Mini Courses NCBI Resources one course at a time

Identification and Correlation of Disease Genes to Phenotypes

This mini-course focuses on the identification of a disease gene and its correlation to a phenotype. It demonstrates how NCBI resources such as the reference human genome assembly combined with literature and expression information can help identify and provide potential functional information for candidate disease genes.

Problem:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype and, if possible, establish a biochemical and structural basis for the disease phenotype.

Outline:

1. Compare ESTs to the human genome (using BLAST).
2. Identify the gene(s) aligning to the ESTs and download their sequences (using MapViewer).
3. Determine what is known about the gene and protein, and identify known SNPs in the gene (using Entrez Gene and dbSNP)
4. Determine whether a mutant HFE gene causes a disease phenotype (using OMIM).
5. Examine the biochemical and structural consequences of the mutation on the HFE protein product (using Blink and Cn3D).

URL:
http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasepheno.html

The following handout includes the screen shots of the exercise demonstrated in the workshop.

Course developed by: Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Instructor: Wayne Matten (matten@ncbi.nlm.nih.gov)

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare their sequences using BLAST with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on the following link:

BLAST (human genome) .

Paste the EST sequence provided below in the query box of the BLAST page, select the database “genome (reference only)”, and start the search by clicking the “Begin Search” button.

Query EST Sequence:

```
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTC
TACGGTGTCGGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGC
TGAAGGATAAGCAGCCAATGGATGCCAAGGAGTTCGAACCTAAAGACGTAT
TGCCCAATGGGGATGGGACCTACCAGGGCTGGATAACCTTGGCTGTACCCC
CTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCCAGGCCTGGATC
AGCCCCTCATTGTGATCTGGG
```

Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

Results of BLAST against the human genome

Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using MapViewer, click on the "Genome View" button near the top of the results page, then on the Map element "NT_007592". Currently, 4 maps should be displayed (Model, RNA, Genes_seq and Contig). Make the Genes_seq map the master map by clicking on the arrow next to the map name. Zoom out using the graphic on the left.
The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene. Display the entire HFE gene sequence by clicking on the "dl" (download) link and then on "Display". Copy the sequence and paste it in the area provided below. You can adjust the sequence coordinates to download

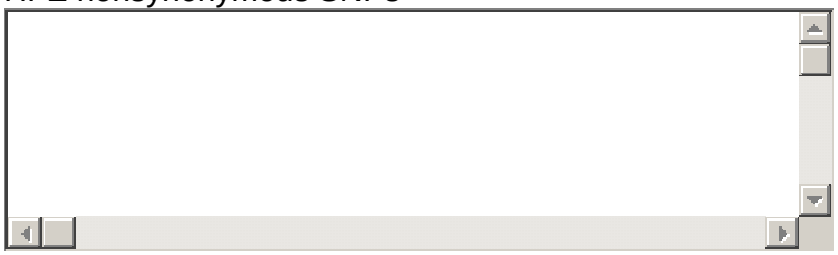
upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" functions.

HFE gene sequence

Step 3. Determine what is known about the HFE gene and protein, and identify known SNPs in the gene:

Click on the HFE gene link from the Genes_seq map, which takes you to the Entrez Gene HFE report.
What is the chromosome location of the HFE gene? How many mRNA variants have been annotated for the HFE gene? Which is the longest isoform? Where is the mouse homolog located? What is the function of the protein? Click on the "GeneView in dbSNP" link from the "Links" menu to get a list of known SNPs in the HFE gene. By default the SNPs in the coding region are shown. All the SNPs can be displayed by clicking on the "in gene region" button. List the nonsynonymous changes with respect to the variant 1 mRNA, NM_000410, in the box provided below. Which nonsynonymous SNP is also observed in the EST sequence that we started with?

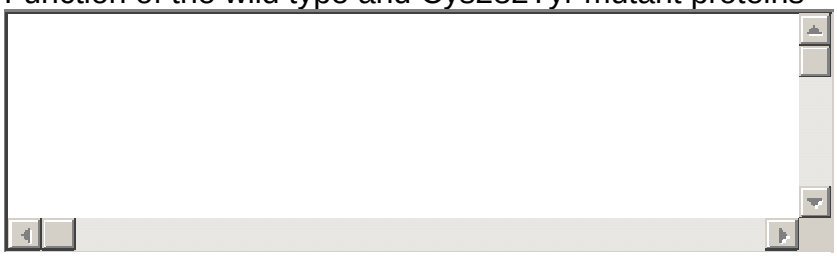
HFE nonsynonymous SNPs



Step 4. Determine whether the mutant HFE gene causes a phenotype:

There is a link to the OMIM database for three of the non-synonymous SNPs noted above in the SNP report. Click on any of the "Yes" links to take us to the OMIM report for the HFE gene. This report details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the "Allelic Variant" link. Is Cys282Tyr listed as causing any phenotype? Use the Links menu to go back to the Gene report. Based on the RefSeq summary and the PubMed articles in the GeneRIFs section, describe the biochemical function of the wild type and Cys282Tyr mutant proteins.

Function of the wild type and Cys282Tyr mutant proteins



Step 5.Elucidate the biochemical and structural basis for the function of the Cys282Tyr mutant protein:

Go back to the Entrez Gene report. Click on the NP_000401 protein link and then on Blink. Click on the 3D structures button. The output contains a list of similar proteins with known 3D structures. The first entry, 1DE4G, represents the G chain of the hemochromatosis protein (complexed with transferrin receptor). Click on the blue dot next to 1DE4G to get the sequence alignment of the query protein to the G chain of 1DE4. Click on the "Get 3D Structure data" button. This downloads the structure of the G chain of 1DE4 and its sequence alignment with the query protein. Zoom in the area of the disulphide bridge (colored tan) by pressing "z" on the keyboard. Select the cysteine residues forming the disulphide bridge by double clicking on them. Mouse over the corresponding cysteine residues on the query line (gi 4504377) in the alignment and view the amino acid number at the bottom left of the window. One of them is the cysteine at position 282. It is the same cysteine which is mutated to tyrosine causing the hemochromatosis phenotype.

You can now easily explain why the Cys282Tyr mutant has an altered function.

Summary:

This mini-course describes steps to identify the gene expressing the ESTs obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene, find SNP-associated phenotypes, and view a 3D structure of HFE to see how the mutation might affect the protein.

Step 1: The query EST sequence was found to align to contig NT_007592.13 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

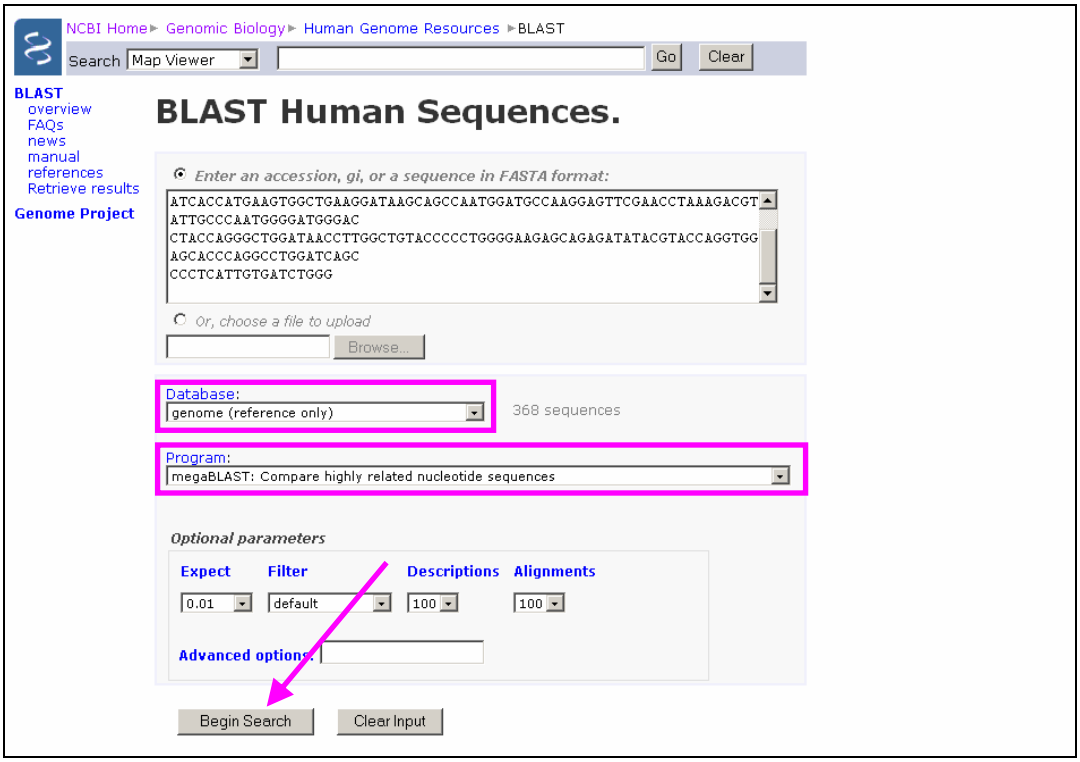
Step 2: The query EST was found to be expressed by the HFE gene.

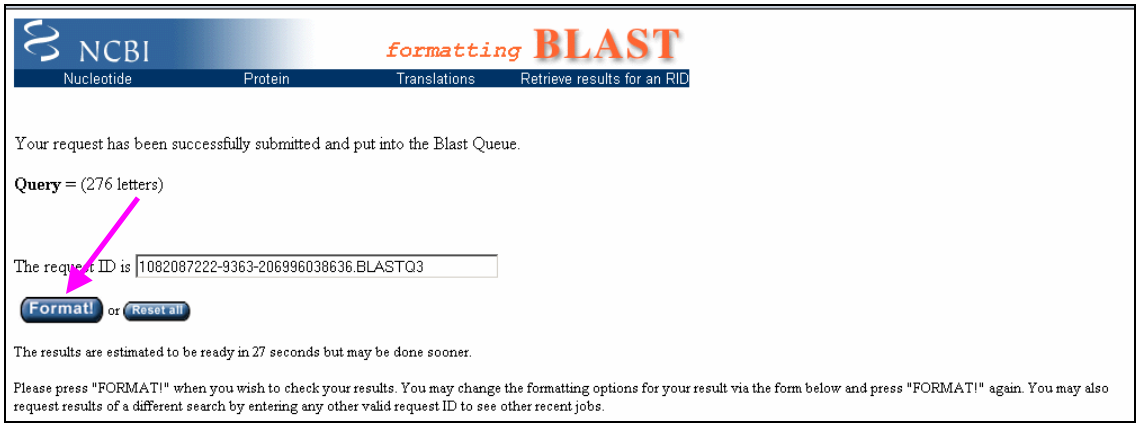
Step 3: The HFE gene has at least 11 alternatively spliced products. The query EST sequence contains a known SNP leading to the change in the translation of the 282nd amino acid from cysteine to tyrosine.

Step 4: Mutations in the HFE gene are associated with hemochromatosis. The HFE protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin, whereas the Cys282Tyr mutant fails to regulate this interaction leading to iron overload.

Step 5: The conserved cysteine 282 in the immunoglobulin constant region domain in the HFE protein is involved in formation of a disulphide bridge. Its mutation to tyrosine will alter the folding of the protein.

