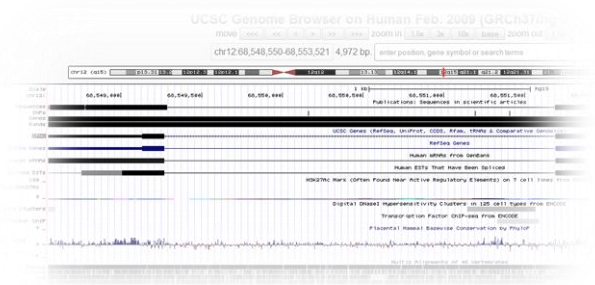
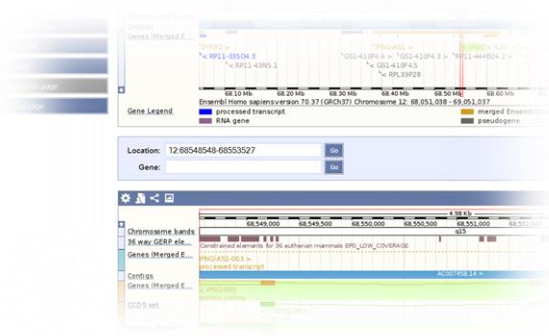
2014

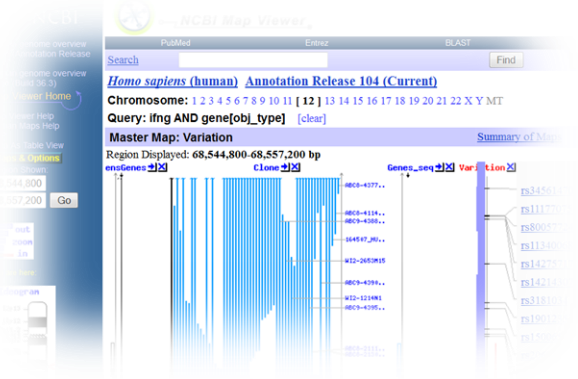
Introduction to Genome Browsers

Rolando Garcia-Milian, *Yale University*



Introduction to Genome Browsers





Rolando Garcia-Milian, MLS, AHIP

Rolando.milian@yale.edu

Curriculum and Research Support Department
Yale Medical Library

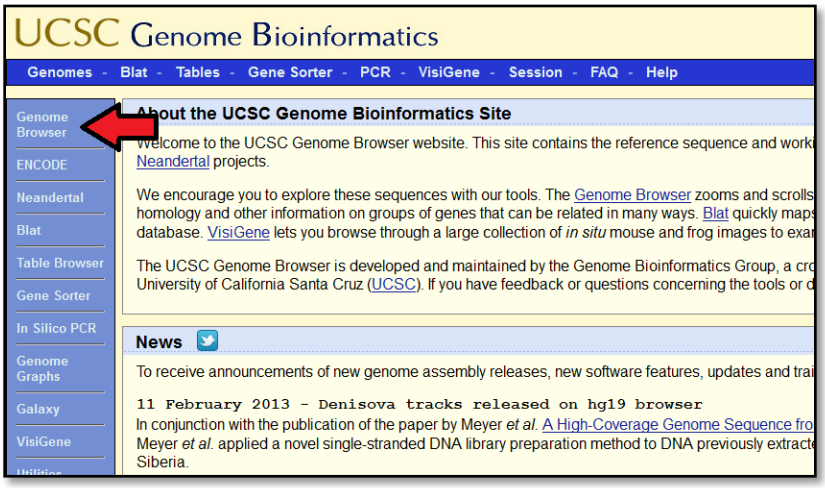
In this workshop we will learn how to navigate the UCSC Genome Browser, Ensembl, NCBI's Map Viewer and Genome Workbench. These browsers are valuable tools when identifying, localizing genes, and looking at their information in the genomic context. By using concrete examples, it will be shown how to locate a human gene, download a gene sequence and its upstream sequence, locate Single Nucleotide Polymorphisms (SNPs) and conserved regions, and use the browsers to download results in a batch. This handout contains screen shots of the demonstration taken on November, 2013

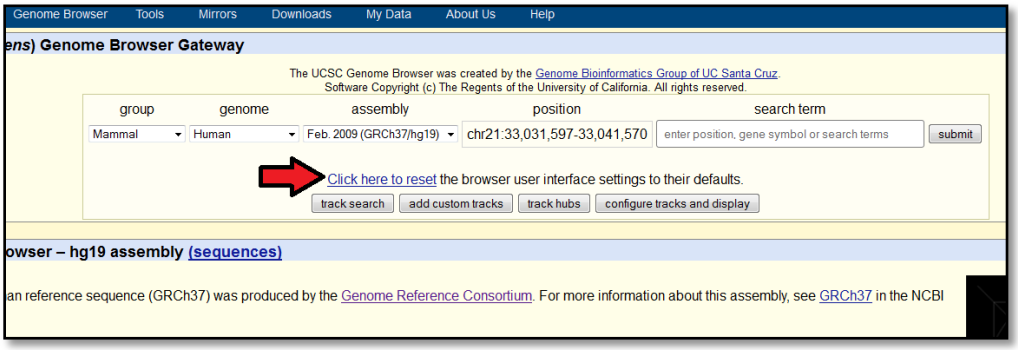
Contents

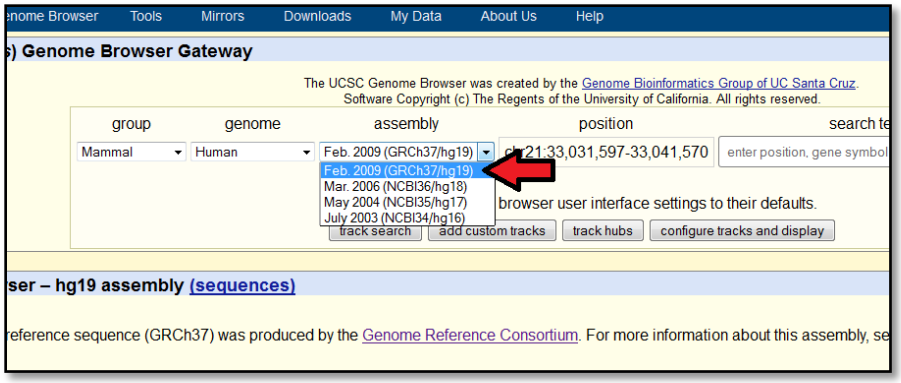
UCSC Genome Browser

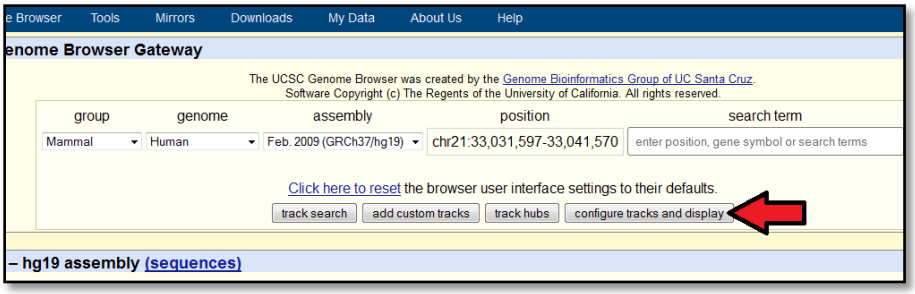
Accessing the gene region and downloading its genomic sequence

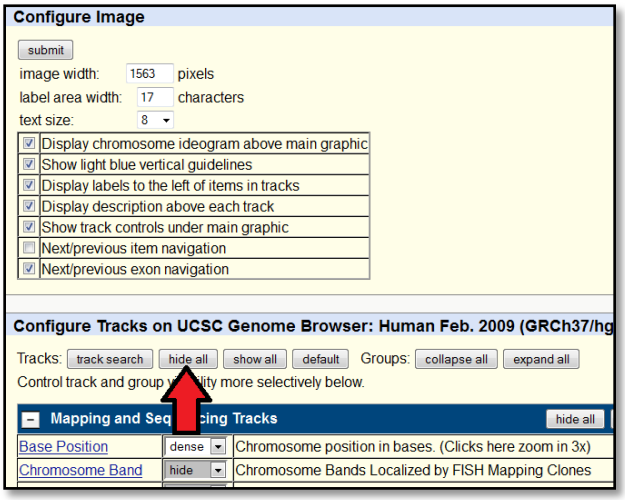
Open the UCSC Genome Browser http://genome.ucsc.edu/ in a web browser and click on "Genome Browser" link. Click on "Click here to reset" link to reset the browser user interface setting to their defaults. Under "assembly", select the preferred assembly. We will use the default one "Feb. 2009 (GRCh37/hg19)". Click on the "configure tracks and display" button. Under "Configure Tracks on UCSC Genome Browser" heading click on the "hide all" button and then, click on the "submit" button.

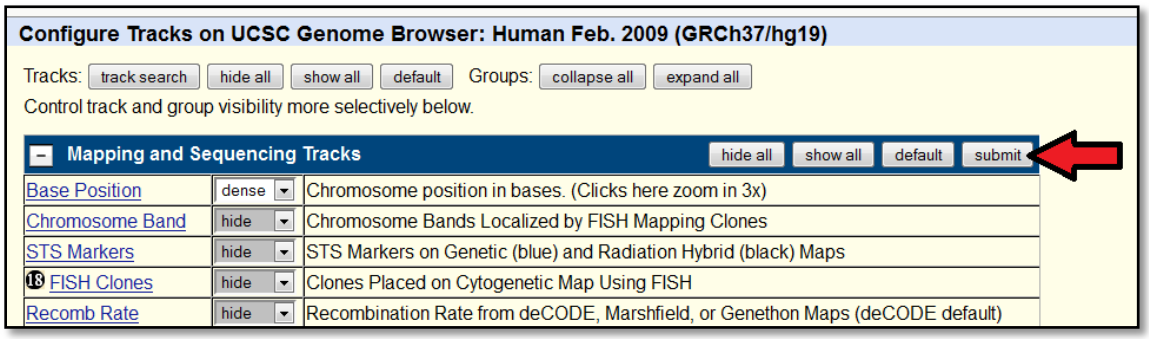




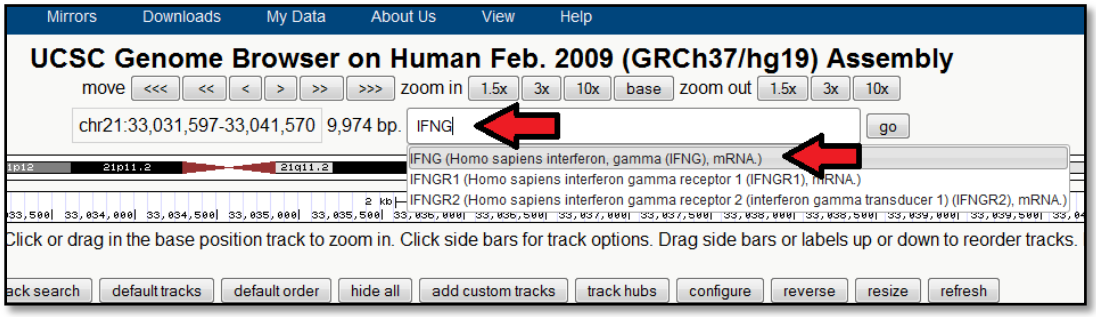

Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
Genome Browser Gateway
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.
group
genome
assembly
position
search te
Mammal
Human
Feb. 2009 (GRCh37/hg19)
Feb. 2009 (GRCh37/hg19)
Mar. 2006 (NCBI36/hg18)
May 2004 (NCBI35/hg17)
July 2003 (NCBI34/hg16)
enter position, gene symbol
browser user interface settings to their defaults.
track search
add custom tracks
track hubs
configure tracks and display
hg19 assembly (sequences)
Genome Reference Consortium

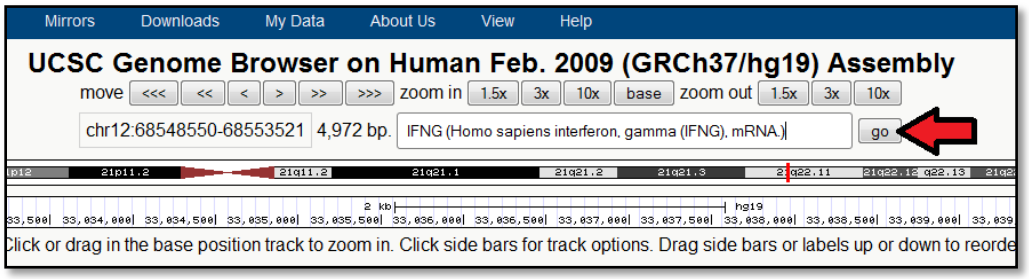

Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
Genome Browser Gateway
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.
group
genome
assembly
position
search term
Mammal
Human
Feb. 2009 (GRCh37/hg19)
chr21:33,031,597-33,041,570
enter position, gene symbol or search terms
Click here to reset the browser user interface settings to their defaults.
track search
add custom tracks
track hubs
configure tracks and display
hg19 assembly (sequences)


Configure Image
submit
image width: 1563 pixels
label area width: 17 characters
text size: 8
Display chromosome ideogram above main graphic
Show light blue vertical guidelines
Display labels to the left of items in tracks
Display description above each track
Show track controls under main graphic
Next/previous item navigation
Next/previous exon navigation
Configure Tracks on UCSC Genome Browser: Human Feb. 2009 (GRCh37/hg
Tracks: track search hide all show all default Groups: collapse all expand all
Mapping and Seq cing Tracks
hide all
Base Position
dense
Chromosome position in bases. (Clicks here zoom in 3x)
Chromosome Band
hide
Chromosome Bands Localized by FISH Mapping Clones

