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 images of the exercise.

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

Instructor: Majda Valjavec-Gratian (**gratian@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)

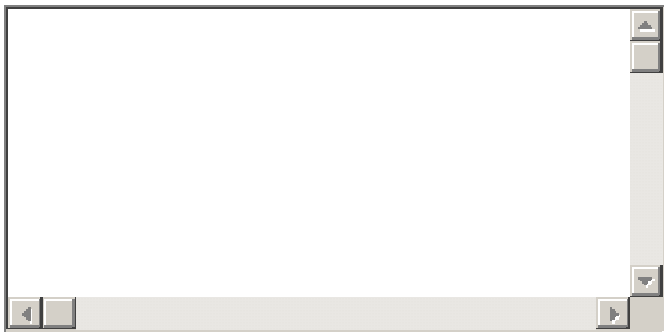
Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

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?

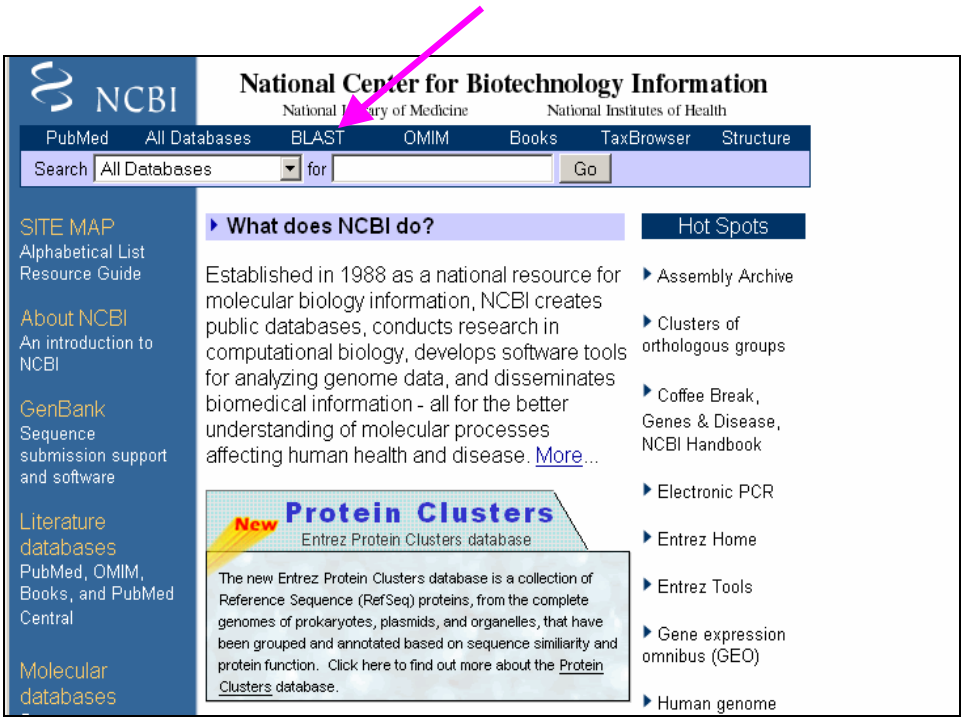
Summary:

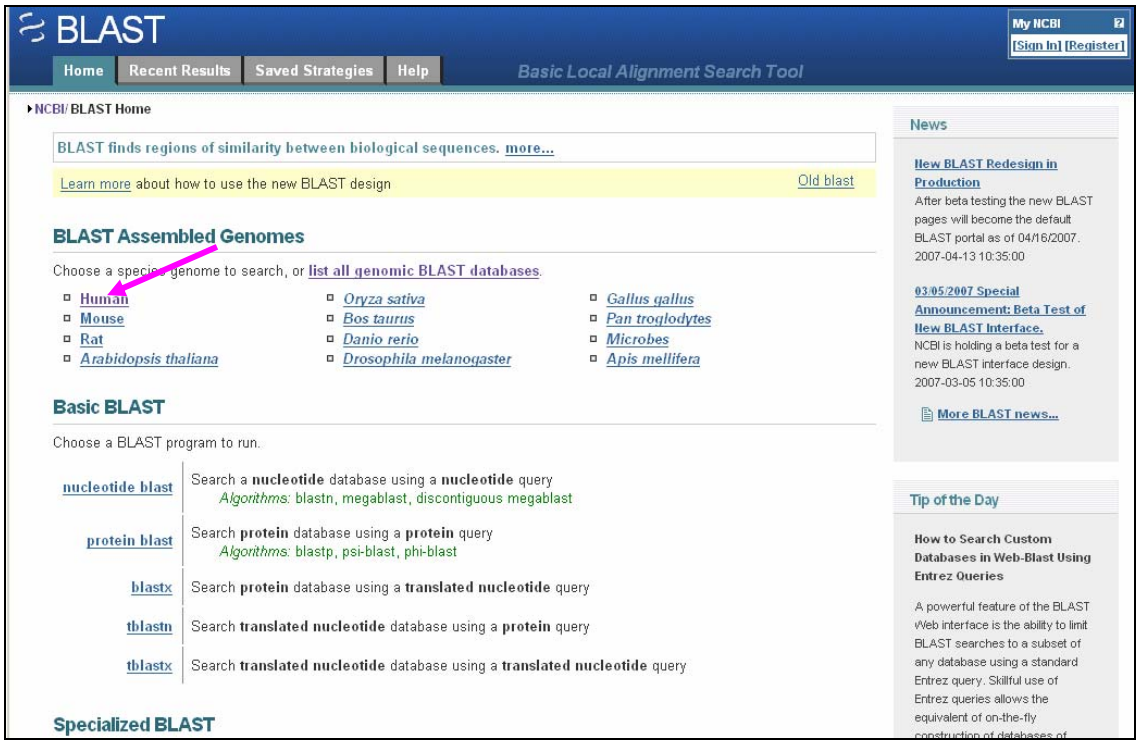
This mini-course describes steps to identify the gene expressing an EST obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

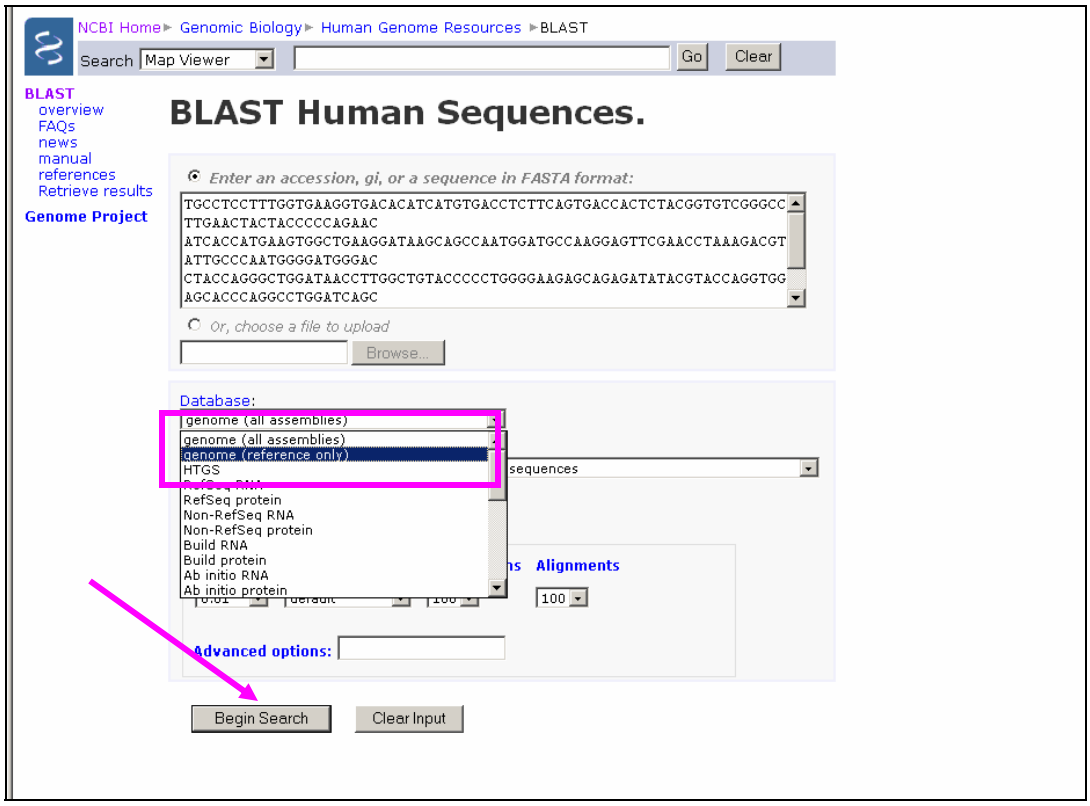
Step 1: The query EST sequence was found to align contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).
Step 2: The query EST was found to be expressed by the HFE gene.
Step 3: The query EST sequence contains a known SNP (G/A with respect to nucleotide 16951392 on contig NT_007592.14).
Step 4: Mutations in the HFE gene are associated with hemochromatosis.

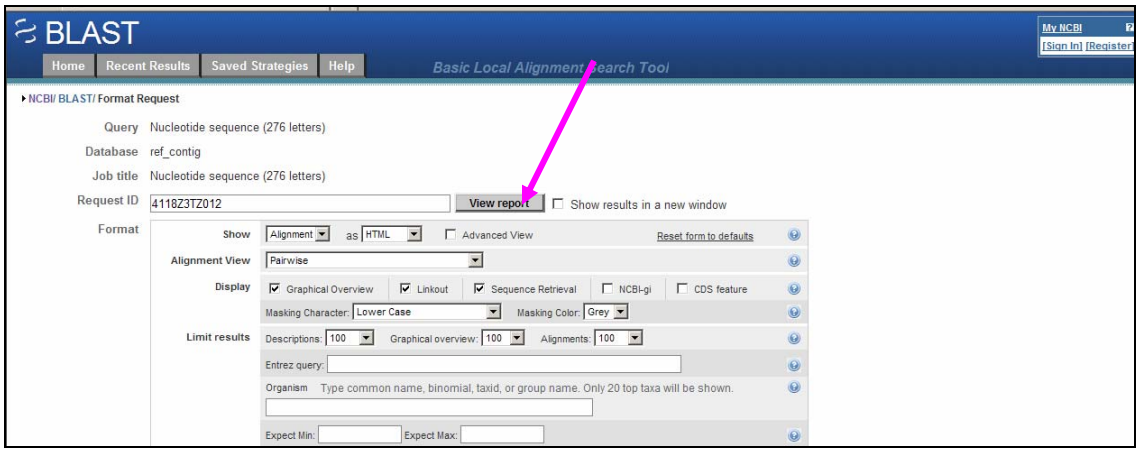
Screen images

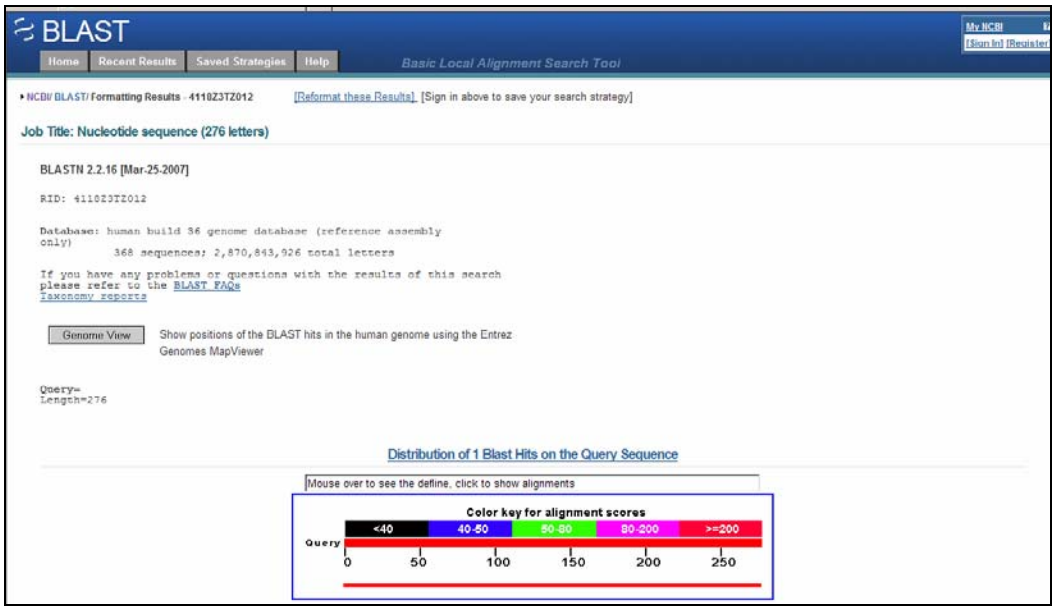
Step 1: Compare EST against the human genome.

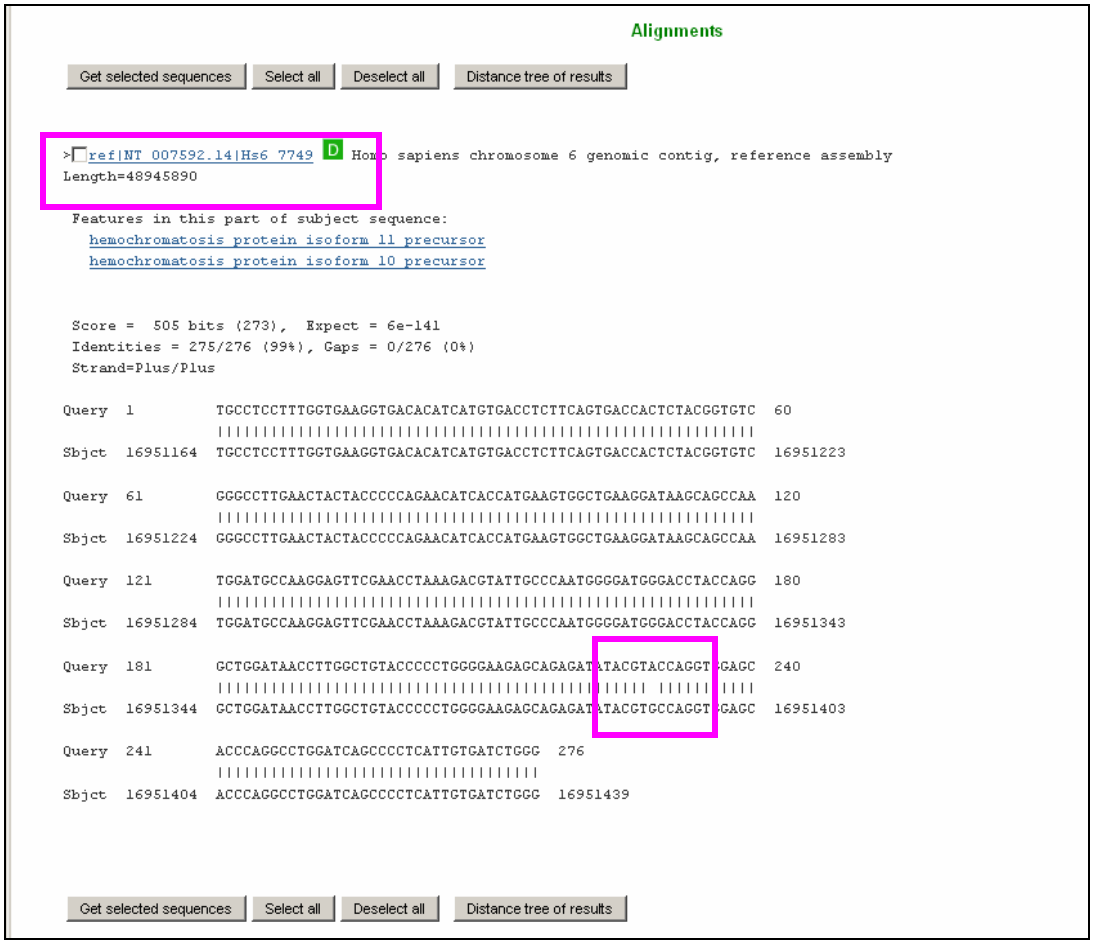





NCBI Home ▸ Genomic Biology ▸ Human Genome Resources ▸ BLAST
Search Map Viewer
Go
Clear
BLAST
overview
FAQs
news
manual
references
Retrieve results
Genome Project
BLAST Human Sequences.
Enter an accession, gi, or a sequence in FASTA format:
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTCGGGCC
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
sequences
Alignments
100
Advanced options:
Begin Search
Clear Input