Step 1. Compare ESTs to the human genome:





NCBI results of BLAST

NCBI Entrez Genome

BLASTN 2.2.14 [Apr-09-2006]

```
RID: 1147199916-12575-117415596313.BLASTQ4


Database: ref_contig
          368 sequences; 2,870,843,926 total letters

If you have any problems or questions with the results of this search
please refer to the BLAST FAQs
```

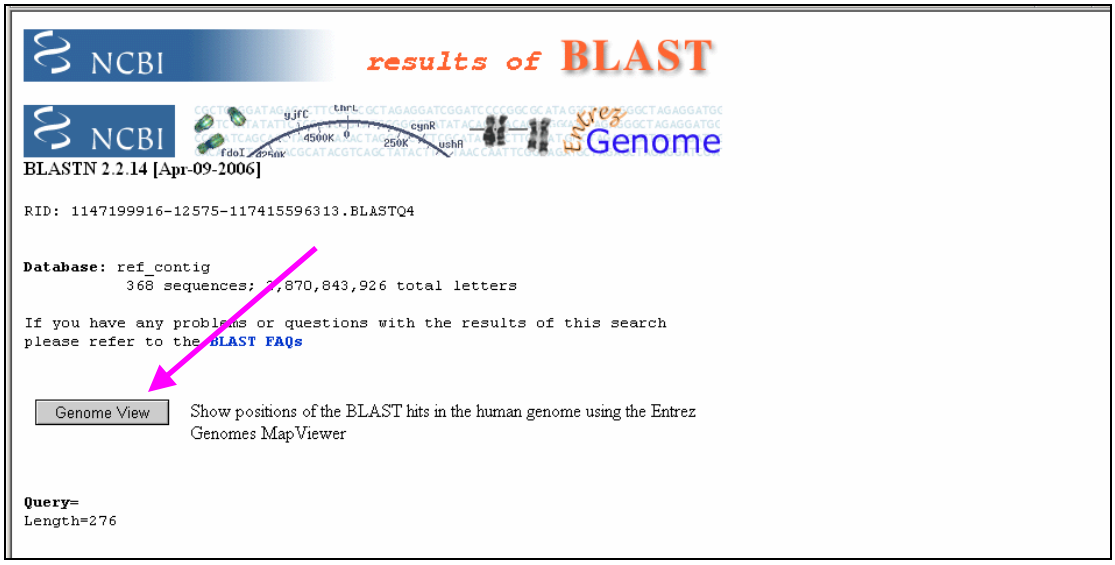
Genome View — Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

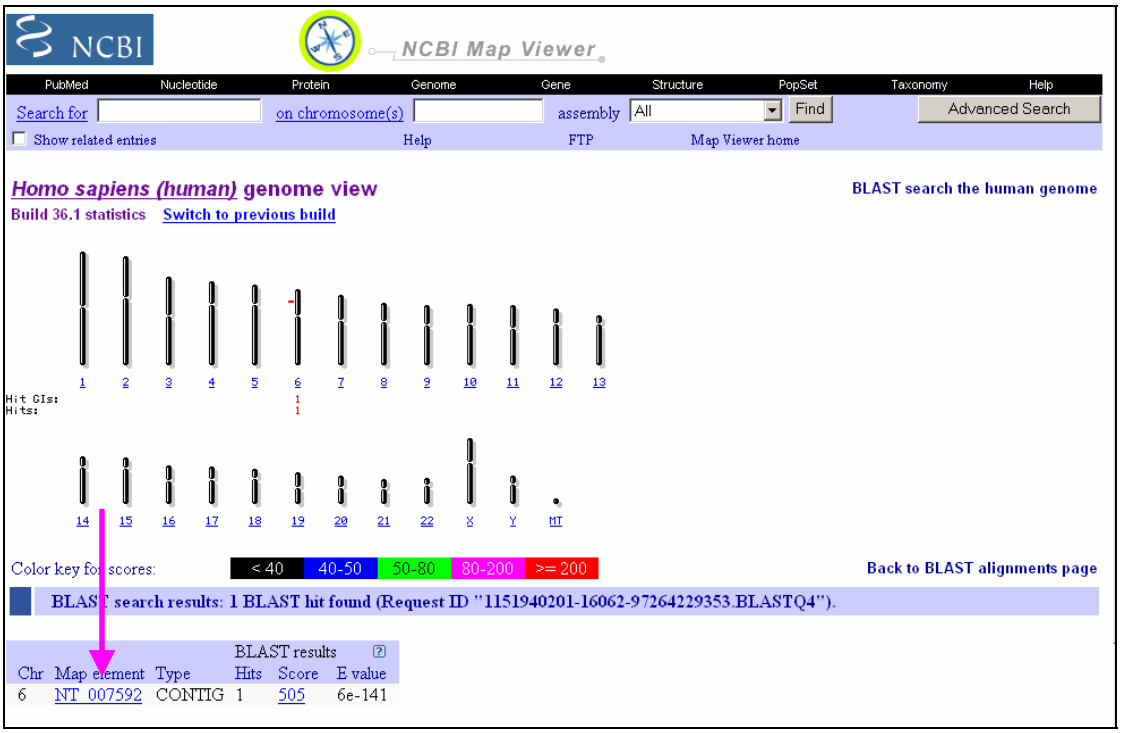
```
Query=
Length=276
```

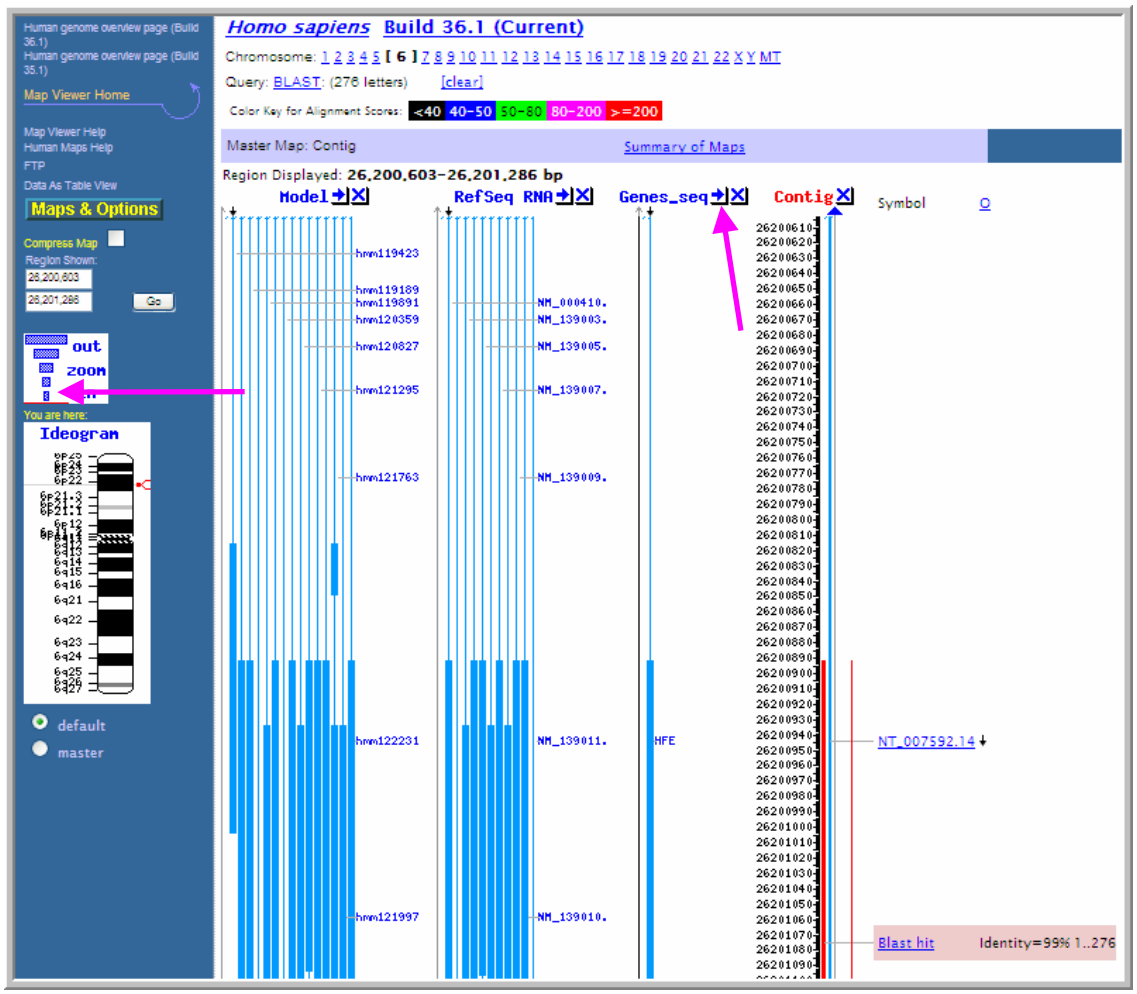
```
>ref|NT_007592.14|Hs6_7749 Homo sapiens chromosome 6 genomic contig
          Length=48945890

 Features in this part of subject sequence:
   hemochromatosis protein isoform 11 precursor
   hemochromatosis protein isoform 10 precursor


 Score =  525 bits (273),  Expect = 4e-147
 Identities = 275/276 (99%), Gaps = 0/276 (0%)
 Strand=Plus/Plus

Query  1         TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16951164  TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC  16951223

Query  61        GGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16951224  GGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA  16951283

Query  121       TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16951284  TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG  16951343

Query  181       GCTGGATAACCTTGGCTGTACCCCCTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGC  240
                 ||||||||||||||||||||||||||||||||||||||||||||| ||| |||| ||||||
Sbjct  16951344  GCTGGATAACCTTGGCTGTACCCCCTGGGGAAGAGCAGAGATATACGTGCCAGGTGGAGC  16951403

Query  241       ACCCAGGCCTGGATCAGCCCCTCATTGTGATCTGGG  276
                 |||||||||||||||||||||||||||||||||||
Sbjct  16951404  ACCCAGGCCTGGATCAGCCCCTCATTGTGATCTGGG  16951439
```

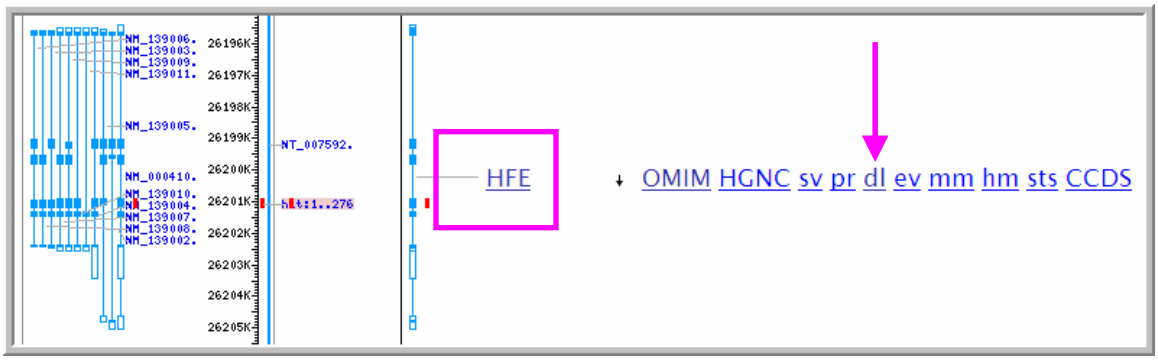
Step 1 result: The query EST sequence aligned to the contig, NT_007592.14, on chromosome 6 with one nucleotide difference (G to A with respect to nucleotide 16951392 on the contig).