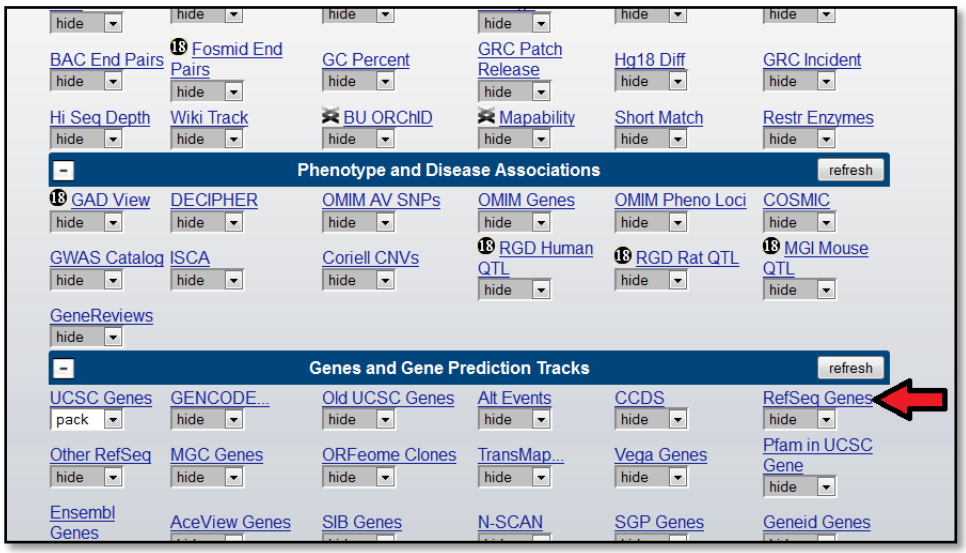

Configure Tracks on UCSC Genome Browser: Human Feb. 2009 (GRCh37/hg19)
Tracks: track search hide all show all default Groups: collapse all expand all
Control track and group visibility more selectively below.
Mapping and Sequencing Tracks
hide all
show all
default
submit
Base Position
dense
Chromosome position in bases. (Clicks here zoom in 3x)
Chromosome Band
hide
Chromosome Bands Localized by FISH Mapping Clones
STS Markers
hide
STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
FISH Clones
hide
Clones Placed on Cytogenetic Map Using FISH
Recomb Rate
hide
Recombination Rate from deCODE, Marshfield, or Genethon Maps (deCODE default)

Enter your query in the search box (e.g. position, gene symbol or search term); in this example we will type IFNG (Homo sapiens interferon gamma). Click on the IFNG mRNA from the autocomplete feature.

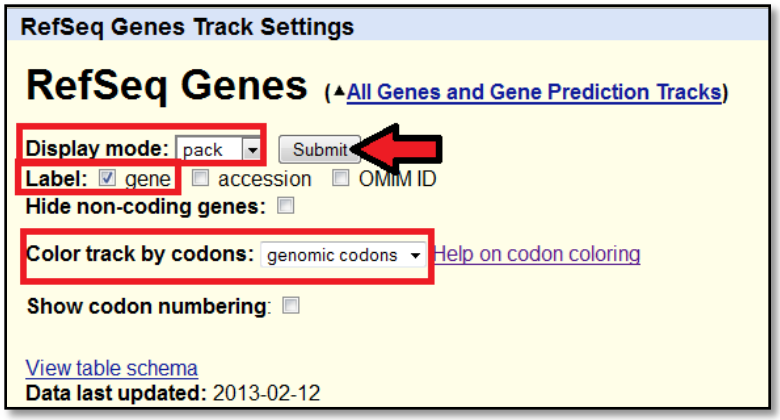




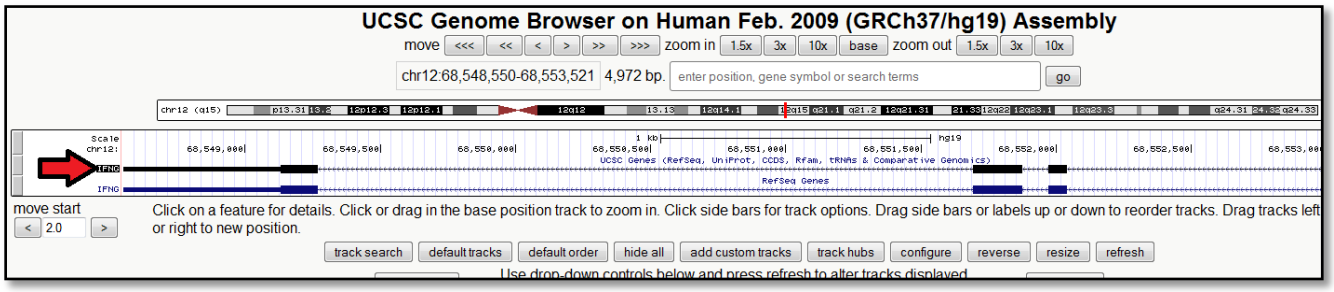
The region corresponding to the IFNG gene on chromosome 12 (chr12: 68,548,550-68,553,521) is displayed. Under the "Genes and Gene Prediction Tracks", click on the "RefSeq Genes" link.



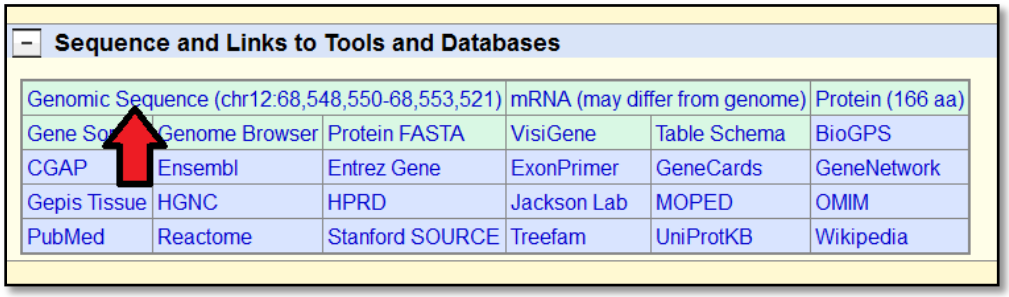
From the "Display mode" select "pack" from the pull down menu, and change the "Color tracks by codons" to "genomic codons". Check the box "Label" by gene symbol and click on the "Submit" button.



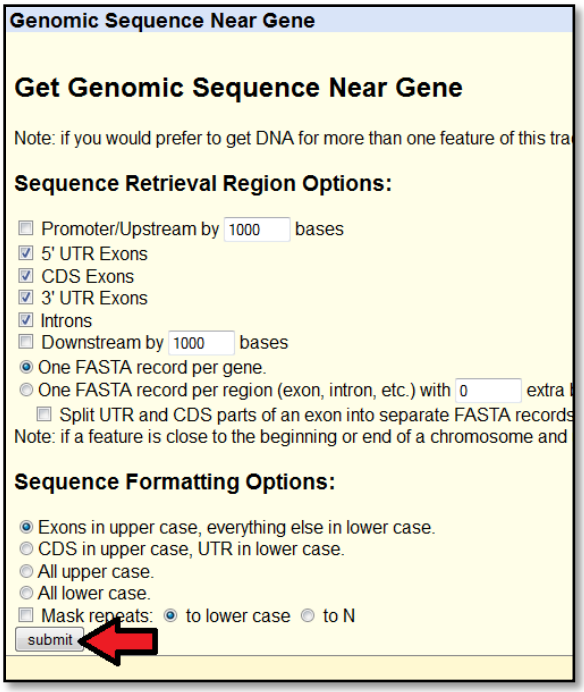
The gene is located on the reverse strand as indicated from the direction of arrows on the gene model. From the TrackMap click on the IFNG symbol label to the left of the UCSC Genes track.

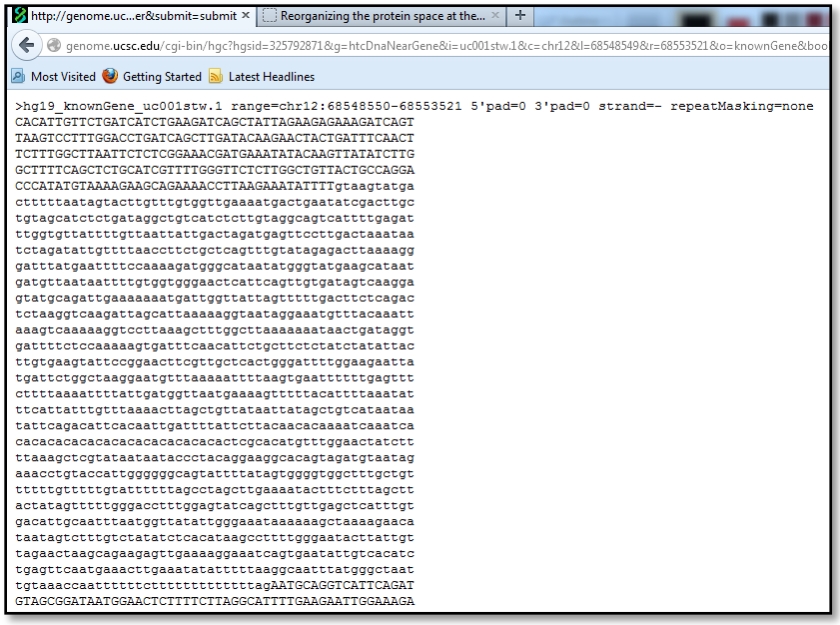


Under “Sequence and Links to Tools and Databases” click on the “Genomic Sequence” Link to download the genomic sequence and/or the upstream and downstream sequence.



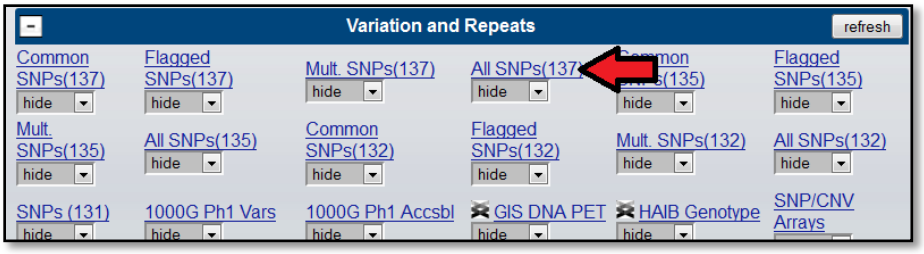
Select your options for retrieving the sequence and click on “Submit” button.

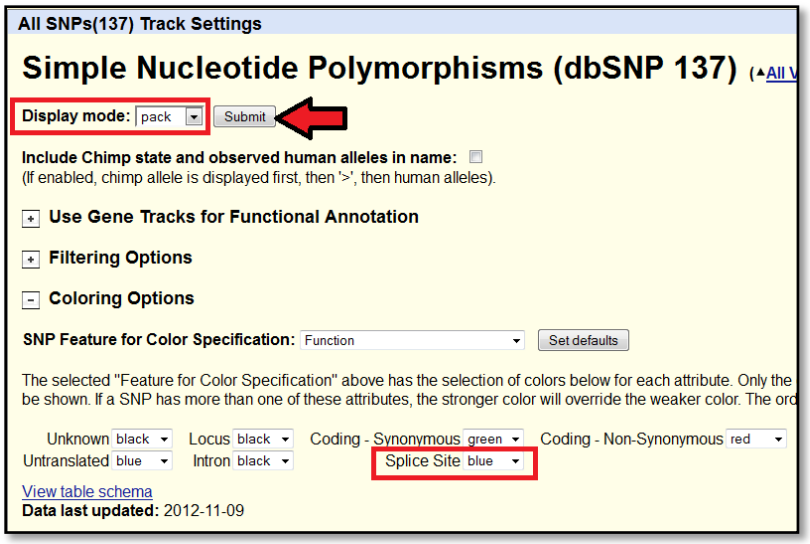


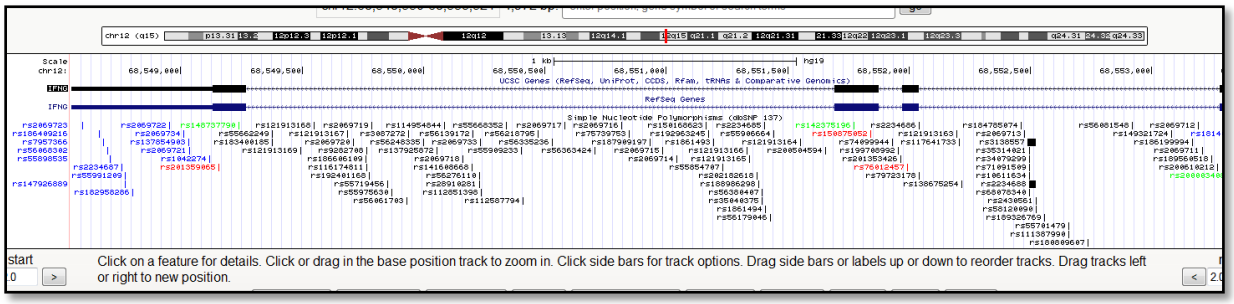


Viewing SNPs annotated on the region

Under the track “Variation and Repeats”, click on the All SNPs(137) link. At the top of the “All SPNs(137) Track Setting”, change the display mode option to “pack”. Expand “Coloring Options” and under “SNP Feature for Color Specification”, change “Splice Site” from red to blue and click on the “Submit” button

