Mini Courses NCBI Resources one course at a time

Identification of Disease Genes

This mini-course focuses on the identification of a disease gene using NCBI's human genome assembly. The reference human genome assembly along with integrated maps, literature, and expression information comprises a powerful discovery system for exploring candidate human disease genes.

Problem: A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

We will follow these steps to solve the problem:

1. Compare an EST to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using MapViewer).
3. Identify whether the EST contains any known SNPs (using dbSNP).
4. Determine whether a mutant form of the gene causes a phenotype (using OMIM).

A web page (**http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene.html**) describes in detail how to perform these steps.

The following handout includes the screen shots of the exercise.

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

Problem 1:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

Outline:

We will follow these steps to solve the problem:
1. Compare the EST from a hemochromatosis patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the ESTs and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

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

One way to identify the genes expressing the EST is to compare the EST sequence 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)

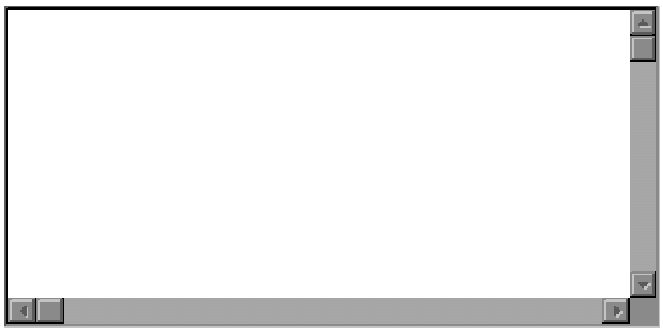
Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

Query EST Sequence:

```
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAG
ATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATG
CTACCAGGGCTGGATAACCTTGGCTGTACCCCCTGGGGAAG
CCCTCATTGTGATCTGGG
```

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_007592". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.
The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Display the entire HFE gene sequence by clicking on the "dl" link and then on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

HFE gene sequence

Step 3. Determine whether the ESTs contain known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs1800562. Click on the link for the SNP. There is an A/G SNP is at the nucleotide position 16951392 on the contig NT_007592 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1? Please note that the SNP results in the Cysteine 282 Tyrosine mutation for the longest protein (expressed by the mRNA NM_000410) as reported under GeneView.

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

Go back to the Map Viewer report. Make the Gene_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HFE gene that details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Cys282Tyr variant mentioned in the list? Which phenotype does it cause?

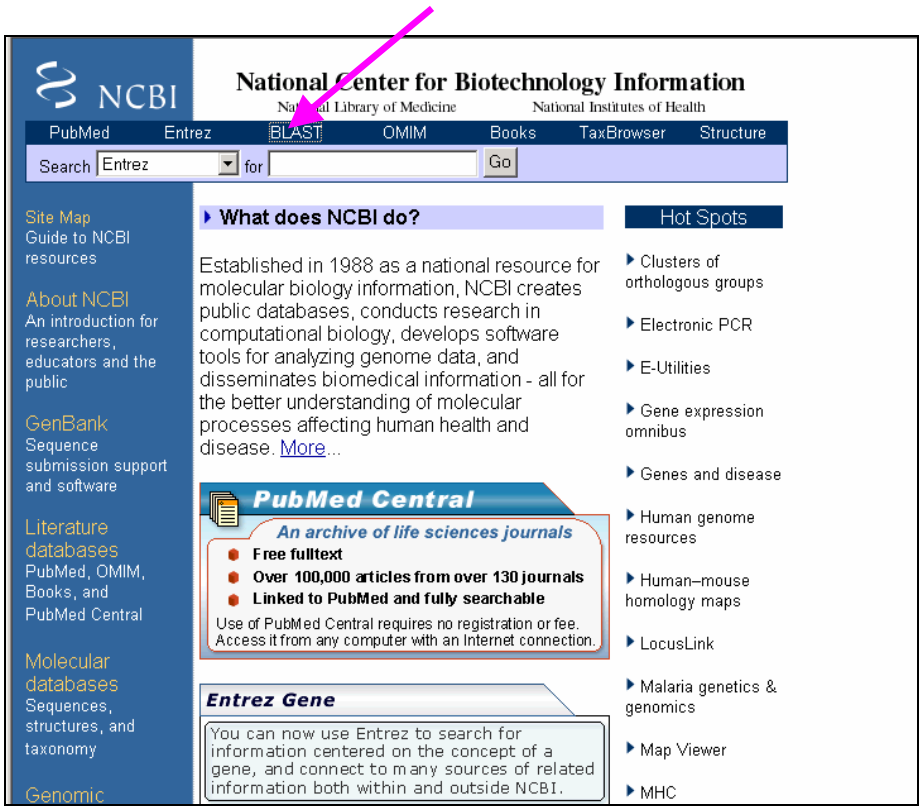
Summary:

This mini-course describes steps to identify the gene expressing an EST 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 contig NT_007592.14 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 query EST sequence contains a known SNP (G/A with respect to nucleotide 16951392 on contig NT_007592.14).
Step 4: Mutations in the HFE gene are associated with hemochromatosis.

Snapshots

Step 1: Compare EST against the human genome.



NCBI → BLAST

Latest news: 5 Mar 2007 : Beta test of new BLAST interface

About

- Getting started
- News
- FAQs

More info

- NAR 2004
- NCBI Handbook
- The Statistics of Sequence Similarity Scores

Software

- Downloads
- Developer info

Other resources

- References
- NCBI Contributors
- Mailing list
- Contact us

The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

New BLAST βETA *Try it at http://ncbi.nlm.nih.gov/blast/beta*

Nucleotide

- Quickly search for highly similar sequences (megablast)
- Quickly search for divergent sequences (discontiguous megablast)
- Nucleotide-nucleotide BLAST (blastn)
- Search for short, nearly exact matches
- Search trace archives with megablast or discontiguous megablast

Protein

- Protein-protein BLAST (blastp)
- Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST)
- Search for short, nearly exact matches
- Search the conserved domain database (rpsblast)
- Protein homology by domain architecture (cdart)

Translated

- Translated query vs. protein database (blastx)
- Protein query vs. translated database (tblastn)
- Translated query vs. translated database (tblastx)