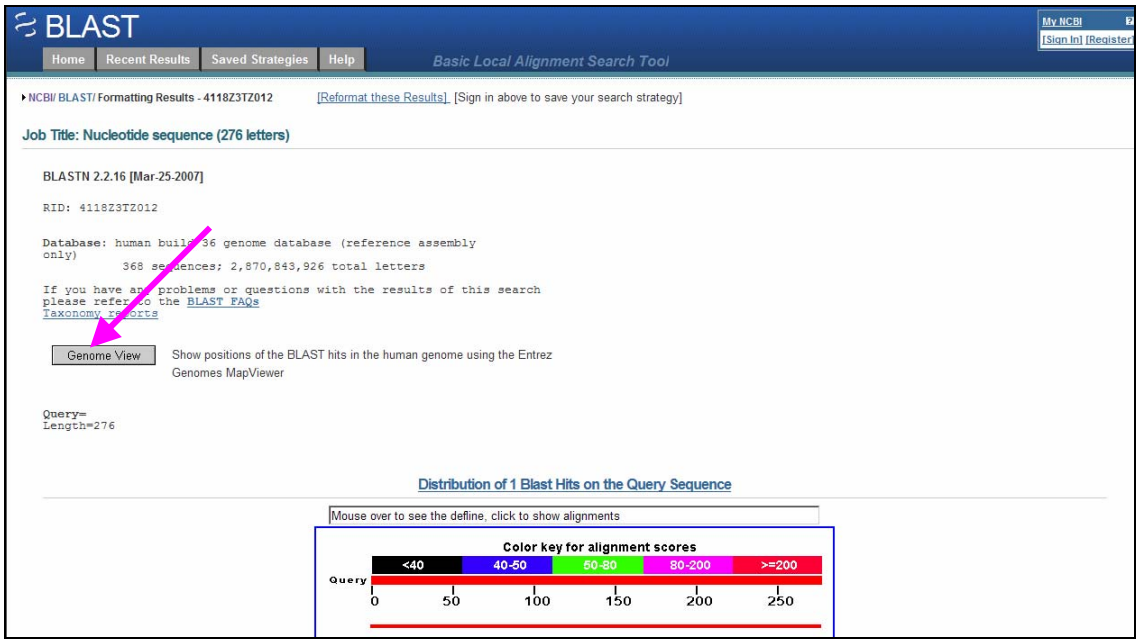

BLAST
Home
Recent Results
Saved Strategies
Help
Basic Local Alignment Search Tool
My NCBI
[Sign In] [Register]
NCBI/ BLAST/ Format Request
Query Nucleotide sequence (276 letters)
Database ref_contig
Job title Nucleotide sequence (276 letters)
Request ID 4118Z3TZ012
View report
Show results in a new window
Format
Show Alignment as HTML
Advanced View
Reset form to defaults
Alignment View Pairwise
Display
Graphical Overview
Linkout
Sequence Retrieval
NCBI-gi
CDS feature
Masking Character: Lower Case
Masking Color: Grey
Limit results
Descriptions: 100
Graphical overview: 100
Alignments: 100
Entrez query:
Organism Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.
Expect Min:
Expect Max:

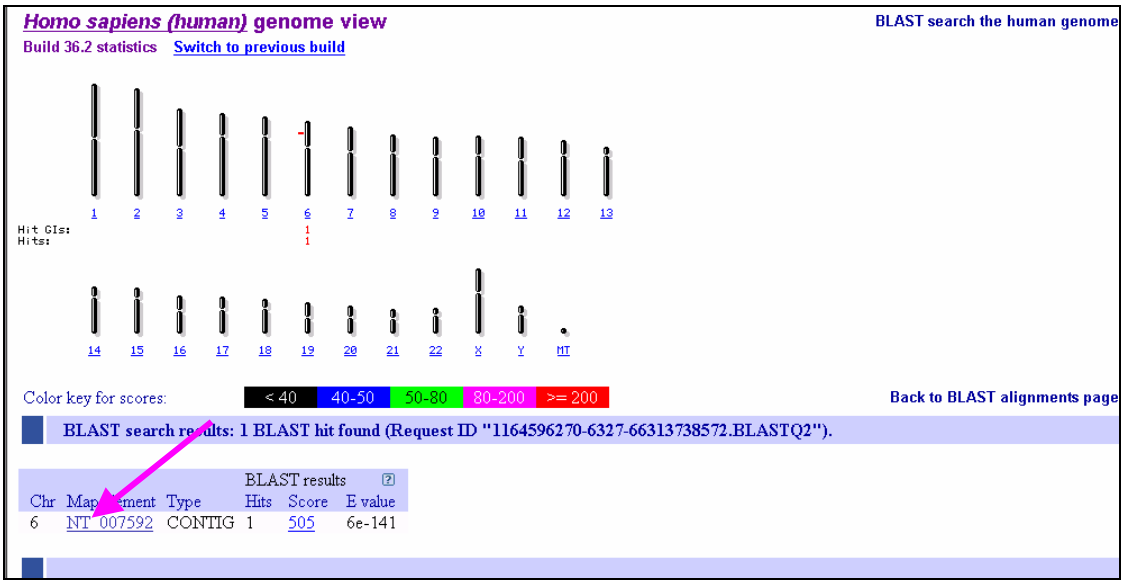


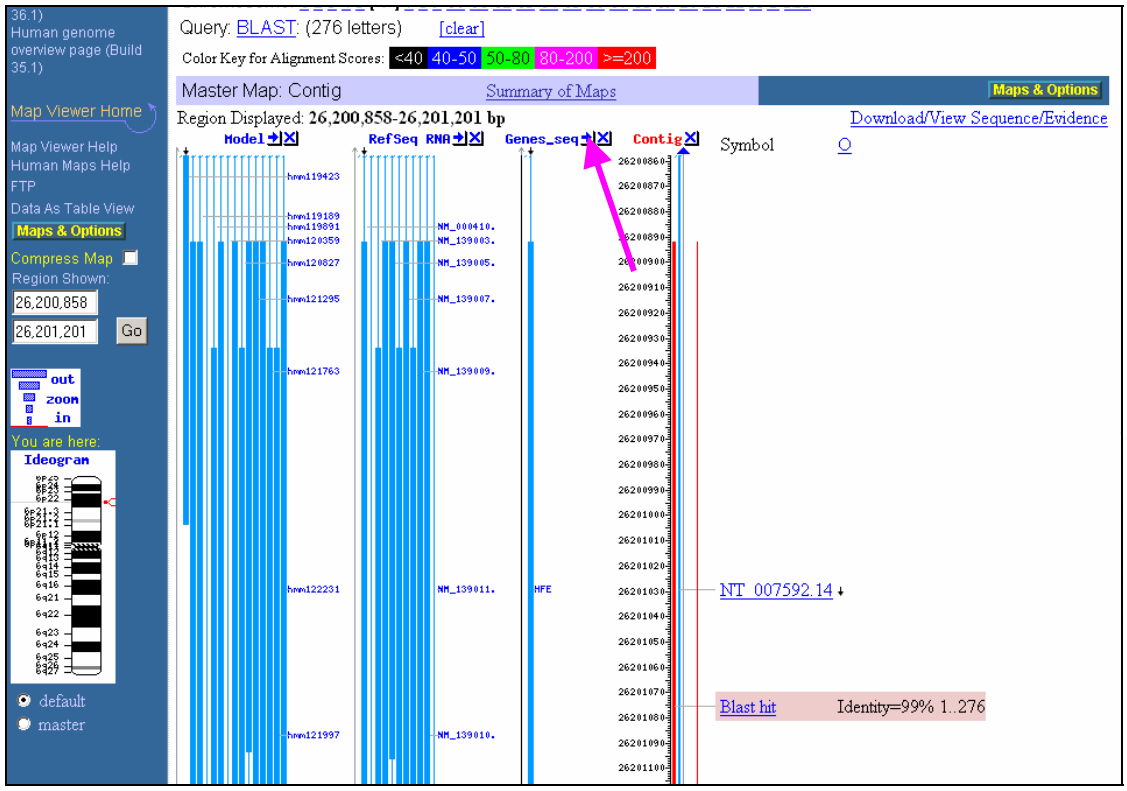


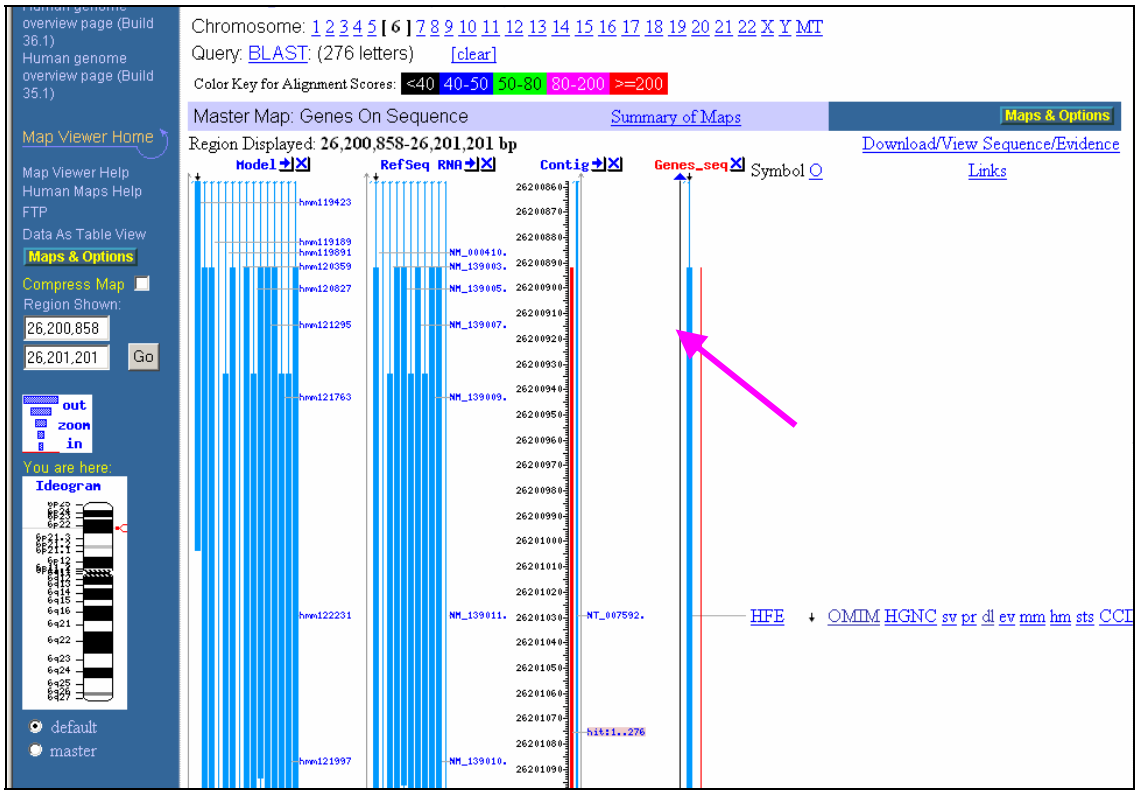
Result: The EST sequence is aligned to the contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

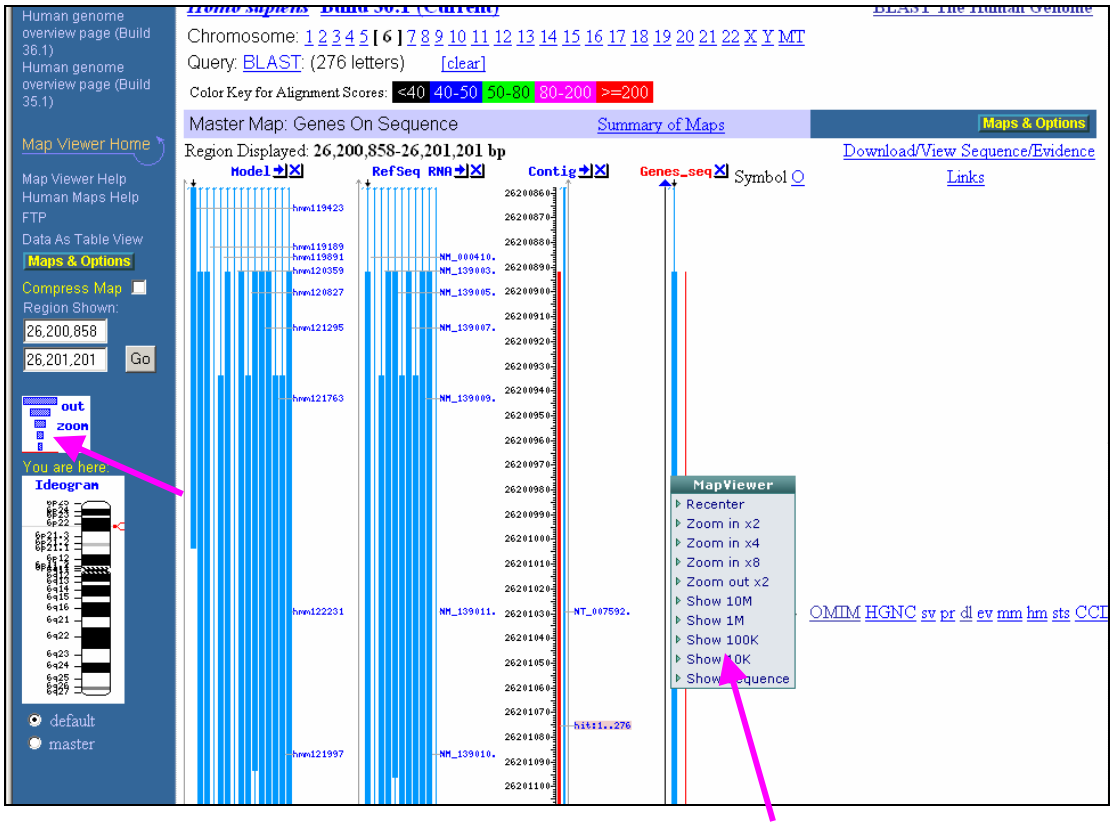
Step 2: Identify the gene(s) expressing the EST and download their sequences

