search. Adding the [sym] search field qualifier to the end of your search term specifies your query so that only those results containing the HFE gene are retrieved.

This close-up shows the search bar at the top of the Map Viewer page. The search term "HFE[sym]" is entered in the search box. The "Find" button is highlighted with a red box. The dropdown menu is set to "on chromosome(s)".

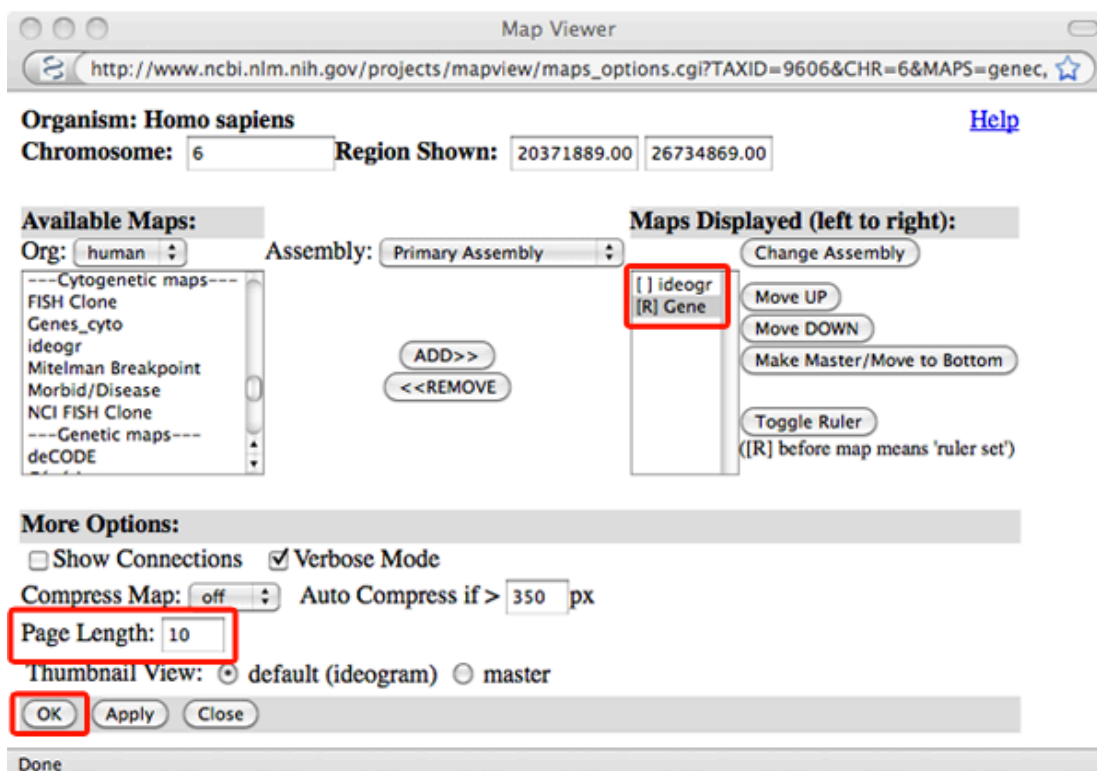
- Red tick marks should be displayed on chromosome 6, indicating the approximate location of the HFE gene in the middle of the short arm of chromosome 6 (see screenshot below). The red number ("61") labeling chromosome 6 indicates the number of objects mapped to different assemblies of the human genome that include the HFE gene.



- Click on the number [6](#) link below the chromosome. This will open a view of chromosome 6 that should look like the screenshot below. In the next step we will modify this view so we can see an ideogram showing the region of chromosome 6 where the HFE gene can be found.



7. To modify the display options, click on the **Maps & Options** button in the upper right corner. This will open a window for customizing map options. Make the following adjustments.
 - Remove all maps listed under **Maps Displayed (left to right)** except the **Gene** map. To remove a map, select it with your mouse and then click the **REMOVE** button.
 - Under **Available Maps** select **ideogr** (you will need to scroll through more than half of the available maps) and then click the **ADD** button. The ideogram map is a graphic showing the banding pattern of a chromosome.
 - The **Maps Displayed** list should look like the screenshot below. The **Gene** map should be designated as your master map. To make a map the master, select it with your mouse and then click the **Make Master/Move to Bottom** button. In the chromosome view, a master map is shown at the right side of the screen along with its details and descriptive text. The **Gene** map includes links for learning more about the genes mapped to a particular region of genomic sequence on a chromosome.
 - Under **More Options** near the bottom of the window, change **Page Length** from 30 to 10. The Page Length option is highlighted in the screenshot below. This will adjust the height of the displayed map.
 - Before you click the **OK** button to submit your changes, the options window should resemble the screenshot below.



8. The new map of chromosome 6 should resemble the screenshot on the next page.

The screenshot displays the NCBI Map Viewer interface. The browser address bar shows the URL: <http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&CHR=6&BE>. The page title is "Map Viewer". The main content area shows "Homo sapiens (human) Build 37.2 (Current)" and "BLAST The Human Genome". A search query "HFE[sym]" is entered. The "Region Displayed" is "26,086K-26,100K bp". The "Master Map: Genes On Sequence" is selected, and the "Maps & Options" button is highlighted. The map shows the HFE gene and its association with OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq 6p21.3 hemochromatosis. A summary of maps section shows "Map 1: Ideogram" and "Map 2: Genes On Sequence" with "Region Displayed: 26,086K-26,100K bp" and "Total Genes On Chromosome: 2054".

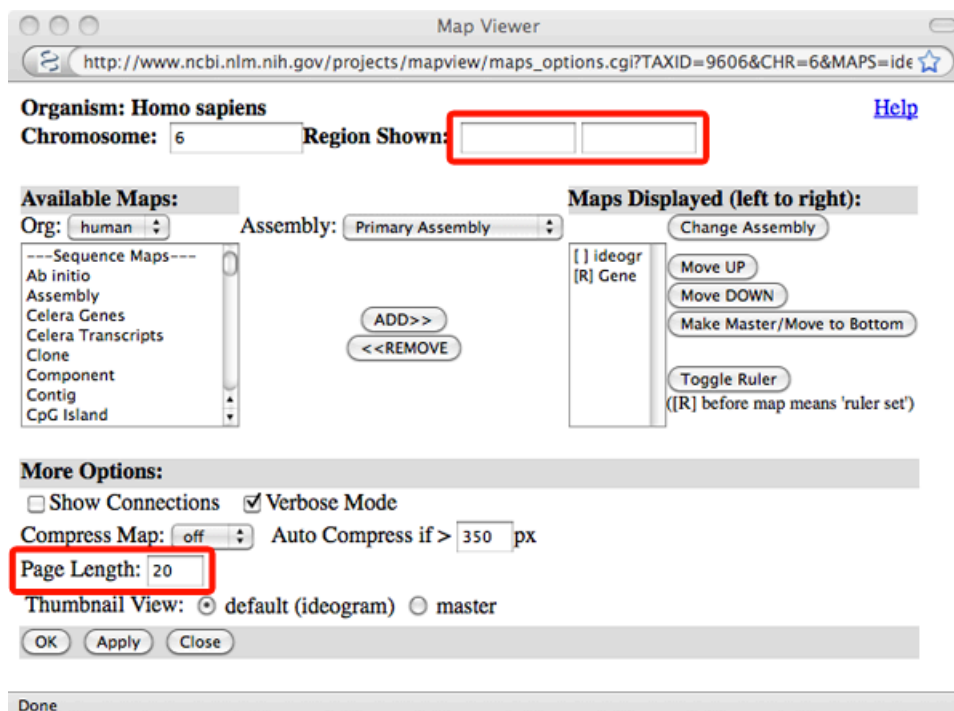
9. Check out some of Map Viewer's features displayed in the screenshot above.

- The portion of chromosome 6 displayed in Map Viewer is highlighted on the ideogram in the blue navigation column on the left. Notice that the red mark indicating the position of the HFE gene lines up with the ideogram at the 6p22 chromosome band, not 6p21.3.
- Rounded to the nearest thousandth, the region of sequence displayed begins at about the 26,086,000th nucleotide and ends at about the 26,100,000th nucleotide of the DNA sequence of chromosome 6. The total DNA sequence for chromosome 6 is about 171 million base pairs long, but this view only shows about 14,000 base pairs.
- Clicking on the [Ideogram](#) or [Genes_seq](#) maps (not the labels) will open a pop-up window with options for zooming in or out on the displayed maps. Map Viewer has zoomed in so much to show the HFE gene, there isn't much of the ideogram map displayed. You can also zoom in and out using the zoom option in the blue navigation column.
- The [Genes_seq](#) map provides links to gene-specific entries in other NCBI databases.
 - [HFE](#) – Links to the HFE entry in the Entrez Gene database, a compendium of genes and mapped phenotypes.
 - [OMIM](#) – Links to the hemochromatosis entry in the Online Mendelian Inheritance in Man (OMIM) database covered in Activity 1.

- [HGNC](#) – Links to the gene symbol report maintained by the HUGO Gene Nomenclature Committee.
- [sv](#) – Links to Sequence Viewer, a graphical interface for investigating the gene's sequence as well as genomic sequence upstream and downstream of the gene.
- [pr](#) – Links to sequence records for the gene's protein product maintained in NCBI's Protein database.
- [dl](#) – Links to a page for downloading the range of sequence data displayed in Map Viewer.
- [ev](#) – Links to Evidence Viewer, a tool for finding biological evidence that supports a particular gene model and for exploring the different types of expressed sequences that align to a particular area within a genome.
- [mm](#) – Links to Model Maker, a tool for building your own version of a gene model by adding or removing exons.
- [hm](#) – Links to Homologene, a resource for comparing genes in homologous segments of DNA from different organisms.
- [sts](#) – Links to UniSTS, a comprehensive database that integrates genetic marker and mapping information. A sequence tagged site (STS) is a short (200 to 500 base pairs) DNA sequence that has a single occurrence in the human genome. Detectable by polymerase chain reaction (PCR), STSs are useful for localizing and orienting the sequence data reported from many different laboratories.
- [CCDS](#) – Links to the CCDS project, an effort to ensure that coding regions within the human genome are consistently annotated.
- [SNP](#) – Links to records for single nucleotide polymorphisms (SNPs) and other areas of sequence variation that have been identified in the selected gene.

10. Let's zoom out to view the entire chromosome using the **Maps & Options** window.

- Click on **Maps & Options** again to open the options window.
- Delete the numbers defining the **Region Shown** at the top of the options window. This will modify the display so it shows the entire chromosome.
- Under **More Options** near the bottom of the window, change **Page Length** from 10 to 20. The Page Length option is highlighted in the screenshot on the next page. This will display 20 labeled genes in the master map and should provide enough space on the screen to view the entire chromosome with readable labels for the chromosome bands.
- Once the Maps & Options window resembles the screenshot on the following page, click the **OK** button to submit your changes.



11. Your view of chromosome 6 should resemble the screenshot on the next page.

- To see a more comprehensive listing of genes on chromosome 6, select the **Data As Table View** link in the blue navigation column on the left. The **Data As Table View** displays 1,000 of the genes on chromosome 6 and shows where genes start and stop in the chromosome's DNA sequence.
- Scroll down to the bottom of the map to examine the **Summary of Maps** section. Use this information and what you have learned about Map Viewer to answer the Questions for Activity 2 on page 51.