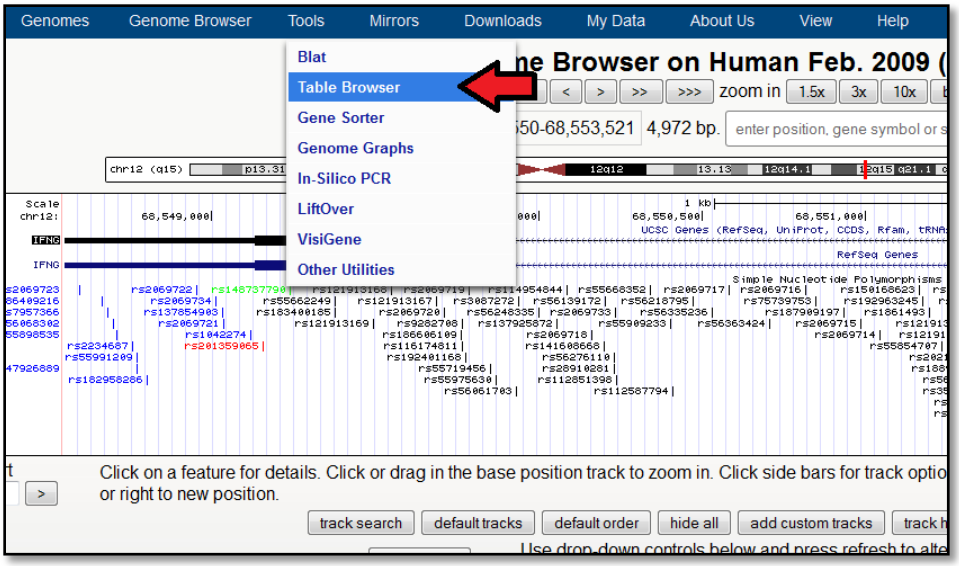




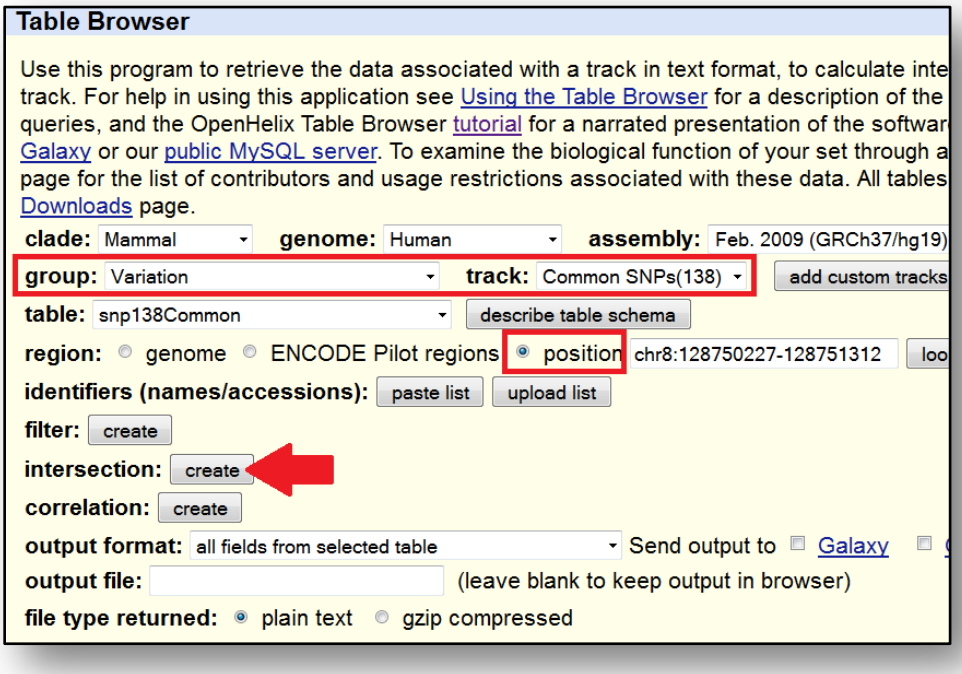


Downloading SNPs in the conserved region

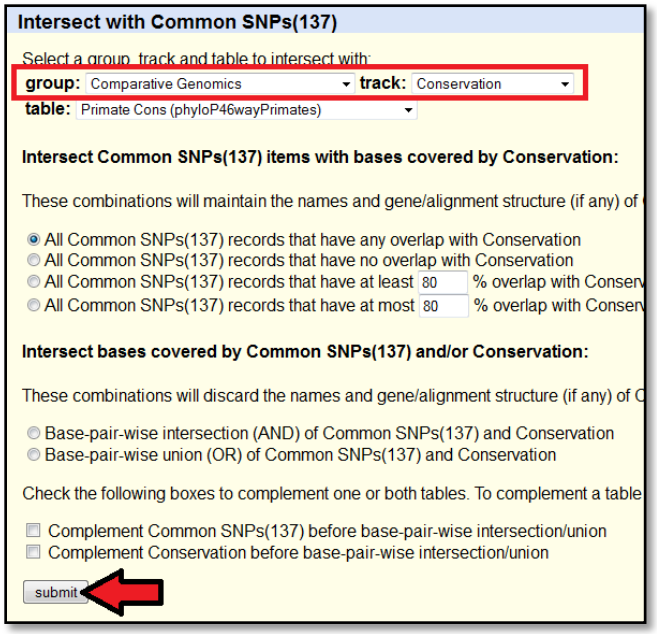
Click on “Tools” at the top of the page. Select/ click on “Table Browser “.



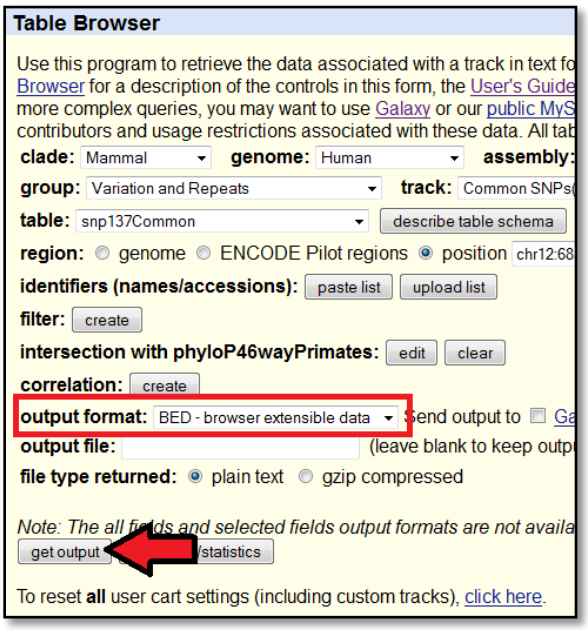
Select the “Variation and Repeats tracks from “group”, and select the “position” radio button. Click on the “create” button next to “intersection”.



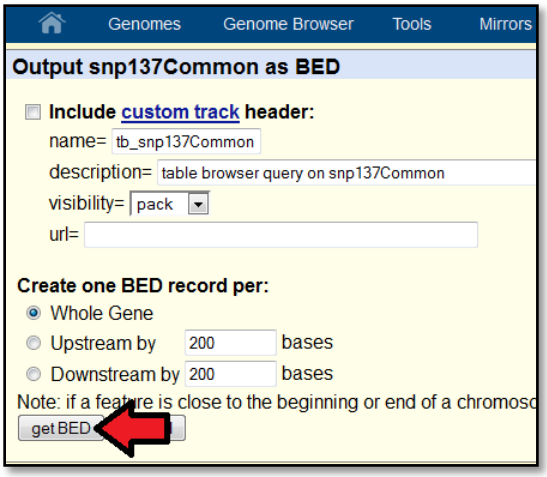
From “group”, select “Comparative Genomics”; and “Conservation” from “track”. Click on the “Submit” button.

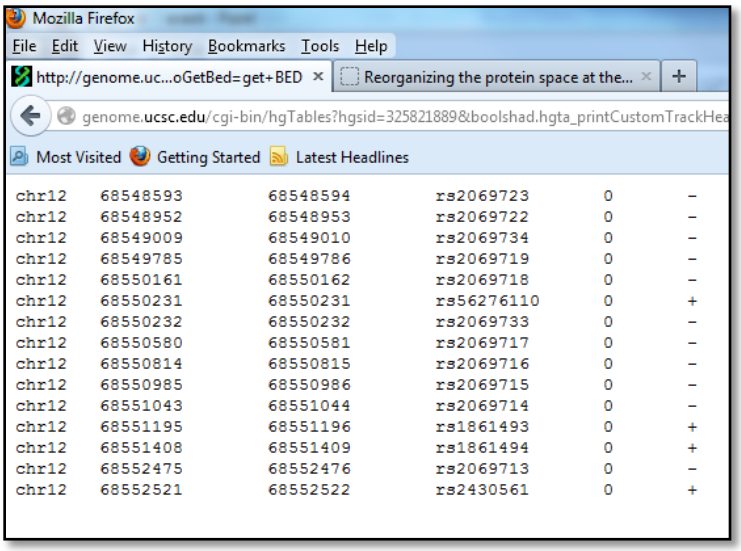


From “output format”, select “BED – browser extensible data” and click on the “get output” button.



Click on “get BED” button. Your results will open in a new window.



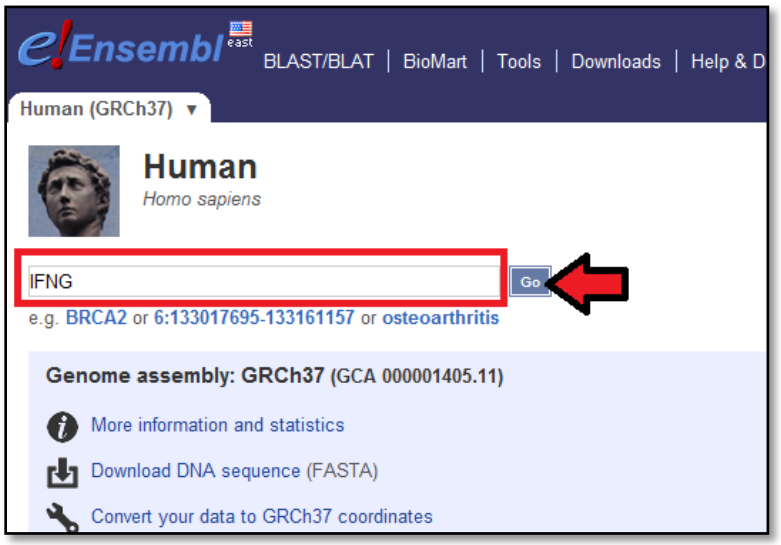


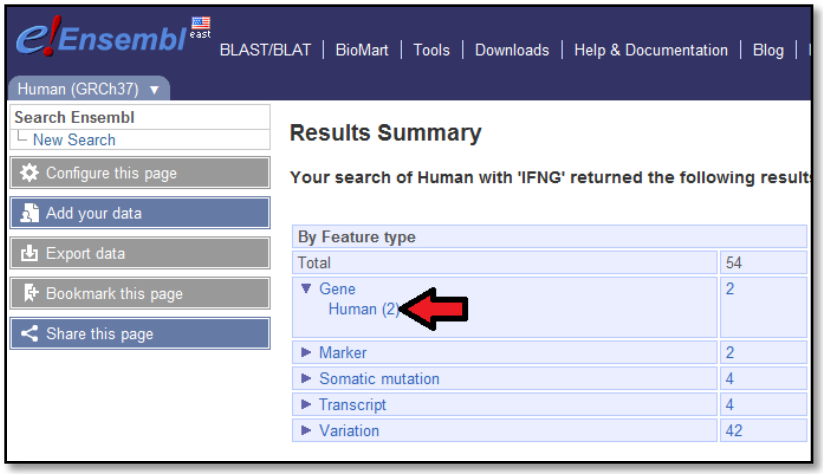

```
chr12   68548593        68548594        rs2069723       0       -
chr12   68548952        68548953        rs2069722       0       -
chr12   68549009        68549010        rs2069734       0       -
chr12   68549785        68549786        rs2069719       0       -
chr12   68550161        68550162        rs2069718       0       -
chr12   68550231        68550231        rs56276110      0       +
chr12   68550232        68550232        rs2069733       0       -
chr12   68550580        68550581        rs2069717       0       -
chr12   68550814        68550815        rs2069716       0       -
chr12   68550985        68550986        rs2069715       0       -
chr12   68551043        68551044        rs2069714       0       -
chr12   68551195        68551196        rs1861493       0       +
chr12   68551408        68551409        rs1861494       0       +
chr12   68552475        68552476        rs2069713       0       -
chr12   68552521        68552522        rs2430561       0       +
```

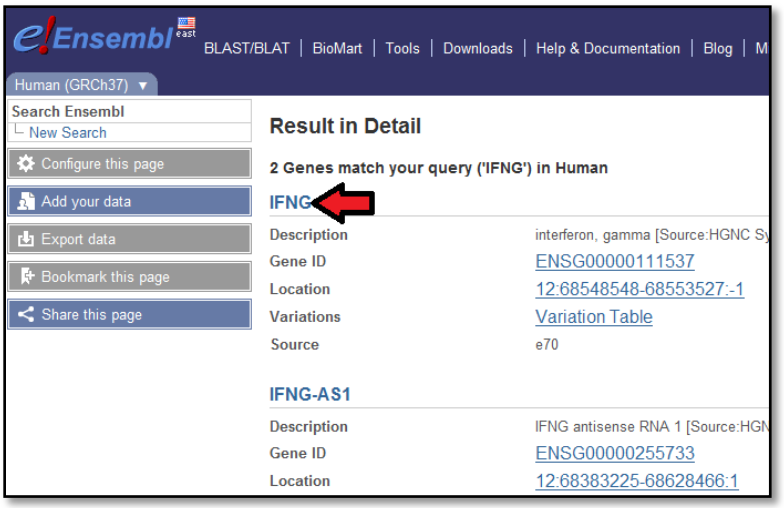
Ensembl

Searching for the IFNG gene and retrieving the cDNA sequence with its amino acid translation.