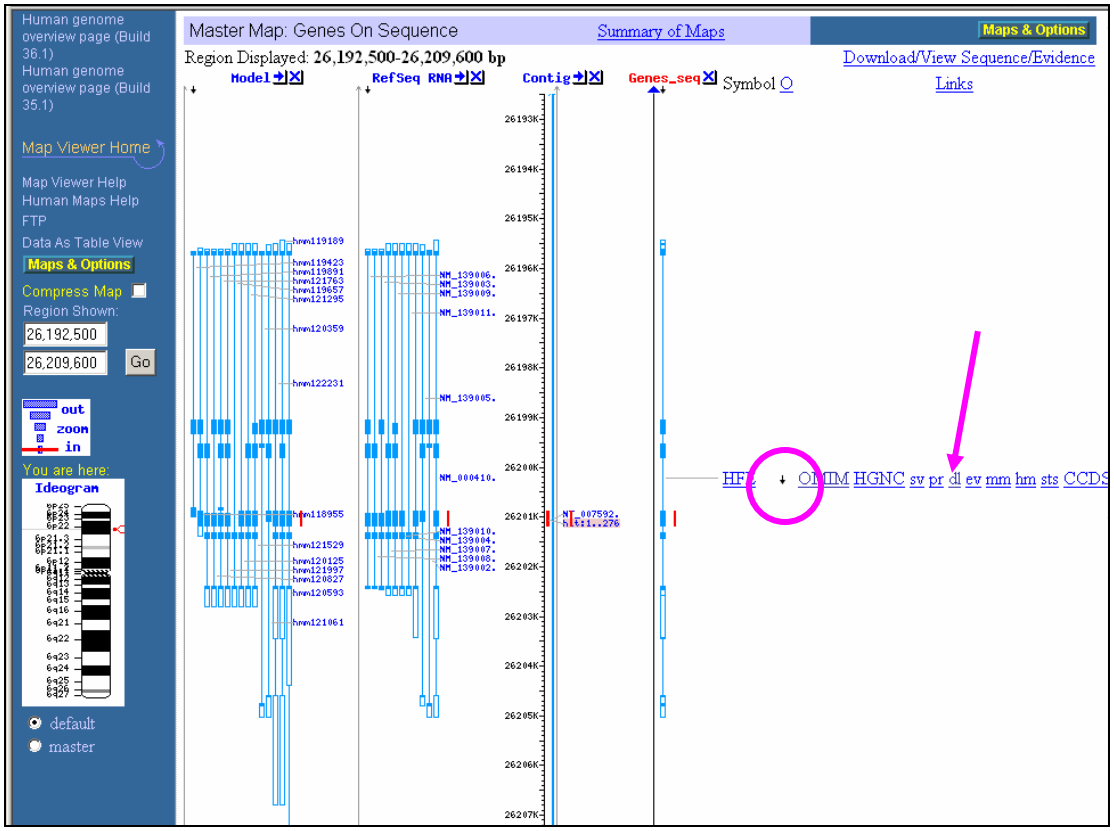


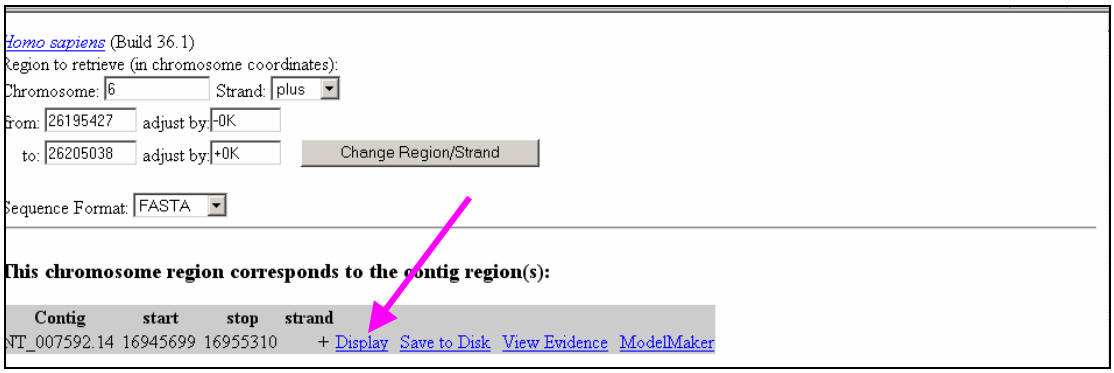


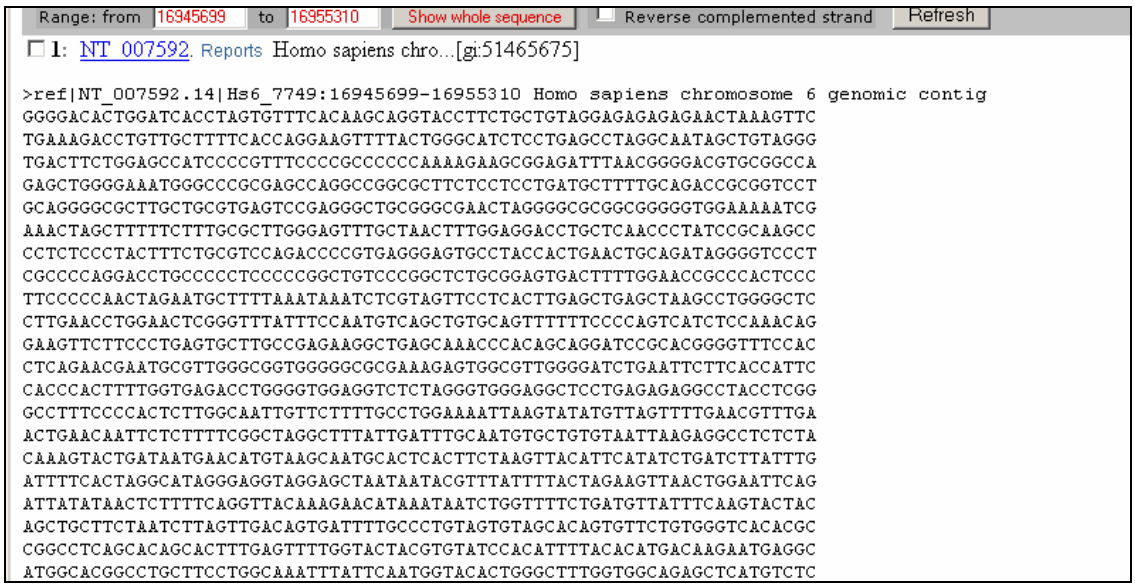

36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom
in
You are here:
Ideogram
default
master
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <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
O
NT_007592.14
Blast hit
Identity=99% 1..276


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 CCI


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
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 10M
Show 1M
Show 100K
OMIM HGNC sv pr dl ev mm hm sts

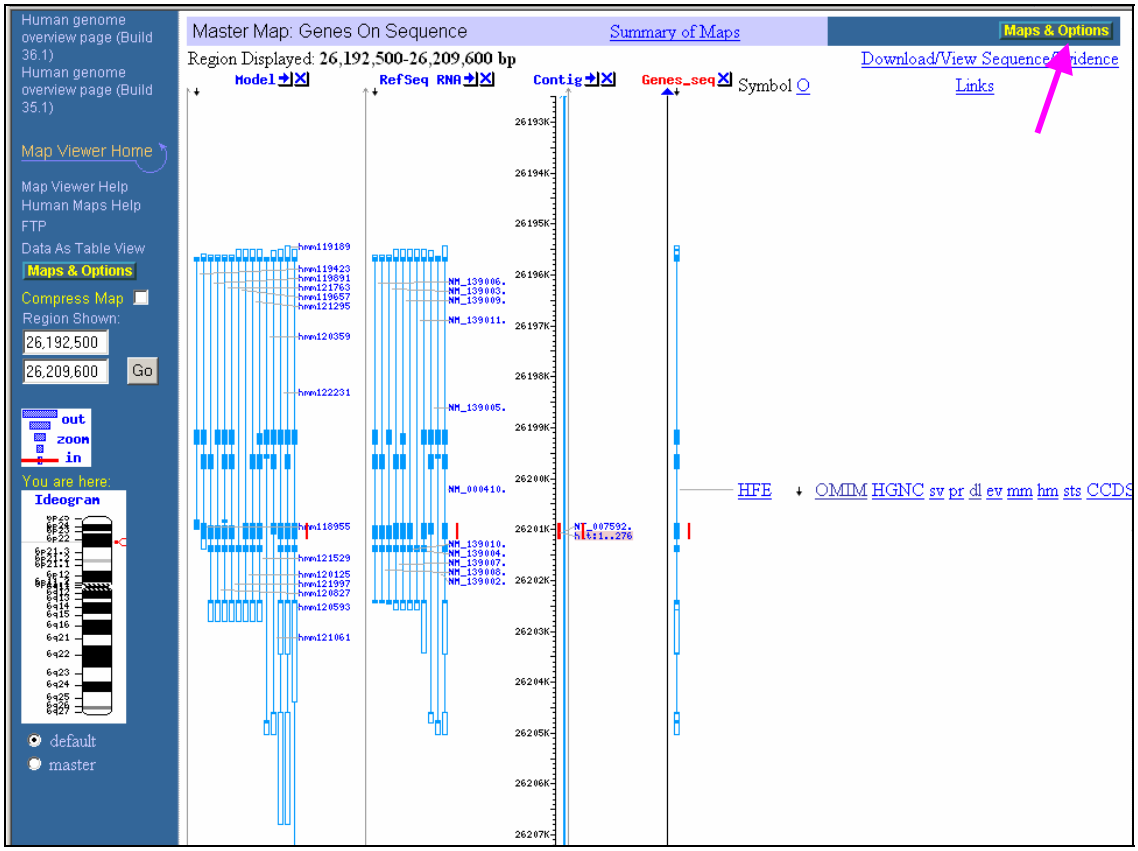

Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,192,500
26,209,600
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Genes On Sequence
Summary 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 O
OMIM HGNC sv pr dl ev mm hm sts CCDS

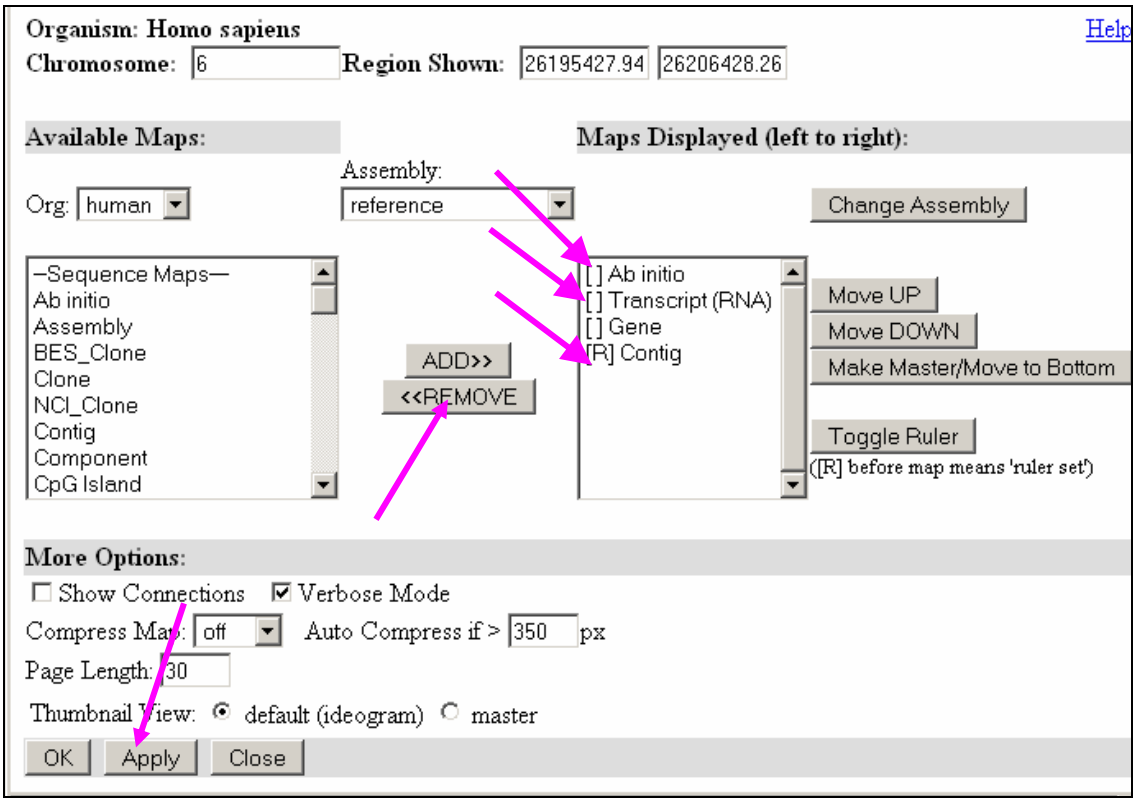


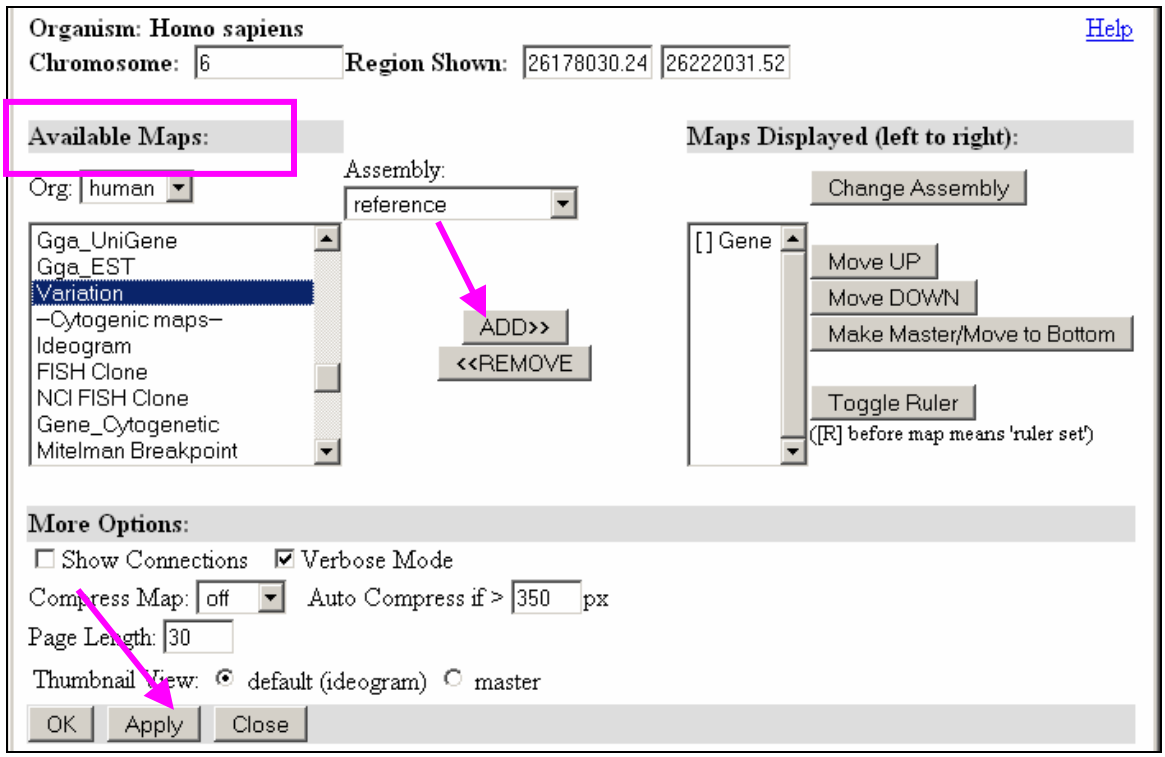


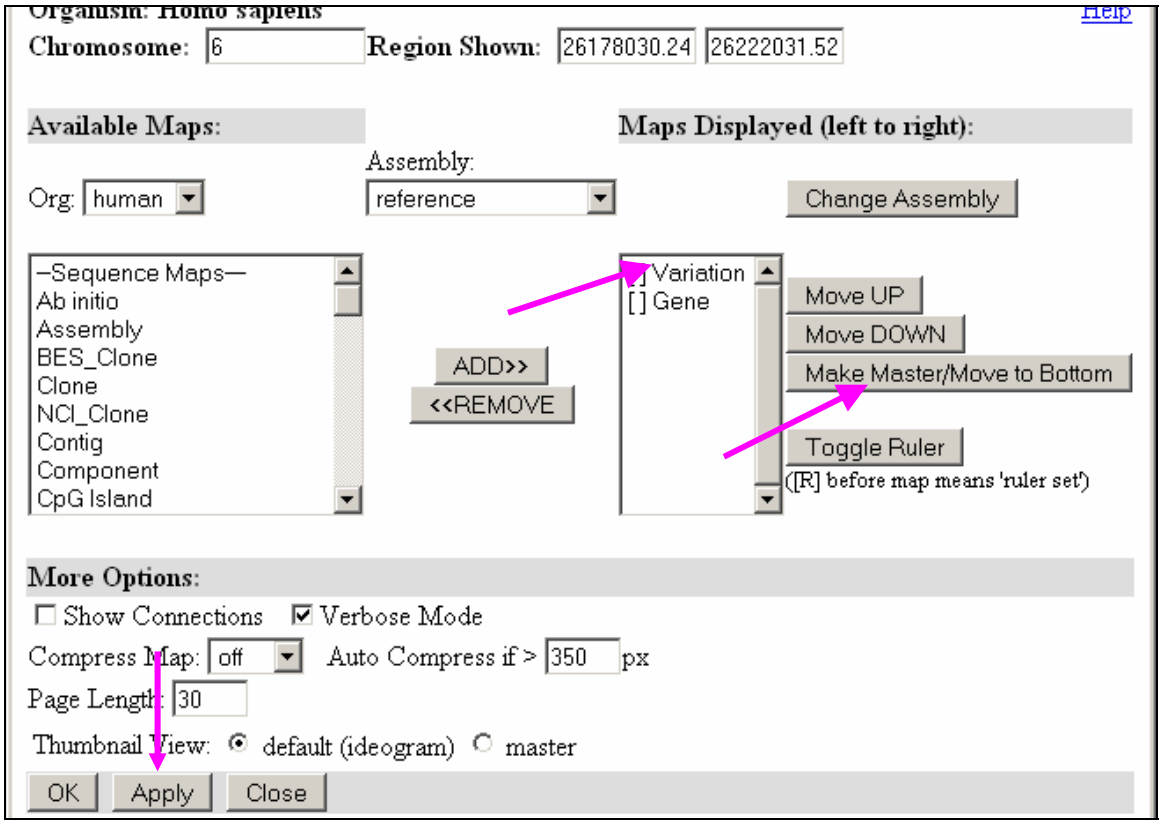
Result: The query EST is expressed by the HFE gene.