Genomes

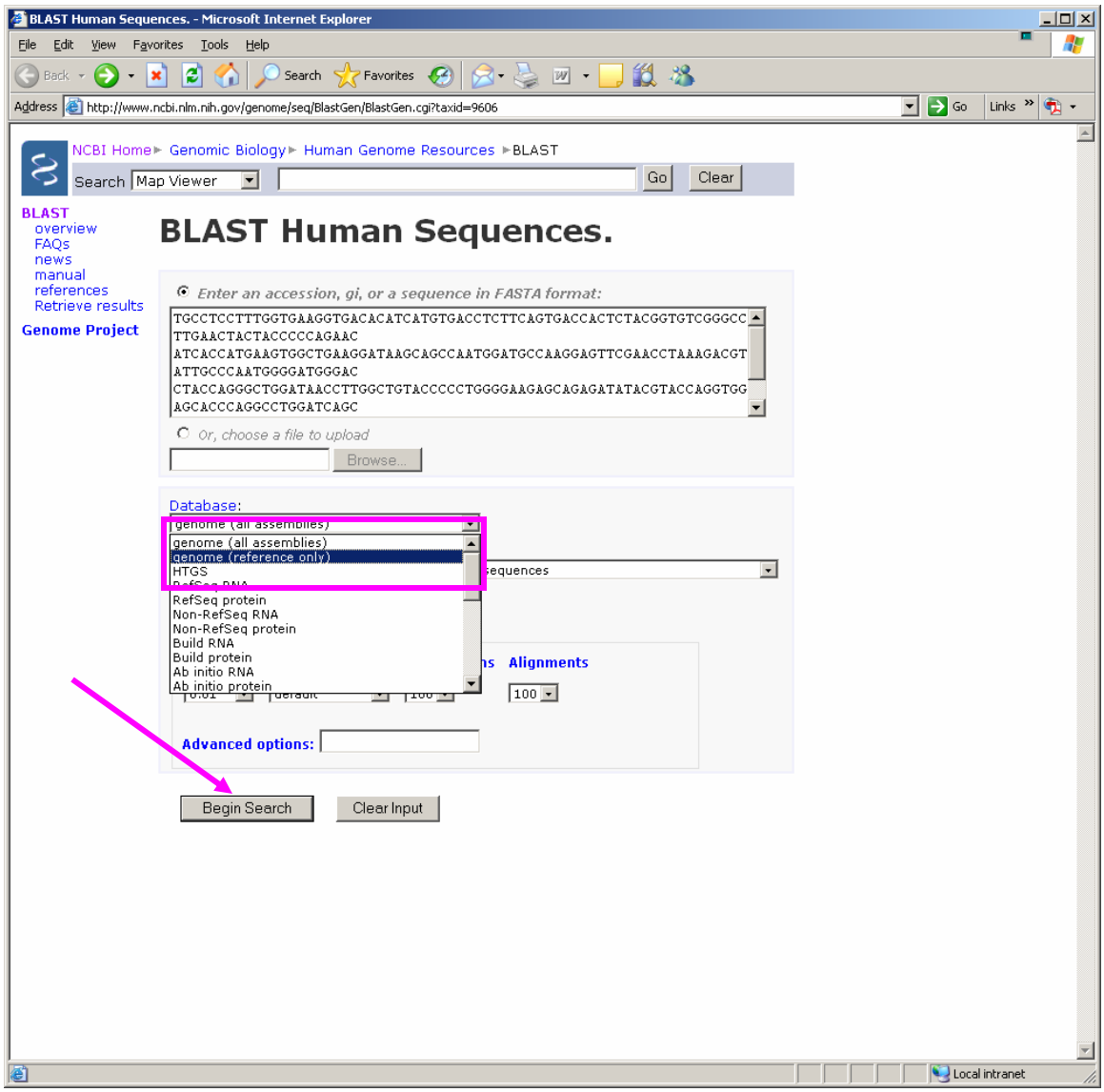
- Human, mouse, rat, chimp, cow, pig, dog, sheep, cat
- Chicken, puffer fish, zebrafish
- Fly, honey bee, other insects
- Microbes, environmental samples
- Plants, nematodes
- Fungi, protozoa, other eukaryotes

Special

- Search for gene expression data (GEO BLAST)
- Align two sequences (bl2seq)
- Screen for vector contamination (VecScreen)
- Immunoglobin BLAST (IgBlast)
- SNP BLAST

Meta

- Retrieve results


BLAST Human Sequences. - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606
Go Links
NCBI Home ▸ Genomic Biology ▸ Human Genome Resources ▸ BLAST
Search Map Viewer Go Clear
BLAST
overview
FAQs
news
manual
references
Retrieve results
Genome Project
BLAST Human Sequences.
Enter an accession, gi, or a sequence in FASTA format:
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTCGGGCC
TTGAACTACTACCCCCAGAAC
ATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATGCCAAGGAGTTCGAACCTAAAGACGT
ATTGCCCAATGGGGATGGGAC
CTACCAGGGCTGGATAACCTTGGCTGTACCCCCTGGGGAAGAGCAGAGATATACGTACCAGGTGG
AGCACCCAGGCCTGGATCAGC
Or, choose a file to upload
Browse...
Database:
genome (all assemblies)
genome (all assemblies)
genome (reference only)
HTGS
RefSeq protein
Non-RefSeq RNA
Non-RefSeq protein
Build RNA
Build protein
Ab initio RNA
Ab initio protein
Alignments
100
Advanced options:
Begin Search
Clear Input
Local intranet

NCBI

formatting **BLAST**

Nucleotide | Protein | Translations | Retrieve results for an RID

New BLAST interface beta release Try it NOW !

Your request has been successfully submitted and put into the Blast Queue.

Query = (276 letters)

The request ID is 1173894413-20411-206785682307.BLASTQ4

Format! or Reset all

NCBI

results of **BLAST**

BLASTN 2.2.16 [Mar-11-2007]

RID: 1173894413-20411-206785682307.BLASTQ4

Database: human build 36 genome database (reference assembly only)
368 sequences; 2,870,843,926 total letters

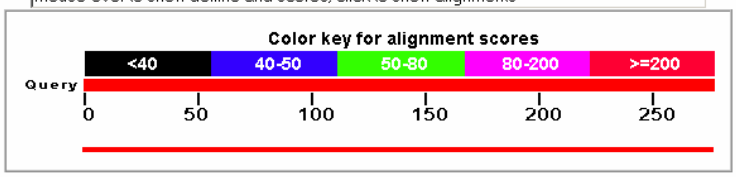
If you have any problems or questions with the results of this search please refer to the **BLAST FAQs**
Taxonomy reports

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

Query=
Length=276

Distribution of 1 Blast Hits on the Query Sequence

Mouse-over to show defline and scores, click to show alignments



Distance tree of results NEW

Sequences producing significant alignments:	Score (Bits)	E Value
ref|NT_007592.14|Hs6_7749 Homo sapiens chromosome 6 genomic cont	505	6e-141

```
                                                    Alignments

>ref|NT_007592.14|Hs6_7749 D 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 =  505 bits (273),  Expect = 6e-141
 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
```

Result: The EST sequence is aligned to the contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

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

NCBI *results of* **BLAST**

BLASTN 2.2.16 [Mar-11-2007]

RID: 1173894413-20411-206785682307.BLASTQ4

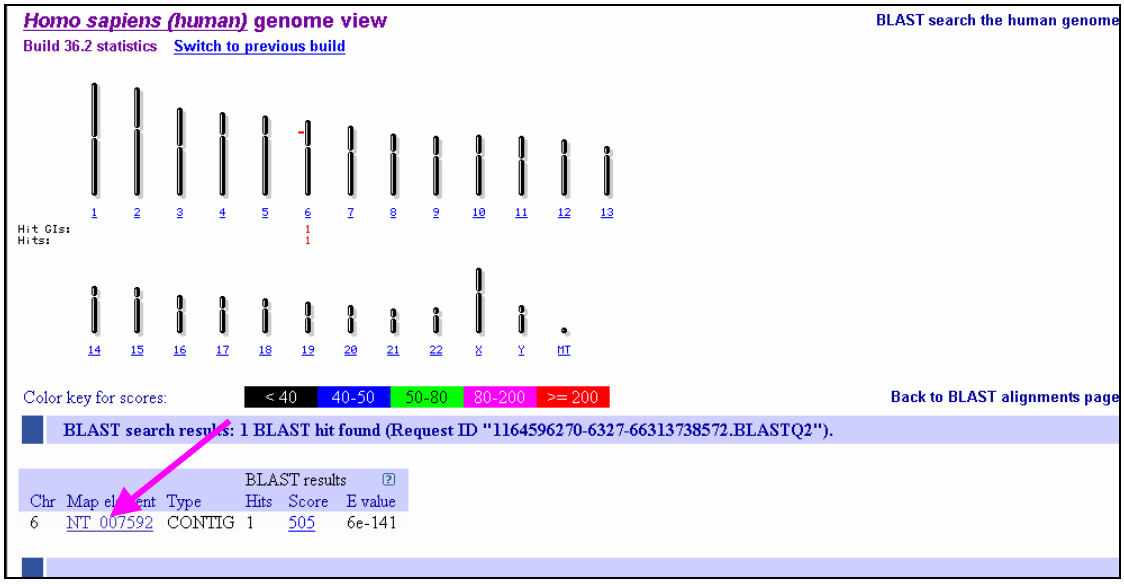
Database: human build 36 genome database (reference assembly only)
368 sequences; 2,870,843,926 total letters