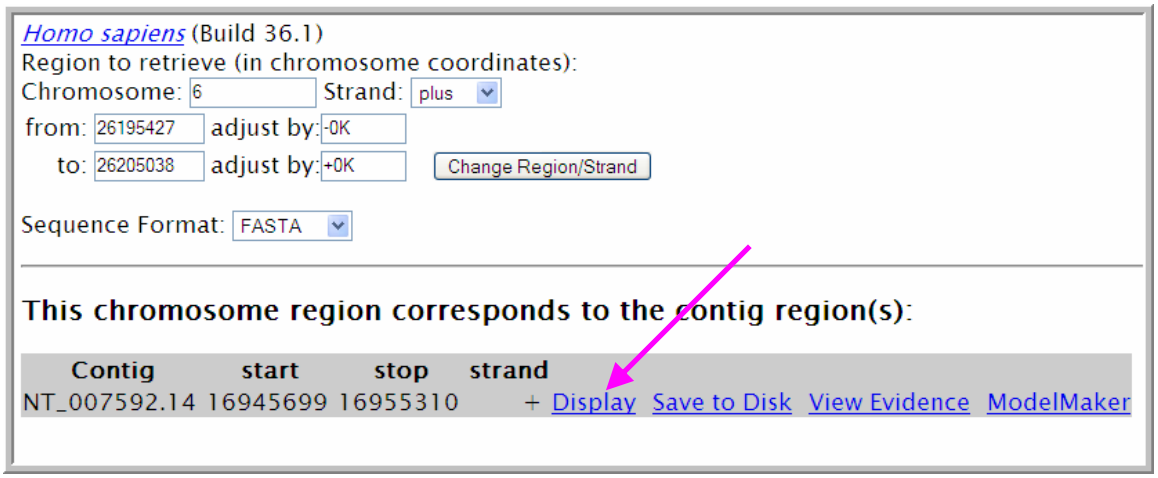
Step 2. Identify the gene(s) expressing the EST and download the sequence:





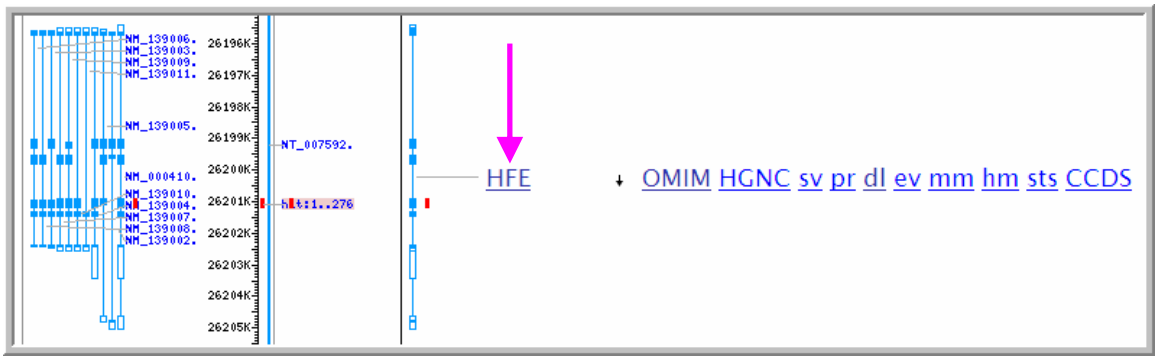

Homo sapiens Build 36.1 (Current)
Chromosome: 1 2 3 4 5 [6] 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Contig
Summary of Maps
Region Displayed: 26,200,603-26,201,286 bp
Model
RefSeq RNA
Genes_seq
Contig
Symbol
Map Viewer Home
Maps & Options
Ideogram
HFE
NT_007592.14
Blast hit
Identity=99% 1..276


HFE
OMIM HGNC sv pr dl ev mm hm sts CCDS



Step 2 results: The query EST was expressed by the HFE gene; the "dl" link can be used to download any sequence on the chromosome.

Step 3: Determine what is known about the gene and protein, and identify known SNPs in the gene



NCBI Entrez Gene

1: HFE hemochromatosis [*Homo sapiens*]

GeneID: 3077 updated 06-Oct-2006

Summary

Official Symbol	HFE
Official Full Name	hemochromatosis
Primary source	HGNC:4886
See related	HPRD:01993; MIM:235200
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; MGC103790; dJ221C16.10.1
Summary	The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (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 eleven alternatively spliced variants have been described for this gene. Additional variants have been found but their full-length nature has not been determined.

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic context

chromosome: 6; Location: 6p21.3

See HFE in MapViewer

Bibliography

Related Articles in Pubmed

PubMed links

GeneRIFs: Gene References Into Function

What's a GeneRIF?

1. Apical distribution of HFE-beta2-microglobulin is associated with inhibition of apical iron uptake in intestinal epithelia cells.
2. determined race-specific frequencies of the HFE mutations, C282Y and H63D
3. Glucose intolerance may be important risk factor for the development of hepatic fibrosis in subjects with the C282Y/H63D HFE genotype.
4. Potential interaction between HFE genotypes and heme iron intake in relation to the risk of type 2 diabetes
5. present data indicate thet there is a link between HFE mutations and idiopathic dilated cardiomyopathy due to a high proportion of patients being either C282Y homozygotes or C28Y/H63D compound heterozygotes
6. The relatively high frequency of HFE H63D allele carriers in Egyptian cases of HCV liver cirrhosis suggests the need to routinely screen for this risk genotype among affected families.
7. a population harboring abnormal iron markers had a different HFE genotype and a higher 282Y allele frequency than the control population
8. HFE mutations do not directly alter erythroid iron handling, but alter the supply of iron to the erythroid tissues

Interactions

Description .....

Product	Interactant	Other Gene	Complex	Source	Pubs
NP_000401.1	Beta 2 microglobulin	B2M		HPRD	PubMed
NP_000401.1	Transferrin receptor 2	TFR2		HPRD	PubMed
NP_000401.1	NP_003225.1	TFRC		HPRD	PubMed

GeneOntology

Provided by GOA

Function	Evidence	
MHC class I receptor activity	IEA	
iron ion binding	IEA	

Process	Evidence	
antigen processing and presentation	IEA	
antigen processing and presentation of endogenous antigen	IEA	
antigen processing and presentation of endogenous peptide antigen via MHC class I	IEA	
ion transport	IEA	
iron ion homeostasis	TAS	Pubmed
iron ion transport	TAS	Pubmed
protein complex assembly	TAS	Pubmed
receptor mediated endocytosis	TAS	Pubmed

Component	Evidence	
MHC class I protein complex	IEA	
cytoplasm	TAS	Pubmed
integral to plasma membrane	TAS	Pubmed
plasma membrane	TAS	Pubmed

General protein information

Names

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

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. Explain

Genomic

1. **NG_001335.1 Reference**
   - Range: 71162..80773
   - Download: GenBank FASTA

mRNA and Protein(s)

1. **NM_000410.3→NP_000401.1 hemochromatosis protein isoform 1 precursor**
   - Description: Transcript Variant: This variant (1) encodes the longest isoform.
   - Source sequence(s): AF115265,AJ249337,U91328
   - Consensus CDS: CCDS4578.1
   - Conserved Domains (2): summary
     - cd00098 — Location:223-298, Blast Score:169 — IGc; Immunoglobulin domain constant region subfamily; members of the IGc subfamily are components of immunoglobulins, T-cell receptors, CD1 cell surface glycoproteins, secretory glycoproteins A/C, and Major Histocompatibility Complex (MHC) class I/II molecules
     - pfam00129 — Location:27-202, Blast Score:314 — MHC_I; Class I Histocompatibility antigen, domains alpha 1 and 2
2. **NM_139002.2→NP_620571.1 hemochromatosis protein isoform 2 precursor**
   - Description: Transcript Variant: This variant (2) lacks a large 3' region including the 3' CDS and UTR but has an alternate 3' exon, as compared to variant 1. The resulting protein (isoform 2) has a unique carboxy terminus.
   - Source sequence(s): AF115264,AJ249337,U91328
   - Conserved Domains (1): summary
     - pfam00129 — Location:27-143, Blast Score:207 — MHC_I; Class I Histocompatibility antigen, domains alpha 1 and 2
3. **NM_139003.2→NP_620572.1 hemochromatosis protein isoform 3 precursor**
   - Description: Transcript Variant: This variant (3) lacks an internal in-frame segment of the coding region, as compared to variant 1, resulting in a shorter protein (isoform 3).
   - Source sequence(s): AF149804,AJ249335,U91328
   - Conserved Domains (2): summary
     - cd00098 — Location:107-192, Blast Score:175 — IGc; Immunoglobulin domain constant region subfamily; members of the IGc subfamily are components of immunoglobulins, T-cell receptors, CD1 cell surface glycoproteins, secretory glycoproteins A/C, and Major Histocompatibility Complex (MHC) class I/II

Related Sequences

Nucleotide		Protein
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	CS187189.1	CAJ42862.1
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
mRNA	AF079408.1	AAC62647.1
mRNA	AF079409.1	AAC62648.1
mRNA	AF109385.1	AAD52104.1
mRNA	AF115264.1	AAG29571.1
mRNA	AF115265.1	AAG29572.1
mRNA	AF144238.1	AAG29573.1
mRNA	AF144239.1	AAG29574.1
mRNA	AF144240.1	AAG29575.1
mRNA	AF144241.1	AAG29576.1
mRNA	AF144242.1	AAG29577.1
mRNA	AF144243.1	AAG29578.1
mRNA	AF144244.1	AAG29579.1

Bibliography

Related Articles in Pubmed

PubMed links

GeneRIFs: Gene References Into Function

What's a GeneRIF?