Map Viewer

http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&CHR=6&BEG=&END=&MAP0=ideog

NCBI Map Viewer

Human genome overview page (Build 37.2)

Human genome overview page (Build 36.3)

Map Viewer Home

Map Viewer Help

Human Maps Help

FTP

Data As Table View

Maps & Options

Compress Map

Region Shown:

Go

out

zoom

in

You are here:

Ideogram

default

master

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Find Find in This View Advanced Search

Homo sapiens (human) Build 37.2 (Current) [BLAST The Human Genome](#)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) **[6](#)** [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [MT](#)

Query: HFE[sym] [\[clear\]](#)

Master Map: Genes On Sequence [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 0-171M bp [Download/View Sequence/Evidence](#)

Ideogram	Genes_seq	Symbol	Q	Links	E	Cyto	Description
		RIPK1	+	OMIM HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p25.2 receptor (TNFRSF)-interacti
		TUBB2A	+	HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p25 tubulin, beta 2A
		F13A1	+	OMIM HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p25.3-p24.3 coagulation factor XIII, A1 p
		RPL21P63	+	HGNC sv dl ev mm		best RefSeq	6p24.2 ribosomal protein L21 pseud
		OR2B6	+	HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p21.3 olfactory receptor, family 2, s
		ZKSCAN4	+	OMIM HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p21 zinc finger with KRAB and S
		UBD	+	OMIM HGNC sv pr dl ev mm hm sts	SNP	best RefSeq	6p21.3 ubiquitin D
		PPP1R10	+	OMIM HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p21.3 protein phosphatase 1, regula
		HSPA1A	+	OMIM HGNC sv pr dl ev mm hm sts	SNP	best RefSeq	6p21.3 heat shock 70kDa protein 1A
		ZBTB9	+	HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p21.32 zinc finger and BTB domain
		LOC285847	+	sv dl ev mm sts		best RefSeq	6p21.31 hypothetical LOC285847
		TFAP2D	+	OMIM HGNC sv pr dl ev mm hm sts	CCDS SNP	best RefSeq	6p12.1 transcription factor AP-2 del
		RPL35P3	+	HGNC sv dl ev mm		best RefSeq	6q21 ribosomal protein L35 pseud
		ATP5LP2	+	HGNC sv dl ev mm		best RefSeq	6q22.31 ATP synthase, H+ transportir
		LOC100130476	+	sv pr dl ev mm sts	SNP	mRNA	6q23.3 similar to hCG2036711
		PBOV1	+	OMIM HGNC sv pr dl ev mm sts	SNP	best RefSeq	6q23-q24 prostate and breast cancer ov
		STXBPS	+	OMIM HGNC sv pr dl ev mm hm sts	SNP	best RefSeq	6q24.3 syntaxin binding protein 5 (to
		LOC100420477	+	sv dl ev mm		best RefSeq	6q25.2 lactate dehydrogenase A-like
		RPL21P69	+	HGNC sv dl ev mm		best RefSeq	6q25.3 ribosomal protein L21 pseud
		LOC441177	+	sv pr dl ev mm		best RefSeq	6q27 hypothetical LOC441177

Summary of Maps:

Map 1: Ideogram
Region Displayed: 6pter-6qter

Map 2: Genes On Sequence [Table View](#)
Region Displayed: 0-171M bp [Download/View Sequence/Evidence](#)

Total Genes On Chromosome: 2054
Genes Labeled: 20 Total Genes in Region: 2054

Done

Activity 3

Online Resources: NCBI Entrez Gene and GenBank

- Examine gene sequence and structure.

NCBI Entrez Gene and GenBank

Entrez Gene is an NCBI resource that serves as a single-query interface for accessing sequence and other biological information for specific genes from a variety of sequenced organisms. GenBank is NCBI's comprehensive repository of annotated DNA sequences.

This activity covers how to use Entrez Gene to access the genomic DNA sequence of the hereditary hemochromatosis (HFE) gene. Then by examining some different features of a GenBank record for the HFE gene, we will learn about the gene's structure (e.g., intron and exon composition, coding sequence).

1. To begin, let's go to the Entrez Gene home page (www.ncbi.nlm.nih.gov/gene). In the search box at the top, enter **HFE[sym] AND human[orgn]** as shown in the screenshot below. Be sure to capitalize any Boolean operator (AND, OR, and NOT) included in your search statements. Then submit your search.

The screenshot shows the NCBI Entrez Gene homepage. The browser address bar displays <http://www.ncbi.nlm.nih.gov/gene>. The search bar contains the query **HFE[sym] AND human[orgn]**, which is highlighted with a red box. Below the search bar, there is a "Welcome to Gene" banner with the text "Gene maintains information about genes from genomes of interest to the RefSeq group." The page is organized into three columns of links: "Using Gene" (Gene Quick Start, FAQ, Download/FTP, RefSeq Mailing List, Gene News), "Gene Tools" (Submit GeneRIFs, Submit Correction, Statistics, BLAST, Genome Workbench, Splotn), and "Other Resources" (HomoloGene, OMIM, RefSeq, RefSeqGene, UniGene, Protein Clusters). At the bottom, there is a navigation bar with categories: GETTING STARTED (NCBI Education), RESOURCES (Chemicals & Bioassays), POPULAR (PubMed), FEATURED (GenBank), and NCBI INFORMATION (About NCBI).

Search Tip: Adding [**sym**] to the end of your query term tells Entrez Gene that you are searching by gene symbol only. If you do not specify that you want to search the gene symbol field, the search will return multiple records that include the query term anywhere within a record's content. Adding [**orgn**] to a search term limits the search to genes from a specific organism. For more information on options for refining your search, see the Search Field Descriptions and Qualifiers section of the Entrez Help Document (www.ncbi.nlm.nih.gov/entrez/query/static/help/Summary_Matrices.html).

2. Submitting this search should retrieve a single result. The HFE record is shown below.

The screenshot displays the NCBI Entrez Gene record for HFE hemochromatosis [Homo sapiens]. The search query used is "HFE[sym] AND human[orgn]". The record includes the following information:

- Official Symbol:** HFE provided by HGNC
- Official Full Name:** hemochromatosis provided by HGNC
- Primary source:** HGNC:4886
- See related:** Ensembl:ENSG0000010704; HPRD:01993; MIM:613609
- Gene type:** protein coding
- RefSeq 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; MVCD7; MGC103790; dJ221C16.10.1; HFE
- 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 (beta2M). 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 haemochromatosis, is a recessive genetic disorder that results from defects in this gene. At least nine alternatively spliced variants have been described for this gene. Additional variants have been found but their full-length nature has not been determined. [provided by RefSeq]

The genomic regions, transcripts, and products section shows the genomic sequence on chromosome 6 (NC_000006) and a gene map with the HFE gene highlighted. The map shows the gene structure with exons and introns, and the HFE gene is located on the positive strand between approximately 26,086,004 and 26,097,430 bases.

3. In the **Summary** section you can find information about the function of the gene's protein product. The HFE protein is thought to have a role in regulating iron transport into cells, and defects in the HFE gene can cause the iron absorption disorder hereditary hemochromatosis. Use information provided in the **Summary** section to answer Question 1 for Activity 3 in the worksheet on page 52.

4. Below the summary section is the **Genomic regions, transcripts, and products** section.
- The sequence viewer box shows a graphic model of the HFE gene consisting of a thin gray line (representing introns that are removed when the mRNA is processed) connected to thicker green boxes (representing exons).
 - The portion of the chromosome 6 sequence included in the sequence viewer box is noted in the upper left corner.
 - Click on the **GenBank** link in the upper right corner to access the GenBank record for the HFE gene sequence that is part of the sequence data generated by the International Human Genome Project. A screenshot of this GenBank record is shown on the next page.

HFE hemochromatosis [Homo sapiens] - Gene result

http://www.ncbi.nlm.nih.gov/gene?term=HFE[sym] AND human[orgn]

Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence NC_000006 chromosome

Go to nucleotide [Graphics](#) [FASTA](#) [GenBank](#)

26,085,991 : 26,096,958 10,968 bases shown, positive strand [Open Full View](#)

Flip Strands [Search](#) [ATG](#) [EEM](#) [Configure](#)

86 K 26,088 K 26,090 K 26,092 K 26,094 K 26,096 K

- Genes

HFE

- SNP

- LSDB or Clinically Associated Variants

- Cited Variants

- NHGRI GWAS Catalog

Genomic context

chromosome: 6; Location: 6p21.3 [See HFE in MapViewer](#)

Chromosome 6 - NC_000006.11

[26445639] HIST1H3C HIST1H1C HFE HIST1H1T [26148364] HIST1H4C

Links to other resources

- [AceView](#)
- [Ensembl](#)
- [Evidence Viewer](#)
- [GeneTests for MIM: 235200](#)
- [GeneTests for MIM: 613609](#)
- [HFE](#)

5. Check out some of the following features in the GenBank record.
 - At the top we see that only a very small portion of chromosome 6 (from 26,087,448th base to 26,097,059th base) is included in this record.
 - The first **Reference** listed for this record identifies the “International Human Genome Sequencing Consortium” as the source for this sequence information. Thus this sequence is a product of the international Human Genome Project.
 - Even after a genome sequence is published in a journal and reported as “complete,” the research community continues to analyze the genome sequence data and improve the annotation that describes different features encoded within the genome sequence. Note that this record was last modified October 25, 2010.

Homo sapiens chromosome 6, GRCh37.p2 primary refer... - Nucleotide result

http://www.ncbi.nlm.nih.gov/nucleotide/NC_000006?report=genbank&from=2608744

NCBI Resources How To My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear

Display Settings: GenBank Send:

Showing 9.61kb region from base 26087448 to 26097059.

Homo sapiens chromosome 6, GRCh37.p2 primary reference assembly

NCBI Reference Sequence: NC_000006.11

[FASTA](#) [Graphics](#)

Go to:

LOCUS NC_000006 9612 bp DNA linear **CON 25-OCT-2010**

DEFINITION Homo sapiens chromosome 6, GRCh37.p2 primary reference assembly.

ACCESSION [NC_000006](#) REGION: 26087448..26097059 GPC_000000030

VERSION NC_000006.11 GI:224589818

DBLINK Project: [168](#)

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 9612)

CONSRM International Human Genome Sequencing Consortium

TITLE Finishing the euchromatic sequence of the human genome

JOURNAL Nature 431 (7011), 931-945 (2004)

PUBMED [15496913](#)

REFERENCE 2 (bases 1 to 9612)

Done

Change region shown

Whole sequence (abbreviated view)

Selected region

from: 26087448 to: 26097059

Update View

Customize view

Analyze this sequence

Run BLAST

Pick Primers

LinkOut to external resources

UCSC Genome Browser [UCSC Genome Browser]

All links from this record

BioAssay by RNA target

Components (Core)

Full text in PMC

6. Scroll down to the **FEATURES** section of this GenBank record (see screenshot on next page).
 - The HFE gene is 9,612 base pairs (bp) long.
 - The information in this GenBank record for the HFE gene was “Derived by automated computational analysis using gene prediction method” as a part of the Human Genome Project.
 - From the multiple entries for “mRNA” listed in this record, we see that more than one mRNA transcript can be generated from the HFE gene. For example, an exon included in one mRNA transcript might be left out in another transcript. Each of these different mRNA transcripts from the same gene is known as a “variant.”

Homo sapiens chromosome 6, GRCh37.p2 primary refer... - Nucleotide result

http://www.ncbi.nlm.nih.gov/nuccore/NC_000006?report=genbank&fr

FEATURES

FEATURES	Location/Qualifiers
source	1..9612 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="6"
gene	1..9612 /gene="HFE" /gene_synonym="dJ221C16.10.1; HFE1; HH; HLA-H; MGC103790; MVCD7" /note="Derived by automated computational analysis using gene prediction method: Curated Genomic." /db_xref="GeneID:3077" /db_xref="HGNC:4886" /db_xref="MIM:235200"
mRNA	join(62..297,3622..3885,4095..4370,5466..5741,5900..6013,6967..8022) /gene="HFE" /gene_synonym="dJ221C16.10.1; HFE1; HH; HLA-H; MGC103790; MVCD7" /product="hemochromatosis, transcript variant 1" /exception="unclassified transcription discrepancy" /note="Derived by automated computational analysis using gene prediction method: Curated Genomic." /transcript_id="NM_000410.3" /db_xref="GI:91718876" /db_xref="GeneID:3077" /db_xref="HGNC:4886" /db_xref="MIM:235200"
mRNA	join(62..297,3622..3885,5466..5741,5900..6013,6967..8022) /gene="HFE" /gene_synonym="dJ221C16.10.1; HFE1; HH; HLA-H; MGC103790; MVCD7" /product="hemochromatosis, transcript variant 4" /note="Derived by automated computational analysis using gene prediction method: Curated Genomic." /transcript_id="NM_139004.2" /db_xref="GI:91718878" /db_xref="GeneID:3077" /db_xref="HGNC:4886" /db_xref="MIM:235200"
mRNA	join(62..297,3622..3885,5508..5741,5900..6013,6967..8022) /gene="HFE" /gene_synonym="dJ221C16.10.1; HFE1; HH; HLA-H; MGC103790; MVCD7" /product="hemochromatosis, transcript variant 3" /note="Derived by automated computational analysis using gene prediction method: Curated Genomic." /transcript_id="NM_139003.2" /db_xref="GI:91718877" /db_xref="GeneID:3077" /db_xref="HGNC:4886" /db_xref="MIM:235200"
mRNA	join(62..297,3691..3885,4095..4370,5466..5741,5900..6013,6967..8022) /gene="HFE"

Done

- Use your browser's "back" button to return to the Entrez Gene page for the human HFE gene.
- Let's access another GenBank record for the HFE gene sequence to see how information can vary in records that come from different sources. As shown in the screenshot on the next page, select the **Related sequences** link in the **Table of contents** box on the right side of the screen.