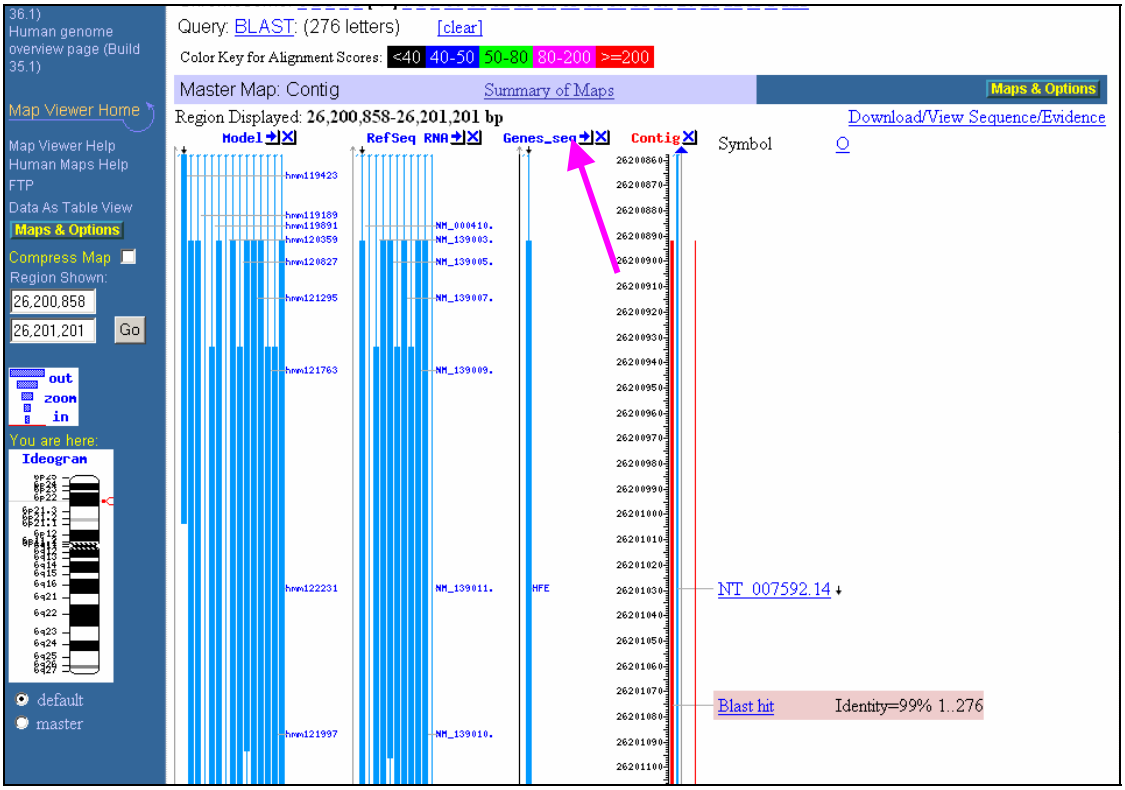
If you have any problems or questions with the results of this search please refer to the **BLAST FAQs**
Taxonomy reports

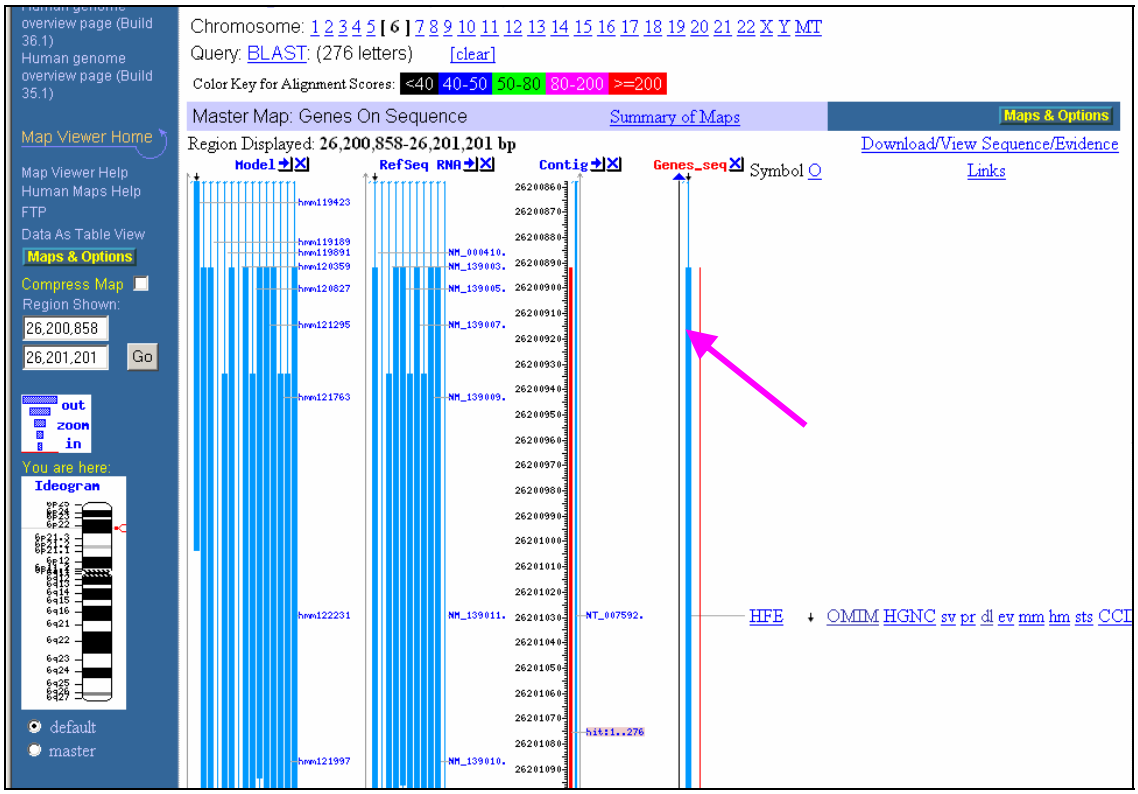
Genome View — Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