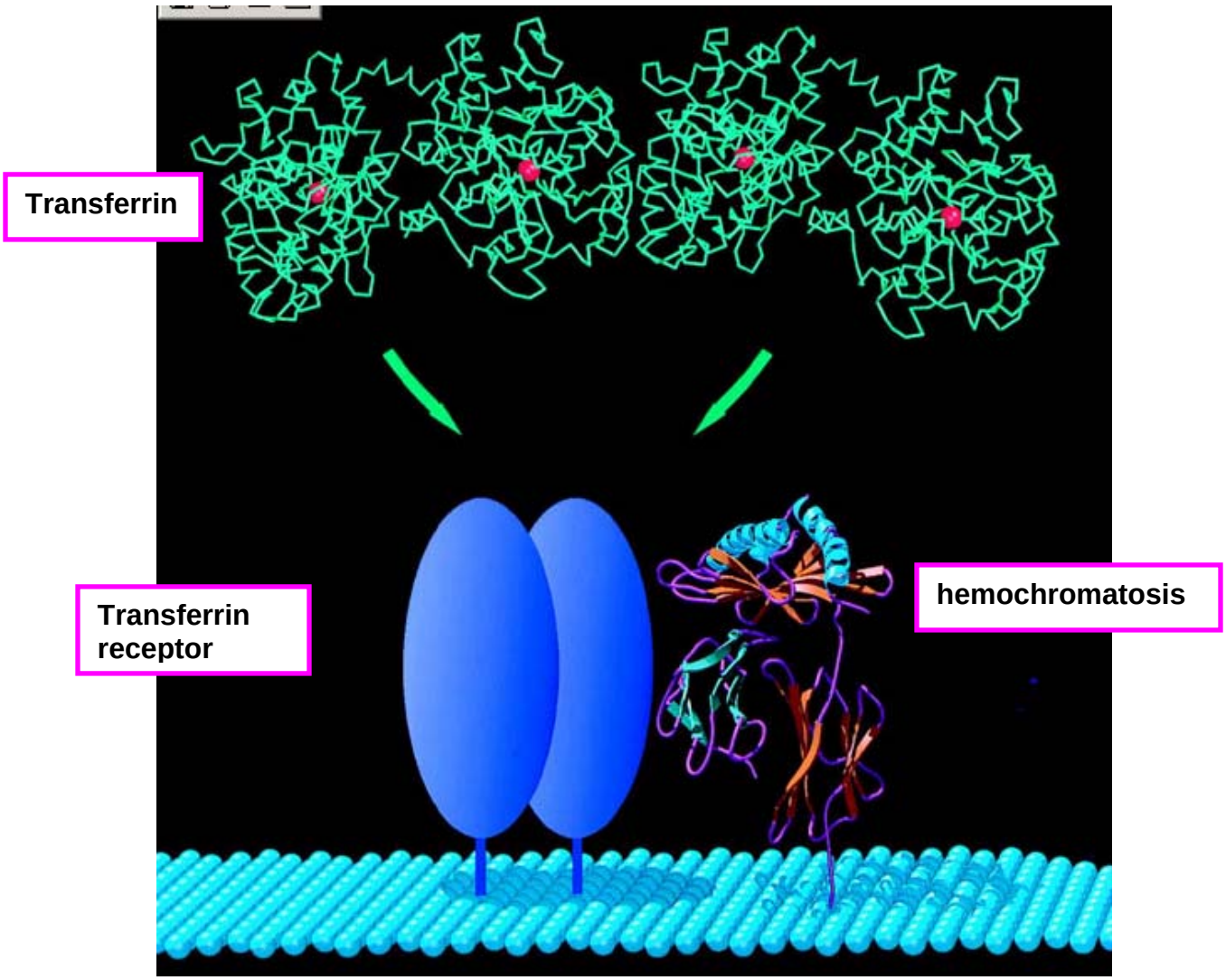
101. Individuals with mutations in the HFE gene show very few hemochromatosis-related symptoms.
102. Association between MHC class I gene HFE polymorphisms and longevity
103. HFE gene implicated in this disorder has been identified on chromosome 6. the most prevalent mutation is a point mutation(histidine to aspartic acid)in iron overload has been controversial.
104. Mutations in HFE protein results in hemochromatosis
105. results suggest that wild-type HFE negatively modulates the endocytic uptake of transferrin
106. genotype and allele frequencies between neonates and referred patients for HFE molecular analysis
107. mutational analysis of the transferrin receptor reveals overlapping HFE and transferrin binding sites
108. REVIEW: C282Y mutant gene product failed to associate with 2-microglobulin and significantly reduced cell surface expression of the HFE-2m complex, thereby affecting the interaction with TfR and its interaction with transferrin.
109. 871 healthy unrelated subjects in Poland were collected to assess the relevant frequencies. Each subject was genotyped for the C282Y and H63D mutations using a PCR-based protocol

Submit: New GeneRIF Correction

Interactions

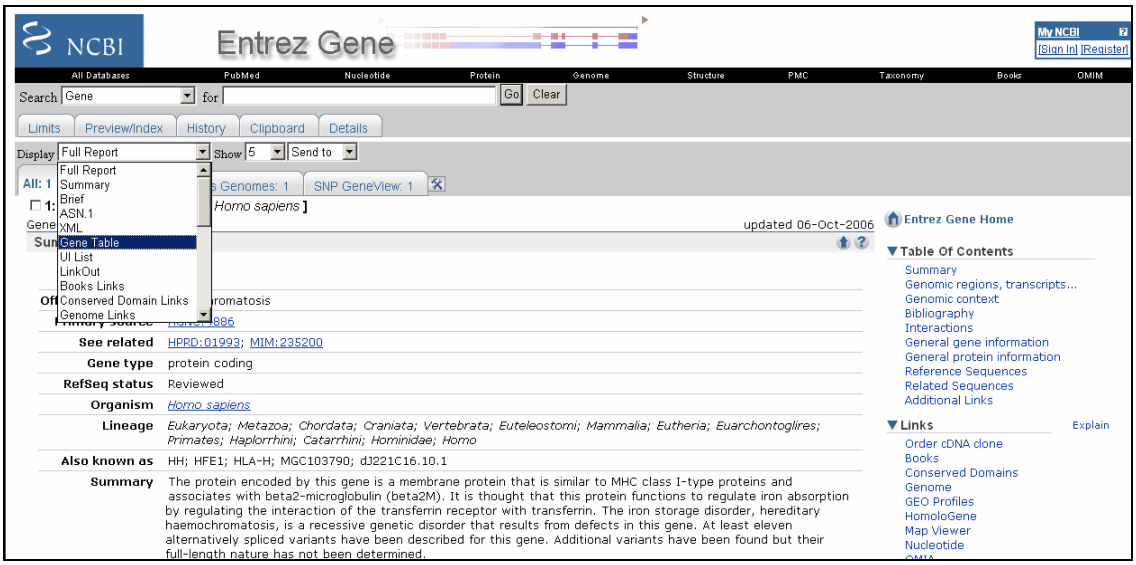
Description .....

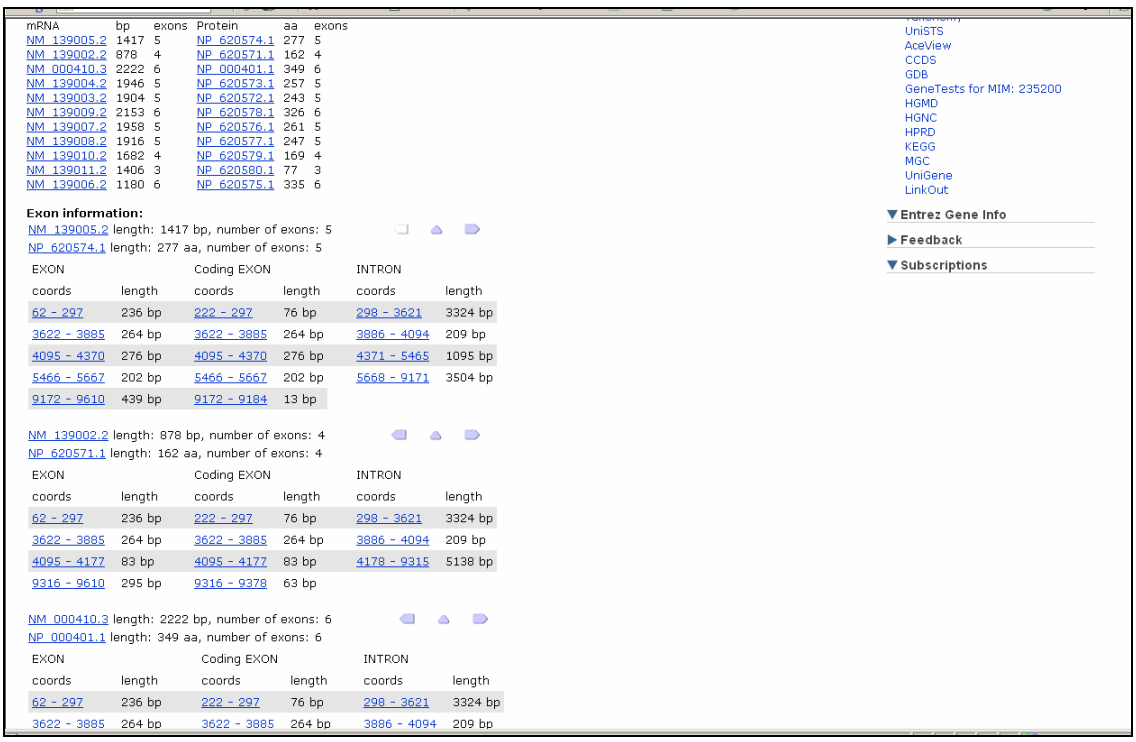
Product	Interactant	Other Gene	Complex	Source	Pubs
NP_000401.1	Beta 2 microglobulin	B2M		HPRD	PubMed



Bacon et al. Gastroenterology, 116:193-207, Figure 4

The hemochromatosis protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin.


NCBI
Entrez Gene
My NCBI
[Sign In] [Register]
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
Books
OMIM
Search Gene for Go Clear
Limits
Preview/Index
History
Clipboard
Details
Display Full Report Show 5 Send to
Full Report
Summary
Brief
ASN.1
XML
Gene Table
UI List
LinkOut
Books Links
Conserved Domain Links
Genome Links
All: 1
Genomes: 1
SNP GeneView: 1
Homo sapiens]
updated 06-Oct-2006
Entrez Gene Home
Table Of Contents
Summary
Genomic regions, transcripts...
Genomic context
Bibliography
Interactions
General gene information
General protein information
Reference Sequences
Related Sequences
Additional Links
Links
Explain
Order cDNA clone
Books
Conserved Domains
Genome
GEO Profiles
HomoloGene
Map Viewer
Nucleotide
See related HPRD:01993; MIM:235200
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; MGC103790; dJ221C16.10.1
Summary The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (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 eleven alternatively spliced variants have been described for this gene. Additional variants have been found but their full-length nature has not been determined.


mRNA bp exons Protein aa exons
NM_139005.2 1417 5 NP_620574.1 277 5
NM_139002.2 878 4 NP_620571.1 162 4
NM_000410.3 2222 6 NP_000401.1 349 6
NM_139004.2 1946 5 NP_620573.1 257 5
NM_139003.2 1904 5 NP_620572.1 243 5
NM_139009.2 2153 6 NP_620578.1 326 6
NM_139007.2 1958 5 NP_620576.1 261 5
NM_139008.2 1916 5 NP_620577.1 247 5
NM_139010.2 1682 4 NP_620579.1 169 4
NM_139011.2 1406 3 NP_620580.1 77 3
NM_139006.2 1180 6 NP_620575.1 335 6
Exon information:
NM_139005.2 length: 1417 bp, number of exons: 5
NP_620574.1 length: 277 aa, number of exons: 5
EXON Coding EXON INTRON
coords length coords length coords length
62 - 297 236 bp 222 - 297 76 bp 298 - 3621 3324 bp
3622 - 3885 264 bp 3622 - 3885 264 bp 3886 - 4094 209 bp
4095 - 4370 276 bp 4095 - 4370 276 bp 4371 - 5465 1095 bp
5466 - 5667 202 bp 5466 - 5667 202 bp 5668 - 9171 3504 bp
9172 - 9610 439 bp 9172 - 9184 13 bp
NM_139002.2 length: 878 bp, number of exons: 4
NP_620571.1 length: 162 aa, number of exons: 4
EXON Coding EXON INTRON
coords length coords length coords length
62 - 297 236 bp 222 - 297 76 bp 298 - 3621 3324 bp
3622 - 3885 264 bp 3622 - 3885 264 bp 3886 - 4094 209 bp
4095 - 4177 83 bp 4095 - 4177 83 bp 4178 - 9315 5138 bp
9316 - 9610 295 bp 9316 - 9378 63 bp
NM_000410.3 length: 2222 bp, number of exons: 6
NP_000401.1 length: 349 aa, number of exons: 6
EXON Coding EXON INTRON
coords length coords length coords length
62 - 297 236 bp 222 - 297 76 bp 298 - 3621 3324 bp
3622 - 3885 264 bp 3622 - 3885 264 bp 3886 - 4094 209 bp
UniSTS
AceView
CCDS
GDB
GeneTests for MIM: 235200
HGMD
HGNC
HPRD
KEGG
MGC
UniGene
LinkOut
Entrez Gene Info
Feedback
Subscriptions

NCBI Entrez Gene

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books | OMIM

Search Gene for Go Clear

Limits | Preview/Index | History | Clipboard | Details

Display Full Report Show 5 Send to

All: 1 | Current Only: 1 | Genes Genomes: 1 | SNP GeneView: 1

1: HFE hemochromatosis [*Homo sapiens*]

GeneID: 3077 updated 06-Oct-2006

Summary

Official Symbol	HFE
Official Full Name	hemochromatosis
Primary source	HGNC:4886
See related	HPRD:01993; MIM:235200
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; MGC103790; dJ221C16.10.1
Summary	The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (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 eleven alternatively spliced variants have been described for this gene. Additional variants have been found but their full-length nature has not been determined.