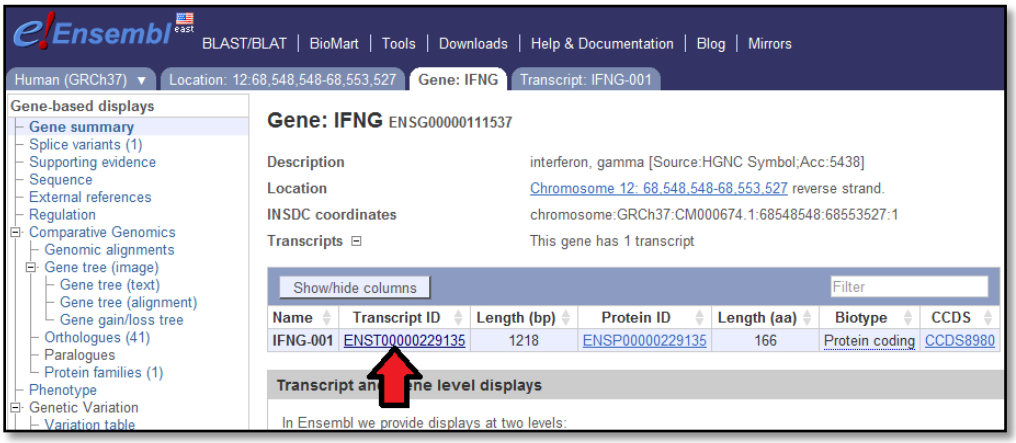
Access the Ensemble genome browser page (http://useast.ensembl.org/index.html). Select "Human" genome under "Popular genomes". Type IFNG in the search box and click on the "Go" button. Under "By Feature type", click on "Gene" and "Human". Click on the IFNG link or its Ensembl Gene ID link (ENSG00000111537).

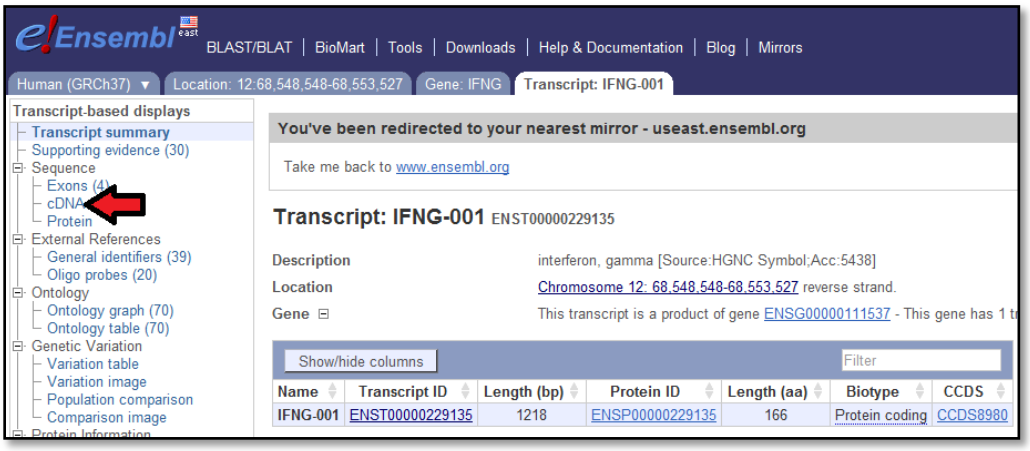




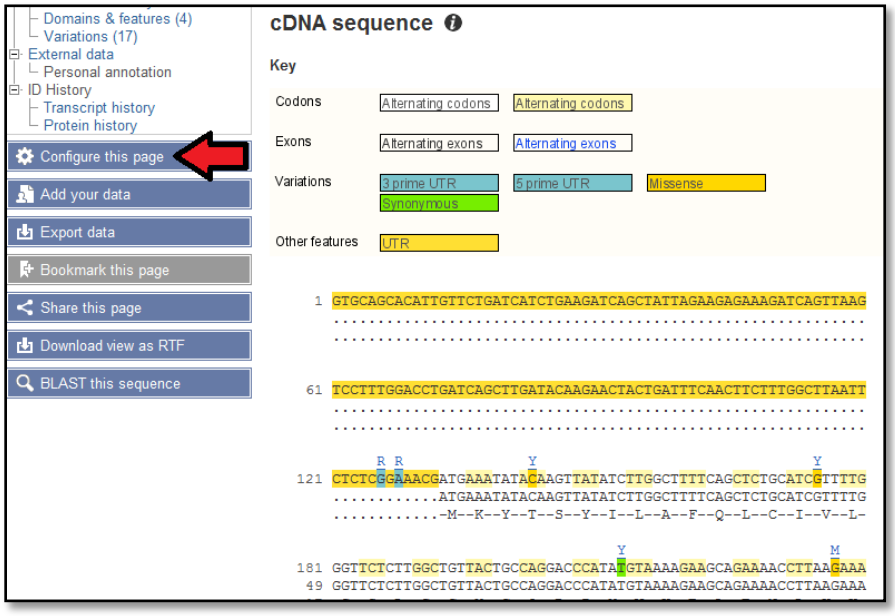


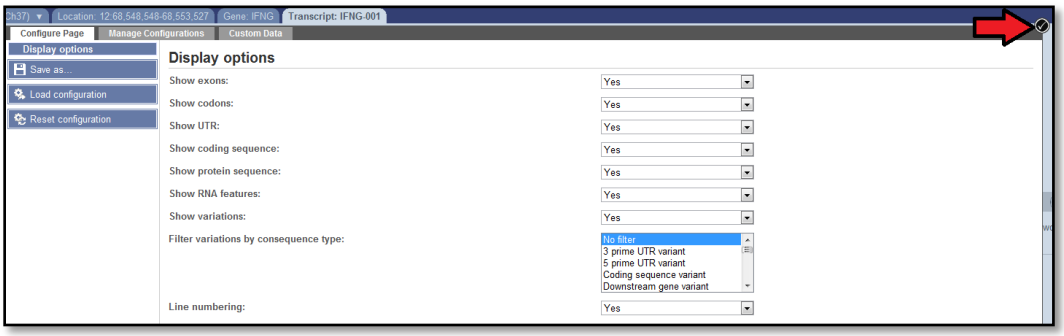
Click on the transcript ID link, it will take you the transcript tab (IFNG-001). Under the "Transcript-based displays" menu (left side of the screen), click on the cDNA link.



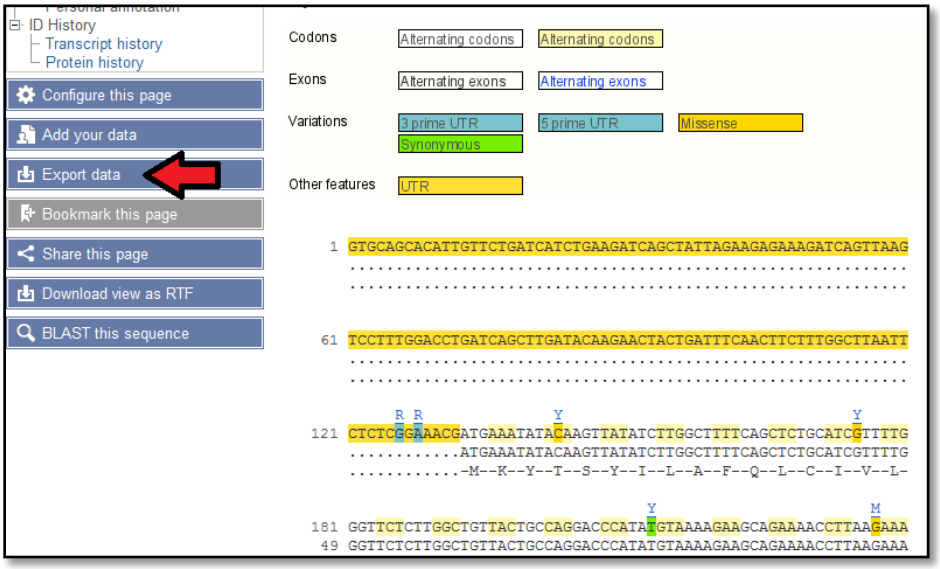


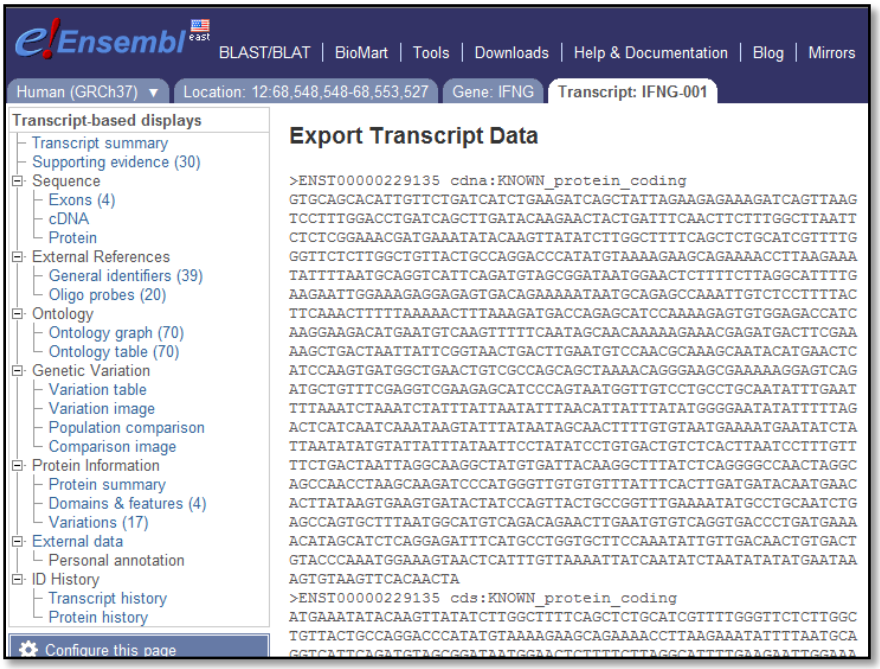
Use “Configure this page” link to change the display settings of the sequence. This will open a pop-up window. Once you have made changes to the settings, just close the window by clicking on the checkmark in the upper right corner.





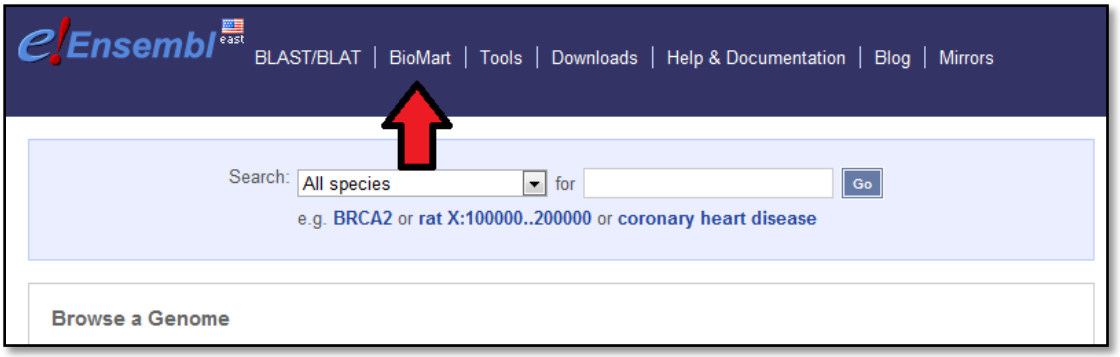
Data can be exported by clicking on the “Export data” link. This opens a pop up window that allows you to change the configuration of the data you will export. Click on the “Next” button to apply your changes. Select your preferred output format for your export. In this example we will click on the HTML link.

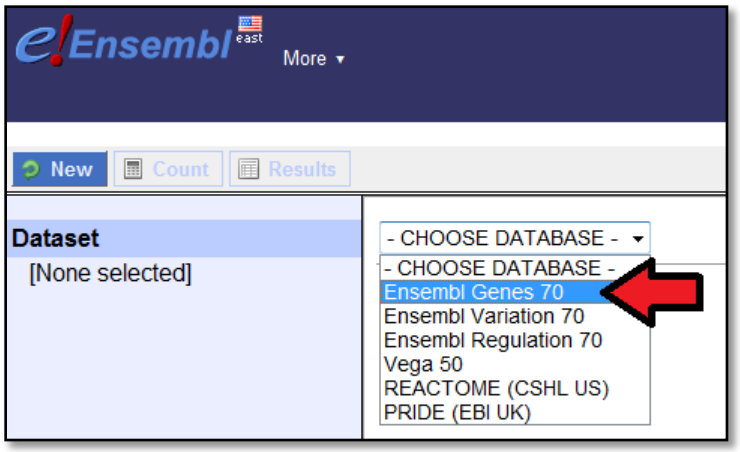


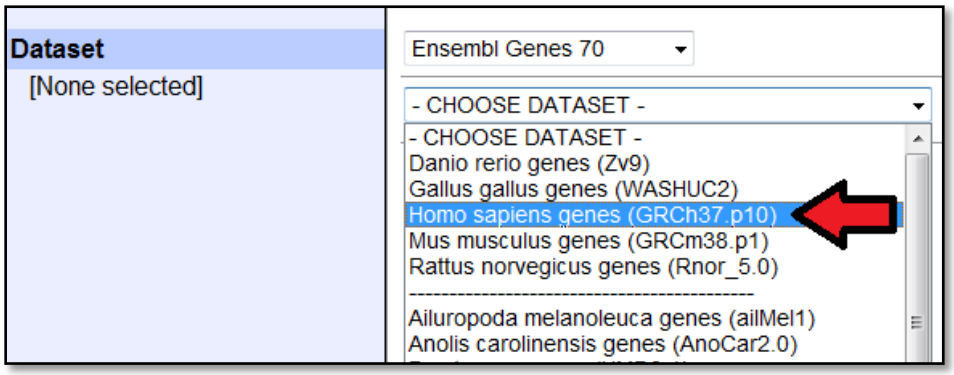


Downloading upstream sequences from multiple human genes in a batch mode

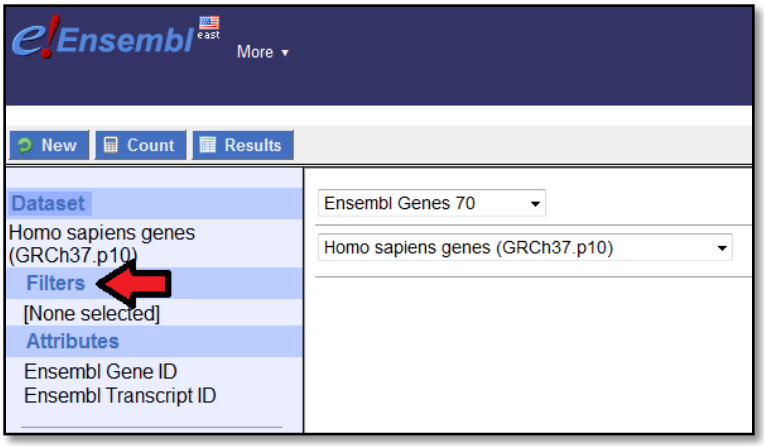
Click on the BioMart link on top of the page. Select “Ensembl Genes 70” from “CHOSE DATABASE” and “Homo sapiens genes GRCh37”