The screenshot shows the NCBI Gene Gateway interface for the HFE hemochromatosis gene in Homo sapiens. The search bar contains 'HFE[sym] AND human[orgn]'. The 'Table of contents' on the right side lists various categories, with 'Related sequences' highlighted in a red box. The 'Summary' section on the left provides details about the gene, including its official symbol, full name, primary source, and a description of its function and associated disorders.

9. In the **Related Sequences** for the HFE gene (see screenshot on the next page), select the genomic sequence record [Z92910.1](#).

How did you know which genomic sequence to select?

The problem with archival sequence databases like NCBI's GenBank is that they usually have multiple sequence records for the same gene. You may need to open each record individually and browse through definition, sequence annotation, and comments to determine how much of the gene's nucleotide sequence is contained within each record.

For example, the [U91328.1](#) record contains the sequence of a genomic segment that not only includes the HFE gene sequence but also sequences for other genes. [Y09801.1](#) contains only sequence information for the HFE promoter and the HFE gene's first exon. Of the genomic records listed, [Z92910.1](#) has the most complete sequence information for the HFE gene.

In sequence databases such as GenBank, "genomic" DNA sequence records for eukaryotic organisms contain both exons and introns, while "mRNA" sequences are intron-free DNA sequences. All sequences in GenBank and similar repositories use the single-letter abbreviations for the DNA bases adenine (A), cytosine (C), guanine (G), and thymine (T) to represent each nucleotide. Even "mRNA" sequence records use A, C, G, and T where T is used to replace each uracil (U) in the mRNA sequence.

Related Sequences for the HFE Gene in Entrez Gene

Heading	Nucleotide	
	Accession and Version	
genomic	AF184234.1	AAF01222.1
genomic	AF204869.1	None
genomic	AF331065.1	AAK16502.1
genomic	AF525359.1	AAM82608.1
genomic	AF525499.1	AAM91950.1
genomic	CH471087.1	EAW55516.1
		EAW55517.1
		EAW55518.1
		EAW55519.1
		EAW55520.1
		EAW55521.1
		EAW55522.1
		EAW55523.1
		EAW55524.1
		EAW55525.1
		EAW55526.1
		EAW55527.1
genomic	CS187189.1	CA_42862.1
genomic	EU523119.1	ACB21042.1
genomic	U80914.1	AAD00449.1
genomic	U91328.1	AAB82083.1
genomic	Y09801.1	CAA70934.1
genomic	Z92910.1	CAB07442.1
mRNA	AF079407.1	AAC62646.1

10. A screenshot of the GenBank record [Z92910.1](#) for the HFE gene is shown on the next page.

- The DNA sequence included in this record is 12,146 base pairs (bp) long. In addition to containing the genomic sequence of the HFE gene, this record also contains several hundred additional base pairs of sequence upstream and downstream of the gene.
- This record was originally submitted by a researcher to GenBank in 1997, so the sequence of the HFE gene was known several years before the Human Genome Project was complete.
- Scroll down to the **FEATURES** section of this record and use this information to answer Questions 2–4 for Activity 3 on page 52. Note that clicking on the [gene](#) link in the FEATURES section shows that the length of the HFE gene is different from what we observed in the GenBank record examined in step 5 of this activity.

GenBank Record Z92910.1 for Human HFE Gene

Homo sapiens HFE gene - Nucleotide result
 http://www.ncbi.nlm.nih.gov/nuccore/1890179

NCBI Resources How To My NCBI Sign In

Nucleotide
 Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear

Display Settings: GenBank Send: Change region shown Customize view

Homo sapiens HFE gene

GenBank: Z92910.1
[FASTA](#) [Graphics](#)

Go to:

LOCUS Z92910 **12146 bp** DNA linear PRI 23-OCT-2008

DEFINITION Homo sapiens HFE gene.

ACCESSION Z92910

VERSION Z92910.1 GI:1890179

KEYWORDS haemochromatosis; HFE gene.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 858)
 AUTHORS Albig,W., Drabent,B., Burmester,N., Bode,C. and Doenecke,D.
 TITLE The haemochromatosis candidate gene HFE (HLA-H) of man and mouse is located in syntenic regions within the histone gene cluster
 JOURNAL J. Cell. Biochem. 69 (2), 117-126 (1998)
 PUBMED [9548560](#)

REFERENCE 2 (bases 1 to 12146)
 AUTHORS Albig,W.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAR-1997) Albig W., Georg-August-Universitaet Goettingen, Biochemie und Molekulare Zellbiologie, Humboldtallee 23, Goettingen, FRG, 37073

FEATURES

source Location/Qualifiers
 1..12146
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:[9606](#)"
 /chromosome="6"
 /map="6p"
 /clone="ICRFy901D1223"
 /clone_lib="ICRF YAC-library"

[gene](#) 1028..10637
 /gene="HFE"

[exon](#) 1028..1324
 /gene="HFE"
 /number=1

[CDS](#) join(1249..1324,4652..4915,5125..5400,6494..6769,6928..7041,7995..8035)
 /gene="HFE"
 /function="iron metabolism"
 /note="haemochromatosis candidate gene"
 /codon_start=1
 /protein_id="CAB07442.1"

Done

Analyze this sequence
 Run BLAST
 Pick Primers
 Find in this Sequence

Articles about the HFE gene
 Identification of a common variant in the TFR2 gene implicated in the phy [Hum Mol Genet. 2011]
 Novel association to the proprotein convertase PCSK7 gene locus reveal [Hum Mol Genet. 2011]
 Decreased clearance of CNS beta-amyloid in Alzheimer's disease. [Science. 2010]
 See all...

Reference sequence information
 RefSeq alternative splicing
 See 9 reference mRNA sequence splice variants for the HFE gene.

More about the HFE gene
 The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin ...
 Also Known As: HFE1, HH, HLA-H, MGC103...

Homologs of the HFE gene
 The HFE gene is conserved in chimpanzee, cow, and mouse.

LinkOut to external resources
 UCSC Genome Browser [UCSC Genome Browser]

11. Some features of the sequence in GenBank record Z92910.1 include

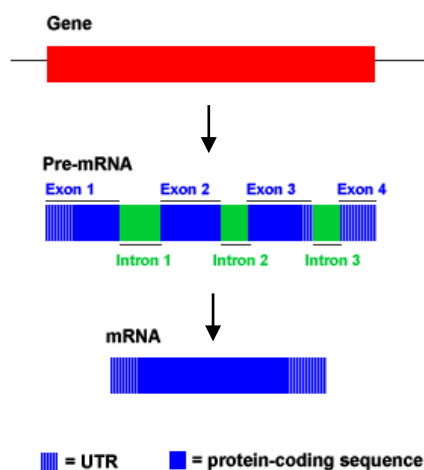
- **source:** Required for every GenBank record, the source provides the entire sequence length and the scientific name of the source organism. Other types of source information may include chromosome number, map location, and clone or strain identification.
- **gene:** This feature provides nucleotide numbers indicating where the gene stops and starts. **This link opens a new sequence record that shows only the gene sequence.**
- **exon:** This feature provides nucleotide numbers indicating where each exon begins and ends. You will see several of these entries as you scroll down. Each exon is a sequence segment that codes for a portion of processed (intron-free) mRNA. The name of the gene

to which the exon belongs and the exon number are provided. An “exon” link opens a new sequence record that shows only the exon sequence.

- **CDS:** The coding sequence (CDS) consists of nucleotides that actually code for amino acids of the protein product. This feature includes the coding sequence's amino acid translation and may also contain gene name, gene product function, a link to protein sequence record, and cross-references to other database entries. A “CDS” link opens a new sequence record that shows only the coding sequence.
- **intron:** This feature provides the nucleotide numbers indicating where each intron begins and ends. An intron is a segment of noncoding sequence that is transcribed but removed from the transcript by splicing together the exons on either side of it. An “intron” link opens a new sequence record that shows only the intron sequence.

What's the difference between exons and coding sequence?

Exons often are described as short segments of protein coding sequence. This is a bit of an oversimplification. Exons are segments of sequence spliced together after introns have been removed from pre-mRNA. Exons carry the coding sequence of a gene, but some exons may contain no coding sequence. Portions of exons or even entire exons may contain sequence that is not translated into amino acids. These are the untranslated regions (UTR) of mRNA. UTRs are found upstream and downstream of the protein-coding sequence. See diagram on right.



12. Sequence information in a GenBank record can also be displayed using graphics in the NCBI Sequence Viewer. To access Sequence Viewer from a GenBank record, click on the **Graphics** link in the upper left corner (as shown below).

NCBI Resources How To My NCBI Sign In

Nucleotide
Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear

Display Settings: GenBank Send

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Find in this Sequence

Articles about the HFE gene

Identification of a common variant in the TFR2

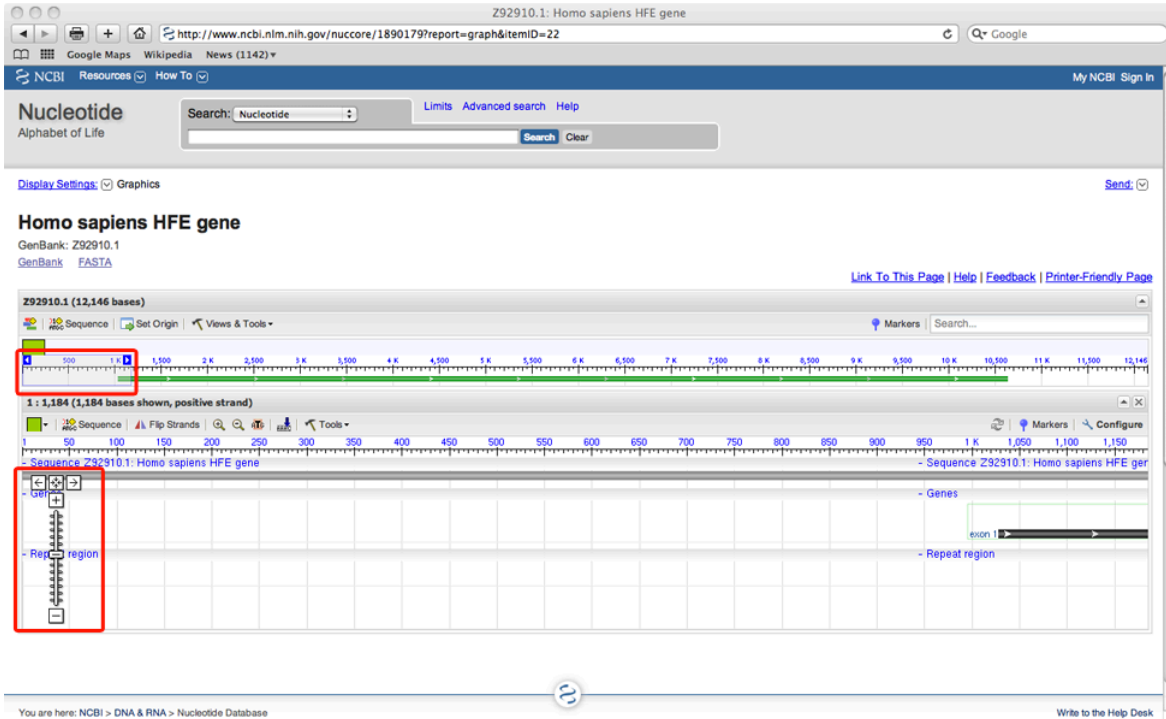
Homo sapiens HFE gene
GenBank: Z92910.1
FASTA Graphics

Go to:

LOCUS	Z92910	12146 bp	DNA	linear	PRI 23-OCT-2008
DEFINITION	Homo sapiens HFE gene.				
ACCESSION	Z92910				
VERSION	Z92910.1	GI:1890179			
KEYWORDS	haemochromatosis; HFE gene.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

13. The Sequence Viewer option for GenBank record Z92910.1 is shown in the screenshot below.

- The top panel displays the entire sequence included in the GenBank record, the green bar represents the HFE gene sequence, and the blue outline of a box with arrows indicates which portion of the sequence is shown in the panel below. Click and drag the arrows on the blue-box outline to change how much of the sequence is displayed in the lower panel.
- You can also use the arrows on the left side of the lower panel to move along the sequence and see where exons and other gene features begin and end. The slider below the arrows can be used to zoom in and out on the sequence.