Genomic regions, transcripts, and products

Go to reference sequence details

NC_000006.10

Entrez Gene Home

Table Of Contents: Summary, Genomic regions, transcripts..., Genomic context, Bibliography, Interactions, General gene information, General protein information, Reference Sequences, Related Sequences, Additional Links

Links (Explain): Order cDNA clone, Books, Conserved Domains, Genome, GEO Profiles, HomoloGene, Map Viewer, Nucleotide, OMIA, OMIM, Full text in PMC, Probe, Protein, PubMed, PubMed (GeneRIF), SNP, SNP: Genotype, SNP: GeneView

NCBI Single Nucleotide Polymorphism

PubMed | Nucleotide | Protein | Genome | Structure | PopSet | Taxonomy | OMIM | Books | SNP

Search SNP for Go

BUILD 126

GENERAL: Contact Us, dbSNP Homepage, SNP Science Primer, Announcements, dbSNP Summary, FTP Download, Build History

SNP SUBMISSION: How to Submit, Handle Request

DOCUMENTATION: Build Release Note, FAQ General, Search FAQ Archive, updated 2006, dbSNP Handbook Overview, RefSNP Summary Info, Database Schema, Database Dictionary NEW, Database Changes NEW, Genotype Schema, Data Formats, Heterozygosity Computation

SEARCH

SNP linked to Gene (geneID:3077)

SNP are linked from gene HFE via the following methods:

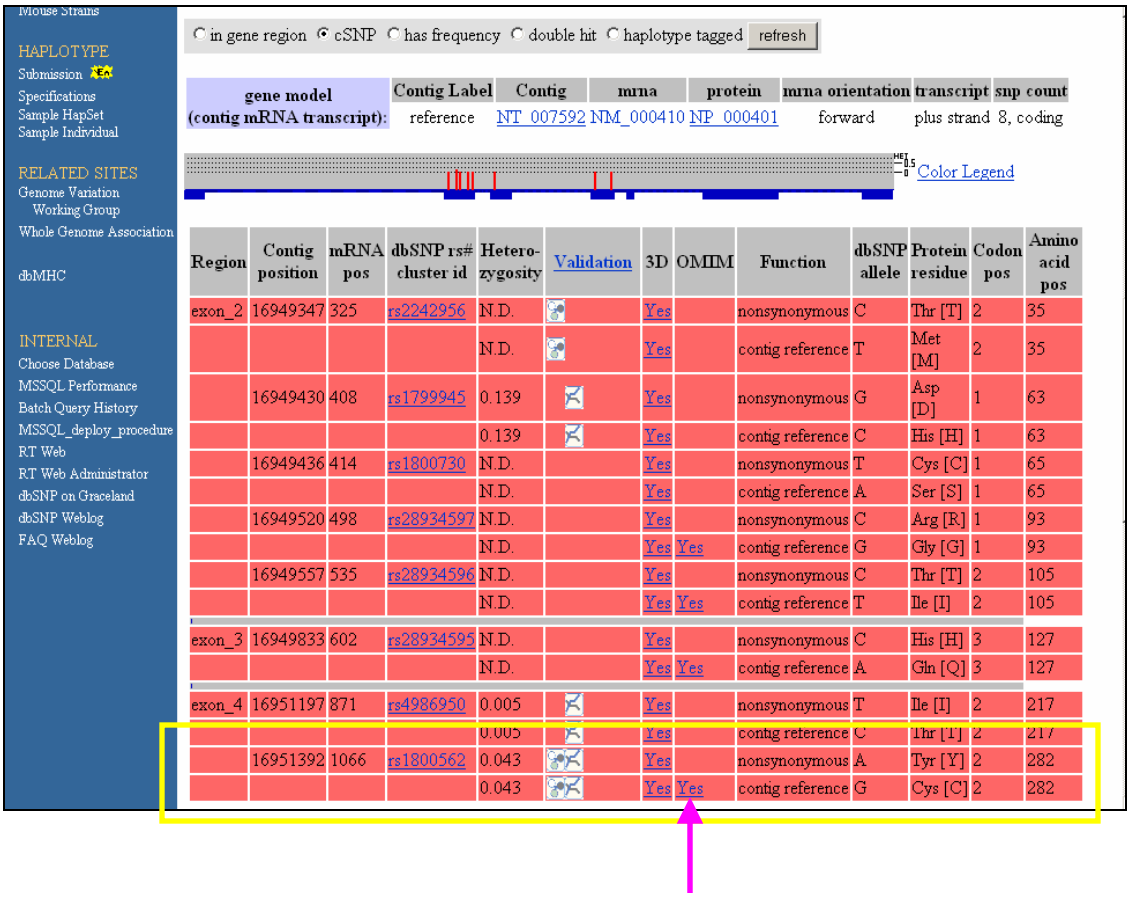
Contig Annotation | GenBank(mrna) Mapping

Send all rs# to Batch Query | Download all rs# to file. | Gene Genotype Report

Gene Model (mRNA alignment) information from genome sequence

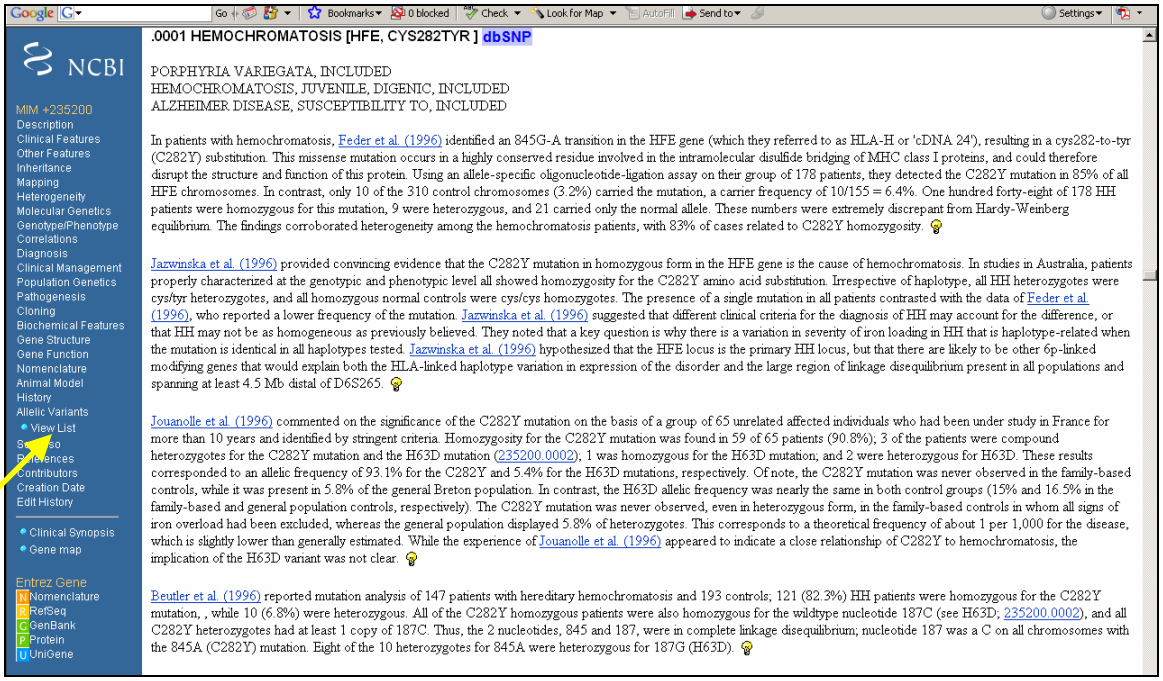
Total gene model (contig mRNA transcript): 22

mrna	transcript	protein	mrna orientation	Contig	Contig Label	snp list
NM_000410	plus strand	NP_000401	forward	NT_007592	reference	currently shown
NM_000410	plus strand	NP_000401	forward	NW_922984	Celera	view
NM_139002	plus strand	NP_620571	forward	NT_007592	reference	view
NM_139002	plus strand	NP_620571	forward	NW_922984	Celera	view
NM_139003	plus strand	NP_620572	forward	NT_007592	reference	view
NM_139003	plus strand	NP_620572	forward	NW_922984	Celera	view
NM_139004	plus strand	NP_620573	forward	NT_007592	reference	view
NM_139004	plus strand	NP_620573	forward	NW_922984	Celera	view
NM_139005	plus strand	NP_620574	forward	NT_007592	reference	view
NM_139005	plus strand	NP_620574	forward	NW_922984	Celera	view
NM_139006	plus strand	NP_620575	forward	NT_007592	reference	view
NM_139006	plus strand	NP_620575	forward	NW_922984	Celera	view
NM_139007	plus strand	NP_620576	forward	NT_007592	reference	view
NM_139007	plus strand	NP_620576	forward	NW_922984	Celera	view



Step 3 result: The HFE gene has at least 11 alternatively spliced products. The query EST sequence contains a known SNP leading to the change in the translation of the 282nd amino acid from cysteine to tyrosine.

Step 4: Determine whether the mutant HFE gene causes phenotypes:



.0001 HEMOCHROMATOSIS [HFE, CYS282TYR] dbSNP

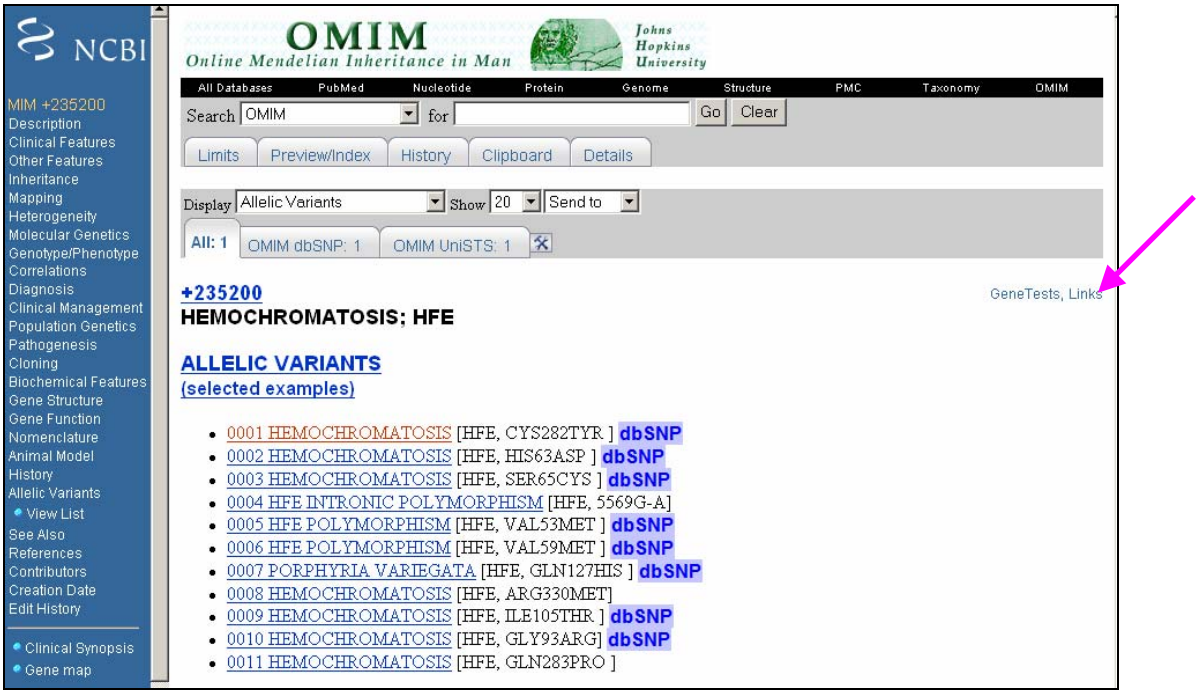
PORPHYRIA VARIEGATA, INCLUDED
HEMOCHROMATOSIS, JUVENILE, DIGENIC, INCLUDED
ALZHEIMER DISEASE, SUSCEPTIBILITY TO, INCLUDED