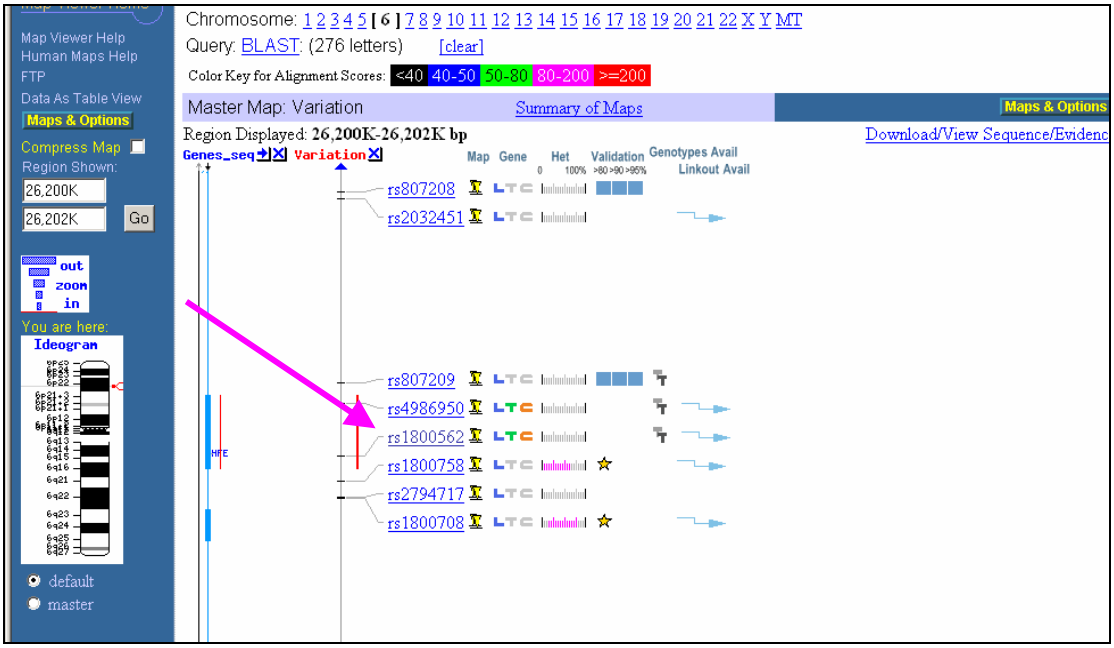
Step 3: Determine whether the EST contains any known SNPs




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
[R] 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		Links , Linkout
Organism:	human (*Homo sapiens*)	Variation Class:	SNP: single nucleotide polymorphism	
Molecule Type:	Genomic	Alleles:	A/G	
Created/Updated in build:	89/126	Ancestral Allele:	G	
Map to Genome Build:	36.2			

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=  |alleles='A/G'|mc=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

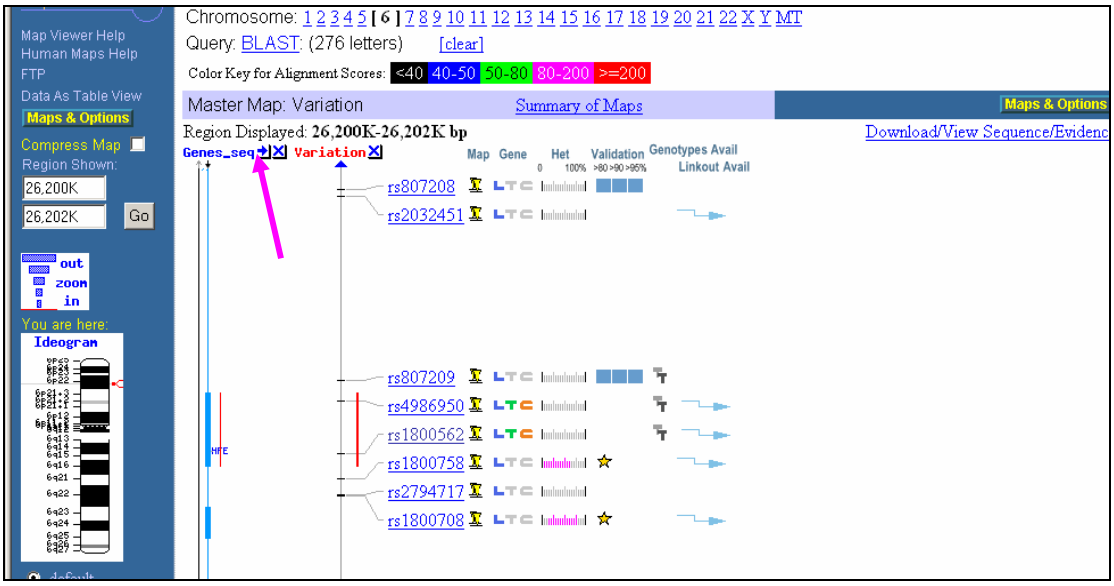
Chromosome	Contig accession	Contig position	Chromosome position	Hit orientation	Contig Allele	Assembly Type	Group label	Contig label	Neighbor SNP	SNP_flank position
6	NT_007592.14	16951392	26201120	plus	G	ref_assembly	reference	reference	view	201
6	NW_922984.1	25717681	27322425	plus	G	alt_assembly_1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

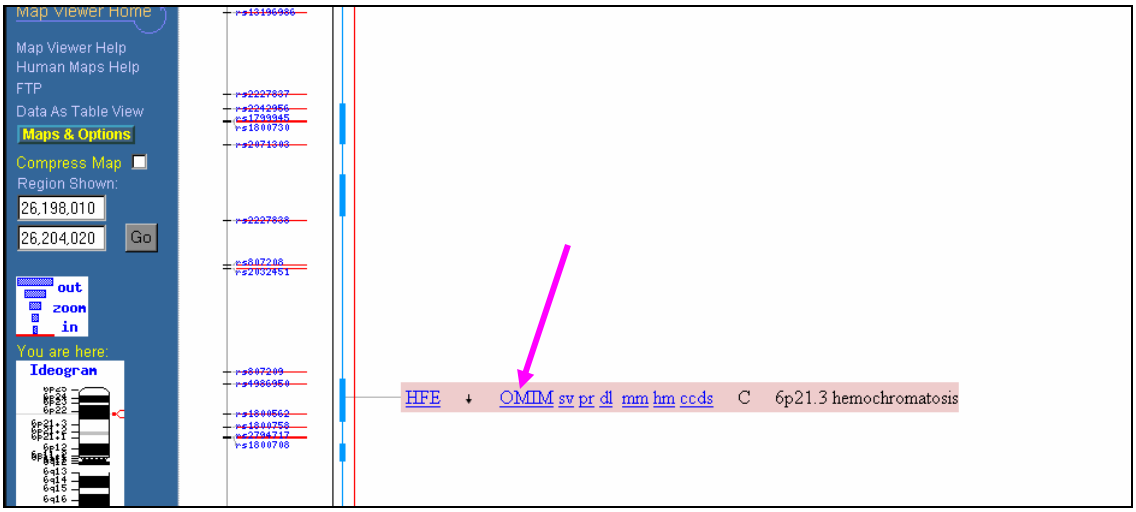
Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) Color Legend
reference	NT_007592->NM_000410	
	sv function	

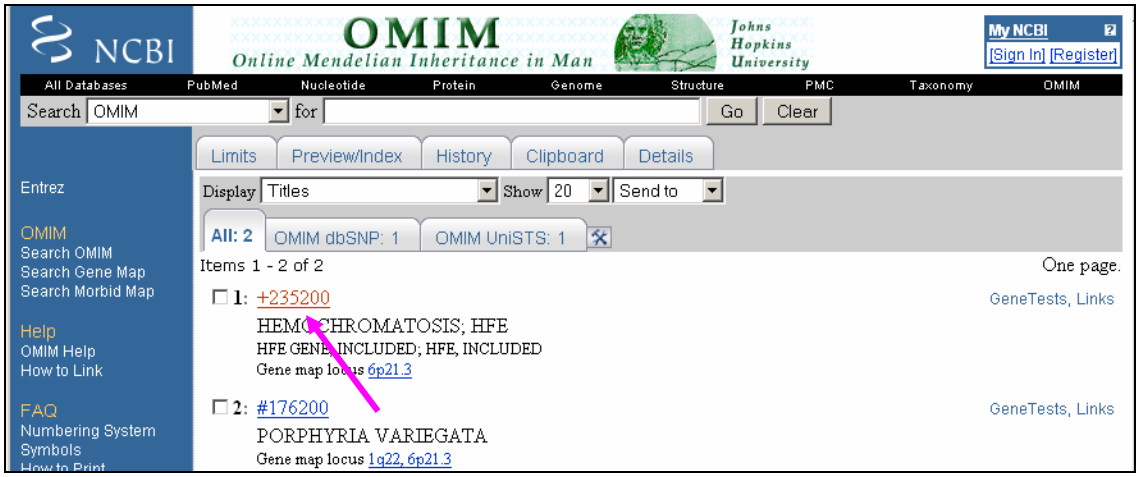
Group label	Contig-->mRNA-->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
reference	NT_007592->NM_000410->NP_000401	16951392	forward	1005	nonsynonymous	A	Tyr [Y]	2	282
					contig reference	G	Cys [C]	2	282

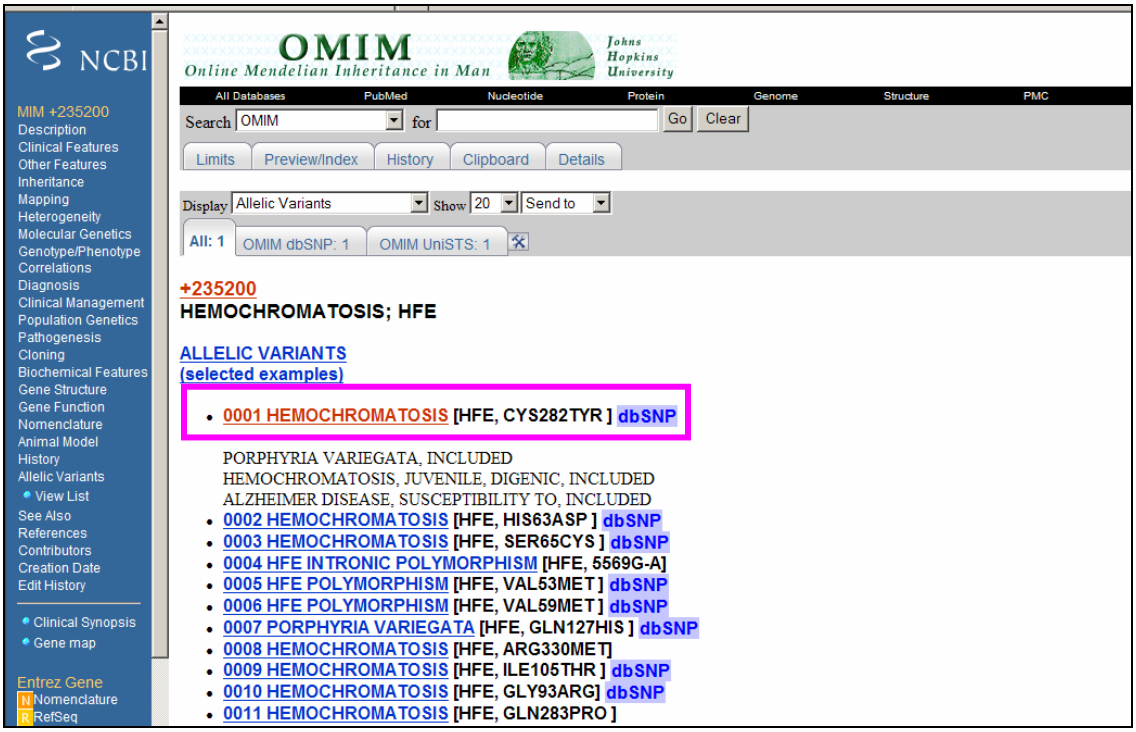
Result: The EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT_007592.14).

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









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

Identification of Disease Genes

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).

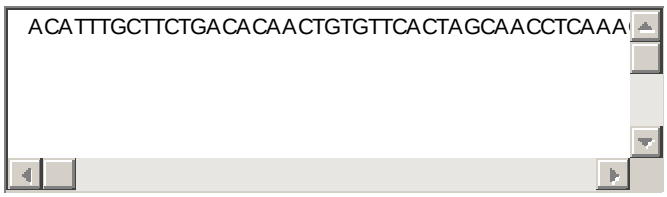
Step 1. Compare ESTs to the human genome (using BLAST):

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 "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 numerous 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?

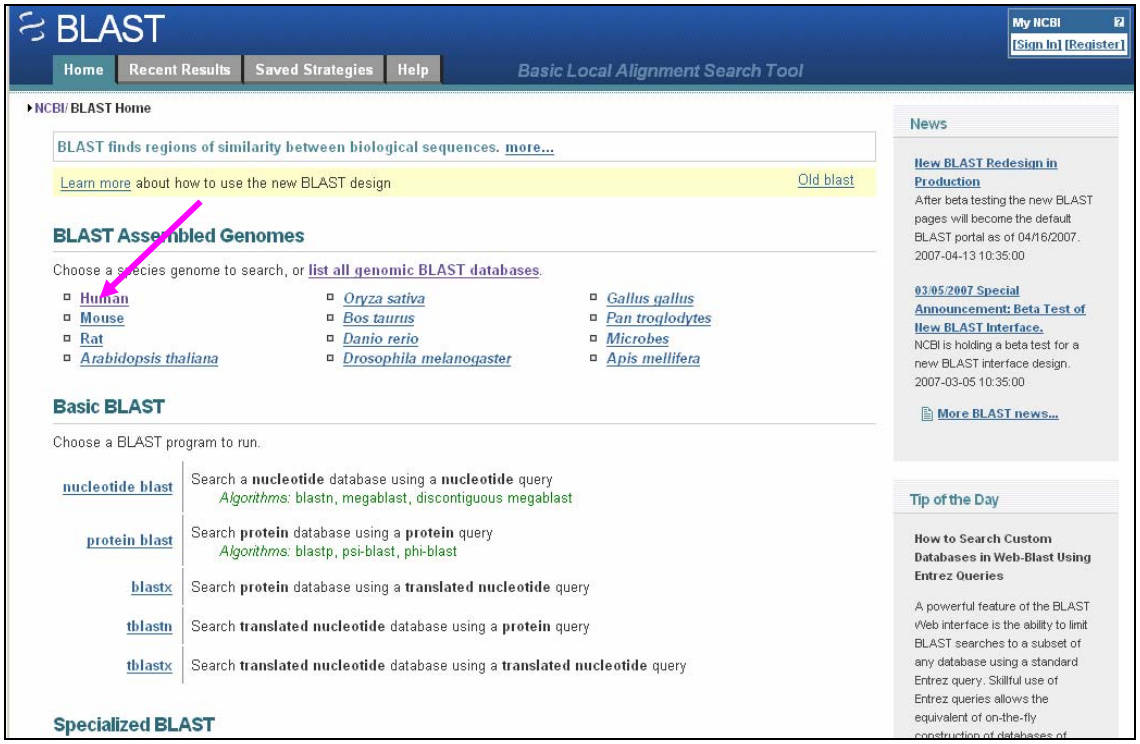
Summary:

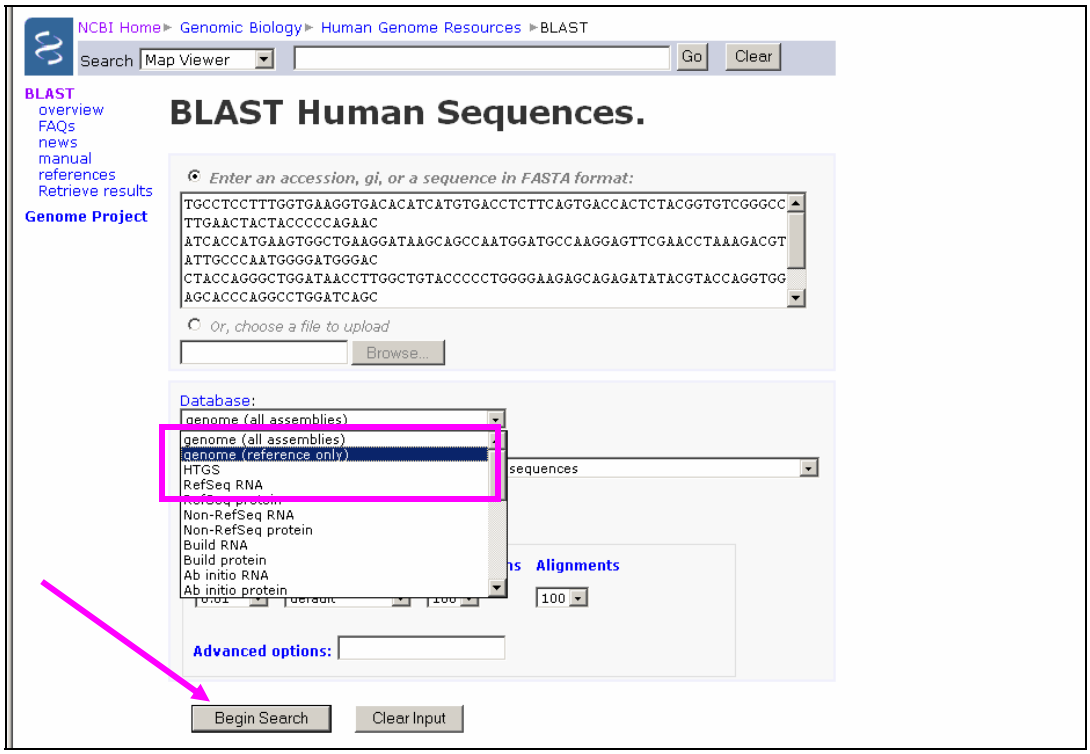
This mini-course describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

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.