Activity 4

Online Resources: UniProt Protein Knowledgebase and BLAST Searching

- Access the amino acid sequence of a gene's protein product.
- Compare the HFE protein sequence with protein sequences of other organisms.

UniProt Protein Knowledgebase and BLAST Searching

The Protein Knowledgebase, which is part of the Universal Protein Resource (UniProt), is a comprehensive, freely accessible database that the scientific community uses to access high-quality protein sequence and functional information. This activity covers how to use UniProt to learn about the amino acid sequence and other features of the hereditary hemochromatosis protein.

1. Go to the UniProt home page (www.uniprot.org), enter **HFE** into the query box as shown in the screenshot below, and then submit your search.

The screenshot shows the UniProt website search interface. The browser address bar displays <http://www.uniprot.org/>. The search bar is set to "Protein Knowledgebase (UniProtKB)" and contains the query "HFE". The "Search" button is highlighted with a red box. Below the search bar, the page content includes a "WELCOME" message, a "What we provide" table, a "Getting started" list, a "NEWS" section with a "UniProt release 2011_02 - Feb 8, 2011" announcement, a "SITE TOUR" section, and a "PROTEIN SPOTLIGHT" section with the title "the twisted way of things February 2011".

What we provide	Description
UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes Complete Proteome Sets .
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords and more .

- [Text search](#)
- [Sequence similarity searches \(BLAST\)](#)
- [Sequence alignments](#)
- [Batch retrieval](#)
- [Database identifier mapping \(ID Mapping\)](#)

2. From the list of results (shown in the screenshot on the next page), notice that some entries have gold stars and others have gray stars. Those with gold stars have descriptions of protein functions and characteristics that have been manually reviewed by experts. Entries with gray stars have descriptions that were automatically generated, and experts have not yet reviewed

these records. Thus selecting a search result with a gold star will provide you with richer, higher quality information about a protein.

- Select accession number **Q30201** for the HFE_HUMAN entry for the hereditary hemochromatosis protein.

HFE in UniProtKB

http://www.uniprot.org/uniprot/?query=HFE&sort=score

UniProtKB

Search Blast Align Retrieve ID Mapping *

Search in Query

Protein Knowledgebase (UniProtKB) HFE Search Clear Advanced Search »

1 - 25 of 74 results for HFE in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway Reduce sequence redundancy to 100%, 90% or 50% | Download

Page 1 of 3 | Next »

Results Customize

Show only reviewed (33) (UniProtKB/Swiss-Prot) or unreviewed (41) (UniProtKB/TrEMBL) entries

Restrict term "hfe" to gene name (42), protein name (10), strain (3), taxonomy (3)

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
Q30201	HFE_HUMAN	★	Hereditary hemochromatosis protein	HFE HLAH	Homo sapiens (Human)	348
Q6ZVN8	RGMC_HUMAN	★	Hemojuvelin	HFE2 HJV RGMC	Homo sapiens (Human)	426
Q6B0J5	Q6B0J5_HUMAN	★	HFE protein	HFE	Homo sapiens (Human)	345
Q8R557	Q8R557_MOUSE	★	HFE protein	Hfe HFE	Mus musculus (Mouse)	34
Q9NP59	S40A1_HUMAN	★	Solute carrier family 40 member 1	SLC40A1 FPN1 IREG1 SLC11A3 MSTP079	Homo sapiens (Human)	571
Q9TQ79	Q9TQ79_HUMAN	★	MHC class I-like protein HFE	HFE	Homo sapiens (Human)	49
P70387	HFE_MOUSE	★	Hereditary hemochromatosis protein homolog	Hfe Mr2	Mus musculus (Mouse)	359
Q9R105	Q9R105_RAT	★	Hemochromatosis gene product HFE splice varia...	Hfe rCG_45141	Rattus norvegicus (Rat)	272
Q1JQE7	Q1JQE7_DANRE	★	Hemochromatosis type 2	hfe2 zgc:136698	Danio rerio (Zebrafish) (Brachydanio rerio)	410
Q9R104	Q9R104_RAT	★	Hemochromatosis gene product HFE splice varia...	Hfe rCG_45141	Rattus norvegicus (Rat)	172
Q7TQ32	RGMC_MOUSE	★	Hemojuvelin	Hfe2 Rgmc	Mus musculus (Mouse)	420
P81172	HEPC_HUMAN	★	Hepcidin	HAMP HEPC LEAP1 UNQ487/PRO1003	Homo sapiens (Human)	84

Done

- The UniProt entry for the HFE protein is shown on the next page. The blue navigation bar at the top of the screen contains links to different parts of the UniProt record for this protein.
 - Make a note of the accession number (Q30201) for this protein. We will use the accession number to search for protein structural information in Activity 5.
 - Scroll down through the record and review the **Protein attributes** and the **General annotation** sections to answer Questions 1–3 for Activity 4 in the worksheet on page 52.
- In the **Protein attributes** section, for “Sequence processing,” note “The displayed sequence is further processed into a mature form.” This means that part of the HFE protein chain needs to be cut off by a proteolytic enzyme to form the “mature” functional protein.

Hereditary hemochromatosis protein precursor – Homo sapiens (Human)

http://www.uniprot.org/uniprot/Q30201

UniProtKB Downloads · Contact · Documentation/Help

Search Blast * Align * Retrieve ID Mapping *

Search in Query
Protein Knowledgebase (UniProtKB) Search Clear Advanced Search »

Q30201 (HFE_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot
Last modified February 8, 2011. Version 129. History..

Contribute
Send feedback
Read comments (0) or add your own

Clusters with 100%, 90%, 50% identity | Documents (6) | Third-party data text xml rdf/xml gff fasta

Names · Attributes · General annotation · Ontologies · Alt products **Sequence annotation** Sequences · References · Web links · Cross-refs · Entry info · Documents · Customize order

Names and origin

Protein names	<i>Recommended name:</i> Hereditary hemochromatosis protein <i>Alternative name(s):</i> HLA-H
Gene names	Name: HFE Synonyms:HLAH
Organism	Homo sapiens (Human) [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota · Metazoa · Chordata · Craniata · Vertebrata · Euteleostomi · Mammalia · Eutheria · Euarchontoglires · Primates · Haplorrhini · Catarrhini · Hominidae · Homo

Protein attributes

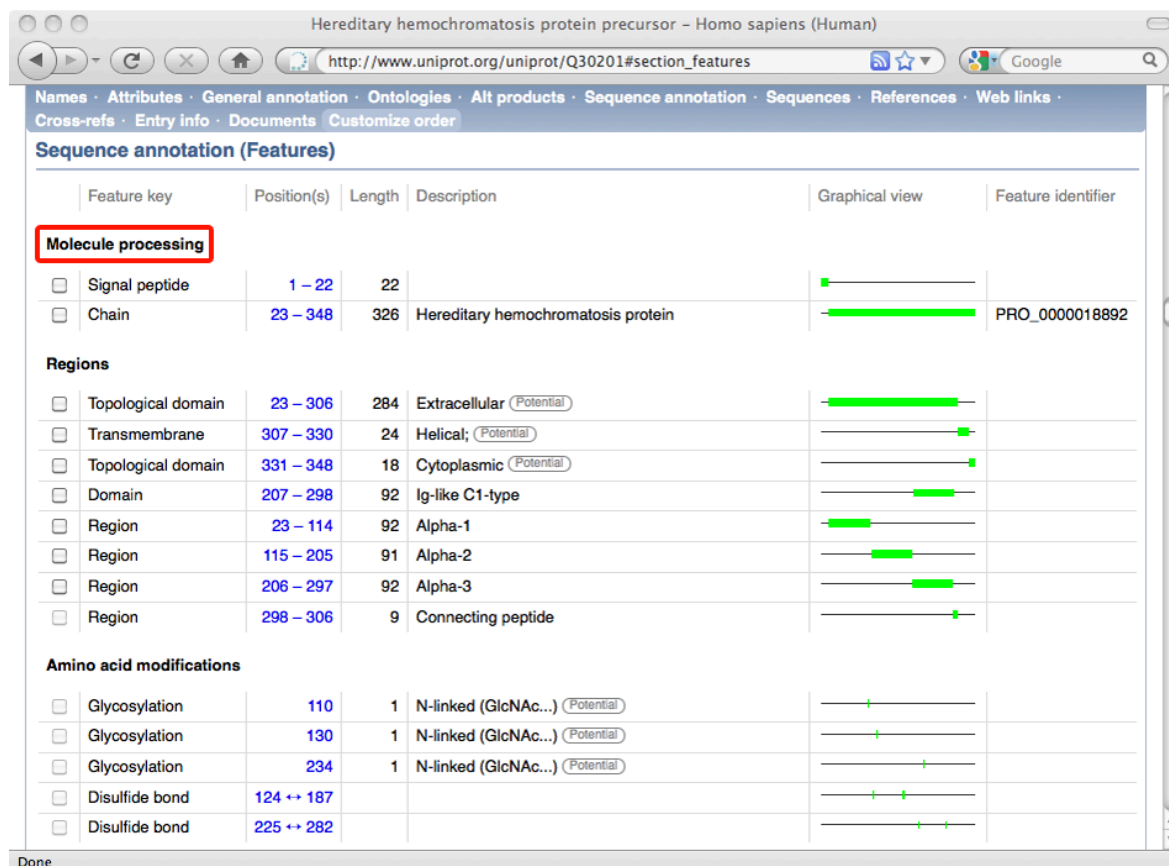
Sequence length	348 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level.

General annotation (Comments)

Function	Binds to transferrin receptor (TFR) and reduces its affinity for iron-loaded transferrin. (Ref.11)
Subunit structure	Binds TFR through the extracellular domain in a pH-dependent manner.
Subcellular location	Membrane; Single-pass type I membrane protein.

Done

- Click on **Sequence annotation** in the blue navigation bar near the top of the record (marked in the screenshot above).
- The **Sequence annotation** section of the HFE protein record is shown in the screenshot on the next page.
 - Under “Molecule processing” in the **Sequence annotation** section, notice that the signal peptide comprises amino acids 1–22. The first 22 amino acids are not associated with any domains (functional units within a protein). This portion is cleaved from the complete protein sequence to make the mature, functional HFE protein, which consists of amino acids 23–348. Clicking on the blue “Position(s)” numbers in the sequence annotation will open a window showing the selected sequence highlighted within the context of the entire protein sequence.
 - In Activity 1 we learned that the cysteine at amino acid position 282 is changed to a tyrosine in a common mutation that causes hemochromatosis. Review the “Regions” and “Amino acid modifications” parts of the **Sequence annotation** section to answer Questions 4–5 for Activity 4 on page 52.