In patients with hemochromatosis, Feder et al. (1996) identified an 845G-A transition in the HFE gene (which they referred to as HLA-H or 'cDNA 24'), resulting in a cys282-to-tyr (C282Y) substitution. This missense mutation occurs in a highly conserved residue involved in the intramolecular disulfide bridging of MHC class I proteins, and could therefore disrupt the structure and function of this protein. Using an allele-specific oligonucleotide-ligation assay on their group of 178 patients, they detected the C282Y mutation in 85% of all HFE chromosomes. In contrast, only 10 of the 310 control chromosomes (3.2%) carried the mutation, a carrier frequency of 10/155 = 6.4%. One hundred forty-eight of 178 HH patients were homozygous for this mutation, 9 were heterozygous, and 21 carried only the normal allele. These numbers were extremely discrepant from Hardy-Weinberg equilibrium. The findings corroborated heterogeneity among the hemochromatosis patients, with 83% of cases related to C282Y homozygosity.

Jazwinska et al. (1996) provided convincing evidence that the C282Y mutation in homozygous form in the HFE gene is the cause of hemochromatosis. In studies in Australia, patients properly characterized at the genotypic and phenotypic level all showed homozygosity for the C282Y amino acid substitution. Irrespective of haplotype, all HH heterozygotes were cys/tyr heterozygotes, and all homozygous normal controls were cys/cys homozygotes. The presence of a single mutation in all patients contrasted with the data of Feder et al. (1996), who reported a lower frequency of the mutation. Jazwinska et al. (1996) suggested that different clinical criteria for the diagnosis of HH may account for the difference, or that HH may not be as homogeneous as previously believed. They noted that a key question is why there is a variation in severity of iron loading in HH that is haplotype-related when the mutation is identical in all haplotypes tested. Jazwinska et al. (1996) hypothesized that the HFE locus is the primary HH locus, but that there are likely to be other 6p-linked modifying genes that would explain both the HLA-linked haplotype variation in expression of the disorder and the large region of linkage disequilibrium present in all populations and spanning at least 4.5 Mb distal of D6S265.

Jouanolle et al. (1996) commented on the significance of the C282Y mutation on the basis of a group of 65 unrelated affected individuals who had been under study in France for more than 10 years and identified by stringent criteria. Homozygosity for the C282Y mutation was found in 59 of 65 patients (90.8%); 3 of the patients were compound heterozygotes for the C282Y mutation and the H63D mutation (235200.0002); 1 was homozygous for the H63D mutation; and 2 were heterozygous for H63D. These results corresponded to an allelic frequency of 93.1% for the C282Y and 5.4% for the H63D mutations, respectively. Of note, the C282Y mutation was never observed in the family-based controls, while it was present in 5.8% of the general Breton population. In contrast, the H63D allelic frequency was nearly the same in both control groups (15% and 16.5% in the family-based and general population controls, respectively). The C282Y mutation was never observed, even in heterozygous form, in the family-based controls in whom all signs of iron overload had been excluded, whereas the general population displayed 5.8% of heterozygotes. This corresponds to a theoretical frequency of about 1 per 1,000 for the disease, which is slightly lower than generally estimated. While the experience of Jouanolle et al. (1996) appeared to indicate a close relationship of C282Y to hemochromatosis, the implication of the H63D variant was not clear.

Beutler et al. (1996) reported mutation analysis of 147 patients with hereditary hemochromatosis and 193 controls; 121 (82.3%) HH patients were homozygous for the C282Y mutation, , while 10 (6.8%) were heterozygous. All of the C282Y homozygous patients were also homozygous for the wildtype nucleotide 187C (see H63D; 235200.0002), and all C282Y heterozygotes had at least 1 copy of 187C. Thus, the 2 nucleotides, 845 and 187, were in complete linkage disequilibrium; nucleotide 187 was a C on all chromosomes with the 845A (C282Y) mutation. Eight of the 10 heterozygotes for 845A were heterozygous for 187G (H63D).



OMIM Online Mendelian Inheritance in Man — Johns Hopkins University

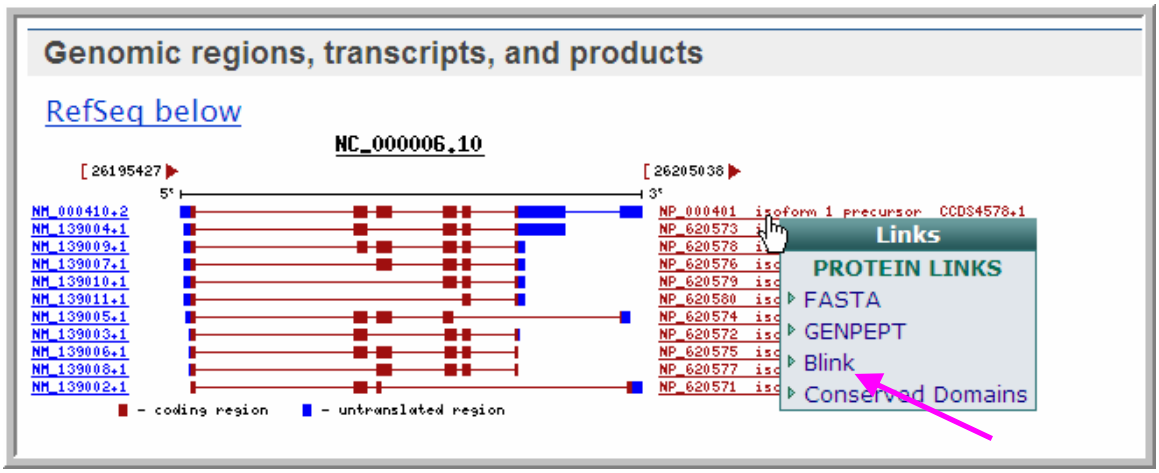
+235200
HEMOCHROMATOSIS; HFE

ALLELIC VARIANTS
(selected examples)

- 0001 HEMOCHROMATOSIS [HFE, CYS282TYR] dbSNP
- 0002 HEMOCHROMATOSIS [HFE, HIS63ASP] dbSNP
- 0003 HEMOCHROMATOSIS [HFE, SER65CYS] dbSNP
- 0004 HFE INTRONIC POLYMORPHISM [HFE, 5569G-A]
- 0005 HFE POLYMORPHISM [HFE, VAL53MET] dbSNP
- 0006 HFE POLYMORPHISM [HFE, VAL59MET] dbSNP
- 0007 PORPHYRIA VARIEGATA [HFE, GLN127HIS] dbSNP
- 0008 HEMOCHROMATOSIS [HFE, ARG330MET]
- 0009 HEMOCHROMATOSIS [HFE, ILE105THR] dbSNP
- 0010 HEMOCHROMATOSIS [HFE, GLY93ARG] dbSNP
- 0011 HEMOCHROMATOSIS [HFE, GLN283PRO]

Step 4 result: Mutations in the HFE gene are associated with hemochromatosis. The HFE protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin, whereas the Cys282Tyr mutant fails to regulate this interaction, leading to iron overload.

Step 5: Elucidate the biochemical and structural basis for the function of the mutant protein:



NCBI

BLAST | Protein | Structure | PubMed | Taxonomy
Genome | Nucleotide | 3D-Domains | Books | Help

Query: gi|4504377 hemochromatosis protein isoform 1 precursor [Homo sapiens]
Matching gi: 109658670, 109658506, 83323630, 80748852, 57114069, 38502807, 29709343, 22854810, 20250786, 15115850, 14100030, 11094315, 2497915, 2370111, 2088551, 1890180, 1469790

Hide identical | Best hits | Common Tree | Taxonomy Report | 3D structures | CDD-Search | GI list | Run BLAST

200 BLAST hits to 25 unique species Sort by taxonomy proximity

0 Archaea | 0 Bacteria | 199 Metazoa | 0 Fungi | 0 Plants | 0 Viruses | 0 Other Eukaryotae

Keep only | Cut-Off 100 | Select | Reset | New search by GI: 4504377 | Go

348 aa

SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
Conserved Domain Database hits				
1870	31	AAI17202	109658670	Hemochromatosis [Homo sapiens]
1870	31	AAI17204	109658506	Hemochromatosis [Homo sapiens]
1870	29	NP_001...	57114069	hemochromatosis protein [Pan troglodytes]
1870	29	P60018	38502807	Hereditary hemochromatosis protein homolog precursor (HLA-H)
1870	29	AAN09793	22854810	hereditary hemochromatosis [Pan troglodytes]
1870	31	AAG29572	11094315	hemochromatosis termination variant terE6; HFE [Homo sapiens]
1870	31	Q30201	2497915	Hereditary hemochromatosis protein precursor (HLA-H)
1870	31	CAA70934	2370111	HFE [Homo sapiens]
1870	31	AAB82083	2088551	hereditary hemochromatosis [Homo sapiens]
1870	31	CAB07442	1890180	HFE [Homo sapiens]
1870	31	AAC51823	1469790	HLA-H
1776	31	AAH74721	50960016	HFE protein [Homo sapiens]
1772	31	NP_620575	21040347	hemochromatosis protein isoform 6 precursor [Homo sapiens]
1772	31	AAC62646	3695107	hemochromatosis splice variant del14E4 [Homo sapiens]
1713	31	NP_620578	21040353	hemochromatosis protein isoform 9 precursor [Homo sapiens]
1713	31	CAC67792	15485419	hemochromatosis protein [Homo sapiens]
1625	26	XP_001...	109069870	PREDICTED: hemochromatosis isoform 6 [Macaca mulatta]
1517	31	1DE4G	6980500	Chain G, Hemochromatosis Protein Hfe Complexed With Transferrin
1517	31	1DE4D	6980497	Chain D, Hemochromatosis Protein Hfe Complexed With Transferrin
1517	31	1DE4A	6980494	Chain A, Hemochromatosis Protein Hfe Complexed With Transferrin

NCBI

BLAST | Protein | Structure | PubMed | Taxonomy
Genome | Nucleotide | 3D-Domains | Books | Help

Query: gi|4504377 hemochromatosis protein isoform 1 precursor [Homo sapiens]
Matching gi: 109658670, 109658506, 83323630, 80748852, 57114069, 38502807, 29709343, 22854810, 20250786, 15115850, 14100030, 11094315, 2497915, 2370111, 2088551, 1890180, 1469790

Get Cn3D Now!