Screen images

Step 1: Compare EST against the human genome.





BLAST — Basic Local Alignment Search Tool

Home | Recent Results | Saved Strategies | Help

NCBI/ BLAST/ Format Request

Query: Nucleotide sequence (232 letters)

Database: ref_contig

Job title: Nucleotide sequence (232 letters)

Request ID: 4KWW22GY012 — View report — Show results in a new window

Format

Show: Alignment as HTML — Advanced View — Reset form to defaults

Alignment View: Pairwise

Display: Graphical Overview | Linkout | Sequence Retrieval | NCBI-gi | CDS feature

Masking Character: Lower Case — Masking Color: Grey

Limit results: Descriptions: 100 — Graphical overview: 100 — Alignments: 100

Entrez query:

Organism — Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.

Expect Min: — Expect Max:

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback on new interface

BLAST — Basic Local Alignment Search Tool

Home | Recent Results | Saved Strategies | Help

My NCBI [Sign In] [Register]

NCBI/ BLAST/ Formatting Results - 4KWW22GY012 [Reformat these Results] [Sign in above to save your search strategy]

Job Title: Nucleotide sequence (232 letters)

BLASTN 2.2.16 [Mar-25-2007]

RID: 4KWW22GY012

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=232

Distribution of 4 Blast Hits on the Query Sequence

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

Color key for alignment scores

<40	40-50	50-80	80-200	>=200

Query: 0 40 80 120 160 200

Alignments

Get selected sequences | Select all | Deselect all | Distance tree of results

```
>ref|NT_009237.17|Hs11_9394 D Homo sapiens chromosome 11 genomic contig, reference assembly
Length=49571094

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position

Features in this part of subject sequence:
  beta globin

 Score =  259 bits (140),  Expect = 4e-67
 Identities = 142/143 (99%), Gaps = 0/143 (0%)
 Strand=Plus/Minus

Query  1        ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4035542  ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC  4035483

Query  61       TGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG  120
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4035482  TGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG  4035423

Query  121      TTGGTGGTGAGGCCCTGGGCAGG  143
                |||||||||||||||||||||||
Sbjct  4035422  TTGGTGGTGAGGCCCTGGGCAGG  4035400

Features in this part of subject sequence:
  delta globin

 Score =  204 bits (110),  Expect = 2e-50
 Identities = 133/144 (92%), Gaps = 2/144 (1%)
 Strand=Plus/Minus

Query  1        ACA-TTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCAT  59
                ||| ||| ||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Sbjct  4042954  ACACTTT-CTTCTGACATAACAGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCAT  4042896
```

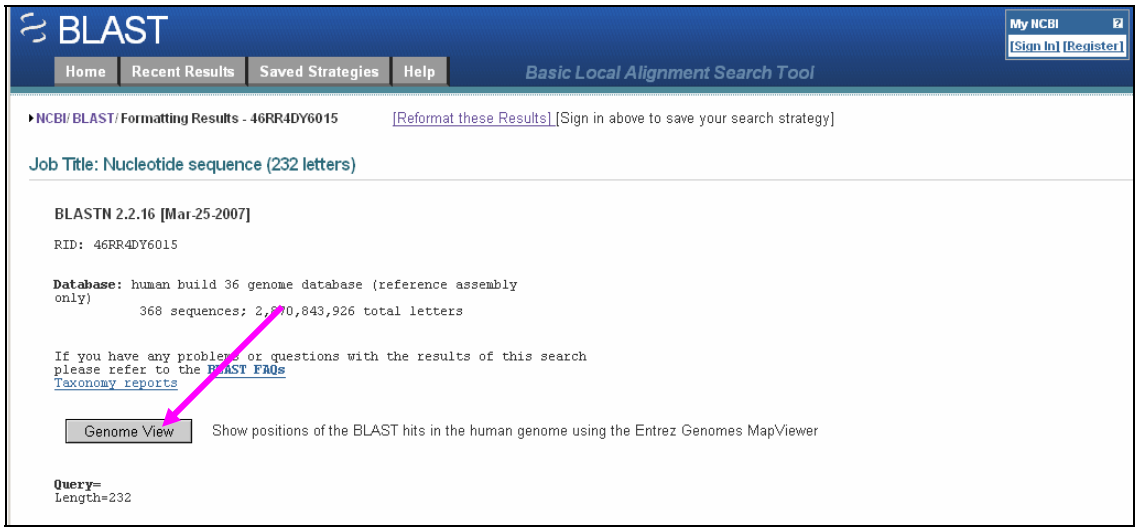
Alignments

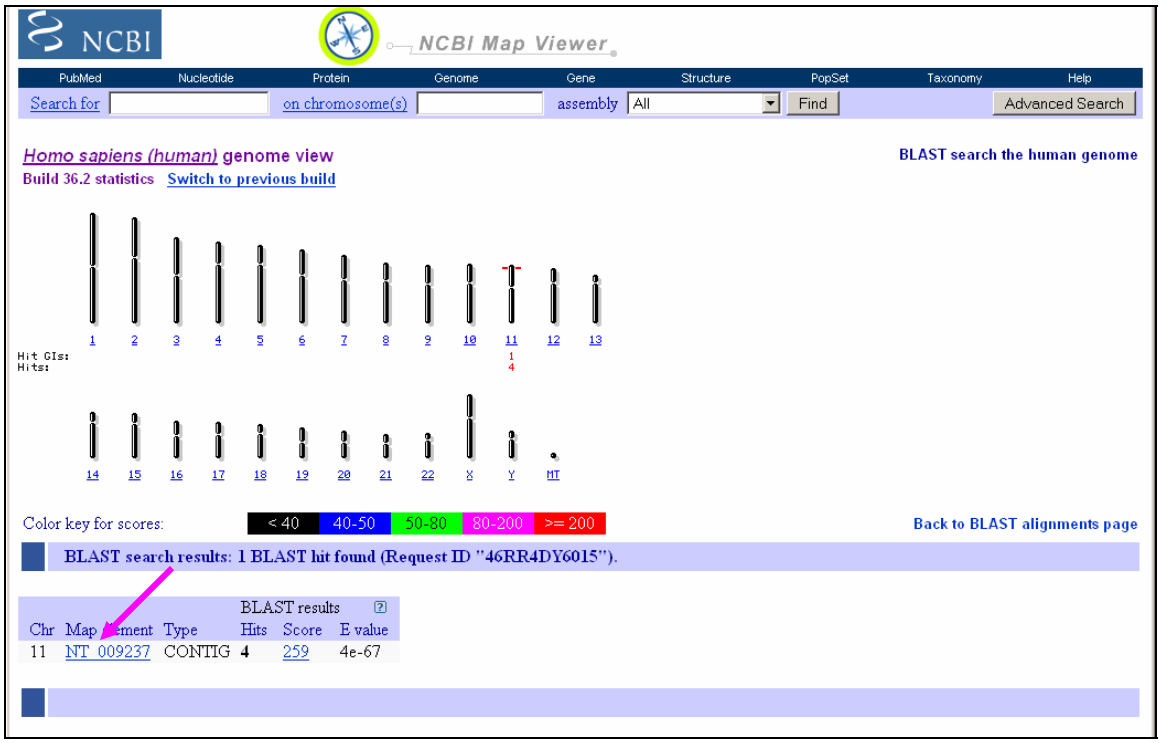
Get selected sequences | Select all | Deselect all | Distance tree of results

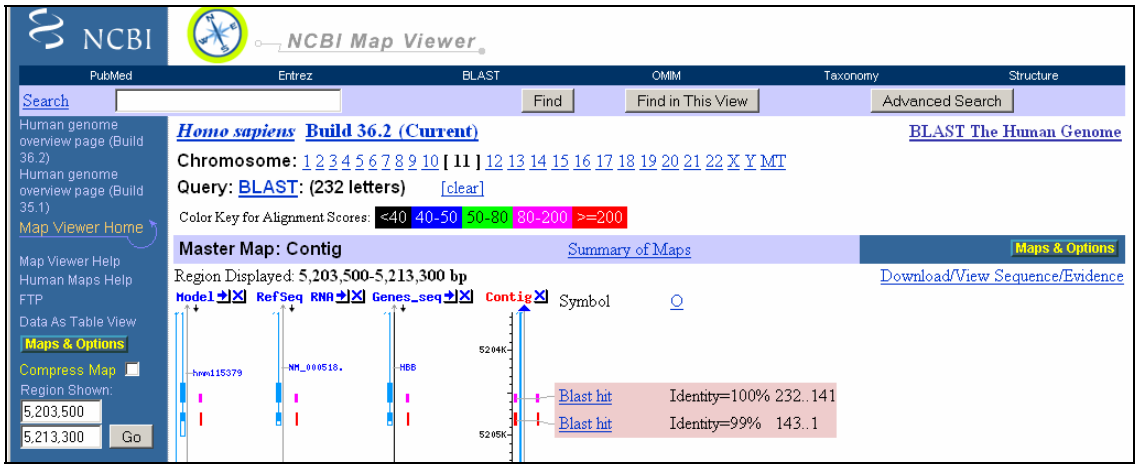
```
>ref|NT_009237.17|Hs11_9394 D Homo sapiens chromosome 11 genomic contig, reference assembly
Length=49571094

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position

Features in this part of subject sequence:
  beta globin

 Score =  171 bits (92),  Expect = 2e-40
 Identities = 92/92 (100%), Gaps = 0/92 (0%)
 Strand=Plus/Minus

Query  141      AGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCC  200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4035272  AGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCC  4035213

Query  201      ACTCCTGATGCTGTTATGGGCAACCCTAAGGT  232
                ||||||||||||||||||||||||||||||||
Sbjct  4035212  ACTCCTGATGCTGTTATGGGCAACCCTAAGGT  4035181

Features in this part of subject sequence:
  beta globin

 Score =  259 bits (140),  Expect = 4e-67
 Identities = 142/143 (99%), Gaps = 0/143 (0%)
 Strand=Plus/Minus

Query  1        ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4035542  ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC  4035483

Query  61       TGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG  120
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4035482  TGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG  4035423

Query  121      TTGGTGGTGAGGCCCTGGGCAGG  143
                |||||||||||||||||||||||
Sbjct  4035422  TTGGTGGTGAGGCCCTGGGCAGG  4035400
```

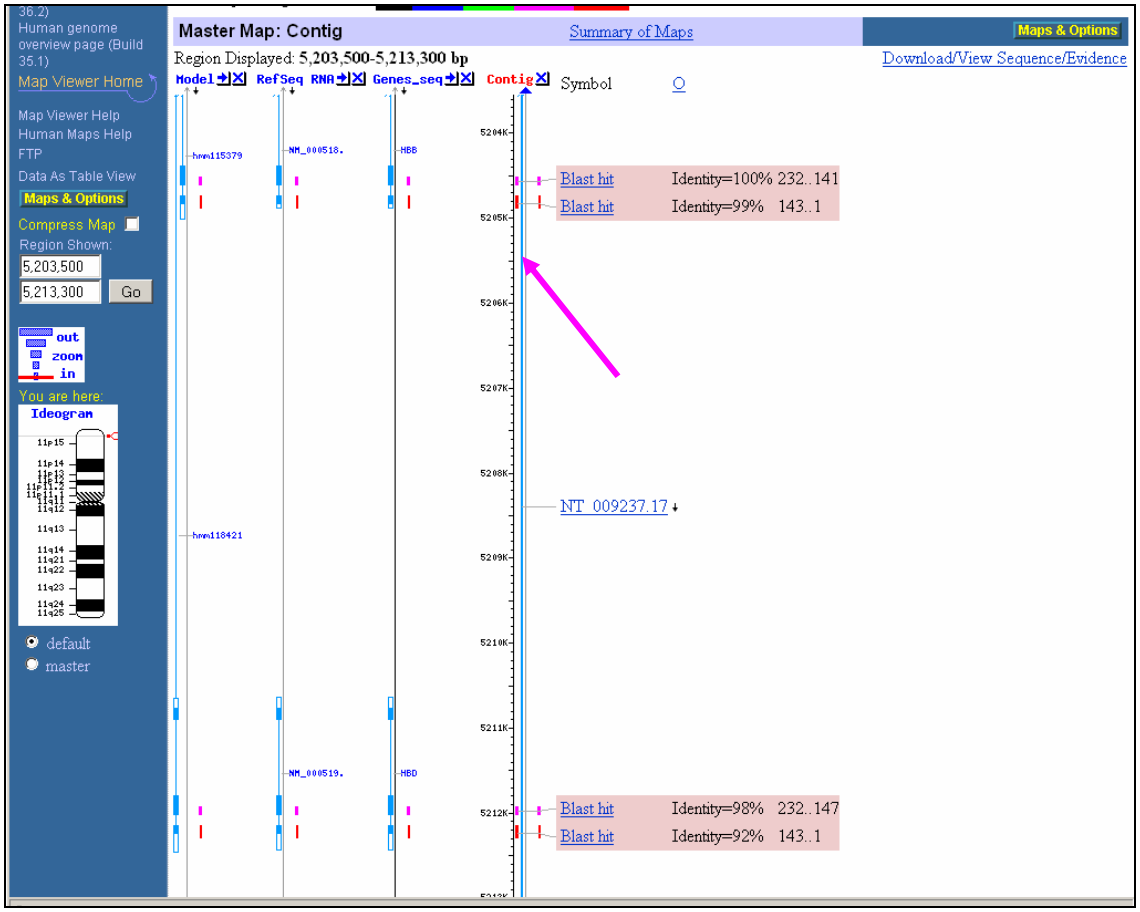
Result: Two segments of the EST sequence aligned to four different locations (corresponding with two individual exons of each- beta globin and delta globin) on the reference contig NT_009237.17 on chromosome 11.
There is one nucleotide difference (T to A) with respect to the contig position 4035473.