Hereditary hemochromatosis protein precursor - Homo sapiens (Human)

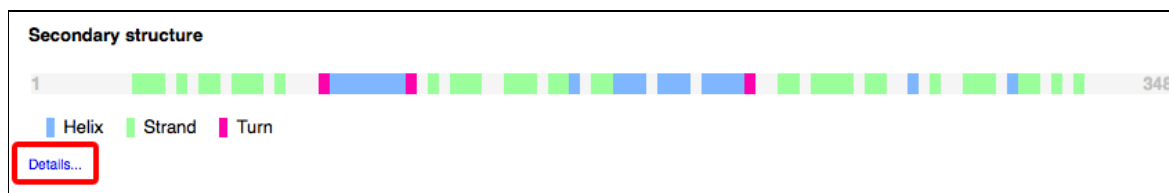
http://www.uniprot.org/uniprot/Q30201#section_features

Names · Attributes · General annotation · Ontologies · Alt products · Sequence annotation · Sequences · References · Web links · Cross-refs · Entry info · Documents · Customize order

Sequence annotation (Features)

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Molecule processing					
<input type="checkbox"/> Signal peptide	1 – 22	22			
<input type="checkbox"/> Chain	23 – 348	326	Hereditary hemochromatosis protein		PRO_0000018892
Regions					
<input type="checkbox"/> Topological domain	23 – 306	284	Extracellular (Potential)		
<input type="checkbox"/> Transmembrane	307 – 330	24	Helical; (Potential)		
<input type="checkbox"/> Topological domain	331 – 348	18	Cytoplasmic (Potential)		
<input type="checkbox"/> Domain	207 – 298	92	Ig-like C1-type		
<input type="checkbox"/> Region	23 – 114	92	Alpha-1		
<input type="checkbox"/> Region	115 – 205	91	Alpha-2		
<input type="checkbox"/> Region	206 – 297	92	Alpha-3		
<input type="checkbox"/> Region	298 – 306	9	Connecting peptide		
Amino acid modifications					
<input type="checkbox"/> Glycosylation	110	1	N-linked (GlcNAc...) (Potential)		
<input type="checkbox"/> Glycosylation	130	1	N-linked (GlcNAc...) (Potential)		
<input type="checkbox"/> Glycosylation	234	1	N-linked (GlcNAc...) (Potential)		
<input type="checkbox"/> Disulfide bond	124 ↔ 187				
<input type="checkbox"/> Disulfide bond	225 ↔ 282				

7. Scroll down to the “Secondary structure” part of the **Sequence annotation** section (shown in image below) and click on **Details** below the colored bar.



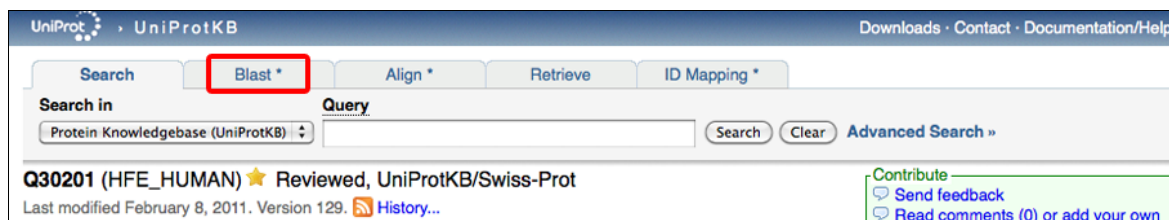
Secondary structure

1 348

■ Helix ■ Strand ■ Turn

[Details...](#)

8. The secondary structure details show which segments of protein sequence make up beta strands, alpha helices, or the turns that form between beta strands and alpha helices. These secondary elements are important in determining the three-dimensional protein structure. Use this secondary structural information to answer Question 6 for Activity 4 on page 52.
9. Return to the top of the HFE protein record by scrolling or by clicking **Names** in the blue navigation bar. Click on the **Blast** tab at the top of the page.



UniProt · UniProtKB

Downloads · Contact · Documentation/Help

Search **Blast *** Align * Retrieve ID Mapping *

Search in Protein Knowledgebase (UniProtKB) Query

Search Clear Advanced Search »

Q30201 (HFE_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified February 8, 2011. Version 129. History...

Contribute

Send feedback

Read comments (0) or add your own

NOTE: BLAST (Basic Local Alignment Search Tool) is a tool used to calculate how similar nucleotide or protein sequences are among the same or different kinds of organisms. Many resources that maintain biological sequence information often support their own BLAST searching capabilities to retrieve and compare sequence data. For more information about BLAST, see *The NCBI Handbook* (www.ncbi.nlm.nih.gov/books/NBK21097/).

Protein sequences are often preferred over nucleotide sequences for BLAST searching because of the greater variability in nucleotide sequences. Remember with the genetic code, different codons of nucleotides can specify the same amino acid. Thus proteins that have similar amino acid sequences may have considerably different nucleotide sequences encoding those proteins.

10. A screenshot of the BLAST search feature for the HFE protein is shown below.

- The amino acid sequence of the complete HFE protein is automatically entered into the text box on the left. The single-letter abbreviations used to represent each amino acid are explained in the Table of Standard Genetic Code on page 50.
- Click on the **Blast** button to compare the amino acid sequence of the HFE protein with all the sequences available from the UniProt Knowledgebase. Be patient. A BLAST search may take several minutes depending on how busy the server is.

11. Once the results are retrieved, scroll down to the **Detailed BLAST results** (see screenshot on next page).

- The **Identity** column on the right provides the percent of each entry's amino acid sequence that is identical to the sequence submitted. To sort all of your results from highest to lowest Identity values, click on the arrows at the top of the **Identity** column.
- To see more results, click **Next** in the upper right corner.
- Use the Detailed BLAST results to answer Question 7 for Activity 4 on page 53.

Detailed BLAST Results for the HFE Protein in UniProt

blast on uniprotkb [completed]

http://www.uniprot.org/blast/uniprot/201102163CS8Z960DH?offset=0&sort=identity

Filter: Overview Results Job information Customize order Page 1 of 10 Next

D2I085 D2I085_AILME Putative uncharacterized protein (Ailuropoda melanoleuca)

Q9R105 Q9R105_RAT Hemochromatosis gene product HFE spli... (Rattus norvegicus)

Detailed BLAST results Customize

Alignments	Accession	Entry name	Status	Protein names	Organism	Length	Identity	Score	E-Value
<input type="checkbox"/>	B4DDZ1	B4DDZ1_HUMAN	★	cDNA FLJ58053, highly similar to Hereditary h...	Homo sapiens (Human)	175	100.0%	597	1.0x10 ⁻⁵⁹
<input type="checkbox"/>	B7ZB37	B7ZB37_HUMAN	★	cDNA, FLJ79401, highly similar to Hereditary ...	Homo sapiens (Human)	175	100.0%	597	1.0x10 ⁻⁵⁹
<input type="checkbox"/>	Q9UK37	Q9UK37_HUMAN	★	Hereditary haemochromatosis protein	Homo sapiens (Human)	129	100.0%	687	5.0x10 ⁻⁷⁰
<input type="checkbox"/>	Q30201-6	HFE_HUMAN	★	Isoform 6	Homo sapiens (Human)	168	100.0%	755	6.0x10 ⁻⁷⁸
<input type="checkbox"/>	Q30201-9	HFE_HUMAN	★	Isoform 9	Homo sapiens (Human)	161	100.0%	767	3.0x10 ⁻⁷⁹
<input type="checkbox"/>	Q86WL1	Q86WL1_HUMAN	★	Hemochromatosis	Homo sapiens (Human)	268	100.0%	1,463	1.0x10 ⁻¹⁶⁰
<input type="checkbox"/>	Q30201-8	HFE_HUMAN	★	Isoform 8	Homo sapiens (Human)	276	100.0%	1,491	1.0x10 ⁻¹⁶³
<input type="checkbox"/>	Q30201	HFE_HUMAN	★	Hereditary hemochromatosis protein	Homo sapiens (Human)	348	100.0%	1,870	0.0
<input type="checkbox"/>	P60018	HFE_PANTR	★	Hereditary hemochromatosis protein homolog	Pan troglodytes (Chimpanzee)	348	100.0%	1,870	0.0
<input type="checkbox"/>	Q9HC69	Q9HC69_HUMAN	★	Hemochromatosis splice variant 861-2305del	Homo sapiens (Human)	116	99.0%	592	5.0x10 ⁻⁵⁹
<input type="checkbox"/>	Q30201-7	HFE_HUMAN	★	Isoform 7	Homo sapiens (Human)	256	99.0%	760	2.0x10 ⁻⁷⁸
<input type="checkbox"/>	Q30201-2	HFE_HUMAN	★	Isoform 2	Homo sapiens (Human)	260	99.0%	1,270	1.0x10 ⁻¹³⁷
<input type="checkbox"/>	B4DV50	B4DV50_HUMAN	★	cDNA FLJ57894, highly similar to Hereditary h...	Homo sapiens (Human)	337	99.0%	1,798	0.0
<input type="checkbox"/>	Q9HC63	Q9HC63_HUMAN	★	Hemochromatosis splice variant delE3,intron3i...	Homo sapiens (Human)	175	98.0%	592	5.0x10 ⁻⁵⁹
<input type="checkbox"/>	Q30201-10	HFE_HUMAN	★	Isoform 10	Homo sapiens (Human)	242	97.0%	694	7.0x10 ⁻⁷¹
<input type="checkbox"/>	Q30201-3	HFE_HUMAN	★	Isoform 3	Homo sapiens (Human)	334	95.0%	1,772	0.0
<input type="checkbox"/>	Q6B0J5	Q6B0J5_HUMAN	★	HFE protein	Homo sapiens (Human)	345	95.0%	1,776	0.0
<input type="checkbox"/>	Q30201-4	HFE_HUMAN	★	Isoform 4	Homo sapiens (Human)	246	93.0%	1,172	1.0x10 ⁻¹²⁶
<input type="checkbox"/>	Q30201-5	HFE_HUMAN	★	Isoform 5	Homo sapiens (Human)	325	93.0%	1,713	0.0
<input type="checkbox"/>	Q9GKZ0	HFE_CERSI	★	Hereditary hemochromatosis protein homolog	Ceratotherium simum (White rhinoceros) (Square-lipped rhinoceros)	348	80.0%	1,495	1.0x10 ⁻¹⁶³
<input type="checkbox"/>	Q9GL41	HFE_RHIUN	★	Hereditary hemochromatosis protein homolog	Rhinoceros unicornis (Greater Indian rhinoceros)	348	79.0%	1,489	1.0x10 ⁻¹⁶³
<input type="checkbox"/>	Q9GL43	HFE_DICBI	★	Hereditary hemochromatosis protein homolog	Diceros bicornis (Black rhinoceros)	348	79.0%	1,492	1.0x10 ⁻¹⁶³
<input type="checkbox"/>	Q9GL42	HFE_DICSU	★	Hereditary hemochromatosis protein homolog	Dicerorhinus sumatrensis (Sumatran rhinoceros)	348	79.0%	1,495	1.0x10 ⁻¹⁶³
<input type="checkbox"/>	Q5EEZ1	Q5EEZ1_BOVIN	★	HFE	Bos taurus (Bovine)	356	76.0%	1,449	1.0x10 ⁻¹⁵⁸
<input type="checkbox"/>	D2I085	D2I085_AILME	★	Putative uncharacterized protein	Ailuropoda melanoleuca (Giant panda)	309	74.0%	1,291	1.0x10 ⁻¹⁴⁰
<input type="checkbox"/>	Q9R105	Q9R105_RAT	★	Hemochromatosis gene product HFE splice varia...	Rattus norvegicus (Rat)	272	70.0%	927	7.0x10 ⁻⁹⁸

Job information

Query sequence >sp|Q30201|HFE_HUMAN Hereditary hemochromatosis protein OS=Homo sapiens GN=HFE PE=1 SV=1
 NGPRARPALLLLMLQTAVLQGRLLRSLSLHYLPMGASEQDLGLSLFALGYDDQLPVF
 YDESRVPEPTPNVSSRISQMLQLSLSLKGNDIMPTVDFWTLMEHNSHESHTLQV
 ILGCEMDENSTEGYWKYGDGQDHFLECPDTLDWRAAEPRAWPTKLEHNSHESHTLQV
 AYLERDCPAQLQLLELGRGVLDDQVPLVHVTHVTSVTLRCLALNYYPNTMKNL
 KDKQPMDAKEFEKDLVLPNGDGTYGWITLAVFPGEGRYTCQVEHPGLDQPLVIMEPS
 FSGTLVIGVISGIAVFPVILFIFGLFILAKRQCSRGAMGHVLAERE

Done

Activity 5

Online Resources: Protein Data Bank

- Explore the sequence and structure of the gene's protein product.