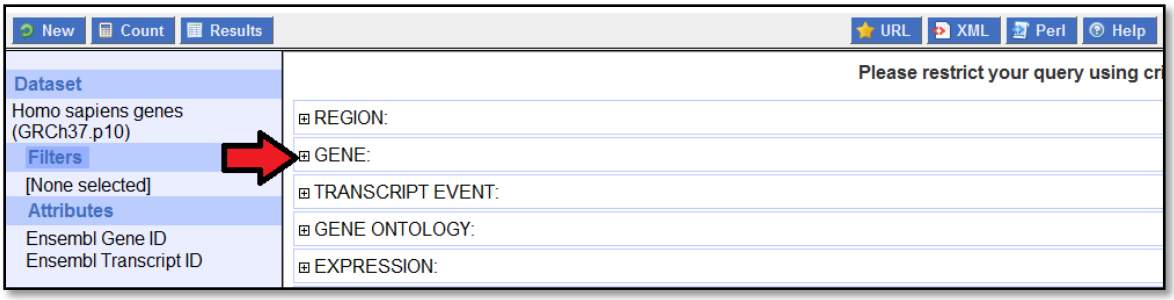


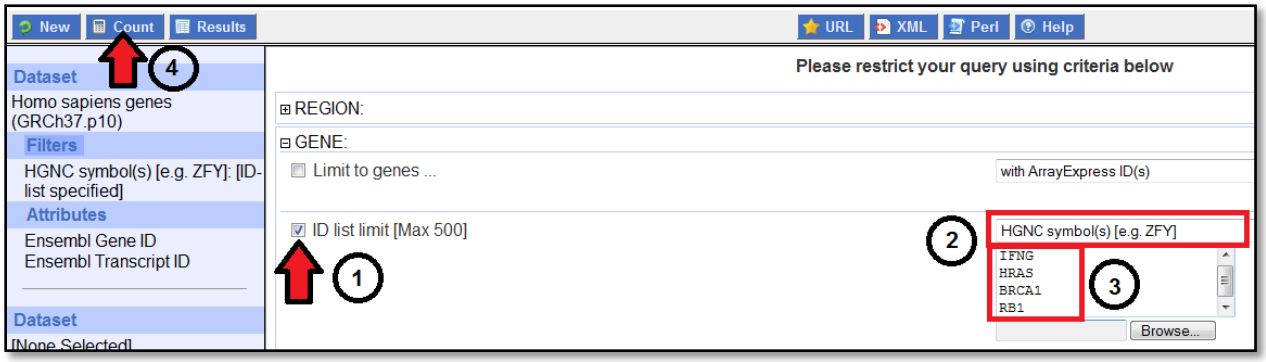


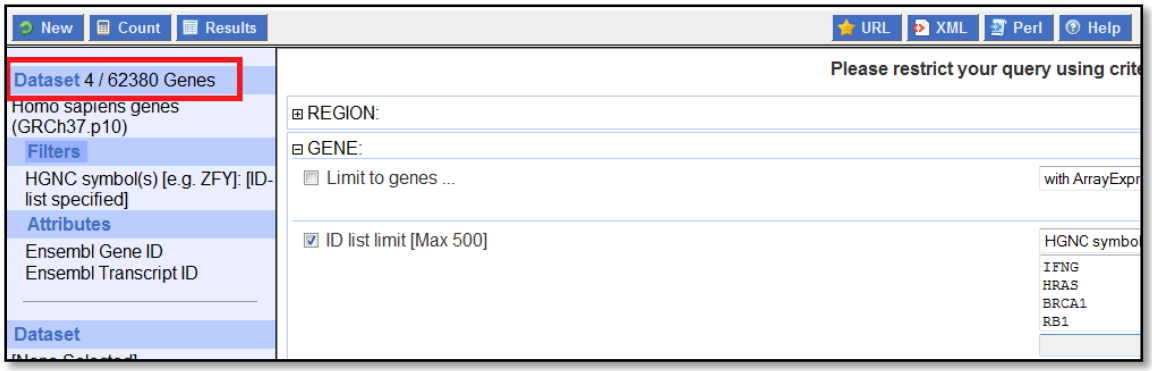


Click on the “Filters” menu. Open the GENE criteria by clicking on the + sign next to it. Check the “ID list limit [Max 500]” box and select the ID type from the dropdown menu. Type the gene symbols of your query in the box below. In this example we will use IFNG, HRAS, BRCA1, RB1. Click on the “Count” button (top of left side) to see how many genes BioMart can find.

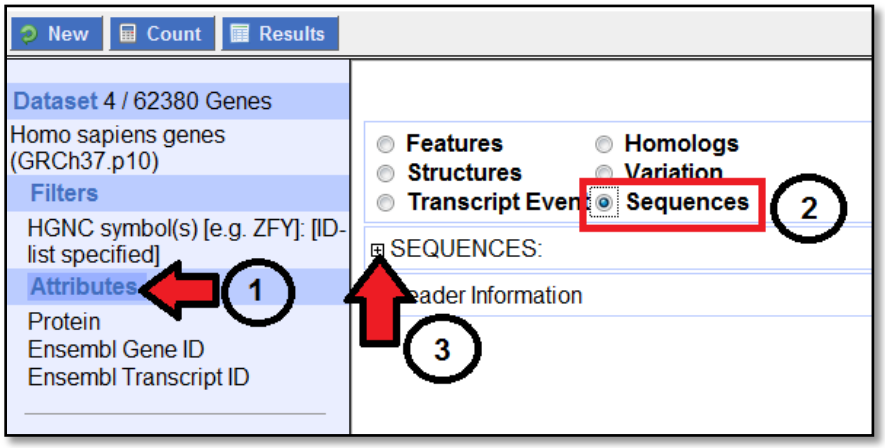

Ensembl east
More
New
Count
Results
Dataset
Homo sapiens genes (GRCh37.p10)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Ensembl Genes 70
Homo sapiens genes (GRCh37.p10)

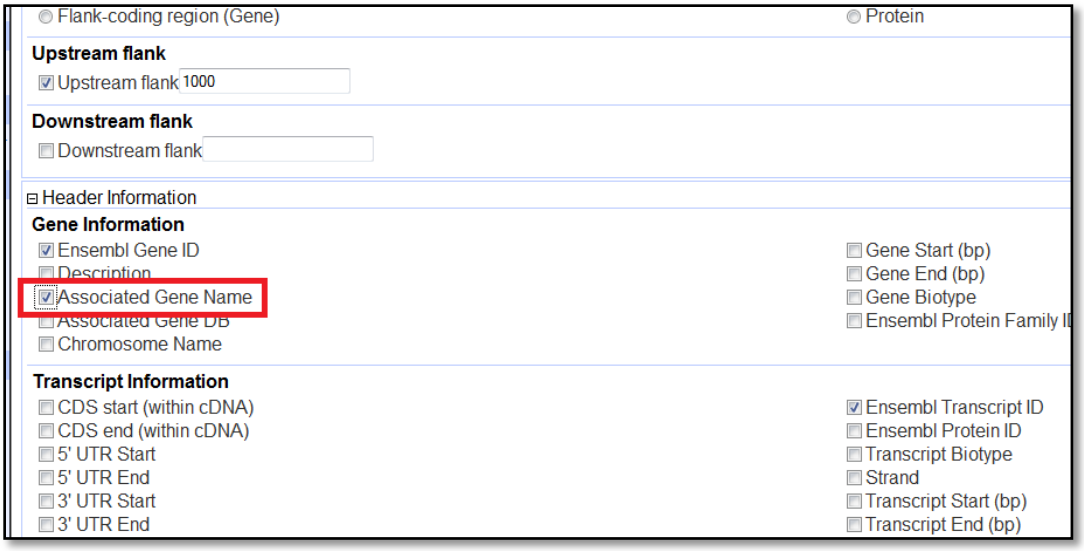

New
Count
Results
URL
XML
Perl
Help
Dataset
Homo sapiens genes (GRCh37.p10)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Please restrict your query using cri
REGION:
GENE:
TRANSCRIPT EVENT:
GENE ONTOLOGY:
EXPRESSION:


New
Count
Results
URL
XML
Perl
Help
4
Dataset
Homo sapiens genes (GRCh37.p10)
Filters
HGNC symbol(s) [e.g. ZFY]; [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Dataset
Please restrict your query using criteria below
REGION:
GENE:
Limit to genes ...
with ArrayExpress ID(s)
ID list limit [Max 500]
1
2
HGNC symbol(s) [e.g. ZFY]
IFNG
HRAS
BRCA1
RB1
3
Browse...

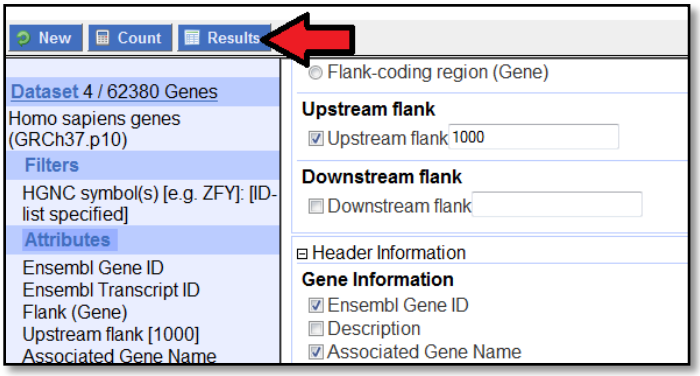

New
Count
Results
URL
XML
Perl
Help
Dataset 4 / 62380 Genes
Homo sapiens genes (GRCh37.p10)
Filters
HGNC symbol(s) [e.g. ZFY]; [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Dataset
Please restrict your query using crit
REGION:
GENE:
Limit to genes ...
ID list limit [Max 500]
IFNG
HRAS
BRCA1
RB1

Click on the “Attributes” option and check the “SEQUENCES” radio button. Open the “SEQUENCES” menu by clicking on the + sign. Check the “Flank (Gene)” radio button. Select the “Upstream flank” and enter the number of nucleotides desired (1000 in this example). Open the “Header information” menu by clicking on the + sign. Select your preferred option such as “Associated Gene Name”, “Description”, etc.





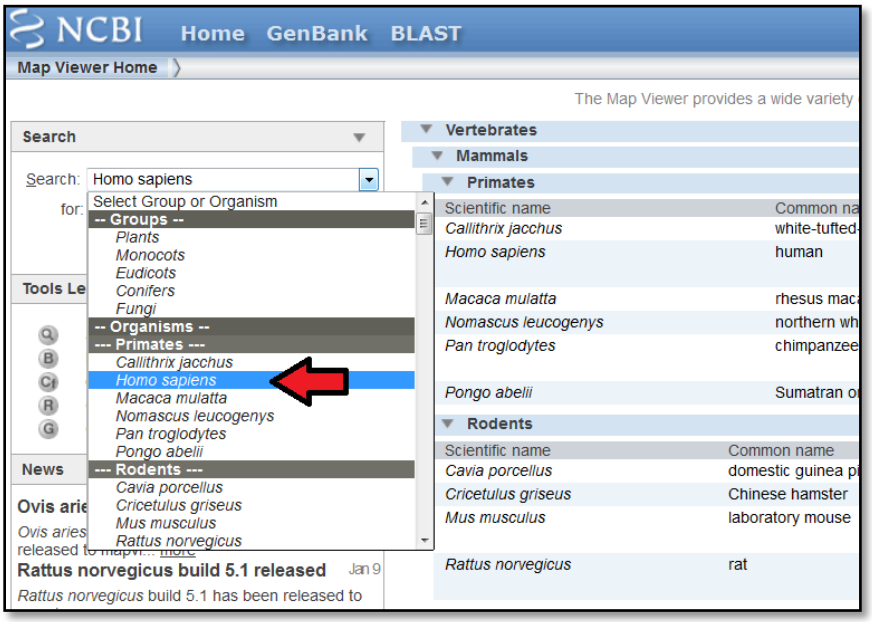
Retrieve the sequences by clicking on the “Results” button at the top.