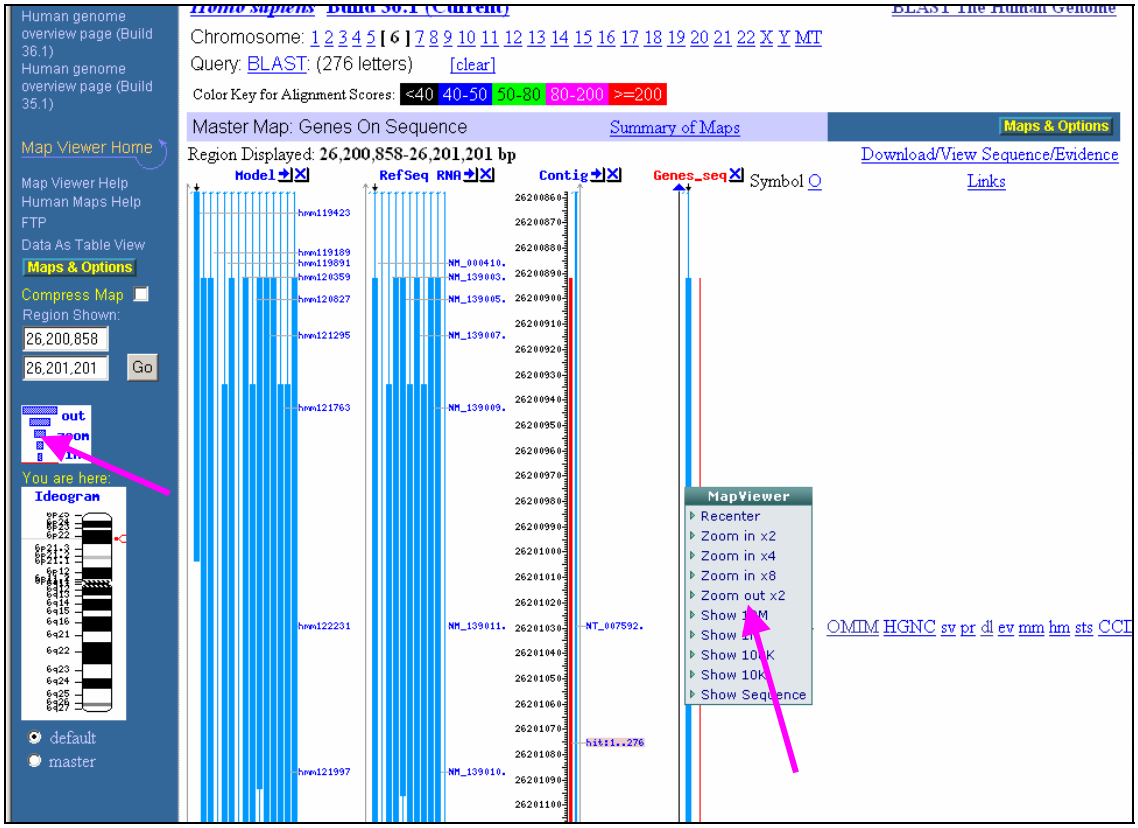
Query=
Length=276

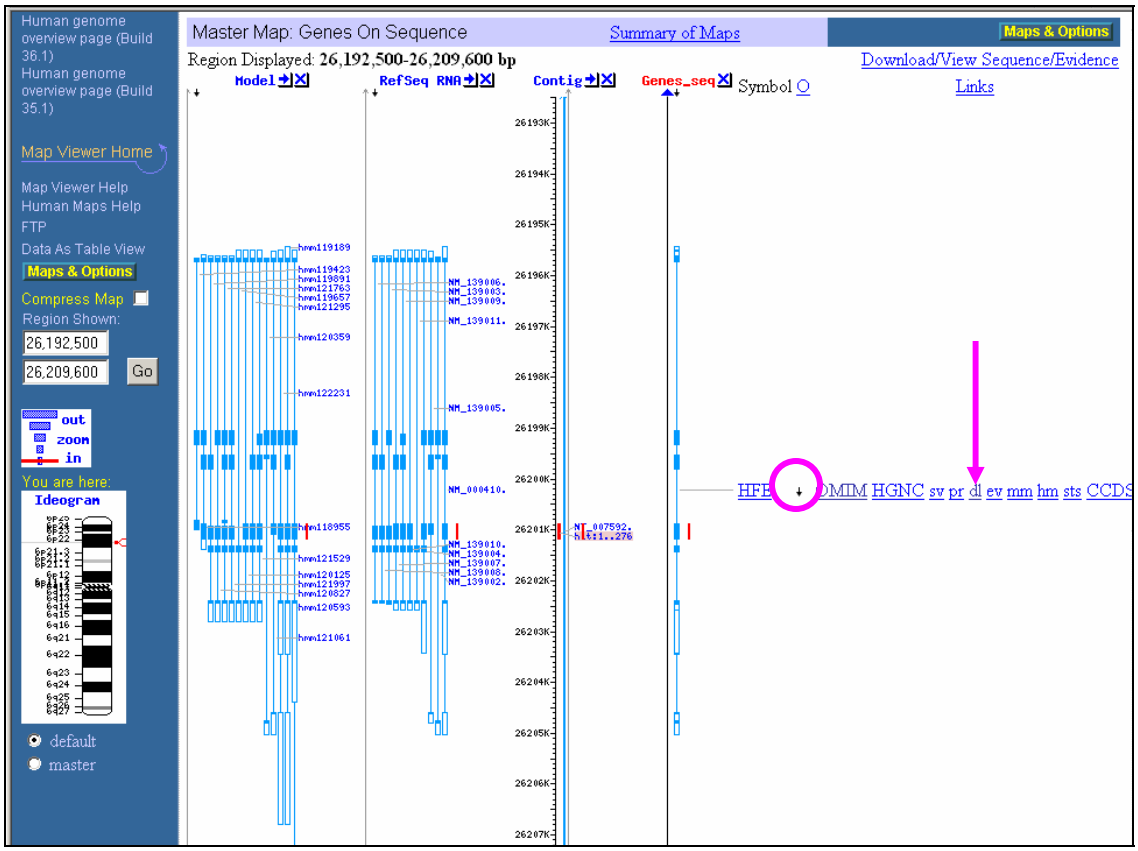
Distribution of 1 Blast Hits on the Query Sequence

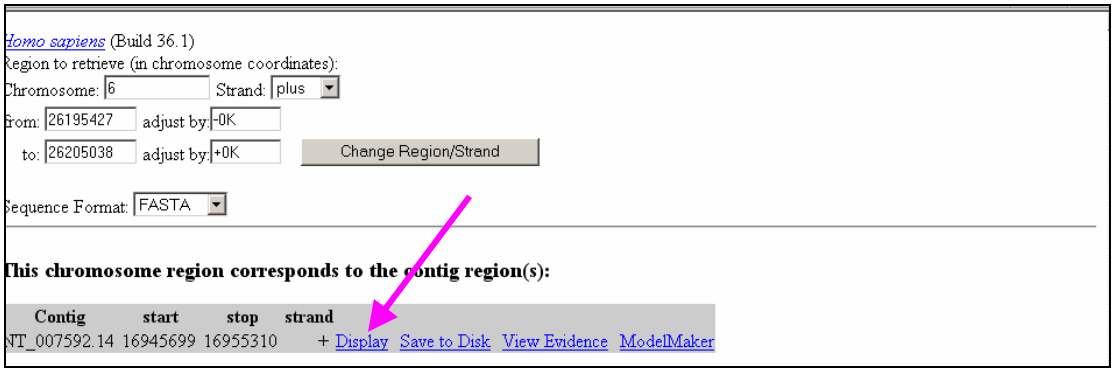

Homo sapiens (human) genome view
BLAST search the human genome
Build 36.2 statistics
Switch to previous build
Hit GIs:
Hits:
Color key for scores:
< 40
40-50
50-80
80-200
>= 200
Back to BLAST alignments page
BLAST search results: 1 BLAST hit found (Request ID "1164596270-6327-66313738572.BLASTQ2").
BLAST results
Chr
Map element
Type
Hits
Score
E value
6
NT_007592
CONTIG
1
505
6e-141


Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom
in
You are here:
Ideogram
default
master
Query: BLAST: (276 letters)
[clear]
Color Key for Alignment Scores:
<40
40-50
50-80
80-200
>=200
Master Map: Contig
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
NT_007592.14
Blast hit
Identity=99% 1..276


Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom
in
You are here:
Ideogram
default
master
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: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol O
HFE
OMIM HGNC sv pr dl ev mm hm sts CCD


Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
You are here:
Ideogram
default
master
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: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol O
MapViewer
Recenter
Zoom in x2
Zoom in x4
Zoom in x8
Zoom out x2
Show 10K
Show 10K
Show Sequence
OMIM HGNC sv pr dl ev mm hm sts CCD


Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,192,500
26,209,600
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,192,500-26,209,600 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol
HFE
OMIM HGNC sv pr dl ev mm hm sts CCDS


Homo sapiens (Build 36.1)
Region to retrieve (in chromosome coordinates):
Chromosome: 6
Strand: plus
From: 26195427
adjust by: -0K
to: 26205038
adjust by: +0K
Change Region/Strand
Sequence Format: FASTA
This chromosome region corresponds to the contig region(s):
Contig start stop strand
NT_007592.14 16945699 16955310 + Display Save to Disk View Evidence ModelMaker