Protein Data Bank

This activity demonstrates how to find and view a protein structure using tools and resources available from the Protein Data Bank (PDB). PDB is an international archive of 3-D structural information for biological macromolecules. PDB records provide access to several interactive molecular graphics programs. This activity also uses FirstGlance in Jmol, a resource that works in most browsers for viewing the major molecular features of a structure with just a few mouse-clicks.

Before You Begin

Many features of the PDB website require newer Web browsers with JavaScript and cookies enabled, and pop-ups should not be blocked. For more information on system requirements see PDB Frequently Asked Questions (www.rcsb.org/pdb/static.do?p=home/faq.html).

Some Protein Structure Basics

- Proteins are created by linking amino acids in a linear fashion to form polypeptide chains. The amino acid sequence of a polypeptide chain is the **primary structure** of a protein. See the Table of Standard Genetic Code on page 50 for single-letter and three-letter abbreviations for the 20 different amino acids.
- Amino acids have different chemical properties. For example, some amino acid residues are strictly hydrophobic ("water fearing") and must be protected from aqueous environments, while other amino acids are hydrophilic ("water loving"). The substitution of just one amino acid for another with very different chemical properties can have serious consequences for a protein's structure and function.
- The folding of regions within the polypeptide chain into alpha helices and beta sheets is a protein's **secondary structure**.
- The packing of the entire polypeptide chain into a three-dimensional globular unit is a protein's **tertiary structure**.
- If a protein molecule is a complex of more than one polypeptide chain, then the complete structure of this molecular complex is called a protein's **quaternary structure**.
- A domain is a discrete portion of a protein with its own function and specific three-dimensional structure. The combination of domains in a single protein determines its overall function.
- Different parts of a polypeptide chain can be linked by disulfide bridges that form between two cysteine residues. Disulfide bridges (or disulfide bonds) stabilize a protein's three-dimensional structure. The loss of a disulfide bridge would be detrimental to a protein's overall structure.

Finding a Structure Record in PDB

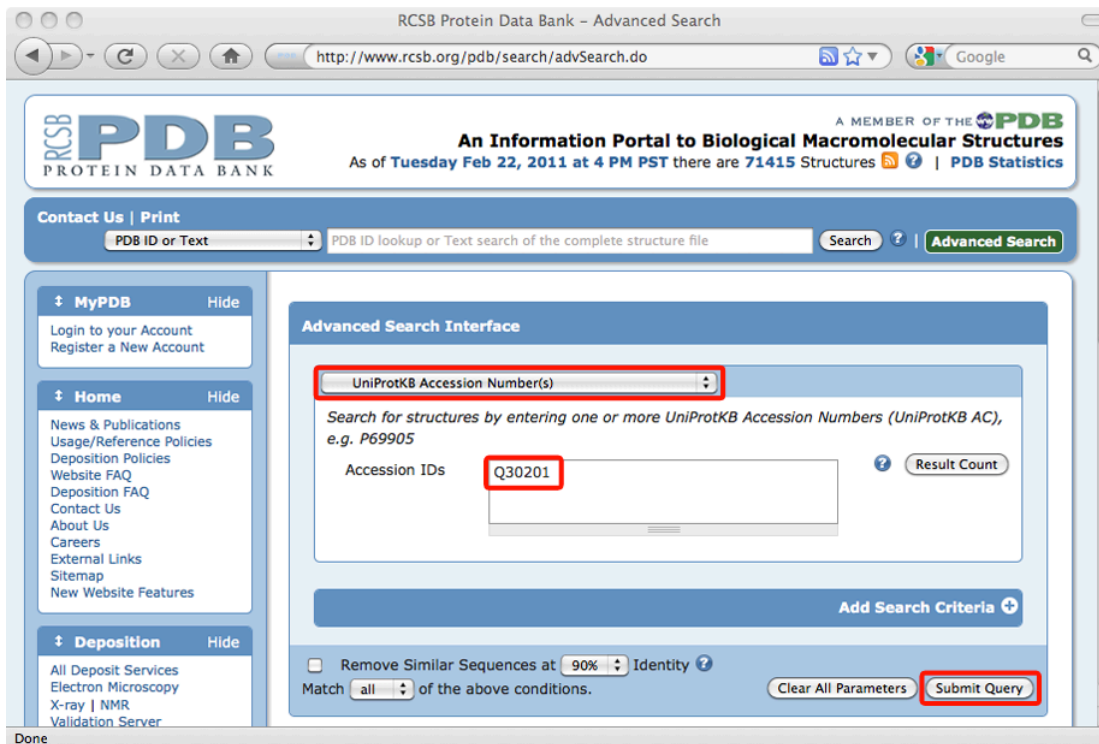
1. To begin, let's go to the Protein Data Bank (www.rcsb.org/pdb/).

The screenshot shows the RCSB Protein Data Bank homepage. At the top, there is a search bar with a dropdown menu for 'PDB ID or Text' and a search button labeled 'Advanced Search'. Below the search bar is a navigation menu with categories like 'MyPDB', 'Home', 'Deposition', 'Search', 'Tools', and 'Education'. The main content area is titled 'A Resource for Studying Biological Macromolecules' and includes a 'Molecule of the Month' section for 'Integrin'. The right sidebar contains sections for 'New Features', 'RCSB PDB News', and 'wwPDB News'. The 'Education' link in the left sidebar is highlighted with a red box.



NOTE: If you are new to PDB, be sure to check out the **Education** links in the light blue column on the left of the screen. Under **Educational Resources** you can find posters, tutorials, activities, and lessons. **Molecule of the Month** is a collection of vignettes, each featuring a different molecular structure and its importance to human welfare.

2. Beside the search box at the top of the PDB home page, select **Advanced Search**.

- On the Advanced Search page, from the **Choose a Query Type** drop box select **UniProtKB Accession Number(s)**. In Activity 4 we accessed the human hemochromatosis protein record Q30201 in the UniProt Protein Knowledgebase. Enter **Q30201** in the search box. The advanced search page should look like the screenshot below. Select the **Submit Query** button to submit your search.



- The search should return two hits. Scroll down the page to see a brief summary of each search result. One record (1DE4) provides structural information on the hemochromatosis protein HFE complexed with a receptor, and the other record (1A6Z) just provides structural information for the HFE protein. Click on **1A6Z HFE (HUMAN) HEMOCHROMATOSIS PROTEIN** to open this PDB record.

<input checked="" type="checkbox"/>	1DE4	HEMOCHROMATOSIS PROTEIN HFE COMPLEXED WITH TRANSFERRIN RECEPTOR
	Chain(s): A,D,G	
	Authors: Bennett, M.J., Lebron, J.A., Bjorkman, P.J.	
	Release Date: 2000-01-19	Classification: Metal Transport Inhibitor/receptor
	Experiment: X-RAY DIFFRACTION with resolution of 2.80 Å	
	Compound: 3 Polymers [Display Full Polymer Details Display for All Results] 3 Ligands [Display Full Ligand Details Display for All Results]	
	Citation: Crystal structure of the hereditary haemochromatosis protein HFE complexed with transferrin receptor. (2000) Nature 403 : 46-53 [Display Full Abstract Display for All Results]	
	Molecule of the Month: Ferritin and Transferrin, Major Histocompatibility Complex, T-Cell Receptor	
<input checked="" type="checkbox"/>	1A6Z	HFE (HUMAN) HEMOCHROMATOSIS PROTEIN
	Chain(s): A,C	
	Authors: Lebron, J.A., Bennett, M.J., Vaughn, D.E., Chirino, A.J., Snow, P.M., Mintier, G.A., Feder, J.N., Bjorkman, P.J.	
	Release Date: 1999-03-23	Classification: Mhc Class I Complex
	Experiment: X-RAY DIFFRACTION with resolution of 2.60 Å	
	Compound: 2 Polymers [Display Full Polymer Details Display for All Results]	
	Citation: Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. (1998) Cell(Cambridge,Mass.) 93 : 111-123 [Display Full Abstract Display for All Results]	

5. The summary tab of the 1A6Z record is shown in the screenshot above.

- Note the **Molecular Description** box in the center of the screenshot. This structure is a complex of four polymer chains: A, B, C, and D. A and C are identical HFE polypeptide chains, and B and D are identical chains of another protein called beta-2-microglobulin.
- Note the **Primary Citation** in the 1A6Z record. The best way to learn about structure details is to access the article listed as the primary citation. Although the full text for some articles may be freely available online, many articles are accessible only by subscription. Some university research libraries may provide public access to their journal collections. The article for this structure has been accessed to reveal the following details:
 - Only the soluble portion of the HFE polypeptide chain is included in the 1A6Z structure. The transmembrane domain is missing, so the HFE protein in this structure has only 275 of the 348 amino acids in the complete HFE protein sequence.
 - The first 22 amino acids of the HFE polypeptide sequence have been excluded because they are not part of the mature, functional protein. Therefore, the first amino acid in this structure is really the 23rd, and cysteine 260 is the cysteine residue involved in the CYS282TYR mutation that we learned about in Activity1.

- Each HFE polypeptide chain is complexed with another polypeptide chain called beta-2-microglobulin.
 - The 1A6Z structure consists of two HFE–beta-2 microglobulin complexes.
6. Select the **Sequence** tab to examine the sequence and secondary structure details for this structure.
7. The Sequence and Structure Details for record 1A6Z are shown in the screenshot below.
- The HFE protein sequence (polypeptide chain A) is presented first. Each letter in the protein sequence represents a different amino acid. C stands for cysteine. See the Table of Standard Genetic Code on page 50 to determine which amino acid is represented by each letter.
 - Secondary structure details are mapped onto sequence details. Different graphical symbols are used to represent extended beta strands, alpha helices, bends, and turns.

RCSB Protein Data Bank – Sequence / Structure Details for 1A6Z – HFE (HUMAN) HEMOCHROMATOSIS PROTEIN

http://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=1A6Z

PDB
PROTEIN DATA BANK

A MEMBER OF THE **PDB**
An Information Portal to Biological Macromolecular Structures
As of Tuesday Feb 22, 2011 at 4 PM PST there are 71415 Structures | PDB Statistics

Contact Us | Print

PDB ID or Text : PDB ID lookup or Text search of the complete structure file Search ? | Advanced Search

MyPDB Hide
Login to your Account
Register a New Account

Home Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

Search Hide
Advanced Search
Latest Release
New Structure Papers
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

Tools Hide
Download: Entries | Ligands
Compare Structures
FTP Services
File Formats
Services: RESTful | SOAP
Widgets

Education Hide
Understanding PDB Data
Molecule of the Month
Educational Resources

Help Hide
Launch Help System
Video Tutorials
Glossary of Terms
PDBMobile FAQ

Summary **Sequence** Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

HFE (HUMAN) HEMOCHROMATOSIS PROTEIN **1A6Z** Display Files ▾
Download Files ▾
Share this Page ▸

Sequence / Structure Details

Redundancy Reduction and Sequence Clustering
View the clustering results for 1A6Z.