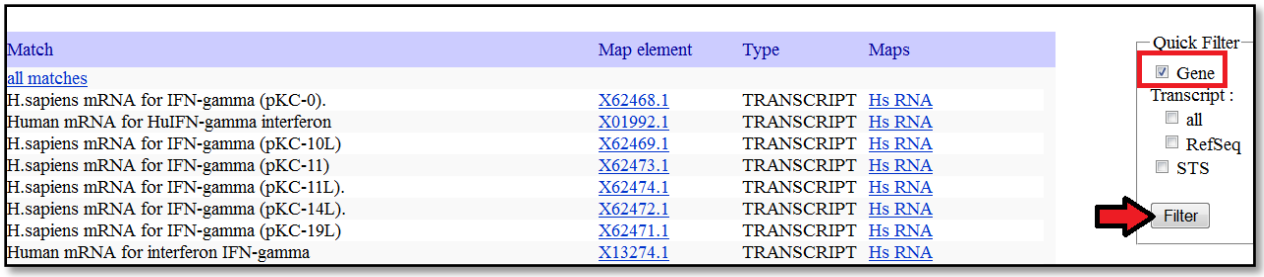


Map Viewer

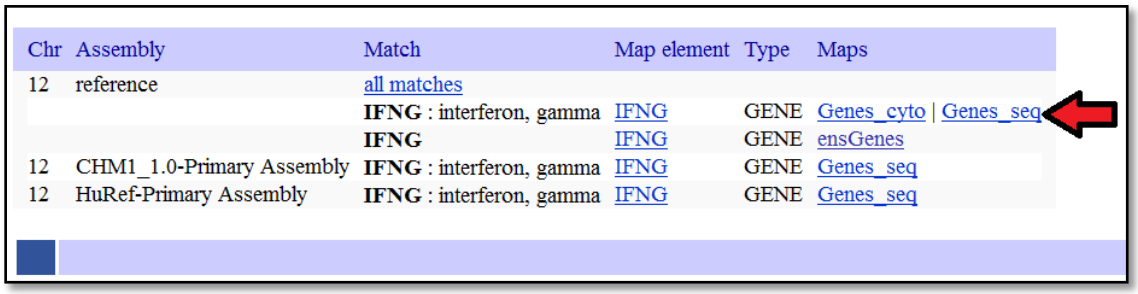
Identifying gene homologs in other organisms

Open NCBI's MapViewer on your internet browser (http://www.ncbi.nlm.nih.gov/mapview/). Select *Homo sapiens* from "Select Group or Organism" drop-down menu. Type "IFNG" in the search box and click on the "Go" button. Select "Gene" from the "Quick Filter" and click on the "Filter" button. Results will show the "reference" assembly.

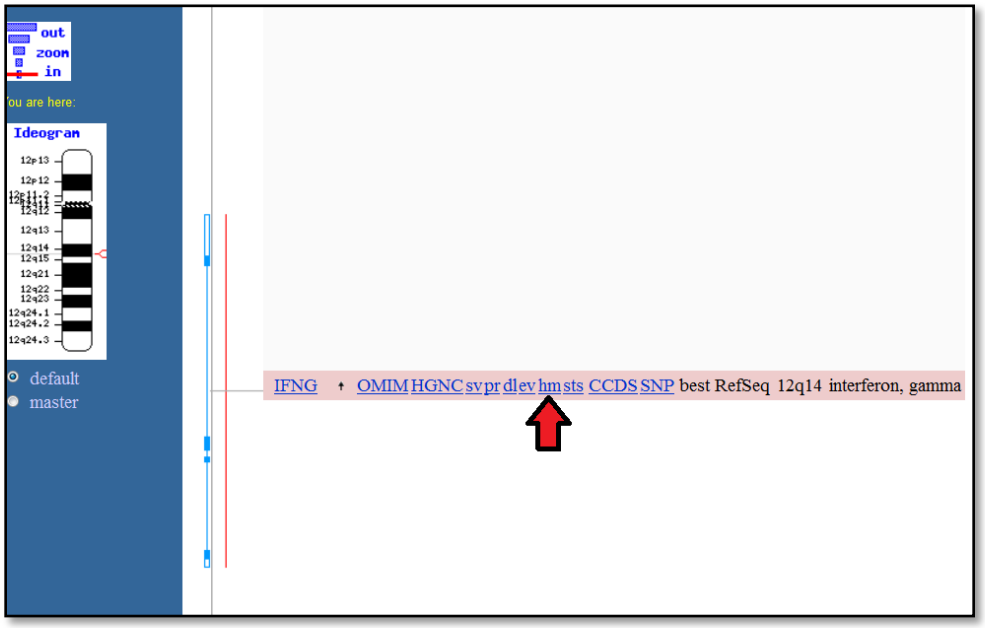


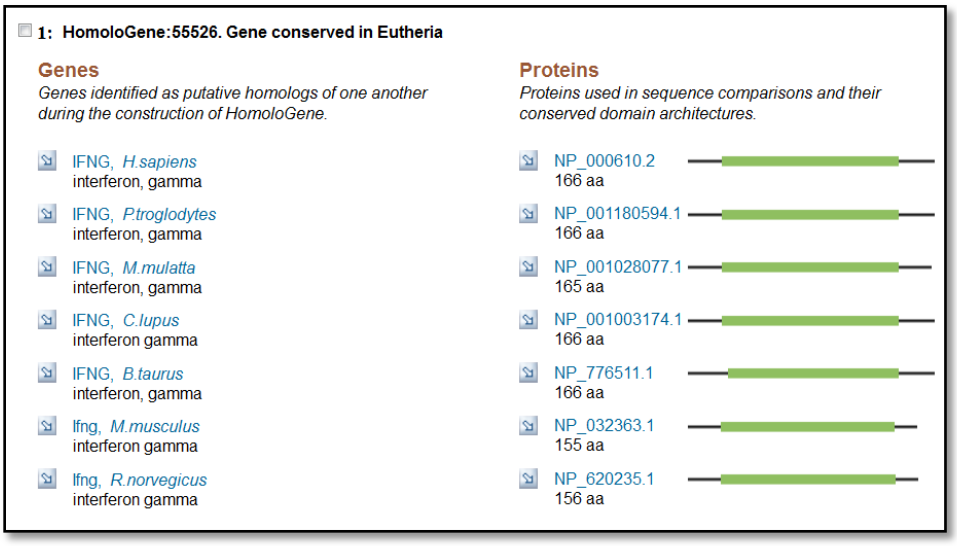


Under "Maps", click on the "Genes_seq" link from the "reference" assembly.



Click on the “hm” link to access the IFGN “HomoloGene” page.



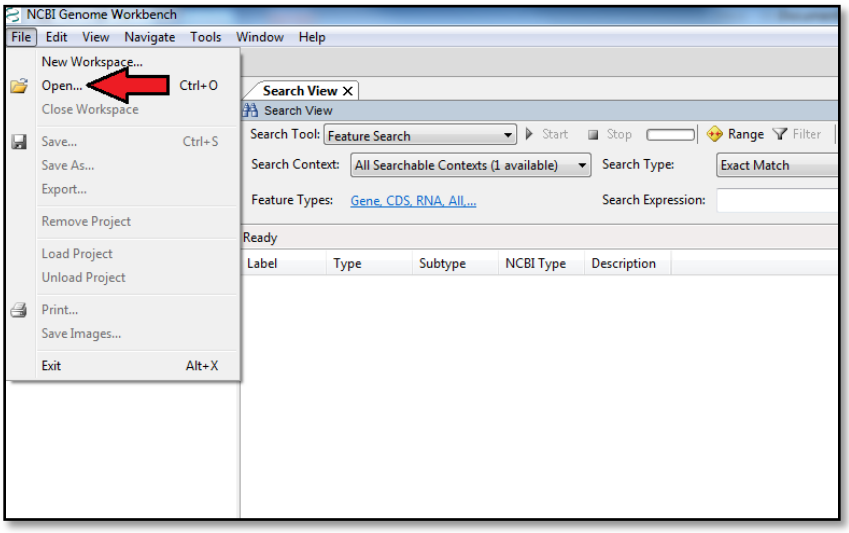


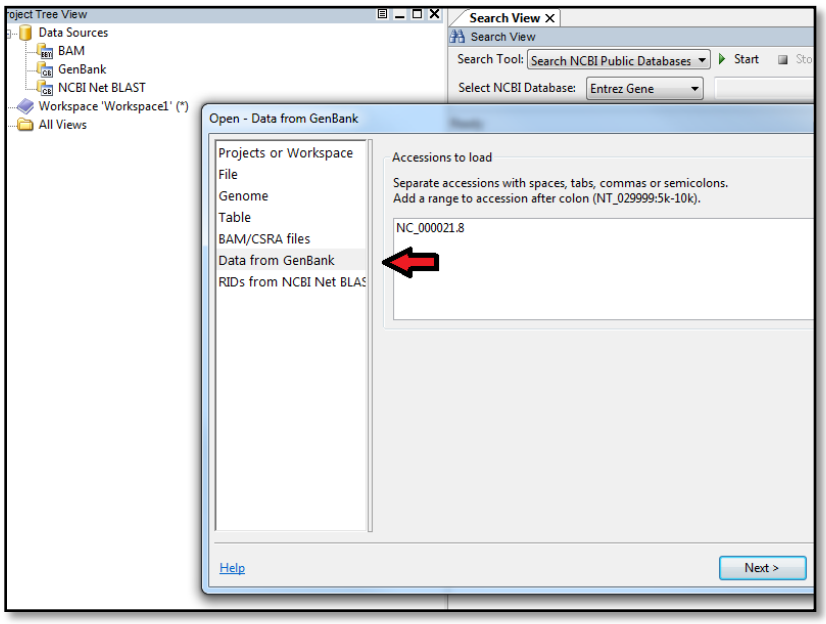
NCBI’s Genome Workbench

Retrieving a sequence from NCBI public database, loading it into genome workbench and displaying it graphically

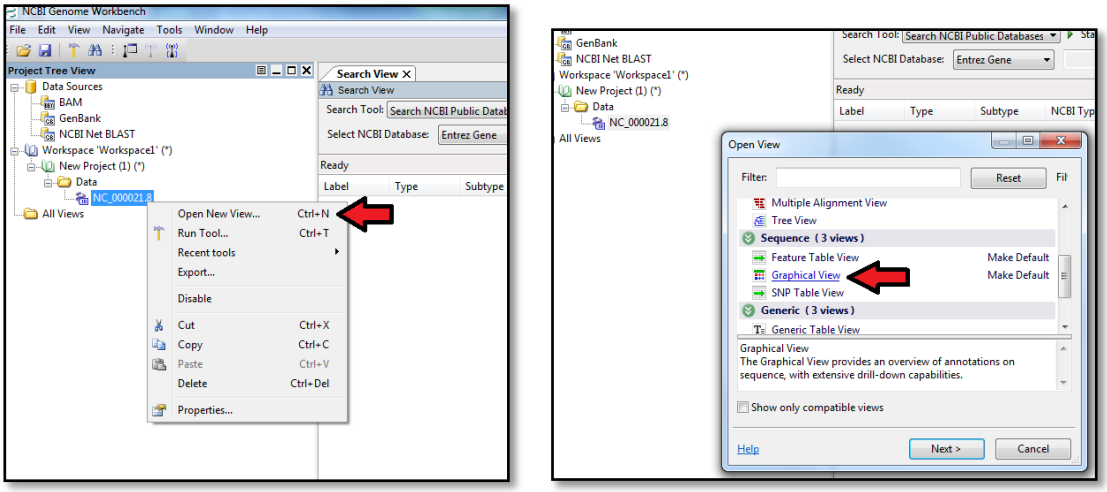
Download and install the Genome Workbench from
http://www.ncbi.nlm.nih.gov/tools/gbench/ . Open the Genome Workbench and click on File ->

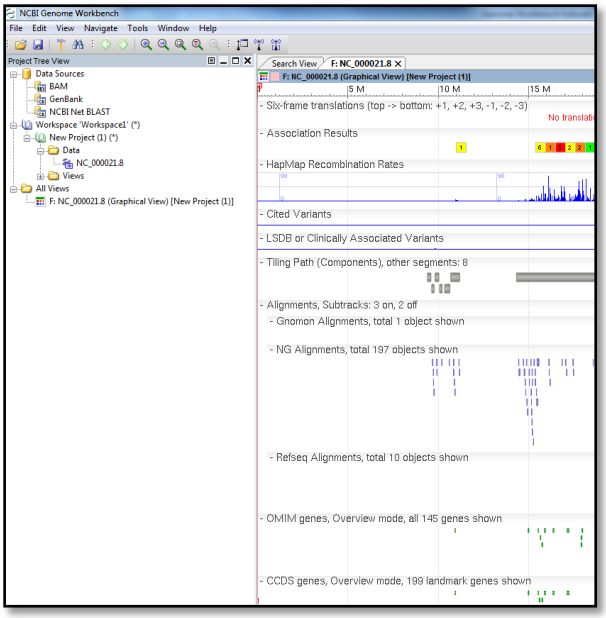
Open -> Data from GenBank. Enter your accession number in the box, in this example: NC_000021.8. Click on “Next” and “Finish”.





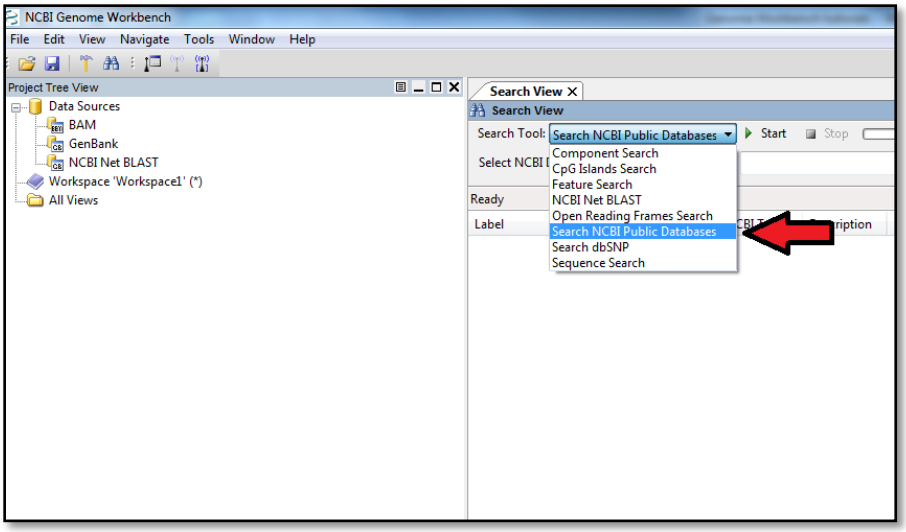
Select your data from the “Data” folder and right-click, and then select “Open New View”. Click on “Graphical View” in the pop-up window. A new window should appear showing the available tracks for your data.