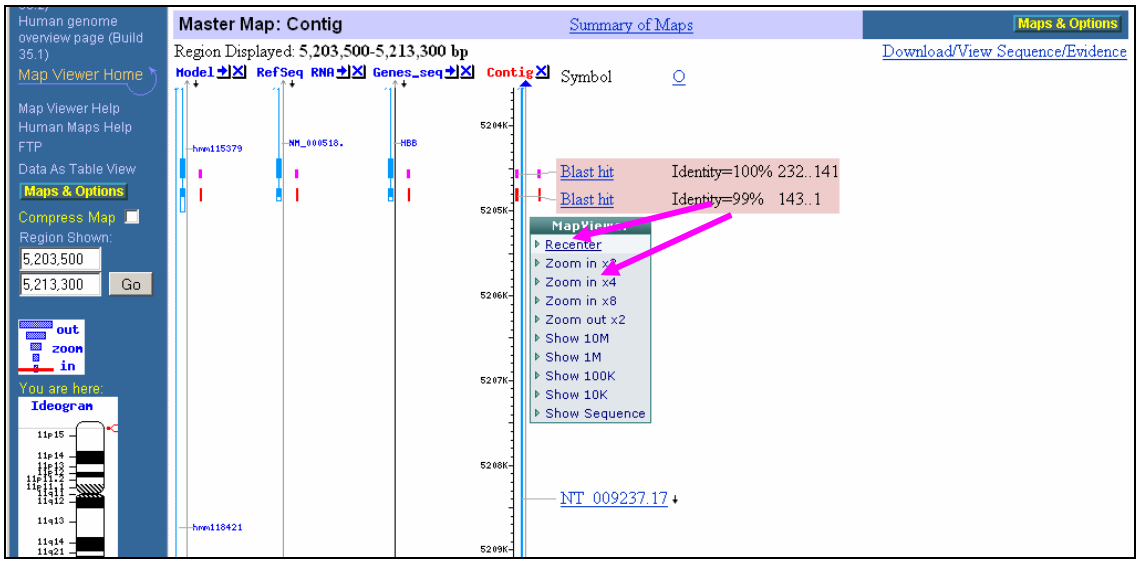
Step 2: Identify the gene(s) expressing the EST and download their sequences

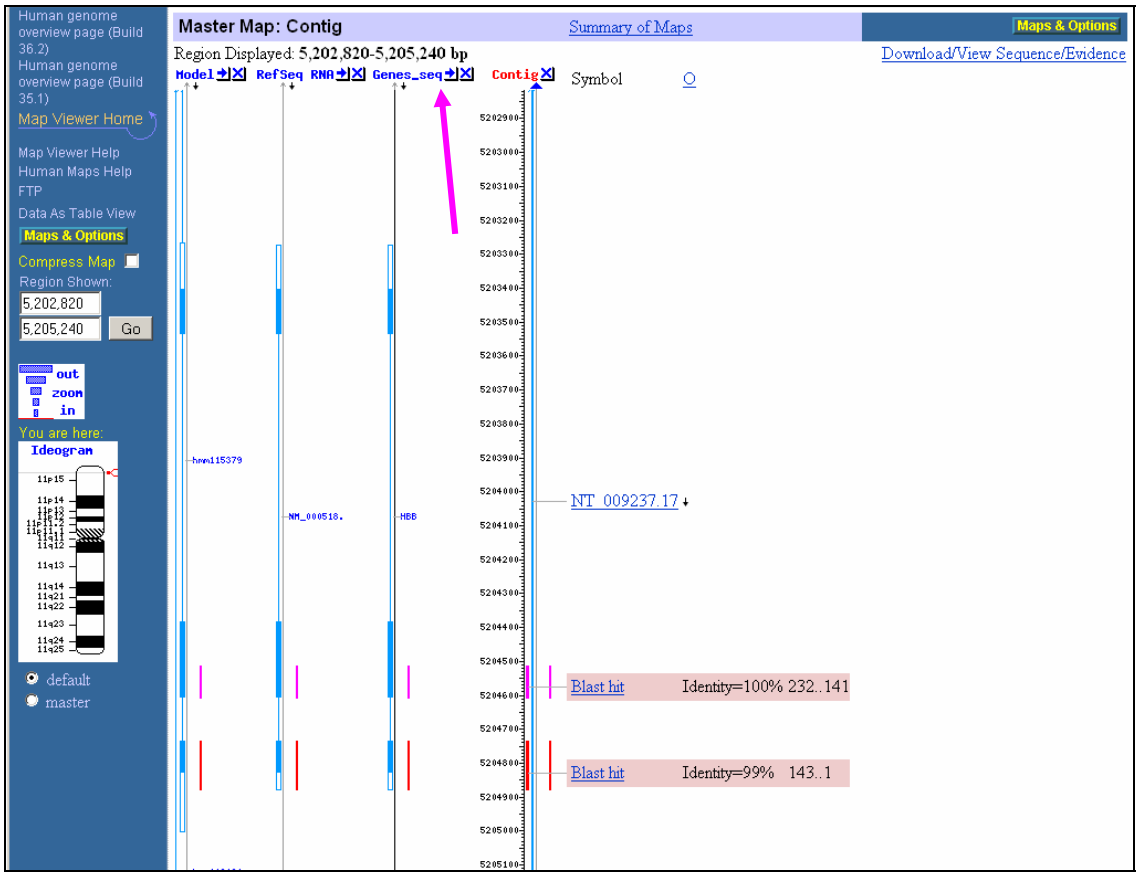


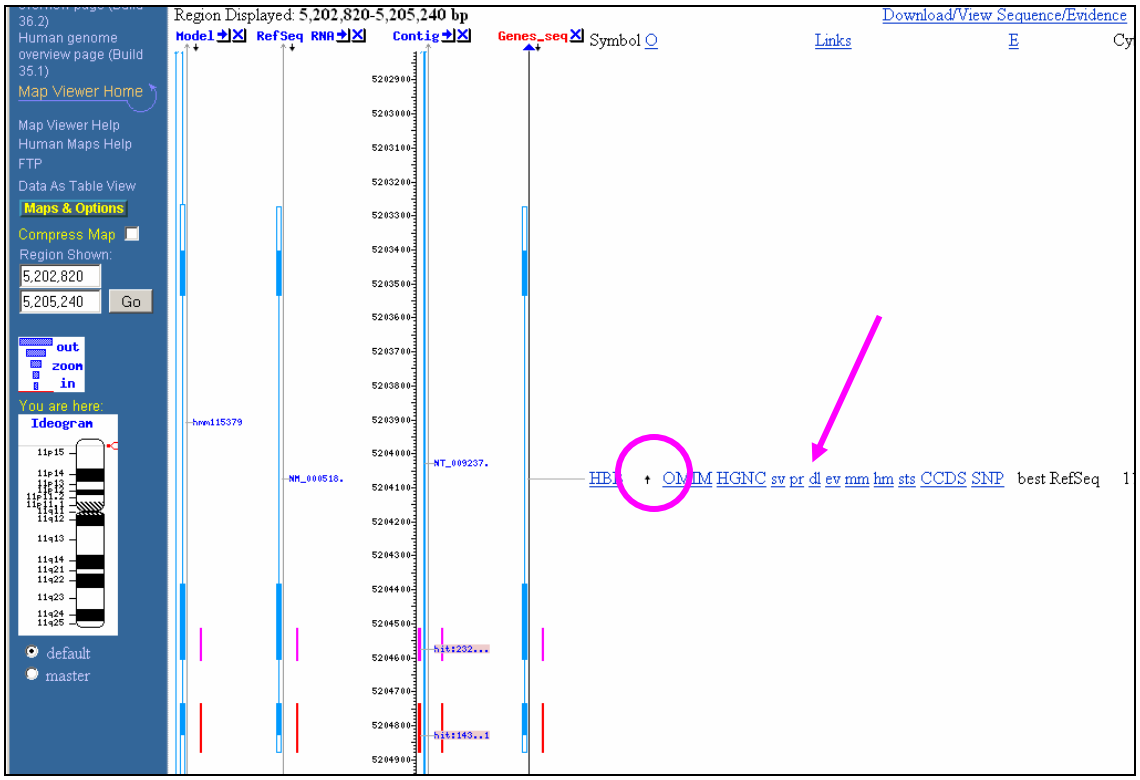


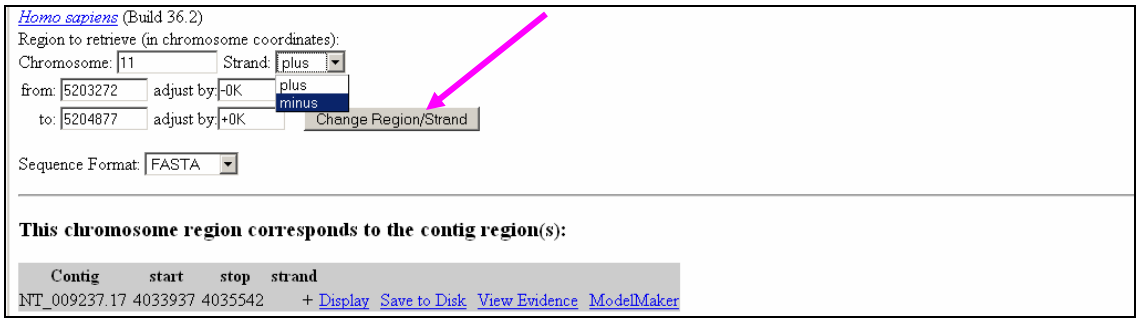

NCBI
NCBI Map Viewer
PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure
Search
Find
Find in This View
Advanced Search
Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
5,203,500
5,213,300
Go
Homo sapiens Build 36.2 (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: BLAST: (232 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: 5,203,500-5,213,300 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
Blast hit Identity=100% 232..141
Blast hit Identity=99% 143..1

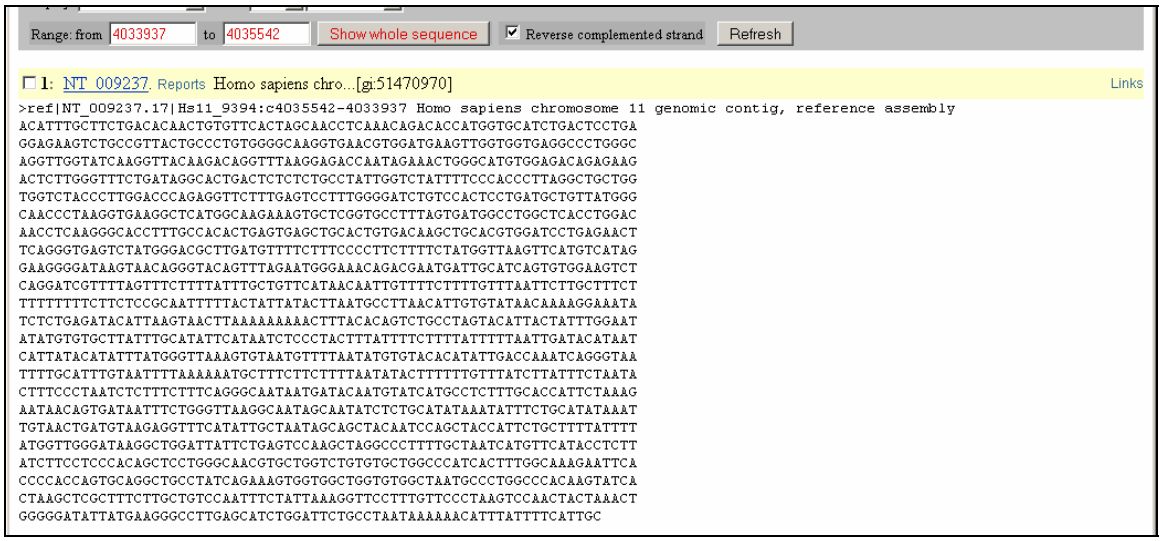

Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
5,203,500
5,213,300
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Contig
Summary of Maps
Maps & Options
Region Displayed: 5,203,500-5,213,300 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
Blast hit Identity=100% 232..141
Blast hit Identity=99% 143..1
NT_009237.17
Blast hit Identity=98% 232..147
Blast hit Identity=92% 143..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:
5,203,500
5,213,300
Go
out
zoom
in
You are here:
Ideogram
Master Map: Contig
Summary of Maps
Maps & Options
Region Displayed: 5,203,500-5,213,300 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
Blast hit
Identity=100% 232..141
Blast hit
Identity=99% 143..1
Recenter
Zoom in x2
Zoom in x4
Zoom in x8
Zoom out x2
Show 10M
Show 1M
Show 100K
Show 10K
Show Sequence
NT_009237.17


Human genome overview page (Build 36.2)
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:
5,202,820
5,205,240
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Contig
Summary of Maps
Maps & Options
Region Displayed: 5,202,820-5,205,240 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
NT_009237.17
Blast hit
Identity=100% 232..141
Blast hit
Identity=99% 143..1

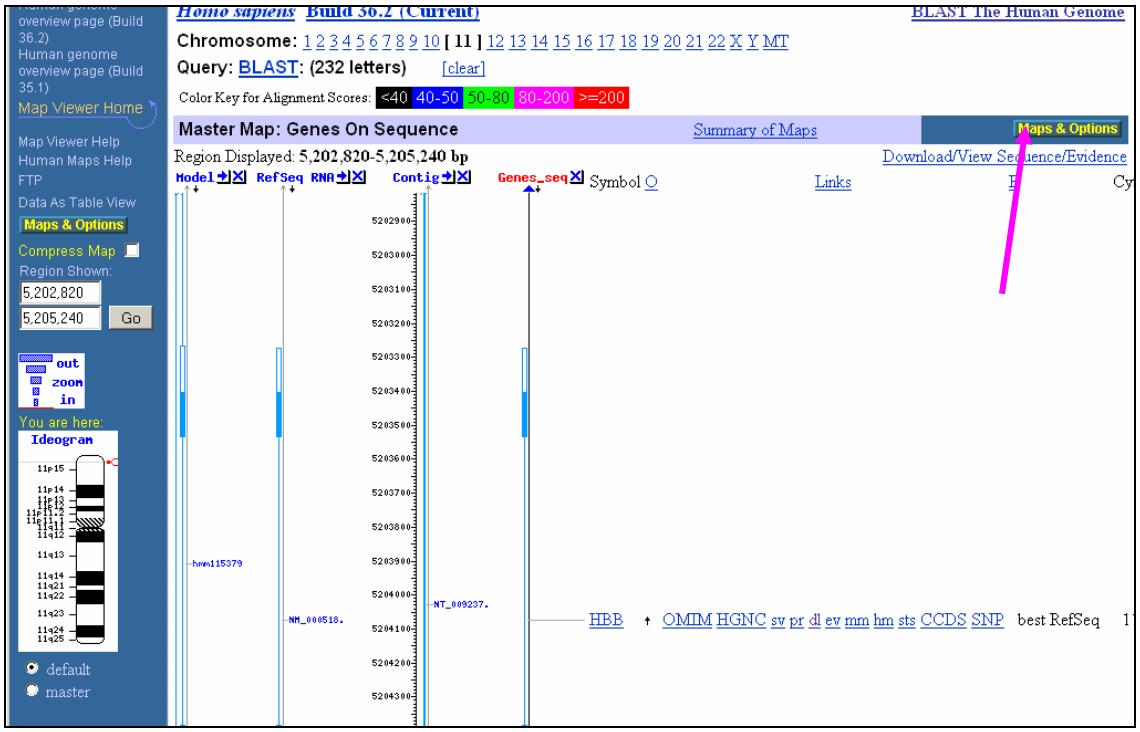


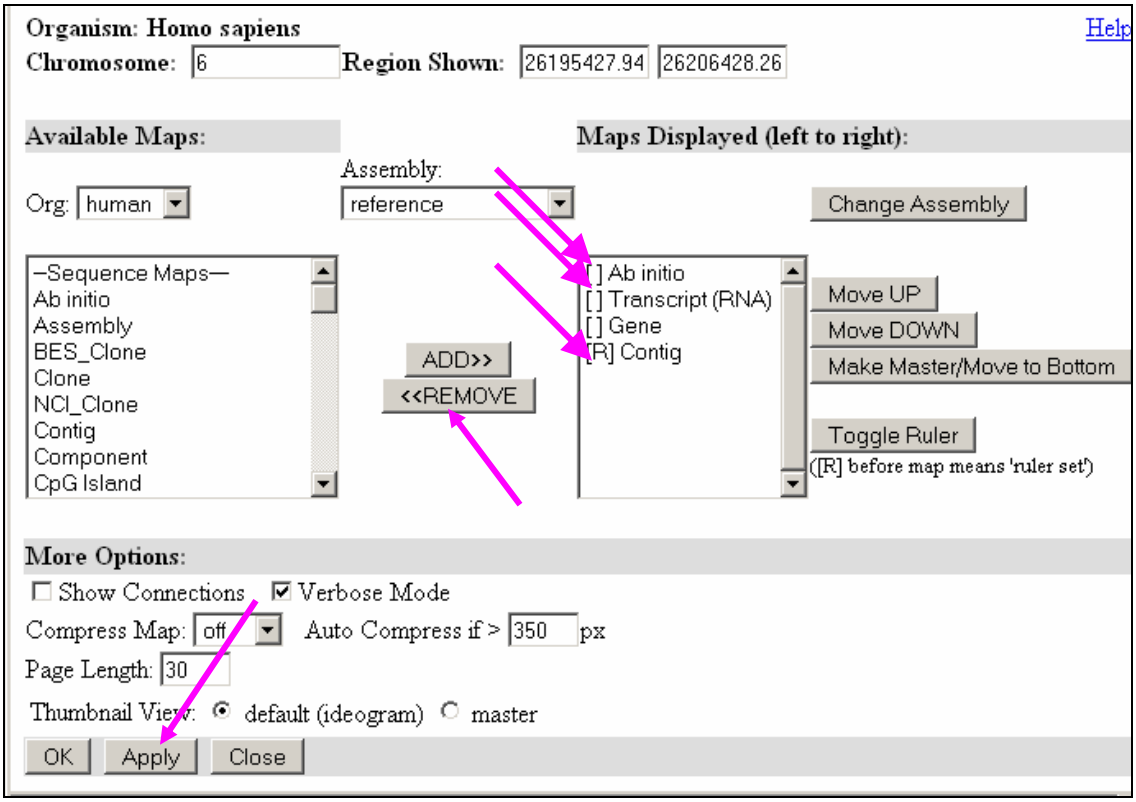




Result: The query EST is expressed by the HBB gene.