Sequence Display ⓘ
The structure **1A6Z** has in total 4 chains. Out of these 2 are sequence-unique.
Currently viewing **unique chains** only. [show all chains] [Show 3D in Jmol]

Chain Display

Chain A (polymer 1) [help] [fasta] [sequence & DSSP]

Description HFE
Identical chains C
[show all chains]

Chain Type polypeptide(L)
UniProtKB reference **Q30201**
Length 275 residues
scop domain assignment **d1a6za1** Hemochromatosis protein Hfe, alpha-3 domain: 94 residues ⓘ
[hide] [reference] **d1a6za2** Hemochromatosis protein Hfe, alpha-1 and alpha-2 domains: 178 residues ⓘ

dssp secondary structure 25% helical (8 helices; 70 residues)
[hide] [reference] 39% beta sheet (20 strands; 109 residues)

More annotations
Select

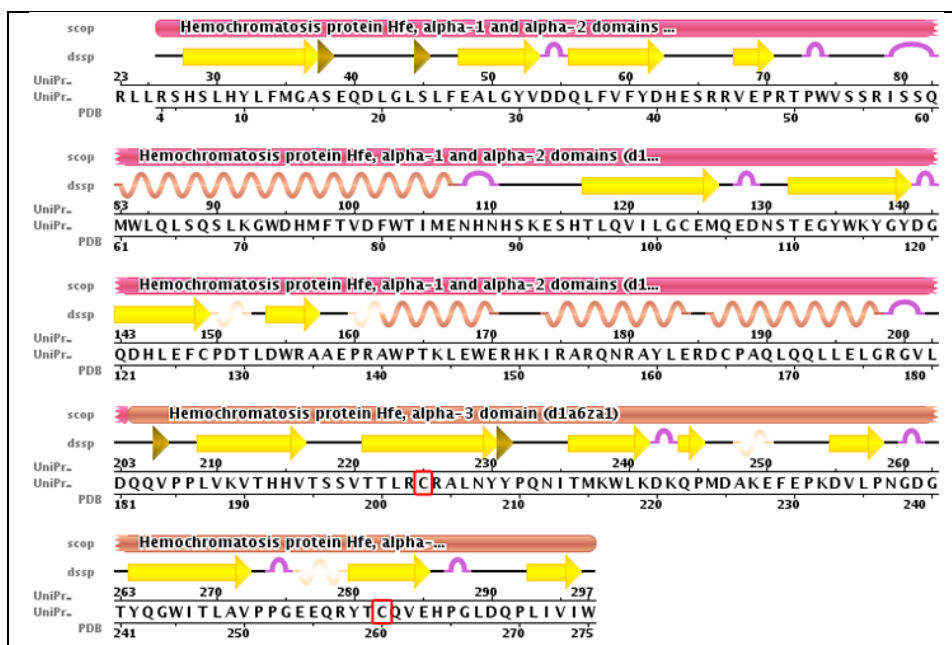
Currently displayed: SEQRES sequence. [display external (UniProtKB) sequence]

Sequence Details [View image]

scop **Hemochromatosis protein Hfe, alpha-1 and alpha-2 domains** ...
dssp
PDB R L L R S H S L H Y L F M G A S E Q D L G L S L F E A L G Y V D D Q L F V F Y D H E S R R V E P R T P W V S S R I S S Q
PDB 4 10 20 30 40 50 60

scop **Hemochromatosis protein Hfe, alpha-1 and alpha-2 domains (d1...**
dssp
PDB M W L Q L S Q S L K G W D H M F T V D F W T I M E N H N H S K E S H T L Q V I L G C E M Q E D N S T E G Y W K Y G Y D G
PDB 61 70 80 90 100 110 120

8. Select the **display external (UniProtKB) sequence** link highlighted in the previous screenshot.
9. The sequence page will reload and display the amino acid numbers for the UniProt HFE protein sequence (that we examined in Activity 4) above the line of single-letter amino acid abbreviations (see screenshot below).
 - Find cysteine 282 in the UniProt sequence. Cysteine 282 is the amino acid that is replaced by tyrosine in the CYS282TYR mutation.
 - You will see that cysteine 282 in the UniProt sequence is at position 260 in the PDB structure sequence. In Activity 4, we learned that cysteine 282 forms a disulfide bond with cysteine 225 in the UniProt HFE protein sequence. In the HFE protein sequence for PDB structure 1A6Z, we see that cysteine 260 forms a disulfide bond with cysteine 203 (which corresponds to cysteine 225 in the UniProt sequence). Disulfide bonds are critical to forming the proper structural arrangement needed to make a functional protein; therefore, the loss of cysteine 260 would be detrimental to protein structure. **Answer the first two questions for Activity 5 in the worksheet on page 54.**



Viewing the Structure

10. Select the **Summary** tab near the top of the page to return to the 1A6Z record summary. In the **Biological Assembly 1** box in the upper right corner of the page there are several options for viewing the molecular structure. Clicking on the **More Images...** link will open a page with options for downloading a still image of the HFE molecular complex 1A6Z. Although PDB provides access to several different molecular viewers for examining a 3-D representation of a molecular complex, many of these options were designed for scientists who specialize in studying molecular structures. In this activity, we will use a molecular viewer called **FirstGlance in Jmol**, which is one of the more user-friendly options for displaying the major structural features of a molecule. FirstGlance in Jmol was developed to work in all popular web browsers without having to download and install anything.

11. To access FirstGlance, first click on the left arrow next to the **Biological Assembly 1** label above the molecular image. This should change the box label to **Asymmetric Unit**.

The screenshot shows the RCSB PDB website for the structure 1A6Z, HFE (HUMAN) HEMOCHROMATOSIS PROTEIN. The page includes a navigation menu on the left with sections for MyPDB, Home, Deposition, and Search. The main content area is divided into several sections: Summary, Primary Citation, Molecular Description, and a molecular image viewer. The molecular image viewer shows a ribbon diagram of the protein structure. The label 'Biological Assembly 1' is highlighted with a red box, indicating the current view. Below the image, there are options to 'View in Jmol' and 'Other Viewers'.

12. By clicking on the arrow next to **Other Viewers**, a drop-down menu will appear. Select **FirstGlance** from the drop-down menu (see screenshot below).

The screenshot shows the molecular image viewer in 'Asymmetric Unit' view. The protein structure is displayed as a ribbon diagram. Below the image, there is a 'View in Jmol' button and an 'Other Viewers' dropdown menu. The dropdown menu is open, showing a list of viewer options: KING, WebMol, QuickPDB, FirstGlance, and Kiosk. The 'FirstGlance' option is highlighted with a red box, indicating it has been selected.

13. A new page should open displaying structure 1A6Z using **FirstGlance in Jmol** (see screenshot below).

- To stop the spinning of the molecule, click the **Spin** box in the upper left.
- To remove the S- labels, uncheck the **Show** box beside **Labels X, S-, ?**.

The screenshot shows the 'FirstGlance in Jmol' web application. The browser address bar shows the URL: <http://molvis.sdsc.edu/fgij/fg.htm?mol=1A6Z>. The interface includes a navigation menu with options like 'Secondary Structure', 'Cartoon', 'N→C Rainbow', 'Composition', 'Hydrophobic/Polar', 'Charge..', 'Contacts..', 'Vines..', 'All Models', 'Hide..', and 'Find..'. There are also checkboxes for 'Ligands+', 'Water..', 'Slab..', 'Background', 'Spin', 'Quality', and 'Zoom'. The 'Spin' checkbox is checked and highlighted with a red box. Below the navigation menu, there are links for 'More Views..', '1A6Z: Key Resources..', 'Center Atom...', 'Troubleshooting', 'Reset', 'Close', and 'New Session..'. At the bottom of the control panel, there are checkboxes for 'Labels X, S-, ?' (checked and highlighted with a red box), 'Front', and 'ID'. The main window displays the protein structure 1A6Z as a ribbon model with four distinct chains colored yellow, pink, blue, and green. The structure is labeled '1A6Z' at the top. At the bottom of the window, there is a visitor count of 309,657 and the version number 1.45.

14. The structure is initially displayed using the **Cartoon** option, which assigns a different color to each molecular chain in the structure. Chains A, B, C, and D should be displayed. Earlier in the activity we learned that chains A and C are identical HFE chains and chains B and D are identical beta-2-microglobulin chains.

- Clicking anywhere on the molecule will generate a label in the lower left corner showing the amino acid residue and the protein chain that you have selected.
- Click on each colored chain to find **Chain A**, which is one of the two HFE protein chains. In the screenshot on the next page, Chain A is the blue chain.
- If you need to rotate the structure, simply click on the structure and drag with your mouse.
- To undo any of the changes you have made and reset the structure to its original configuration, click **Reset** in the upper left corner.

By clicking on the blue chain, the label in the lower left indicates that the blue chain is Chain A.

1A6Z: FirstGlance in Jmol

Secondary Structure [Cartoon](#) [N→C Rainbow](#)
[Composition](#) [Hydrophobic/Polar Charge](#)
[Contacts](#) [Yines](#) [All Models](#) [Hide](#) [Find](#)

Ligands+ Background
 Water Spin Quality
 Slab Zoom

[More Views](#) [1A6Z: Key Resources](#)
[Center Atom](#) [Troubleshooting](#) [Reset](#) [Close](#)
[New Session](#)

Labels [X, S, ?](#): Show Front ID

Introduction [español](#)

FirstGlance in Jmol is a simple, free tool for macromolecular visualization. The initial display is [Cartoon plus Ligands+](#). (Unusual moieties, when present, are labeled [X, S, or ?](#).) Click on the links and buttons above to see different aspects of the molecular structure.

- **Rotate** the molecule by dragging near it with the mouse.
- **Identify** any atom by clicking on it. Its identity will be displayed to the lower left of the molecule (and in the browser status bar). [More](#). If you don't recognize the abbreviation, go to [Proteopedia](#), and click on the green link for the ligand abbreviation. The full name will be displayed, and the group will be highlighted in the molecular scene. There are also [other methods](#).
- **Center** a region of interest using [Center Atom](#) above. You can then inspect details by [zooming in](#). Regions distant from the centered moiety can be [hidden](#) by toggling [Slab](#) on. (When finished, [Center Atom](#) offers the option to re-center the entire molecule.)
- **Specific oligomers** or single molecules can be visualized using [Key External Resources](#).

GLN 261 **Chain=A** Atom=CA (Carbon Alpha, 1st)

[Can't see the molecule?](#) [FirstGlance in Jmol](#) (ver. 1.45)

309,680 Visitors
Since February 8, 2006

Jmol script terminated

15. Let's hide all the chains except Chain A. Click on the **Hide..** link in the upper left corner and then click on each chain **except Chain A**. Your screen should look like the screenshot below.

1A6Z: FirstGlance in Jmol

Secondary Structure [Cartoon](#) [N→C Rainbow](#)
[Composition](#) [Hydrophobic/Polar Charge](#)
[Contacts](#) [Yines](#) [All Models](#) [Hide](#) [Find](#)

Ligands+ Background
 Water Spin Quality
 Slab Zoom

[More Views](#) [1A6Z: Key Resources](#)
[Center Atom](#) [Troubleshooting](#) [Reset](#) [Close](#)
[New Session](#)

Labels [X, S, ?](#): Show Front ID

Hide: Click on any [Exit Hiding](#)

Chain† Residue/Group Atom
 Range of residues in one chain (2 clicks) to **hide** it, or **hide**

Protein* DNA* Hydrogen
 Carbohydrate RNA*

Currently hidden:
Chains D, C, B

These will remain hidden in **all views**, until re-displayed: [Re-Display](#) the most recently hidden **Chain**.
[Re-Display Everything](#).

[Center Visible Chains](#) [Center by Clicking](#)
 Invert Hiding

To see only one (or a few) small parts of a molecule, first hide them, and then check [Invert Hiding](#) (immediately above).

† [Ligands+](#) are not hidden when you hide a chain, even if they belong to that chain. Use the [Residue/Group](#) option, or the [Ligands+](#) button to hide them.
Nonstandard residues are not always hidden with the chain, but can be hidden as [Residues](#).
 * "Protein", "DNA", or "RNA", as the terms are used here, exclude amino acids or

Chain B Hidden

[Can't see the molecule?](#) [FirstGlance in Jmol](#) (ver. 1.45)

309,680 Visitors
Since February 8, 2006

Jmol script terminated

16. Click on the **Center Visible Chains** link (highlighted in screenshot above) to place Chain A in the center of the display panel.

17. Once Chain A is centered, use the **Zoom** tool to enlarge Chain A. In addition to using the Zoom arrows in the upper left corner, you can zoom in and out by clicking on the background of the structure and then using the wheel on your mouse. Alternatively, you can also hold down the Shift key and drag the mouse up and down over the molecule to zoom in and out. Your screen should look something like the screenshot below.