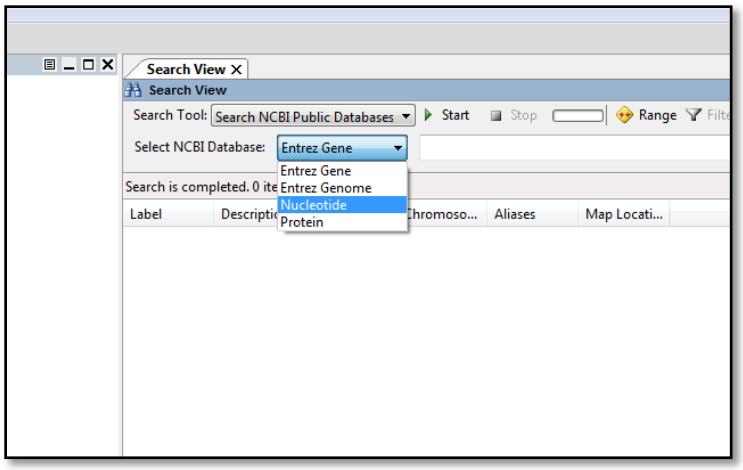




Retrieving a sequence from NCBI and loading it into genome workbench. (if you DO NOT know the accession number)

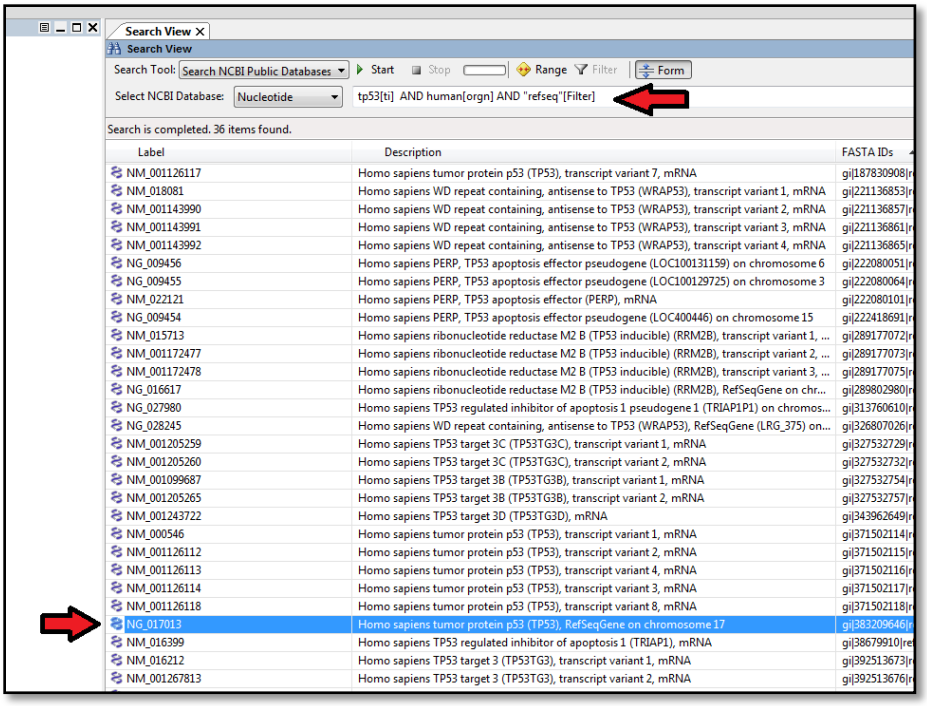
On the “Search View” window, select “Search NCBI Public Databases” from the “Search Tools” drop-down menu. Select one of the NCBI databases from the “Select NCBI Database” drop-down menu; in this example, we will use the “Nucleotide” database.

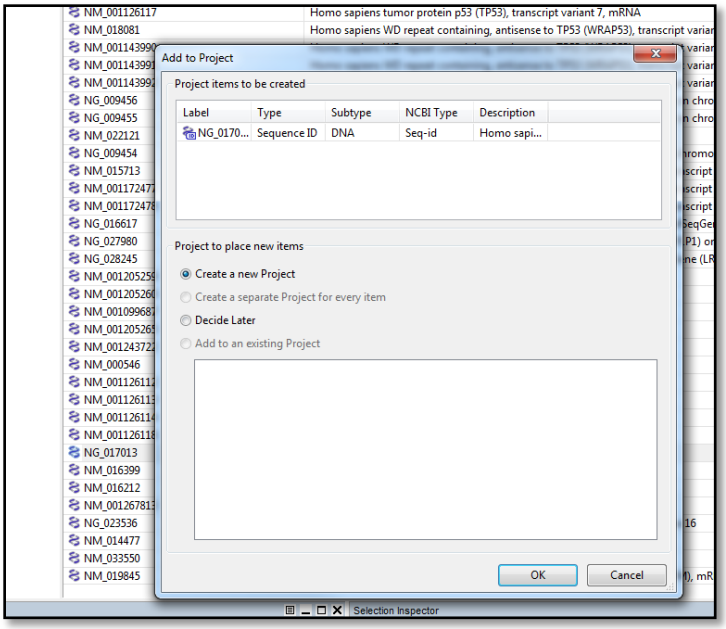




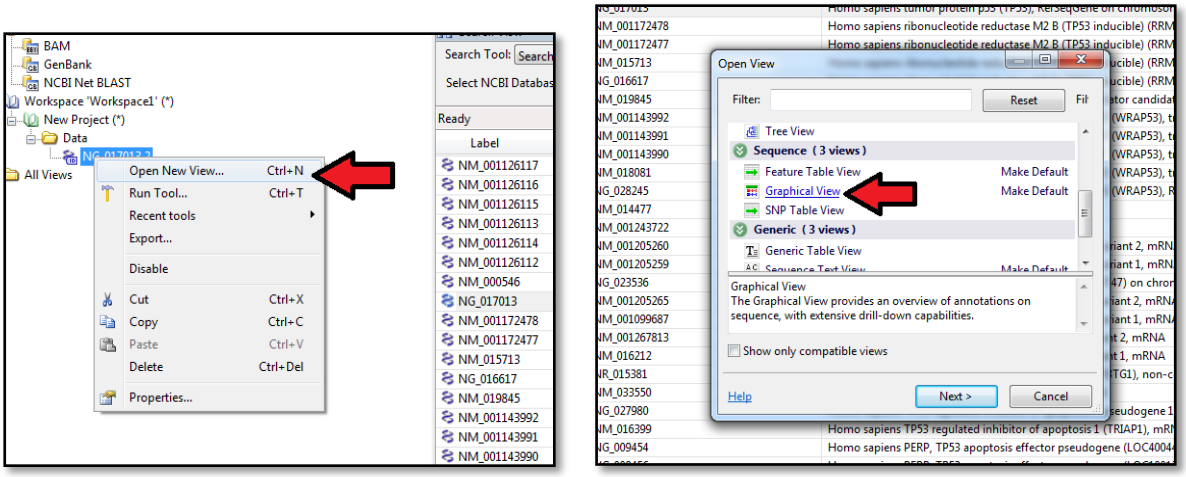
Enter your search query. You can make sophisticated Entrez queries in order to obtain a set of sequences through the Genome Workbench. In this example: *tp53[ti] AND human[orgn] AND "refseq"[Filter]*.

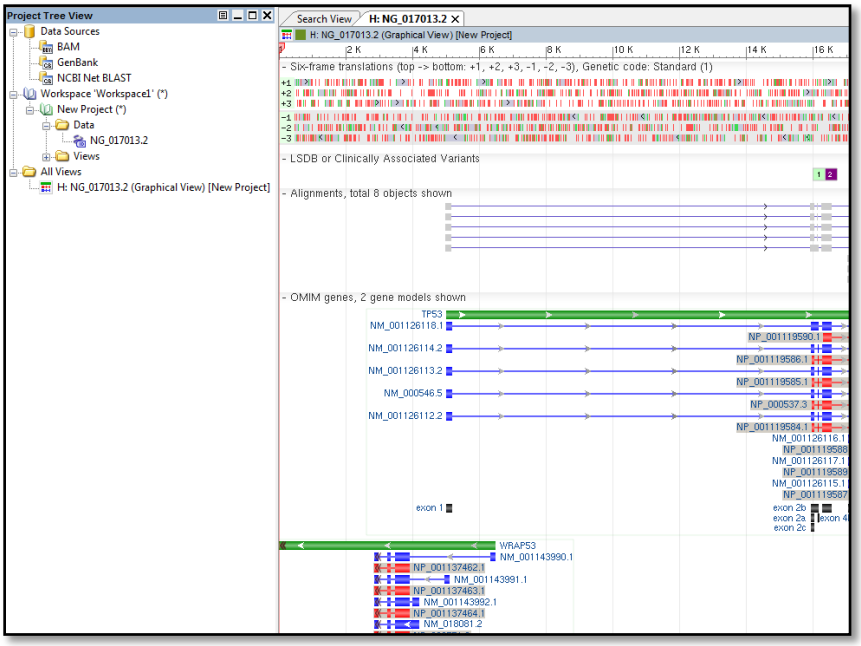
In the search results, double-click on your sequence (e.g. NG_017013), decide the project to place the new item (e.g. add to an existing project, create a new project, decide later) and click on the “OK” button.





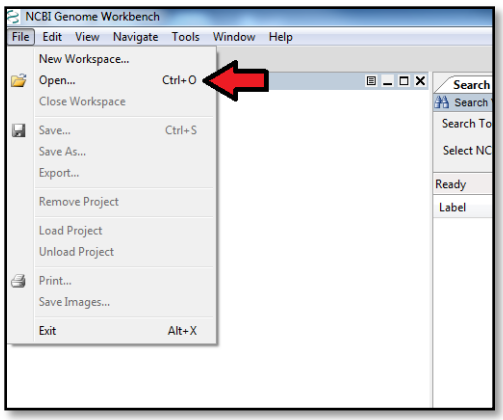
Select your data from the “Data” folder and right-click, and then select “Open New View”. Click on “Graphical View” in the new window. A new view window should appear showing the available tracks for your data.

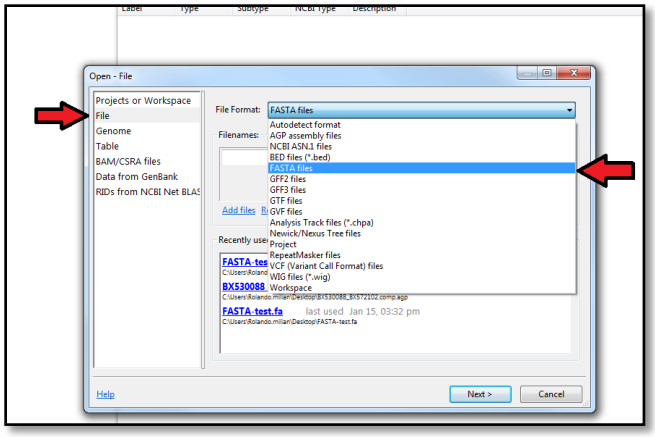


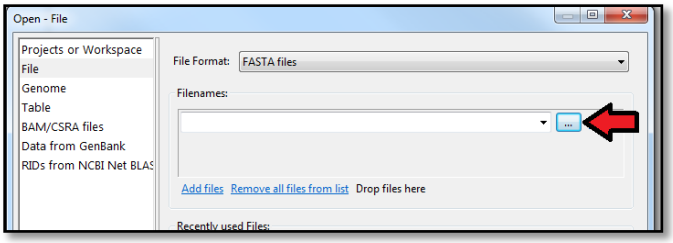


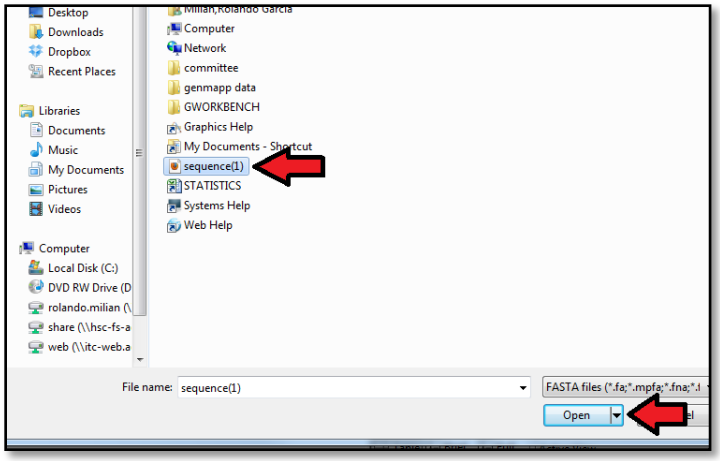
You can also import your data into the Genome Browser.