Step 3: Determine whether the EST contains any known SNPs





Organism: Homo sapiens
Help
Chromosome: 6 Region Shown: 26178030.24 26222031.52

Available Maps:
Org: human
Assembly: reference
Gga_UniGene
Gga_EST
Variation
–Cytogenic maps–
Ideogram
FISH Clone
NCI FISH Clone
Gene_Cytogenetic
Mitelman Breakpoint
ADD>>
<<REMOVE

Maps Displayed (left to right):
Change Assembly
[ ] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')

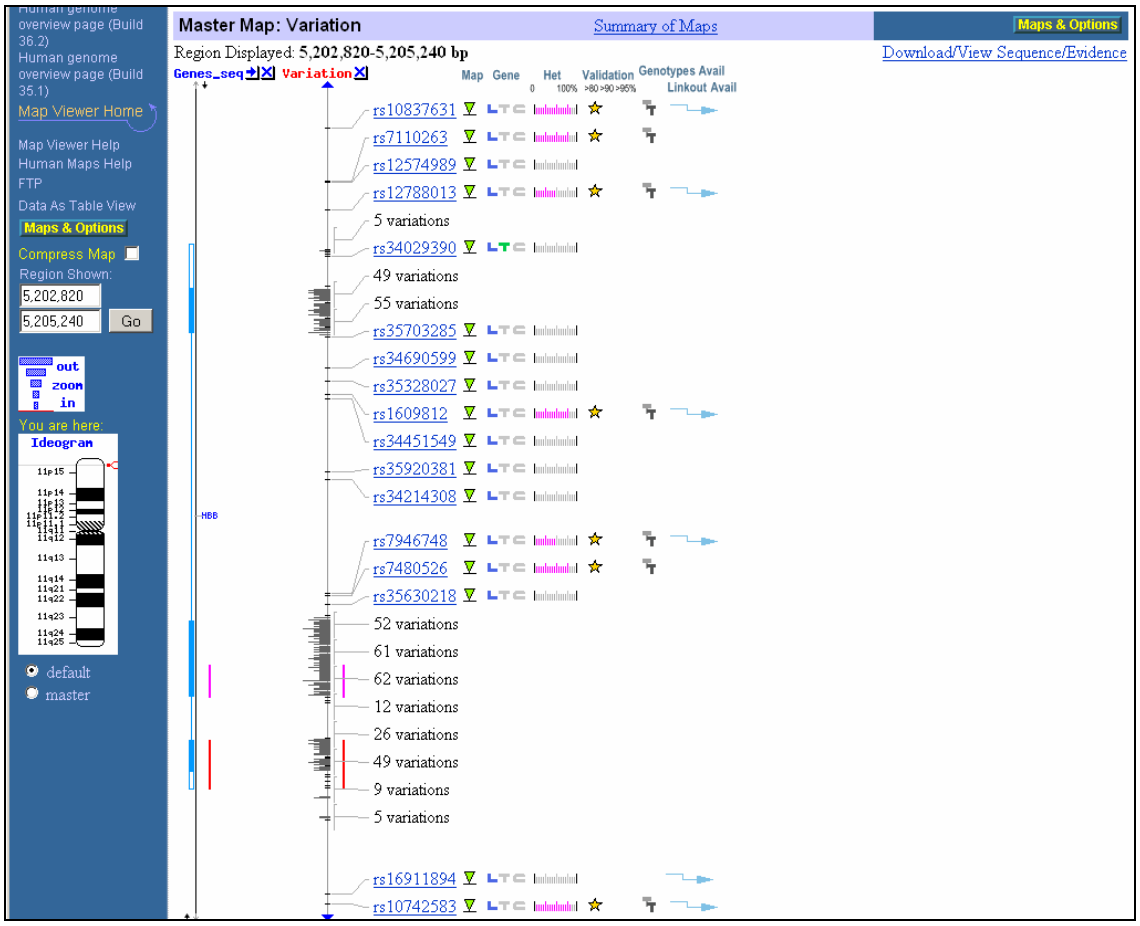
More Options:
☐ Show Connections ☑ Verbose Mode
Compress Map: off Auto Compress if > 350 px
Page Length: 30
Thumbnail View: ◉ default (ideogram) ○ master
OK Apply Close

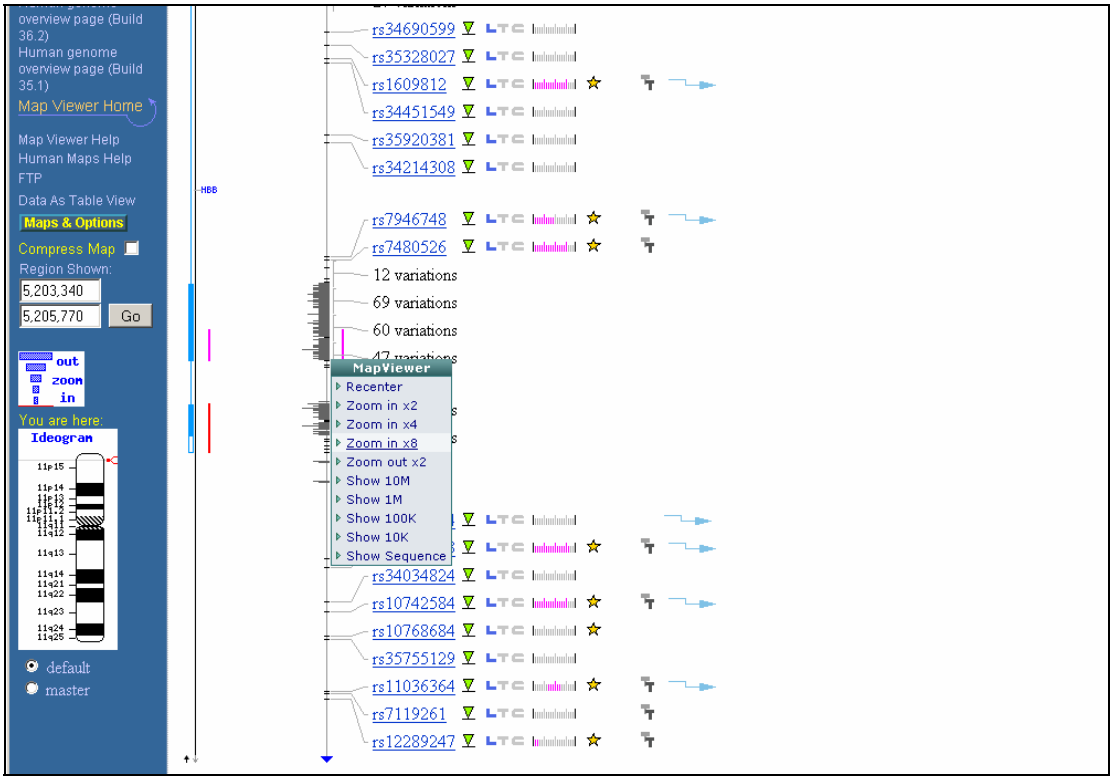
Organism: Homo sapiens
Help
Chromosome: 6 Region Shown: 26178030.24 26222031.52

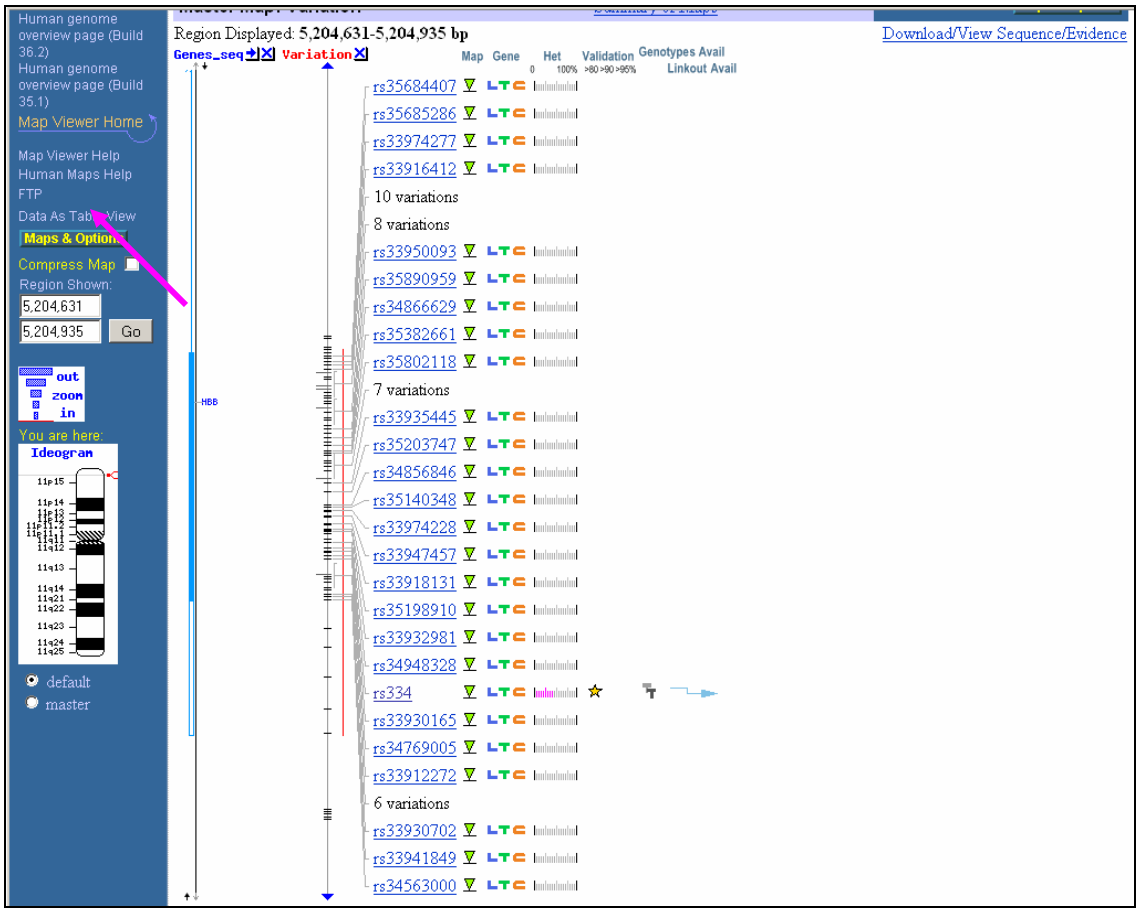
Available Maps:
Org: human
Assembly: reference
–Sequence Maps–
Ab initio
Assembly
BES_Clone
Clone
NCI_Clone
Contig
Component
CpG Island
ADD>>
<<REMOVE

Maps Displayed (left to right):
Change Assembly
[ ] 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


Human genome overview page (Build 36.2)
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:
5,202,820
5,205,240
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Variation
Summary of Maps
Maps & Options
Region Displayed: 5,202,820-5,205,240 bp
Download/View Sequence/Evidence
Genes_seq
Variation
Map
Gene
Het
Validation
Genotypes Avail
Linkout Avail
rs10837631
rs7110263
rs12574989
rs12788013
5 variations
rs34029390
49 variations
55 variations
rs35703285
rs34690599
rs35328027
rs1609812
rs34451549
rs35920381
rs34214308
HBB
rs7946748
rs7480526
rs35630218
52 variations
61 variations
62 variations
12 variations
26 variations
49 variations
9 variations
5 variations
rs16911894
rs10742583


Human genome overview page (Build 36.2)
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:
5,203,340
5,205,770
Go
out
zoom
in
You are here:
Ideogram
default
master
rs34690599
rs35328027
rs1609812
rs34451549
rs35920381
rs34214308
HBB
rs7946748
rs7480526
12 variations
69 variations
60 variations
MapViewer
Recenter
Zoom in x2
Zoom in x4
Zoom in x8
Zoom out x2
Show 10M
Show 1M
Show 100K
Show 10K
Show Sequence
rs34034824
rs10742584
rs10768684
rs35755129
rs11036364
rs7119261
rs12289247



5204796	5204796	rs33947457	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
start	stop		Map	Gene	Het	Validation	Genotypes Avail / Linkout Avail												
5204797	5204797	rs34548294	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204799	5204799	rs33918131	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204800	5204800	rs35699606	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204801	5204801	rs35198910	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204802	5204802	rs33932981	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204802	5204803	rs35497102	0					1	1	1				0	0	0	0	0	0
5204803	5204803	rs33926764	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204805	5204805	rs34387455	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204806	5204806	rs34948328	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204808	5204808	rs33[illegible]	0					1	1	1	HBB	NM_000518	NP_000509	0.068	0.171	0	1	1	1
5204809	5204809	rs33930165	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204810	5204811	rs34889882	0					1	1	1				0	0	0	0	0	0
5204811	5204811	rs34769005	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204812	5204812	rs33912272	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204818	5204818	rs34058656	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204818	5204818	rs34126315	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204819	5204819	rs713040	0					1	1	1	HBB	NM_000518	NP_000509	0.272	0.249	0	1	1	1
5204820	5204820	rs33983205	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204821	5204821	rs35906307	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0
5204823	5204823	rs33949930	0					1	1	1	HBB	NM_000518	NP_000509	0	0	0	0	0	0

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

GENERAL
Contact Us
Site Map
dbSNP Homepage
Announcements
dbSNP Summary
FTP Download
SNP SUBMISSION
DOCUMENTATION
SEARCH
HAPLOTYPE
RELATED SITES

Reference SNP(refSNP) Cluster Report: rs334

refSNP ID: rs334	
Organism:	human (*Homo sapiens*)
Molecule Type:	Genomic
Created/Updated in build:	36/126
Map to Genome Build:	36.2

Allele	
Variation Class:	MIXED: cluster contains submissions from 2 or more alleleic classes
Alleles:	-/A/C/T
Ancestral Allele:	Not available

Links , Linkout

SNP Details are organized in the following sections:

Submission | Fasta | Resource | GeneView | Map | Diversity | Validation

Submitter records for this RefSNP Cluster

The submission **ss48419811** has the longest flanking sequence of all cluster members and was used to instantiate sequence for **rs334** during BLA

NCBI Assay ID	Handle|Submitter ID	Validation Status	Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Se
ss335	EXAMPLE|sickle01		fwd/T	A/T	aaacagacaccatggtgcacctgactcctg	ggagaagtctgccgtt
ss1536049	LEE|665712		fwd/T	A/T	aaacagacaccatggtgcacctgactcctg	ggagaagtctgccgtt

Fasta sequence (Legend)

>gnl|dbSNP|rs334|allelePos=301|totalLen=601|taxid=9606|snpclass=7|alleles='-/A/C/T'|mol=Genomic|build=126

```
GCAATTTGTA CTGATGGTAT GGGGCCAAGA GATATATCTT AGAGGGAGGG CTGAGGGTTT
GAAGTCCAAC TCCTAAGCCA GTGCCAGAAG AGCCAAGGAC AGGTACGGCT GTCATCACTT
AGACCTCACC CTGTGGAGCC ACACCCTAGG GTTGGCCAAT CTACTCCCAG GAGCAGGGAG
GGCAGGAGCC AGGGCTGGGC ATAAAAGTCA GGGCAGAGCC ATCTATTGCT TACATTTGCT
TCTGACACAA CTGTGTTCAC TAGCAACCTC AAACAGACAC CATGGTGCAT CTGACTCCTG
N
GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG TTGGTGGTGA
GGCCCTGGGC AGGTTGGTAT CAAGGTTACA AGACAGGTTT AAGGAGACCA ATAGAAACTG
GGCATGTGGA GACAGAGAAG ACTCTTGGGT TTCTGATAGG CACTGACTCT CTCTGCCTAT
TGGTCTATTT TCCCACCCTT AGGCTGCTGG TGGTCTACCC TTGGACCCAG AGGTTCTTTG
AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG GTGAAGGCTC
```

GeneView

GeneView via analysis of contig annotation: HBB hemoglobin, beta

Click to see [all] [cSNP] [has frequency] [double hit] [haplotype tagged] variations associated with this gene.

Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) Color Legend
reference	NT_009237->NM_000518 sv function	
Celera	NW_925006->NM_000518 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_009237->NM_000518->NP_000509	4035473	reverse	70	nonsynonymous	C	Ala [A]	2	7
					nonsynonymous	T	Val [V]	2	7
					contig reference	A	Glu [E]	2	7

GeneView: no link established by BLAST analysis of mRNA sequences

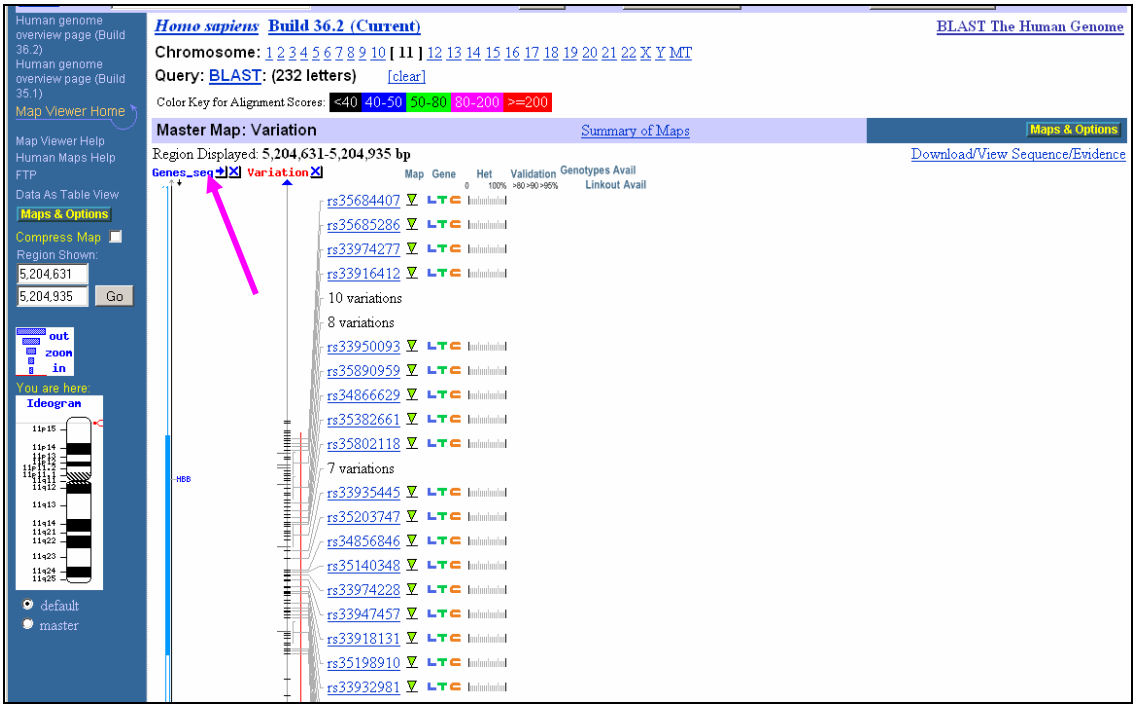
Integrated Maps:

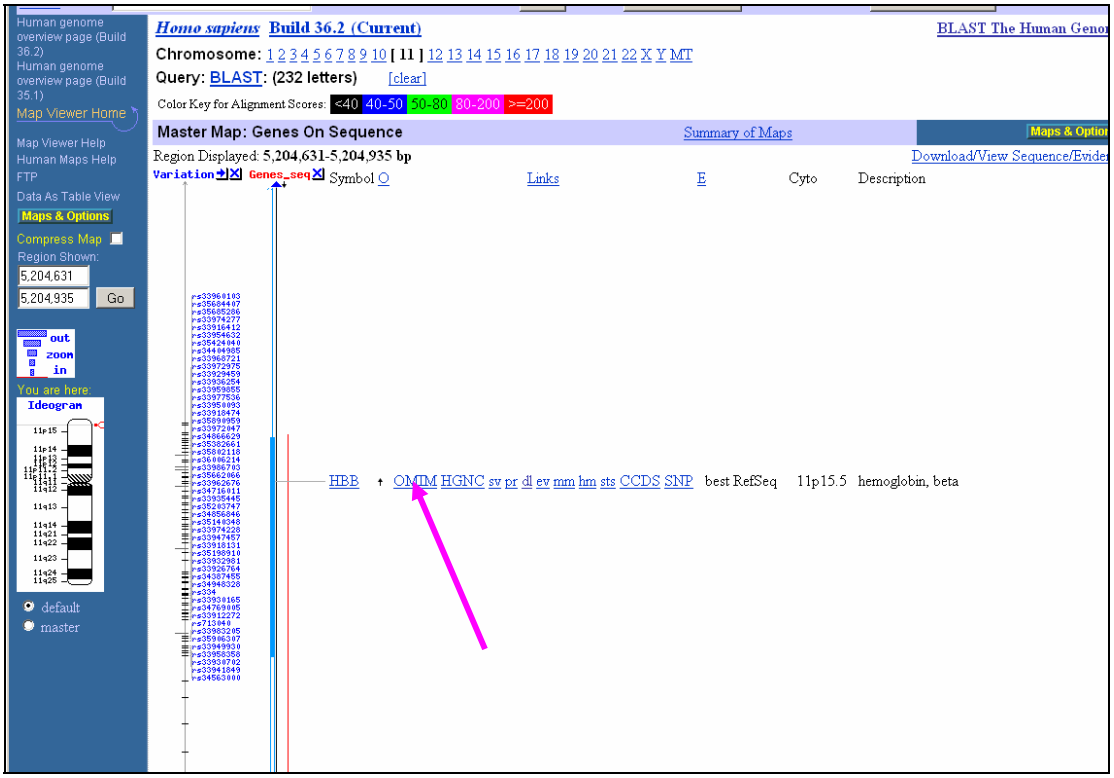
NCBI MapViewer: rs334 maps exactly once on NCBI human chromosome 11

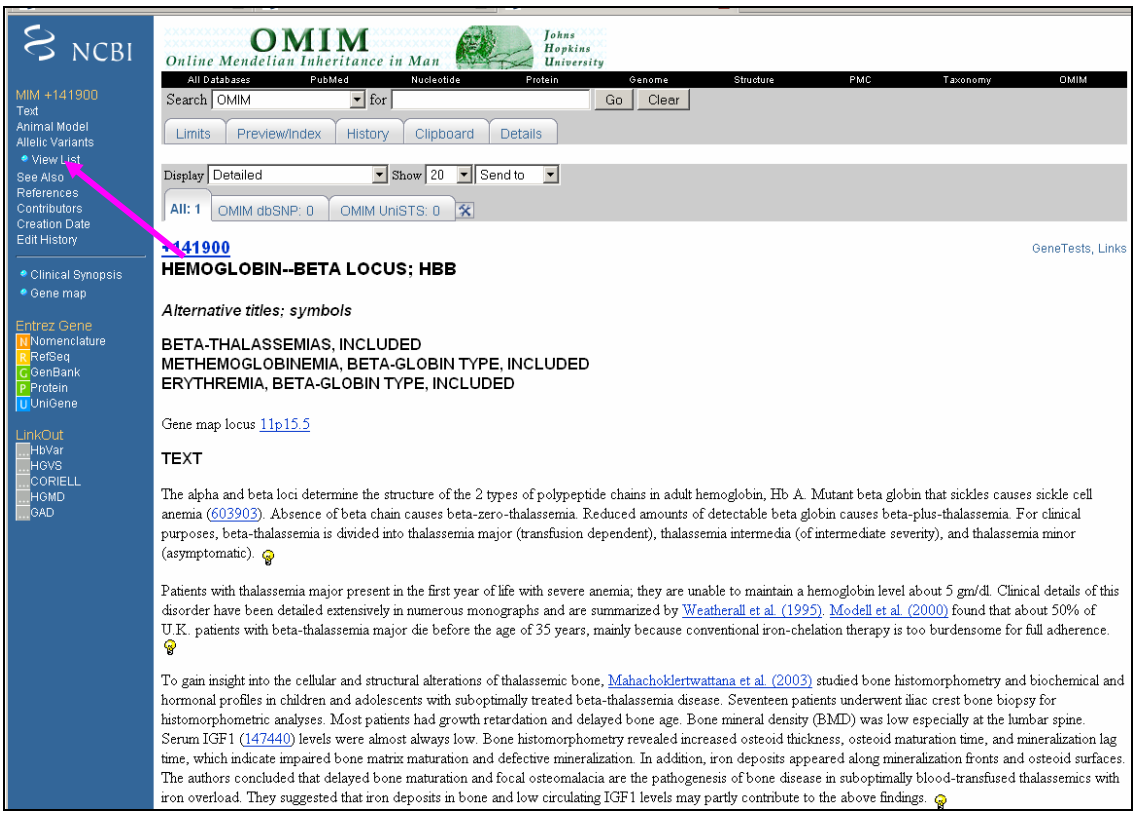
Chromosome	Contig accession	Contig position	Chromosome position	Hit orientation	Contig Allele	Assembly Type	Group label	Contig label	Neighbor SNP	SNP_flank position
11	NT_009237.17	4035473	5204808	minus	T	ref_assembly	reference	reference	view	300
11	NW_925006.1	867227	5366903	minus	T	alt_assembly_1	Celera	Celera	view	300

The query EST sequence contains a known SNP (T/A with respect to the nucleotide 4035473 on contig NT_009237.17).

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







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

MIM +141900
Text
Animal Model
Allelic Variants
View List
See Also
References
Contributors
Creation Date
Edit History

Clinical Synopsis
Gene map

Entrez Gene
Nomenclature
RefSeq
GenBank
Protein
UniGene

LinkOut
HbVar
HGVS
CORIELL
HGMD
GAD

- 0224 HEMOGLOBIN PORTO ALEGRE [HBB, SER9CYS]
- 0225 HEMOGLOBIN POTOMAC [HBB, GLU101ASP]
- 0226 HEMOGLOBIN PRESBYTERIAN [HBB, ASN108LYS]
- 0227 HEMOGLOBIN PROVIDENCE [HBB, LYS82ASX]
- 0228 HEMOGLOBIN PYRGOS [HBB, GLY83ASP]
- 0229 HEMOGLOBIN QUIN-HAI [HBB, LEU78ARG]
- 0230 HEMOGLOBIN RADCLIFFE [HBB, ASP99ALA]
- 0231 HEMOGLOBIN RAHERE [HBB, LYS82THR]
- 0232 HEMOGLOBIN RAINIER [HBB, TYR145CYS]
- 0233 HEMOGLOBIN RALEIGH [HBB, VAL1ALA]
- 0234 HEMOGLOBIN RANDWICK [HBB, TRP15GLY]
- 0235 HEMOGLOBIN REGINA [HBB, LEU96VAL]
- 0236 HEMOGLOBIN RICHMOND [HBB, ASN102LYS]
- 0237 HEMOGLOBIN RIO GRANDE [HBB, LYS8THR]
- 0238 HEMOGLOBIN RIVERDALE-BRONX [HBB, GLY24ARG]
- 0239 HEMOGLOBIN RIYADH [HBB, LYS120ASN]

  HEMOGLOBIN KARATSU
- 0240 HEMOGLOBIN ROSEAU-POINTE A PITRE [HBB, GLU90GLY]
- 0241 HEMOGLOBIN ROTHSCHILD [HBB, TRP37ARG]
- 0242 HEMOGLOBIN RUSH [HBB, GLU101GLN]
- 0243 HEMOGLOBIN S [HBB, GLU6VAL]
- 0244 HEMOGLOBIN S (ANTILLES) [HBB, GLU6VAL AND VAL23ILE]
- 0245 HEMOGLOBIN S (OMAN) [HBB, GLU6VAL AND GLU121LYS]

  HEMOGLOBIN S/O (ARAB)
- 0246 HEMOGLOBIN S (PROVIDENCE) [HBB, GLU6VAL AND LYS82ASX]
- 0247 HEMOGLOBIN S (TRAVIS) [HBB, GLU6VAL AND ALA142VAL]
- 0248 HEMOGLOBIN SABINE [HBB, LEU91PRO]
- 0249 HEMOGLOBIN SAINT JACQUES [HBB, ALA140THR]
- 0250 HEMOGLOBIN SAITAMA [HBB, HIS117PRO]
- 0251 HEMOGLOBIN SAKI [HBB, LEU14PRO]
- 0252 HEMOGLOBIN SAN DIEGO [HBB, VAL109MET]

MIM +141900
Text
Animal Model
Allelic Variants
View List
See Also
References
Contributors
Creation Date
Edit History

Clinical Synopsis
Gene map

Entrez Gene
Nomenclature
RefSeq
GenBank
Protein
UniGene

LinkOut
HbVar
HGVS
CORIELL
HGMD
GAD

PSEUDOGENES

The eta locus is 1 of 5 ancient beta-related globin genes linked in a cluster, 5-prime--epsilon--gamma--eta--delta--beta--3-prime, that arose from tandem duplications (Koop et al., 1986). The eta locus was embryonically expressed in early eutherians and persisted as a functional gene in artiodactyls (e.g., goat), but became a pseudogene in proto-primates and was lost from rodents and lagomorphs. Sequence studies show that the goat eta gene is orthologous to the pseudogene located between the gamma and delta loci of primates and called psi-beta-1. (The Hb beta-1 pseudogene (psi-beta-1) can be symbolized HBBP or HBBP1.)

(In the allelic variants that follow, as well as in the allelic variants listed under the other globin genes, the codon count begins with the first amino acid of the mature protein because a large portion of the variants were characterized on the basis of a protein rather than the gene itself. It is more customary for the count to begin with the methionine initiator codon as number one. Thus, the Hb S mutation (141900.0243) is designated glu6-to-val; in the gene based system of counting now used, it would be designated glu7-to-val. Some inconsistency is represented by the fact that some initiator mutations in the globin genes are indicated by a system counting from the initiator methionine; e.g., beta-thalassemia due to met1-to-ile (141900.0430).)

ANIMAL MODEL

Ciavatta et al. (1995) created a mouse model of beta-zero-thalassemia by targeted deletion of both adult beta-like globin genes, beta(maj) and beta(min), in mouse embryonic stem cells. Heterozygous animals derived from the targeted cells were severely anemic with dramatically reduced hemoglobin levels, abnormal red cell morphology, splenomegaly, and markedly increased reticulocyte counts. Homozygous animals died in utero; however, heterozygous mice were fertile and transmitted the deleted allele to progeny. The anemic phenotype was completely rescued in progeny derived from mating beta-zero-thalassemic animals with transgenic mice expressing high levels of human hemoglobin A. The authors suggested that beta-zero-thalassemic mice could be used to test genetic therapy for beta-zero-thalassemia and could be bred with transgenic mice expressing high levels of hemoglobin S to produce an improved mouse model of sickle cell disease.

Hemoglobin disorders were among the first to be considered for gene therapy. Transcriptional silencing of genes transferred into hematopoietic stem cells, however, posed one of the most significant challenges to its success. If the transferred gene is not completely silenced, a progressive decline in gene expression as mice age