The screenshot shows the Jmol web interface for protein 1A6Z. The main window displays a ribbon diagram of the protein structure. The left sidebar contains various navigation and control options, including 'Secondary Structure', 'Cartoon', 'N→C Rainbow', 'Composition', 'Hydrophobic/Polar', 'Charge', 'Contacts', 'Yines', 'All Models', 'Hide', and 'Find..'. The 'Find..' link is highlighted in red. Below the sidebar, there is a 'Zoom' section with instructions on how to use the mouse wheel and Shift key for zooming. At the bottom right, there is a visitor count of 309,680 and a date of February 8, 2006.

18. Let's find cysteine 260 and cysteine 203 (the cysteine residues that form the disulfide bond involved in the CYS282TYR mutation). Click on the **Find..** link (highlighted in screenshot above).
19. The **Find** option (shown in the screenshot on the next page) allows you to search for particular residues within a molecule. The locations of the residues are indicated using yellow dots. The background color automatically changes to black when you select **Find**. A black background makes the yellow dots easier to see. You can toggle between black and white background colors by clicking on the **Background** box in the upper left corner.
- Type **CYS260, CYS203** into the text box.
 - Press the Enter key on your keyboard to submit your search.
 - Yellow dots should indicate where these two residues are in the protein chain. You may need to rotate the structure by clicking and dragging your mouse over the molecule so that you can obtain a good view of the yellow dots. Note that the yellow dots surround a thin gold bar. This thin gold bar represents a disulfide bond. You can see that a bond between cysteines 203 and 260 would create a strong connection between two different strands within the protein.

Finding Cysteine Residues in the HFE Protein

1A6Z: FirstGlance in Jmol

http://molvis.sdsc.edu/fgij/fg.htm?mol=1A6Z

RCSB Protein Data Bank - Struc... 1A6Z: FirstGlance in Jmol

Secondary Structure Cartoon N→C Rainbow
Composition Hydrophobic/Polar Charge,
Contacts, Vines, All Models Hide, Find,

Ligands+... Background
 Water... Spin Quality
 Slab... Zoom

More Views... 1A6Z: Key Resources...
Center Atom... Troubleshooting Reset Close
New Session...

Labels X, S, ? Show Front ID

Portions Hidden (Re-Display...) Halos around: CYS260, CYS203.

Find: CYS260, CYS203 Clear Halos
See All Halos Easily Undo

In the slot above, enter one or more (separated by commas):

- Sequence number(s) (such as 10,51).
- Range of sequence numbers (such as 18-23).
- Amino acid 3-letter names (such as PRO,HIS).
- Nucleotide 1-letter names (A,C,G,T,U). More...
- Name-sequence combinations (SER41, C9).
- To restrict one of the above to a specific chain, add colon plus the chain letter (such as 41:A, C9:D).
- Full name(s) of elements (zinc, selenium).
- More...

After pressing Enter, yellow dot-halos will appear around the atoms in the residues listed in the slot above (even if the atoms themselves are not visible). If no halos appear, either the listed residues don't exist in this molecule, or your list is not formatted correctly following the above

Jmol script terminated

1A6Z

Can't see the molecule? FirstGlance in Jmol (ver. 1.45)

309,680 Visitors
Since February 8, 2006

20. To obtain a better view of the disulfide bonds in the HFE protein, click on the **More Views..** link in the upper left corner, and then click on the **Disulfide Bonds: Show All** link. The page should change so that it looks like the screenshot below. The backbone of the protein chain is modified to a thin line (which is difficult to see in the screenshot), and the disulfide bonds become thicker and easier to see. The cysteine residues are also labeled. **Answer Questions 3–4 for Activity 5 in the worksheet on page 54.**

1A6Z: FirstGlance in Jmol

http://molvis.sdsc.edu/fgij/fg.htm?mol=1A6Z

RCSB Protein Data Bank - Struc... 1A6Z: FirstGlance in Jmol

Secondary Structure Cartoon N→C Rainbow
Composition Hydrophobic/Polar Charge,
Contacts, Vines, All Models Hide, Find,

Ligands+... Background
 Water... Spin Quality
 Slab... Zoom

More Views... 1A6Z: Key Resources...
Center Atom... Troubleshooting Reset Close
New Session...

Labels X, S, ? Show Front ID

Portions Hidden (Re-Display...) Halos around: CYS260, CYS203. (Change)

Disulfide Bonds

Label Cysteines (Sequence Number:Chain)
 Show Backbones
 Connect Backbones
 Hide Cysteines Not Disulfide Bonded
 Color Disulfide Bonds By Chain

All disulfide bonds are shown as thick rods connecting protein backbone traces. If no thick rods are visible, there are no disulfide bonds (example: 5ccy which contains no cys residues). Cysteines (not participating in disulfide bonds) are shown as spacefilled alpha carbons.

An example containing disulfide bonds (cystines), and also cysteines not disulfide bonded, and also methionines is 3kwb. An example containing cysteines and no disulfide bonds is 3h8e.

Jmol script terminated

1A6Z

Can't see the molecule? FirstGlance in Jmol (ver. 1.45)

309,680 Visitors
Since February 8, 2006

21. Now that you are familiar with a few options for modifying a molecular structure using **FirstGlance**, you may want to **Reset** the structure and practice what you have learned. In addition to the display options in the upper left corner of the screen, you can also use pop-up menus to modify the structure by clicking on **Jmol** in the lower right corner of the display panel (highlighted in the previous screenshot).
22. If you are interested in copying or saving a particular view of a structure that you have created, check out the **Presenting Molecular Views from FirstGlance in Jmol** page (molvis.sdsc.edu/fgjj/slides.htm).

Protein Structure and Hereditary Hemochromatosis Development

By examining the HFE protein's sequence and structure, we discover that the cysteine lost in the CYS282TYR mutation has an important role in establishing the correct three-dimensional HFE structure. In this mutation, a cysteine residue is replaced by another amino acid, tyrosine, and the disulfide bond between two cysteines in the polypeptide chain is lost. This is detrimental to the protein's structure. As a result, the HFE protein can no longer perform its normal function of regulating iron uptake, and cells become overloaded with iron. This buildup of iron in cells, if untreated, can lead to organ damage and other complications.

Table of Standard Genetic Code for Translating DNA Sequence Records

	T	C	A	G
T	TTT Phe (F) TTC Phe (F) TTA Leu (L) TTG Leu (L)	TCT Ser (S) TCC Ser (S) TCA Ser (S) TCG Ser (S)	TAT Tyr (Y) TAC TAA STOP TAG STOP	TGT Cys (C) TGC TGA STOP TGG Trp (W)
C	CTT Leu (L) CTC Leu (L) CTA Leu (L) CTG Leu (L)	CCT Pro (P) CCC Pro (P) CCA Pro (P) CCG Pro (P)	CAT His (H) CAC His (H) CAA Gln (Q) CAG Gln (Q)	CGT Arg (R) CGC Arg (R) CGA Arg (R) CGG Arg (R)
A	ATT Ile (I) ATC Ile (I) ATA Ile (I) ATG Met (M) START	ACT Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T)	AAT Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K)	AGT Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R)
G	GTT Val (V) GTC Val (V) GTA Val (V) GTG Val (V)	GCT Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A)	GAT Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E)	GGT Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G)

Key to the Table of Standard Genetic Code

Alanine	ALA	A	Arginine	ARG	R
Asparagine	ASN	N	Aspartic acid	ASP	D
Cysteine	CYS	C	Glutamic acid	GLU	E
Glutamine	GLN	Q	Glycine	GLY	G
Histidine	HIS	H	Isoleucine	ILE	I
Leucine	LEU	L	Lysine	LYS	K
Methionine	MET	M	Phenylalanine	PHE	F
Proline	PRO	P	Serine	SER	S
Threonine	THR	T	Tryptophan	TRP	W
Tyrosine	TYR	Y	Valine	VAL	V

START = Initiation Signal (signifies the beginning of a polypeptide chain)

STOP = Termination Signal (signifies the end of a polypeptide chain)

Hereditary Hemochromatosis Worksheet

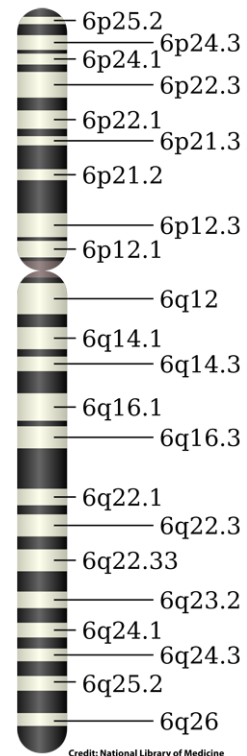
This worksheet provides questions to be answered as you complete the activities in the Gene Gateway Workbook.

Questions for Activity 1

- 1) What is the official gene symbol of the hereditary hemochromatosis gene?
- 2) Which allelic variant (genetic mutation) most commonly causes hereditary hemochromatosis?
- 3) What are some characteristics of hereditary hemochromatosis? How is it treated?

Questions for Activity 2

- 1) On the diagram to the right, mark the general region where the HFE gene can be found on chromosome 6.
- 2) About how many genes are on chromosome 6?
- 3) How long is the DNA sequence for chromosome 6?



Questions for Activity 3

- 1) Using the summary from the Entrez Gene record for the HFE gene, briefly describe the function of the gene's protein product.

Use the GenBank sequence record Z92910.1 to answer questions 2–4.

- 2) In the Features section of record Z92910.1, select the [gene](#) link. How many base pairs (bp) are in the genomic sequence of the HFE gene?
- 3) Scroll through the Features section of the [gene](#) sequence in Z92910.1. How many exons have been identified in this sequence?
- 4) Return to the main record Z92910.1. Select the [CDS](#) link. How many base pairs are in the coding sequence?

Questions for Activity 4

- 1) How many amino acids (AA) are in the complete HFE protein?
- 2) In what part of the cell is the HFE protein located?
- 3) What type of tissue does not express the HFE protein?
- 4) Is cysteine 282 found on the extracellular or cytoplasmic side of the HFE protein?
- 5) What is the number of the cysteine residue that forms a disulfide bond with cysteine 282?
- 6) What kind of secondary structural element contains cysteine 282: alpha helix, turn, or beta strand?

- 7) Using the BLAST search results, list the first 10 non-human organisms that have proteins similar to the human HFE protein sequence. Include the percent identity score with each organism you list, and order the list from highest to lowest identity score. Skip any human entries, and do not list any organism more than once.

Organism Name	Identity %
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	

Questions for Activity 5

1. Examine the amino acid sequence for the human HFE protein from the UniProt Protein Knowledgebase (shown below). Find cysteine 282, the amino acid that is replaced by tyrosine in the CYS282TYR mutation. Refer to the Table of Standard Genetic Code on Page 50 for help with the single-letter amino acid abbreviations.

10	20	30	40	50	60
MGPRARPALL	LLMLLQTAVL	QGRLLRSHSL	HYLFMGASEQ	DLGLSLFEAL	GYVDDQLFVF
70	80	90	100	110	120
YDHESRRVEP	RTPWVSSRIS	SQMWLQLSQS	LKGWDHMFTV	DFWTIMENHN	HSKESHTLQV
130	140	150	160	170	180
ILGCEMQEDN	STEGYWKYGY	DGQDHLEFCP	DTLDWRAAEP	RAWPTKLEWE	RHKIRARQNR
190	200	210	220	230	240
AYLERDCPAQ	LQQLLELGRG	VLDQQVPPLV	KVTHHVTSSV	TTLRCRALNY	YPQNITMKWL
250	260	270	280	290	300
KDKQPMDAKE	FEPKDVLPNG	DGTYQGWITL	AVPPGEEQRY	TCQVEHPGLD	QPLIVIWEPS
310	320	330	340		
PSGTLVIGVI	SGIAVFFVIL	FIGILFIILR	KRQGSRGAMG	HYVLAERE	

2. Compare the amino acid sequence above with the HFE sequence details provided for PDB structure 1A6Z. In question 1, underline the portion of the amino acid sequence included in the PDB structure.
3. How many disulfide bonds are present in the hereditary hemochromatosis protein?
4. Why is the cysteine residue affected in the CYS282TYR mutation important?