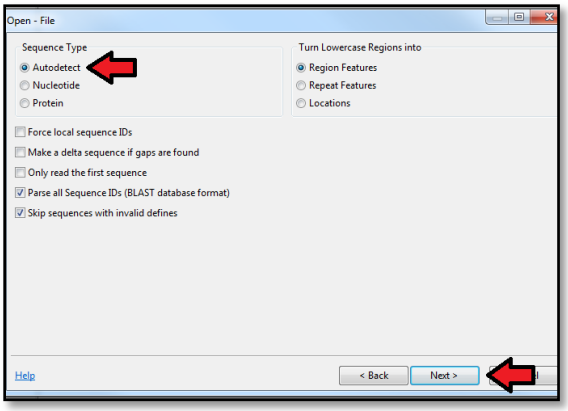
Click on File -> Open. Select “File” from the “Open- File” pop-up window, and then select the File Format, in this example FASTA files. Select and open the file of interest. Click on “Next”


NCBI Genome Workbench
File Edit View Navigate Tools Window Help
New Workspace...
Open... Ctrl+O
Close Workspace
Save... Ctrl+S
Save As...
Export...
Remove Project
Load Project
Unload Project
Print...
Save Images...
Exit Alt+X
Search
Search To
Select NC
Ready
Label


Open - File
Projects or Workspace
File
Genome
Table
BAM/CSRA files
Data from GenBank
RIDs from NCBI Net BLAS
File Format:
FASTA files
Filenames:
Autodetect format
AGP assembly files
NCBI ASN.1 files
BED files (*.bed)
FASTA files
GFF2 files
GFF3 files
GTF files
GVF files
Analysis Track files (*.chpa)
Newick/Nexus Tree files
Project
RepeatMasker files
VCF (Variant Call Format) files
WIG files (*.wig)
Workspace
Add files
Recently used
FASTA-test.fa
last used Jan 15, 03:32 pm
Help
Next >
Cancel

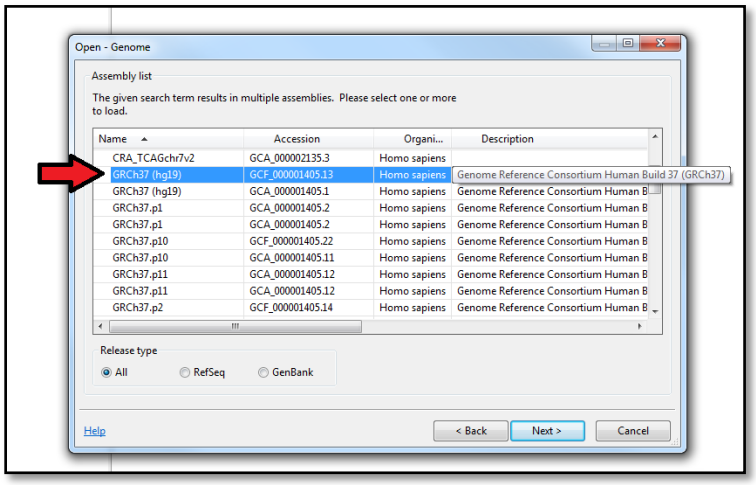

Open - File
Projects or Workspace
File
Genome
Table
BAM/CSRA files
Data from GenBank
RIDs from NCBI Net BLAS
File Format:
FASTA files
Filenames:
...
Add files
Remove all files from list
Drop files here
Recently used Files:

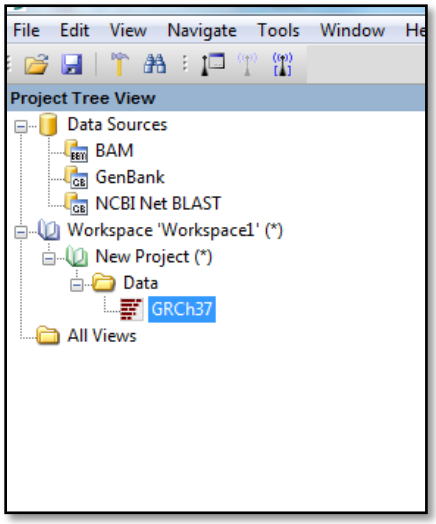

Desktop
Downloads
Dropbox
Recent Places
Libraries
Documents
Music
My Documents
Pictures
Videos
Computer
Local Disk (C:)
DVD RW Drive (D
rolando.milian (\
share (\\hsc-fs-a
web (\\itc-web.a
Computer
Network
committee
genmapp data
GWORKBENCH
Graphics Help
My Documents - Shortcut
sequence(1)
STATISTICS
Systems Help
Web Help
File name: sequence(1)
FASTA files (*.fa;*.mpfa;*.fna;*.f
Open


Open - File
Sequence Type
Autodetect
Nucleotide
Protein
Turn Lowercase Regions into
Region Features
Repeat Features
Locations
Force local sequence IDs
Make a delta sequence if gaps are found
Only read the first sequence
Parse all Sequence IDs (BLAST database format)
Skip sequences with invalid defines
Help
< Back
Next >

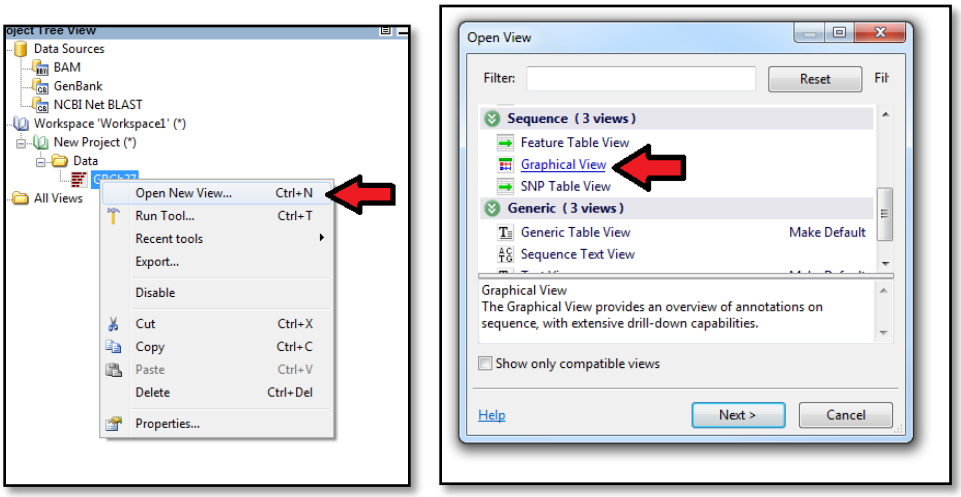
Loading a genome

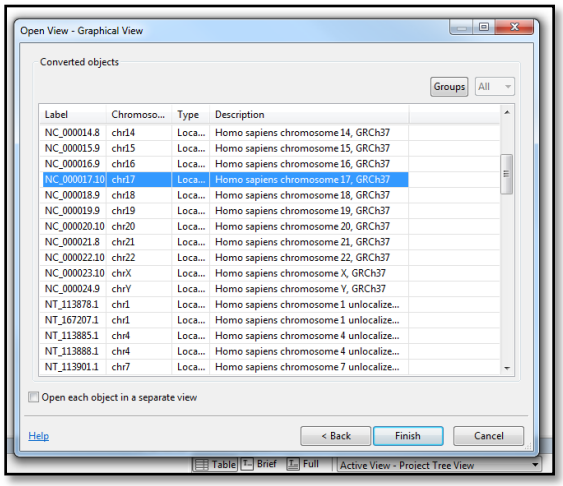
Go to File -> Open. Select “Genome” and type your search term in the search box (e.g. type “human”). Click on “Next” and select one or more assemblies from the table. Select “GRCh37 (hg19)” (Genome Reference Consortium Built 37) and Click on “Next”. Select “Create a new Project” and Click on “Finish”. The human genome build should be visible under a New Project.

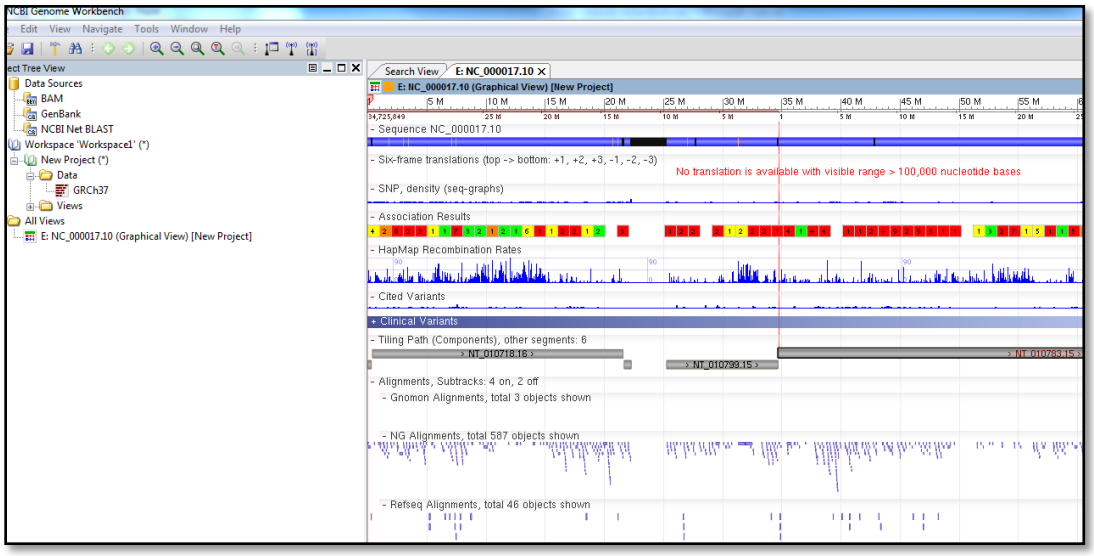




Once the genome is loaded, you can use the Genome workbench tools with it. In this example we will visualize Chromosome 17. Righ-click on the GRCh37 and select “Open New View”. Select “Graphical View”. A new window will open; select the chromosome you would like to view (NC_000017.10) and click “Finish”


Data Sources
BAM
GenBank
NCBI Net BLAST
Workspace 'Workspace1' (*)
New Project (*)
Data
All Views
Open New View... Ctrl+N
Run Tool... Ctrl+T
Recent tools
Export...
Disable
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Delete Ctrl+Del
Properties...
Open View
Filter:
Reset
Sequence (3 views)
Feature Table View
Graphical View
SNP Table View
Generic (3 views)
Generic Table View
Make Default
Sequence Text View
Graphical View
The Graphical View provides an overview of annotations on sequence, with extensive drill-down capabilities.
Show only compatible views
Help
Next >
Cancel


Open View - Graphical View
Converted objects
Groups
All
Label
Chromoso...
Type
Description
NC_000014.8 chr14 Loca... Homo sapiens chromosome 14, GRCh37
NC_000015.9 chr15 Loca... Homo sapiens chromosome 15, GRCh37
NC_000016.9 chr16 Loca... Homo sapiens chromosome 16, GRCh37
NC_000017.10 chr17 Loca... Homo sapiens chromosome 17, GRCh37
NC_000018.9 chr18 Loca... Homo sapiens chromosome 18, GRCh37
NC_000019.9 chr19 Loca... Homo sapiens chromosome 19, GRCh37
NC_000020.10 chr20 Loca... Homo sapiens chromosome 20, GRCh37
NC_000021.8 chr21 Loca... Homo sapiens chromosome 21, GRCh37
NC_000022.10 chr22 Loca... Homo sapiens chromosome 22, GRCh37
NC_000023.10 chrX Loca... Homo sapiens chromosome X, GRCh37
NC_000024.9 chrY Loca... Homo sapiens chromosome Y, GRCh37
NT_113878.1 chr1 Loca... Homo sapiens chromosome 1 unlocalize...
NT_167207.1 chr1 Loca... Homo sapiens chromosome 1 unlocalize...
NT_113885.1 chr4 Loca... Homo sapiens chromosome 4 unlocalize...
NT_113888.1 chr4 Loca... Homo sapiens chromosome 4 unlocalize...
NT_113901.1 chr7 Loca... Homo sapiens chromosome 7 unlocalize...
Open each object in a separate view
Help
< Back
Finish
Cancel
Table
Brief
Full
Active View - Project Tree View


NCBI Genome Workbench
Edit View Navigate Tools Window Help
Data Sources
BAM
GenBank
NCBI Net BLAST
Workspace 'Workspace1' (*)
New Project (*)
Data
GRCh37
Views
All Views
E: NC_000017.10 (Graphical View) [New Project]
Search View
E: NC_000017.10
- Sequence NC_000017.10
- Six-frame translations (top -> bottom: +1, +2, +3, -1, -2, -3)
No translation is available with visible range > 100,000 nucleotide bases
- SNP, density (seq-graphs)
- Association Results
- HapMap Recombination Rates
- Cited Variants
+ Clinical Variants
- Tiling Path (Components), other segments: 6
- Alignments, Subtracks: 4 on, 2 off
- Gnomon Alignments, total 3 objects shown
- NG Alignments, total 587 objects shown
- Refseq Alignments, total 46 objects shown

References

Ensembl tutorials and worked examples.
http://www.ensembl.org/info/website/tutorials/index.html

Fernandez-Suarez XM, Schuster MK (2010) Unit 1.15Using the Ensembl Genome Server to Browse Genomic Sequence Data
http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0115s30/pdf

Furey, TS. (2006) Comparison of human (and other) genome browsers. Human Genomics 2(4): 266–270.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2834533/

Karolchik D, Hinrichs AS, Kent J (2012) Unit 1.4 The UCSC genome browser. Current Protocols in Bioinformatics.
http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0104s40/abstract

Karolchik D, Hinrichs AS, Kent J (2011) The UCSC Genome Browser. Unit 18.6 The UCSC genome browser. Current Protocols in Human Genetics.

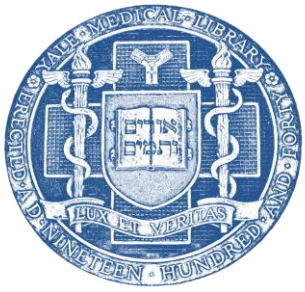
NCBI, Entrez Map Viewer help document
http://www.ncbi.nlm.nih.gov/projects/mapview/static/MapViewerHelp.html

NCBI, Genome Workbench basic operation

http://www.ncbi.nlm.nih.gov/tools/gbench/tutorial1/

UCSC genome browser user guide
http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html

Human Interferon gamma P01579 (IFNG_HUMAN) Reviewed, UniProtKB/Swiss-Prot
http://www.uniprot.org/uniprot/P01579



Curriculum and Research Support Department
Yale Medical Library