Hide identical | Best hits | Common Tree | Taxonomy Report | 3D structures | CDD-Search | GI list | Run BLAST

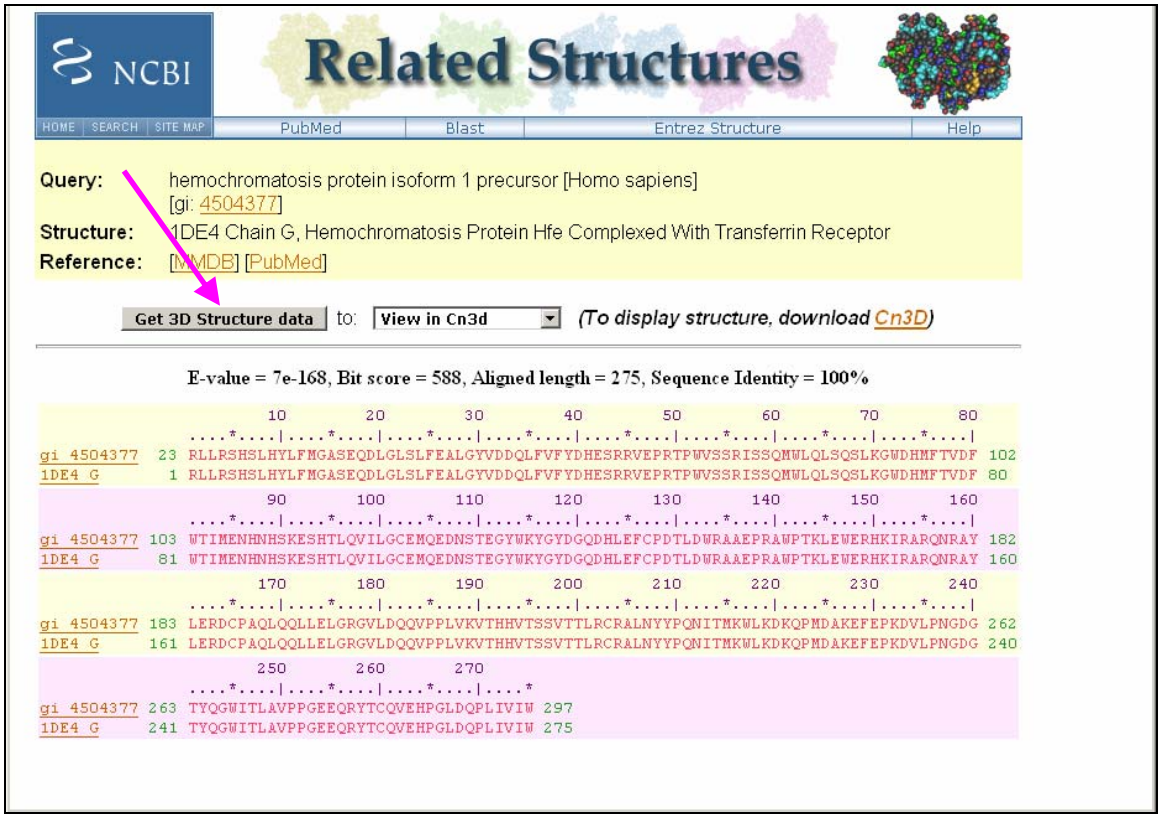
200 BLAST hits to 3 unique species Sort by taxonomy proximity

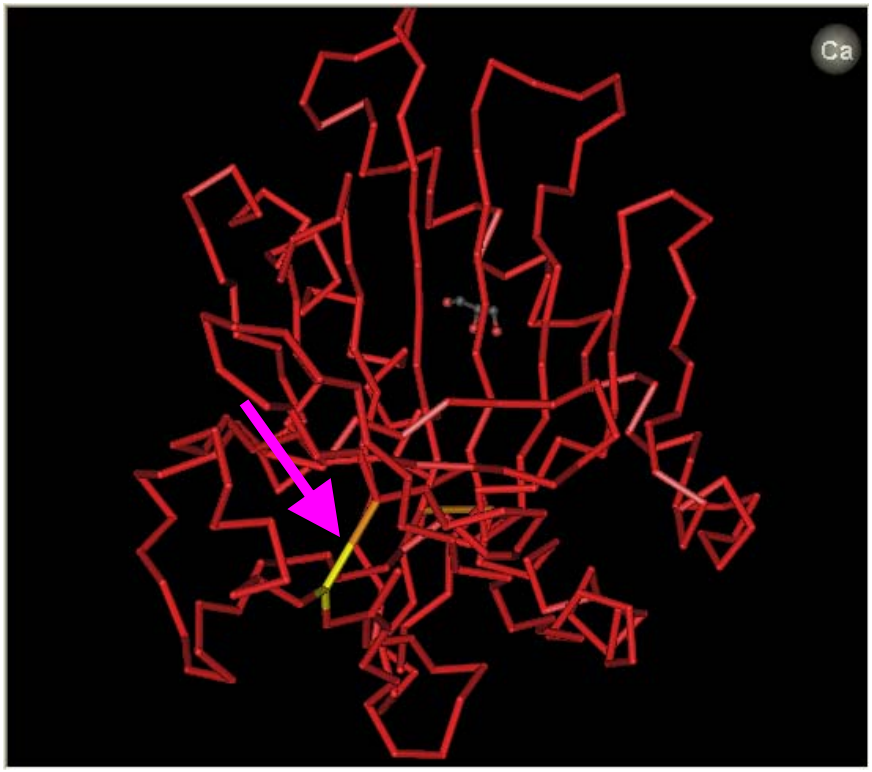
0 Archaea | 0 Bacteria | 200 Metazoa | 0 Fungi | 0 Plants | 0 Viruses | 0 Other Eukaryotae

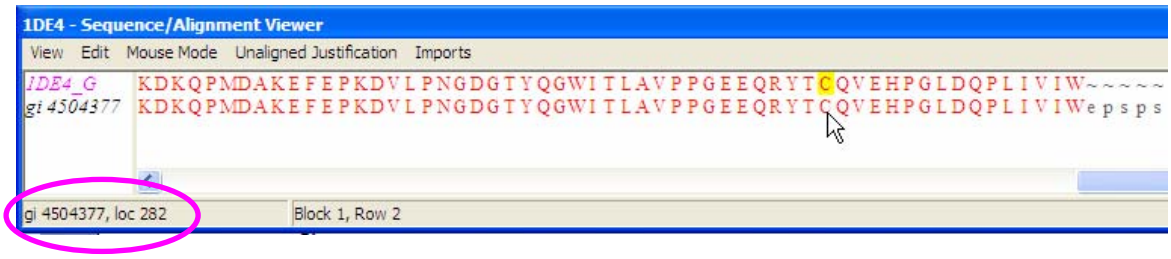
Keep only | Cut-Off 100 | Select | Reset | New search by GI: 4504377 | Go

348 aa

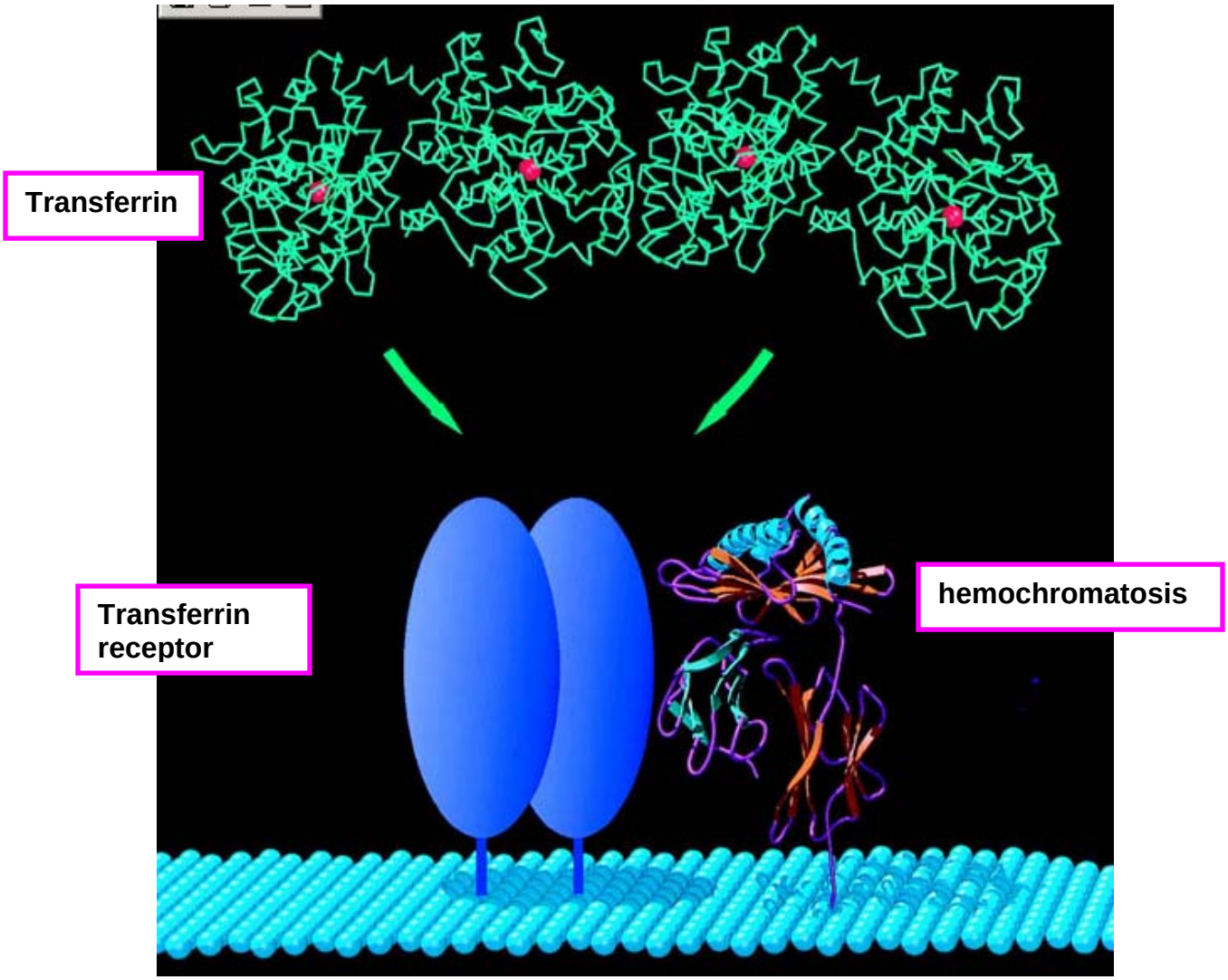
SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
Conserved Domain Database hits				
1517	●	1DE4G	6980500	Chain G, Hemochromatosis Protein Hfe Complexed With Transferrin F
1517	●	1DE4D	6980497	Chain D, Hemochromatosis Protein Hfe Complexed With Transferrin F
1517	●	1DE4A	6980494	Chain A, Hemochromatosis Protein Hfe Complexed With Transferrin F
1517	●	1A6ZC	4699712	Chain C, Hfe (Human) Hemochromatosis Protein
1517	●	1A6ZA	4699710	Chain A, Hfe (Human) Hemochromatosis Protein
525	●	1BIIA	3891929	Chain A, The Crystal Structure Of H-2dd Mhc Class I In Complex Wi
507	●	1S7TD	48425604	Chain D, Crystal Structures Of The Murine Class I Major Histocomp
507	●	1S7TA	48425601	Chain A, Crystal Structures Of The Murine Class I Major Histocomp
507	●	1S7SA	48425598	Chain A, Crystal Structures Of The Murine Class I Major Histocomp
507	●	1S7RD	48425595	Chain D, Crystal Structures Of The Murine Class I Major Histocomp
507	●	1S7RA	48425592	Chain A, Crystal Structures Of The Murine Class I Major Histocomp
507	●	1S7QA	48425589	Chain A, Crystal Structures Of The Murine Class I Major Histocomp
502	●	2BCKD	88192434	Chain D, Crystal Structure Of Hla-A2402 Complexed With A Telomere
502	●	2BCKA	88192431	Chain A, Crystal Structure Of Hla-A2402 Complexed With A Telomere
502	●	1X7QA	73535522	Chain A, Crystal Structure Of Hla-A1101 With Sars Nucleocapsid Pe
502	●	1QVOD	49258587	Chain D, Structures Of Hla-A1101 In Complex With Immunodominant N
502	●	1QVOA	49258584	Chain A, Structures Of Hla-A1101 In Complex With Immunodominant N
502	●	1Q94D	49258567	Chain D, Structures Of Hla-A1101 In Complex With Immunodominant N