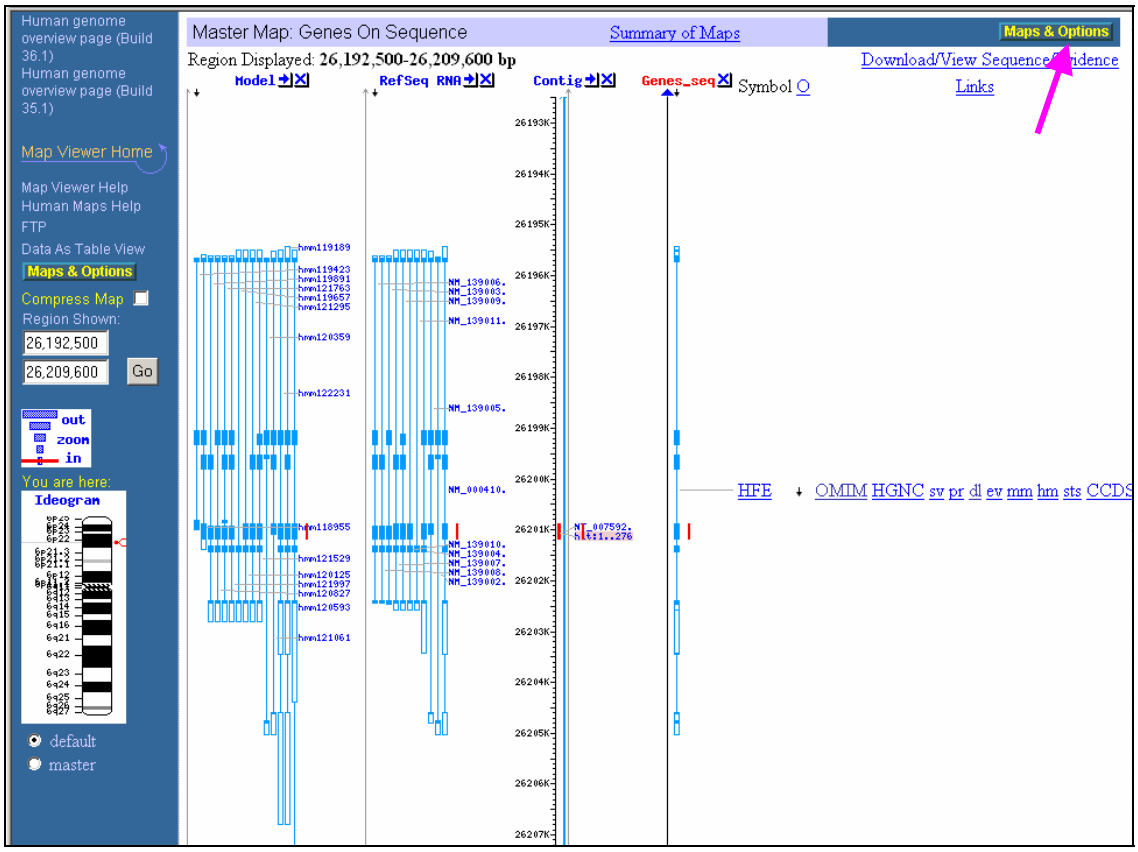
Range: from 16945699 to 16955310 Show whole sequence Reverse complemented strand Refresh

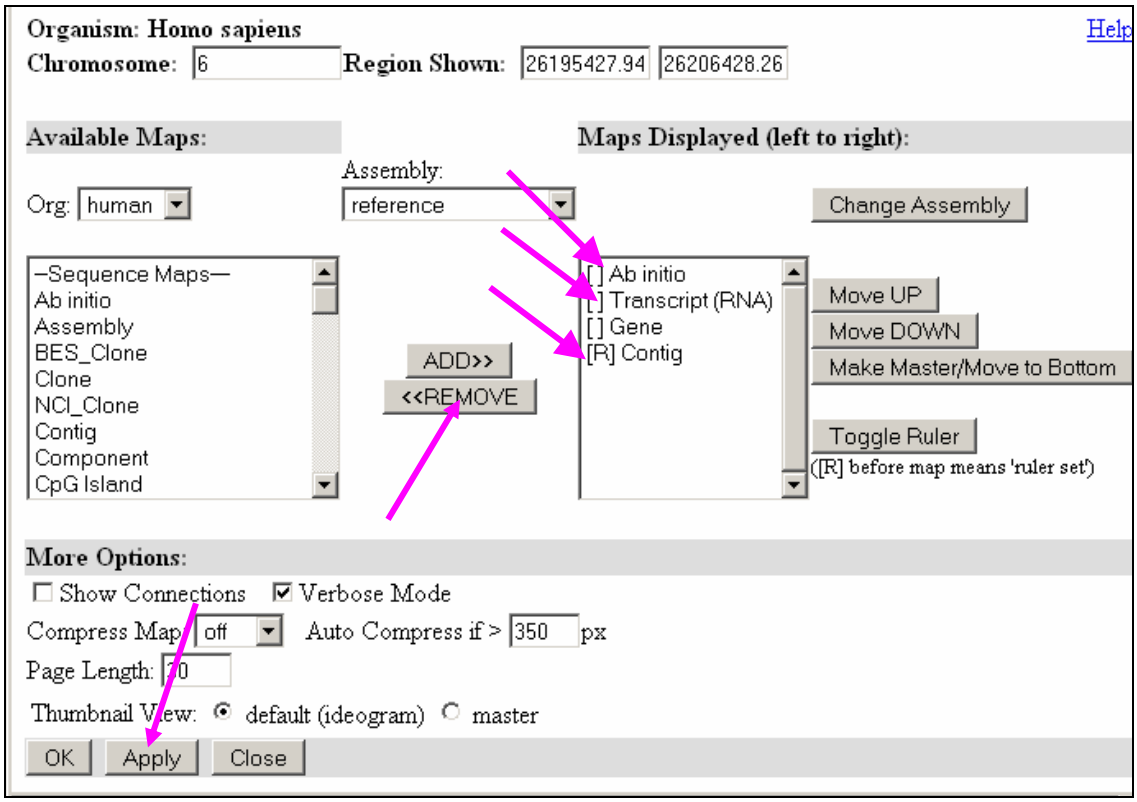
1: NT_007592. Reports Homo sapiens chro...[gi:51465675]