NCBI
Related Structures
HOME SEARCH SITE MAP
PubMed
Blast
Entrez Structure
Help
Query: hemochromatosis protein isoform 1 precursor [Homo sapiens]
[gi: 4504377]
Structure: 1DE4 Chain G, Hemochromatosis Protein Hfe Complexed With Transferrin Receptor
Reference: [MMDB] [PubMed]
Get 3D Structure data
to:
View in Cn3d
(To display structure, download Cn3D)
E-value = 7e-168, Bit score = 588, Aligned length = 275, Sequence Identity = 100%
gi 4504377 23 RLLRSHSLHYLFMGASEQDLGLSLFEALGYVDDQLFVFYDHESRRVEPRTPWVSSRISSQMWLQLSQSLKGWDHMFTVDF 102
1DE4 G 1 RLLRSHSLHYLFMGASEQDLGLSLFEALGYVDDQLFVFYDHESRRVEPRTPWVSSRISSQMWLQLSQSLKGWDHMFTVDF 80
gi 4504377 103 WTIMENHNHSKESHTLQVILGCEMQEDNSTEGYWKYGYDGQDHLEFCPDTLDWRAAEPRAWPTKLEWERHKIRARQNRAY 182
1DE4 G 81 WTIMENHNHSKESHTLQVILGCEMQEDNSTEGYWKYGYDGQDHLEFCPDTLDWRAAEPRAWPTKLEWERHKIRARQNRAY 160
gi 4504377 183 LERDCPAQLQQLLELGRGVLDQQVPPLVKVTHHVTSSVTTLRCRALNYYPQNITMKWLKDKQPMDAKEFEPKDVLPNGDG 262
1DE4 G 161 LERDCPAQLQQLLELGRGVLDQQVPPLVKVTHHVTSSVTTLRCRALNYYPQNITMKWLKDKQPMDAKEFEPKDVLPNGDG 240
gi 4504377 263 TYQGWITLAVPPGEEQRYTCQVEHPGLDQPLIVIW 297
1DE4 G 241 TYQGWITLAVPPGEEQRYTCQVEHPGLDQPLIVIW 275


Ca



Step 5 results: The conserved cysteine 282 in the HFE protein is involved in formation of a disulphide bridge. Its mutation to tyrosine will alter the folding of the protein.



Bacon et al. Gastroenterology, 116:193-207, Figure 4

Problem 2:

http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasepheno2.html

A laboratory has generated an EST library from a sickle cell anemia patient and wants to identify the gene(s) causing the phenotype. Sickle cell anemia is a disease in which the red blood cells are curved in shape, and which causes pain and fever.

Outline:

We will follow these steps to solve the problem:
1. Compare ESTs to the human genome (using BLAST).
2. Identify the gene(s) aligning to the ESTs and download their sequences (using MapViewer).
3. Determine what is known about the gene and protein, and identify known SNPs in the gene (using Entrez Gene and dbSNP)
4. Determine whether the mutant gene causes phenotypes (using OMIM).
5. Elucidate the biochemical and structural basis for the function of the mutant protein (using Blink and Cn3D).

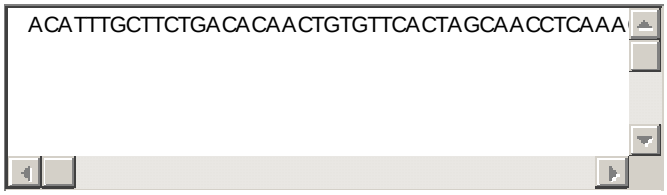
Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare their sequences using BLAST with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

BLAST (human genome)

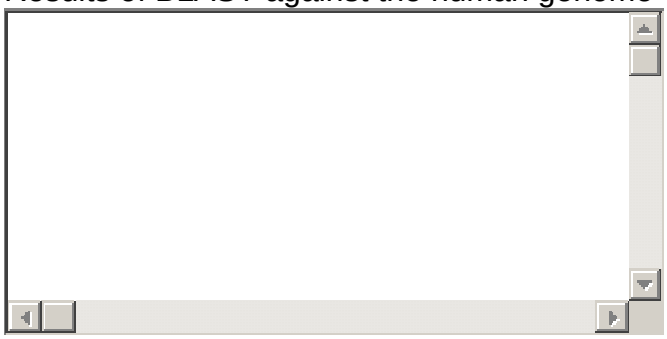
Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

Query EST Sequence:



Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

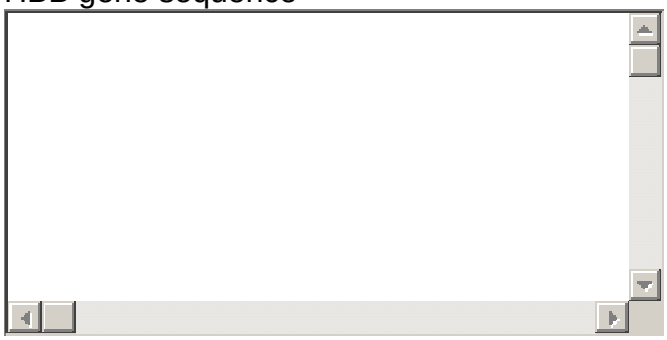
Results of BLAST against the human genome



Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_009237". Currently, 4 maps should be displayed (Model, RNA, Genes_seq and Contig). Make the Genes_seq map as the master map by clicking on the arrow at the top of the map.
The best BLAST hits, indicated by the red bars, are in the region of two exons of the HBB gene annotated on the human genome. Use the Recenter option by clicking on the grey line of the Genes_seq map to keep only the HBB gene in the current display. Note that the gene is annotated on the minus strand. To display the entire HBB gene sequence, click on the "dl" link, choose minus strand from the pull down menu, click on "Change Region/Strand" and display the sequence by clicking on "Display". Copy the sequence and paste it in the area provided below. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

HBB gene sequence



Step 3. Determine what is known about the HBB gene and protein, and identify known SNPs in the gene:

Go back to the Map Viewer report. Click on the HBB gene link from the Genes_seq map which takes you to the Entrez Gene HBB report.

What is the chromosome location of the HBB gene? How many mRNA variants have been annotated for the HBB gene? Which is the longest isoform? What is the function of the protein? Click on the "GeneView in dbSNP" link from the "Links" menu to get a list of known SNPs in the HBB gene. By default the SNPs in the coding region are shown. The information about all the SNPs can be obtained by clicking on the "in gene region" button. List the nonsynonymous changes with respect to the mRNA NM_000518 in the box provided below.

HBB nonsynonymous SNPs

Step 4. Determine whether the mutant HBB gene causes a phenotype:

Click on the OMIM link from the SNP report. The OMIM report details how mutations in the HBB gene are associated with the phenotype of sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks an initiator methionine. Hence, the allelic variants in the OMIM report are off by one amino acid compared to the precursor protein in NP_000509. Click on the Allelic Variants “View list” to get information about the mutant proteins from patients. Is the Glu6Val variant mentioned in the list? Which phenotype does it cause? What is the name of the mutant hemoglobin (hemoglobin S).

Phenotypes caused by the HBB mutants

Step 5. Elucidate the biochemical and structural basis for the function of the wild type and Glu7Val mutant proteins:

Go to the Gene report through the Links menu. Based on the RefSeq summary and the PubMed articles, describe the biochemical functions of beta hemoglobin and hemoglobin S. PubMed articles in the Entrez Gene report indicate that the 3-D structure of hemoglobin S is available.
Let us first take a look at the structure of the wild type protein.

Click on the NP_000509 protein link in the section, 'Genomic regions, transcripts, and products', then on BLink. Click on the 3D structures button. This output contains a list of similar proteins with solved 3D structures. The first entry, 1DXTD, represents a structure of deoxyhemoglobin chain D. Click on the blue dot next to 1DXTD to get the sequence alignment of the query protein with the D chain of 1DXT.

To get information about the chains in this structure, we'll go to the PDB record. Click on the MMDB link next to 'Reference'. That takes us to the MMDB structure summary page for 1DXT. Access the PDB entry by clicking on 1DXT. Note that the chains A and C in the structure represent alpha chains, and B and D represent beta chains (see the SCOP Classification section).

Go back to the MMDB summary page. View the deoxyhemoglobin tetramer by clicking on the "View 3D Structure" button. To search for the structure of the mutant (deoxyhemoglobin S) in the structure database, follow the link to "Structure" on the MMDB Summary page (in the black, horizontal bar), and enter the query term "deoxyhemoglobin s" (without quotes). Two entries, 1HBS and 2HBS, are retrieved. Click on the 2HBS link, then click on the PubMed link (in the Reference line). The abstract indicates that the mutated valine residue of the beta chain contacts with another hemoglobin tetramer molecule to form hemoglobin polymers, which are building blocks for the sickle cell fiber.

Show the side chains of the mutant residue and view its interaction with another hemoglobin molecule:

Download the structure 2HBS by clicking on View 3D Structure. For easier viewing, remove the helix objects using Style--Edit global style by unchecking the box next to Helix objects. Highlight valine 6 from the H chain (a beta chain) by clicking on the "v" at location 6 in the Sequence/Alignment viewer.

To show the side chains of the residue, use Style--Annotate--new. Give a name to this annotation, such as "valine", then click on Edit Style. Change the protein backbone Rendering to "Space Fill", Color Scheme to "charge" or

"hydrophobicity". Repeat these steps for the Protein Sidechains row and click the Protein Sidechains on. To show the amino acid number, choose the Labels panel, and change the Protein Backbone spacing to 1. Click on the Done, OK and Done buttons, respectively. The valine interacts with a pocket between the two helices on another tetramer. To identify the residues on other chains that are within 4 angstroms of the valine, use Show/Hide--Select by distance--other molecules. To remove highlighted residues, click in any white space in the Sequence/Alignment viewer.

Summary:

This mini-course describes steps to identify the genes expressing the ESTs obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align to contig NT_009237.17 on chromosome 11 with one nucleotide difference (A to T with respect to the nucleotide 4035473 on the contig).

Step 2: The query EST was found to be expressed by the HBB gene.

Step 3: The query EST sequence contains a known SNP, leading to a change in the translation of the 7th amino acid from glutamic acid to valine.

Step 4: Mutations in the HBB gene are associated with sickle cell anemia.

Step 5: HBB encodes beta hemoglobin; hemoglobin is a tetramer of 2 beta and 2 alpha chains. An HBB mutation at the 7th (in the genomic annotation) negatively charged amino acid, glutamic acid, to hydrophobic valine leads to polymerization of hemoglobin into a sickle fiber that changes the shape of red blood cells and leads to sickle cell anemia.