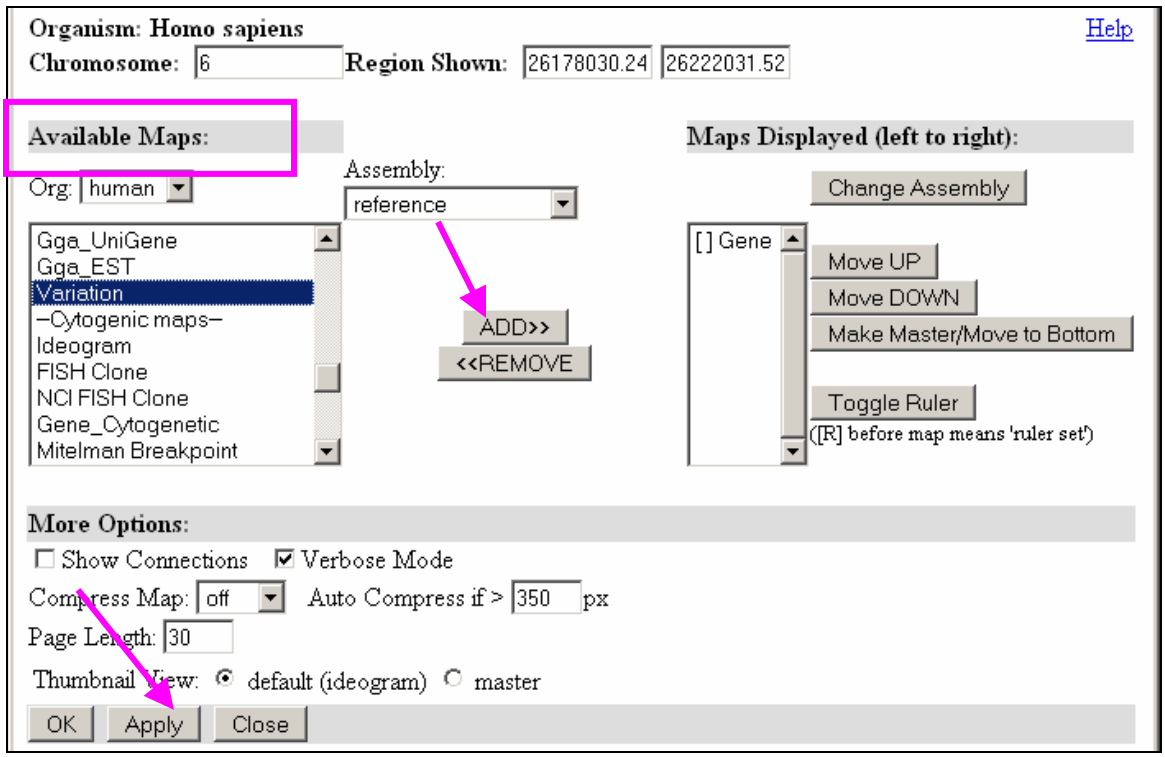
```
>ref|NT_007592.14|Hs6_7749:16945699-16955310 Homo sapiens chromosome 6 genomic contig
GGGGACACTGGATCACCTAGTGTTTCACAAGCAGGTACCTTCTGCTGTAGGAGAGAGAGAACTAAAGTTC
TGAAAGACCTGTTGCTTTTCACCAGGAAGTTTTACTGGGCATCTCCTGAGCCTAGGCAATAGCTGTAGGG
TGACTTCTGGAGCCATCCCCGTTTCCCCGCCCCCCAAAAGAAGCGGAGATTTAACGGGGACGTGCGGCCA
GAGCTGGGGAAATGGGCCCGCGAGCCAGGCCGGCGCTTCTCCTCCTGATGCTTTTGCAGACCGCGGTCCT
GCAGGGGCGCTTGCTGCGTGAGTCCGAGGGCTGCGGGCGAACTAGGGGCGCGGCGGGGGTGGAAAAATCG
AAACTAGCTTTTTCTTTGCGCTTGGGAGTTTGCTAACTTTGGAGGACCTGCTCAACCCTATCCGCAAGCC
CCTCTCCCTACTTTCTGCGTCCAGACCCCGTGAGGGAGTGCCTACCACTGAACTGCAGATAGGGGTCCCT
CGCCCCAGGACCTGCCCCCTCCCCCGGCTGTCCCGGCTCTGCGGAGTGACTTTTGGAACCGCCCACTCCC
TTCCCCCAACTAGAATGCTTTTAAATAAATCTCGTAGTTCCTCACTTGAGCTGAGCTAAGCCTGGGGCTC
CTTGAACCTGGAACTCGGGTTTATTTCCAATGTCAGCTGTGCAGTTTTTTCCCCAGTCATCTCCAAACAG
GAAGTTCTTCCCTGAGTGCTTGCCGAGAAGGCTGAGCAAACCCACAGCAGGATCCGCACGGGGTTTCCAC
CTCAGAACGAATGCGTTGGGCGGTGGGGGCGCGAAAGAGTGGCGTTGGGGATCTGAATTCTTCACCATTC
CACCCACTTTTGGTGAGACCTGGGGTGGAGGTCTCTAGGGTGGGAGGCTCCTGAGAGAGGCCTACCTCGG
GCCTTTCCCCACTCTTGGCAATTGTTCTTTTGCCTGGAAAATTAAGTATATGTTAGTTTTGAACGTTTGA
ACTGAACAATTCTCTTTTCGGCTAGGCTTTATTGATTTGCAATGTGCTGTGTAATTAAGAGGCCTCTCTA
CAAAGTACTGATAATGAACATGTAAGCAATGCACTCACTTCTAAGTTACATTCATATCTGATCTTATTTG
ATTTTCACTAGGCATAGGGAGGTAGGAGCTAATAATACGTTTATTTTACTAGAAGTTAACTGGAATTCAG
ATTATATAACTCTTTTCAGGTTACAAAGAACATAAATAATCTGGTTTTCTGATGTTATTTCAAGTACTAC
AGCTGCTTCTAATCTTAGTTGACAGTGATTTTGCCCTGTAGTGTAGCACAGTGTTCTGTGGGTCACACGC
CGGCCTCAGCACAGCACTTTGAGTTTTGGTACTACGTGTATCCACATTTTACACATGACAAGAATGAGGC
ATGGCACGGCCTGCTTCCTGGCAAATTTATTCAATGGTACACTGGGCTTTGGTGGCAGAGCTCATGTCTC
```

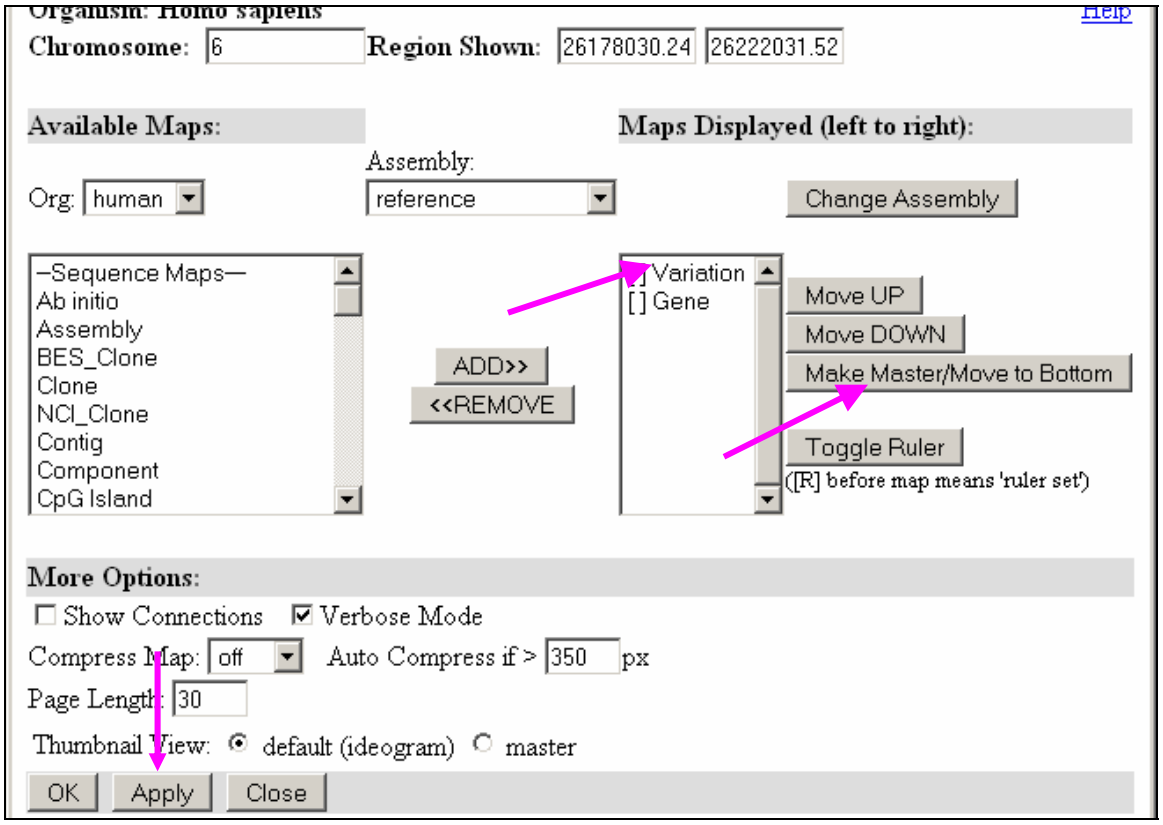
Result: The query EST is expressed by the HFE gene.

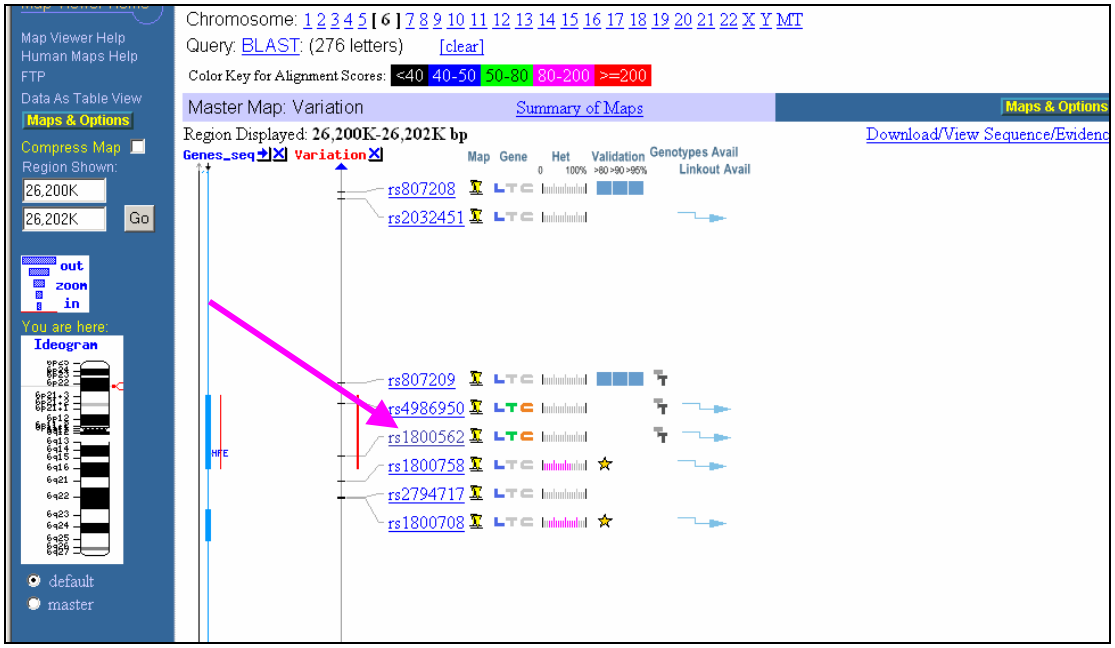
Step 3: Determine whether the EST contains any known SNPs




Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26195427.94 26206428.26
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
–Sequence Maps–
Ab initio
Assembly
BES_Clone
Clone
NCI_Clone
Contig
Component
CpG Island
ADD>>
<<REMOVE
[ ] Ab initio
[ ] Transcript (RNA)
[ ] Gene
[R] Contig
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close


Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26178030.24 26222031.52
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
Gga_UniGene
Gga_EST
Variation
–Cytogenic maps–
Ideogram
FISH Clone
NCI FISH Clone
Gene_Cytogenetic
Mitelman Breakpoint
ADD>>
<<REMOVE
[ ] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close


Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26178030.24 26222031.52
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
–Sequence Maps–
Ab initio
Assembly
BES_Clone
Clone
NCI_Clone
Contig
Component
CpG Island
ADD>>
<<REMOVE
[] Variation
[] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close


Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200K
26,202K
Go
out
zoom
in
You are here:
Ideogram
default
master
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: Variation
Summary of Maps
Maps & Options
Region Displayed: 26,200K-26,202K bp
Download/View Sequence/Evidenc
Genes_seq
Variation
Map Gene Het Validation Genotypes Avail
Linkout Avail
rs807208
rs2032451
rs807209
rs4986950
rs1800562
rs1800758
rs2794717
rs1800708

NCBI **Single Nucleotide Polymorphism**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for Go

BUILD 127

Have a question about dbSNP? Try searching the SNP FAQ Archive!

Go

Reference SNP(refSNP) Cluster Report: rs1800562

refSNP ID: rs1800562		Allele	
Organism:	human (*Homo sapiens*)	Variation Class:	SNP: single nucleotide polymorphism
Molecule Type:	Genomic		
Created/Updated in build:	89/126	Alleles:	A/G
Map to Genome Build:	36.2	Ancestral Allele:	G

Links , Linkout

SEARCH
Entrez SNP
Blast SNP
Batch Query
By Submitter
New Batches
Method
Population
Detail
Class
Publication
Chromosome Report
Locus Information
STS Markers
Free Form Search

Fasta sequence (Legend)

```
>gnl|dbSNP|rs1800562|allelePos=202|totalLen=450|taxid=9606|snpclass=1|alleles='A/G'|molc=Genomic|build=113

ATGTGAYCTC TTCAGTGACC ACTCTACGGT GTCGGGCCTT GAACTACTAC CCCCAGAACA
TCACCATGAA GTGGCTGAAG GATAAGCAGC CAATGGATGC CAAGGAGTTC GAACCTAAAG
ACGTATTGCC CAATGGGGAT GGGACCTACC AGGGCTGGAT AACCTTGGCT GTACCCCCTG
GGGAAGAGCA GAGATATACG T
R
CCAGGTGGAG CACCCAGGCC TGGATCAGCC CCTCATTGTG ATCTGGGGTA TGTGACTGAT
GAGAGCCAGG AGCTGAGAAA ATCTATTGGG GGTTRAGAGG AGTGCCTGAG GAGGTAATTA
TGGCAGTGAG ATGAGGATCT GCTCTTTGTT AGGGGGTGGG CTGAGGGTGG CAATCAAAGG
CTTTAACTTG CTTTTTCTGT TTTAGAGCCC TCACCGTCTG GCACCCTAGT CATTGGAGTC
ATCAGTGG
```

Integrated Maps:

NCBI MapViewer: rs1800562 maps exactly once on NCBI human chromosome 6

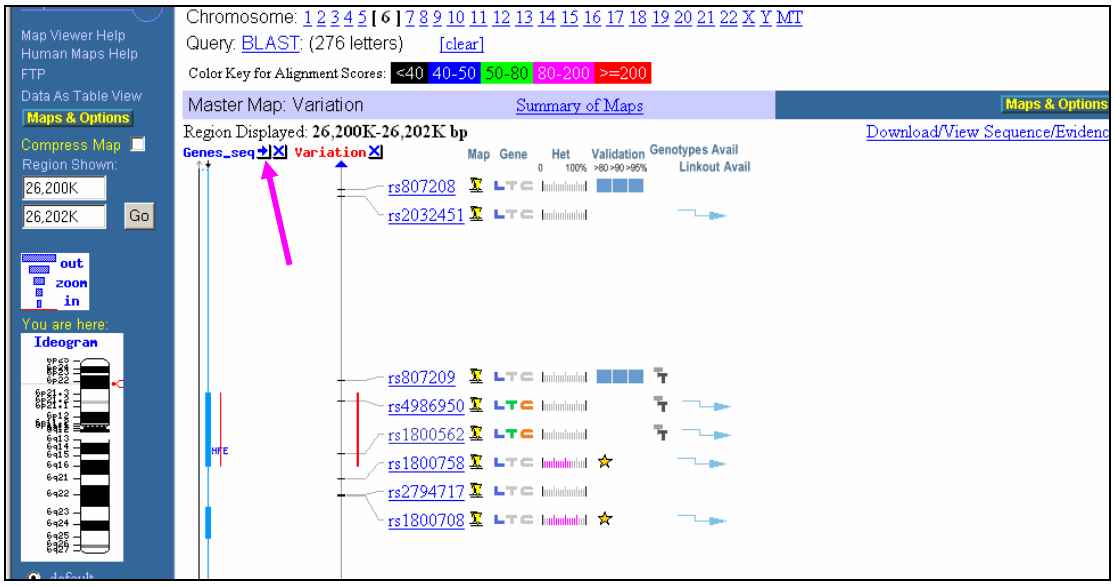
Chromosome	Contig accession	Contig position	Chromosome position	Hit orientation	Contig Allele	Assembly Type	Group label	Contig label	Neighbor SNP	SNP_flank position
6	NT_007592.14	16951392	26201120	plus	G	ref_assembly	reference	reference	view	201
6	NW_922984.1	25717681	27322425	plus	G	alt_assembly_1	Celera	Celera	view	201

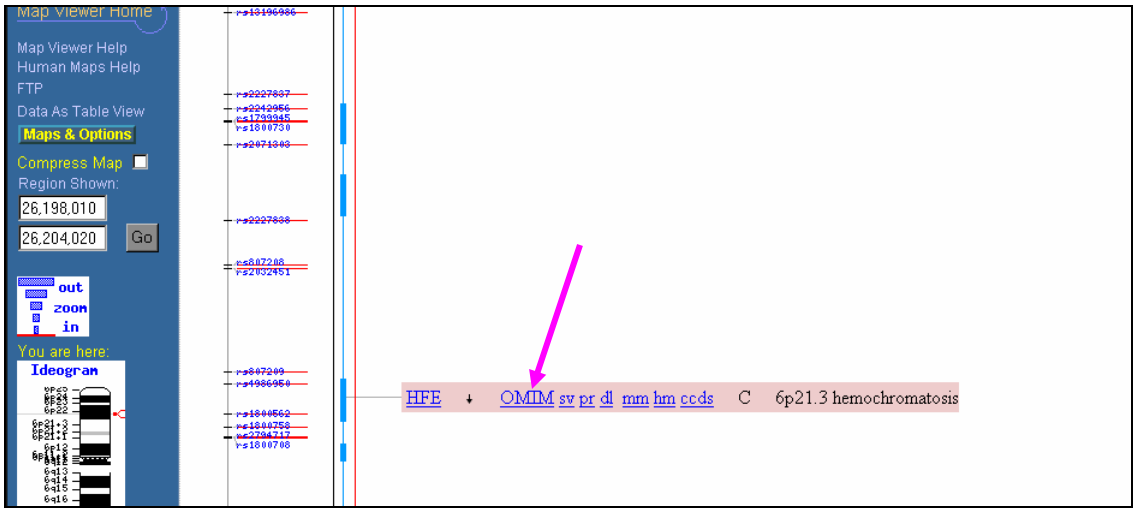
Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) Color Legend
reference	NT_007592->NM_000410	
	sv function	

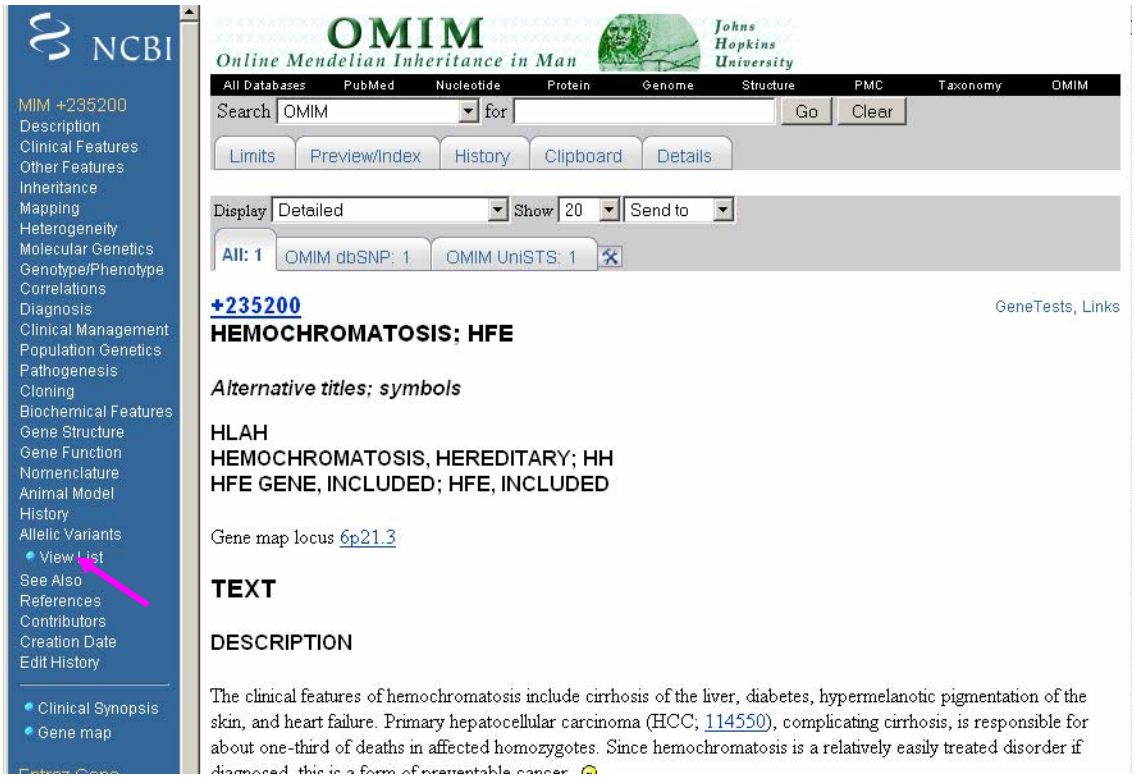
Group label	Contig-->mRNA-->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
reference	NT_007592->NM_000410->NP_000401	16951392	forward	1066	nonsynonymous	A	Tyr [Y]	2	282
					contig reference	G	Cys [C]	2	282

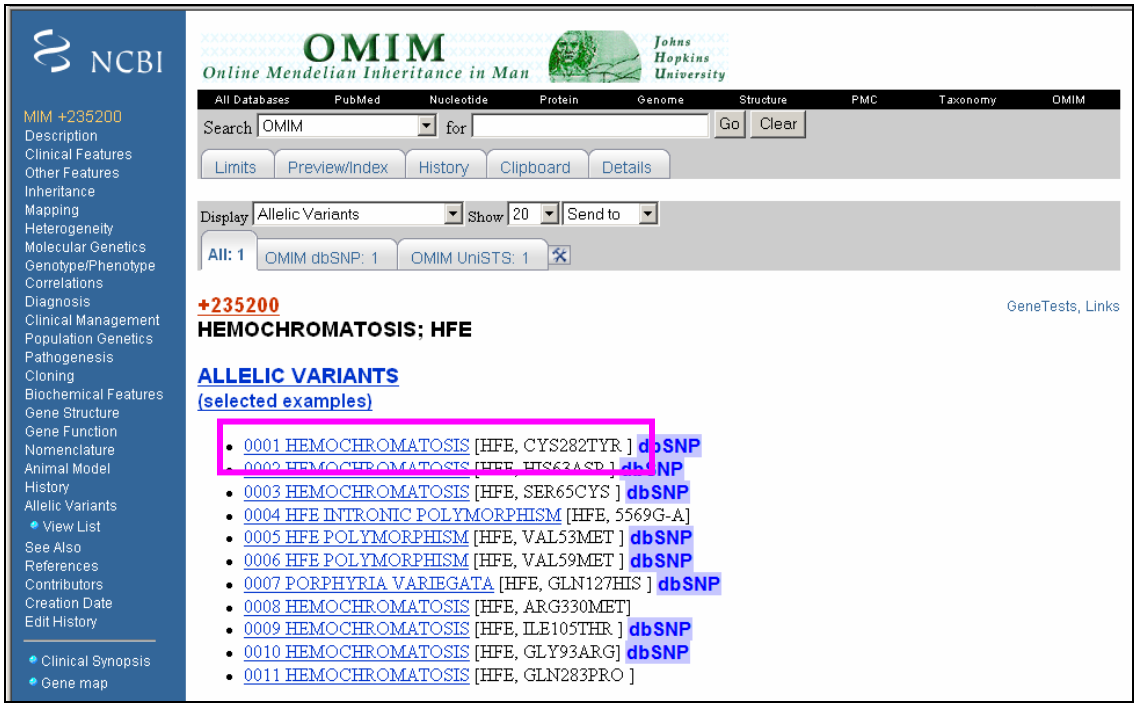
Result: The EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT_007592.14).

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









Result: Mutations in the HFE gene are associated with hemochromatosis disease.

Problem 2:

http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene2.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 an EST from a sickle cell anemia patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

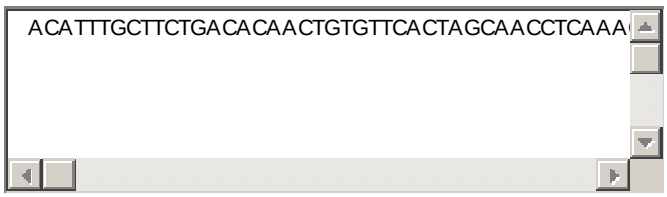
Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare the EST sequence 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. Note that the similarity is on the minus strand of genome. 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 EST 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 (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.
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. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. 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 on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. 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 whether the EST contains known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and the master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs334. Click on the link for the SNP. There is an A/T SNP is at the nucleotide position 4035473 on the contig NT_009237 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1?

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

Go back to the Map Viewer report. Make the Gene_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HBB gene that details how mutations in the HBB gene are associated with a phenotype, sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks initiator methionine. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Glu6Val variant mentioned in the list? Which phenotype does it cause?

Summary:

This mini-course describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia 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 (T to A 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 (T/A with respect to the nucleotide 4035473 on contig NT_009237.17).
Step 4: Mutations in the HBB gene are associated with sickle cell anemia.