

# NCBI HANDOUT

Dr. Umesh P., Department of Computational Biology and Bioinformatics, University of Kerala

The National Center for Biotechnology Information (NCBI) advances science and health by providing access to biomedical and genomic information. It is one of the world's premier Web sites for biomedical and bioinformatics research. NCBI hosts many databases used by biomedical and research professionals. The services include PubMed, the bibliographic database; GenBank, the nucleotide sequence database; and the BLAST algorithm for sequence comparison, among many others.

It is very important to access sequence data of your interest before you start your Bioinformatics experiment. Also please note that you will get the exact information that you are looking for only if you have basic idea about the NCBI web site and retrieval procedure.

Let us explore NCBI

1. Go to web page [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) See Figure 1

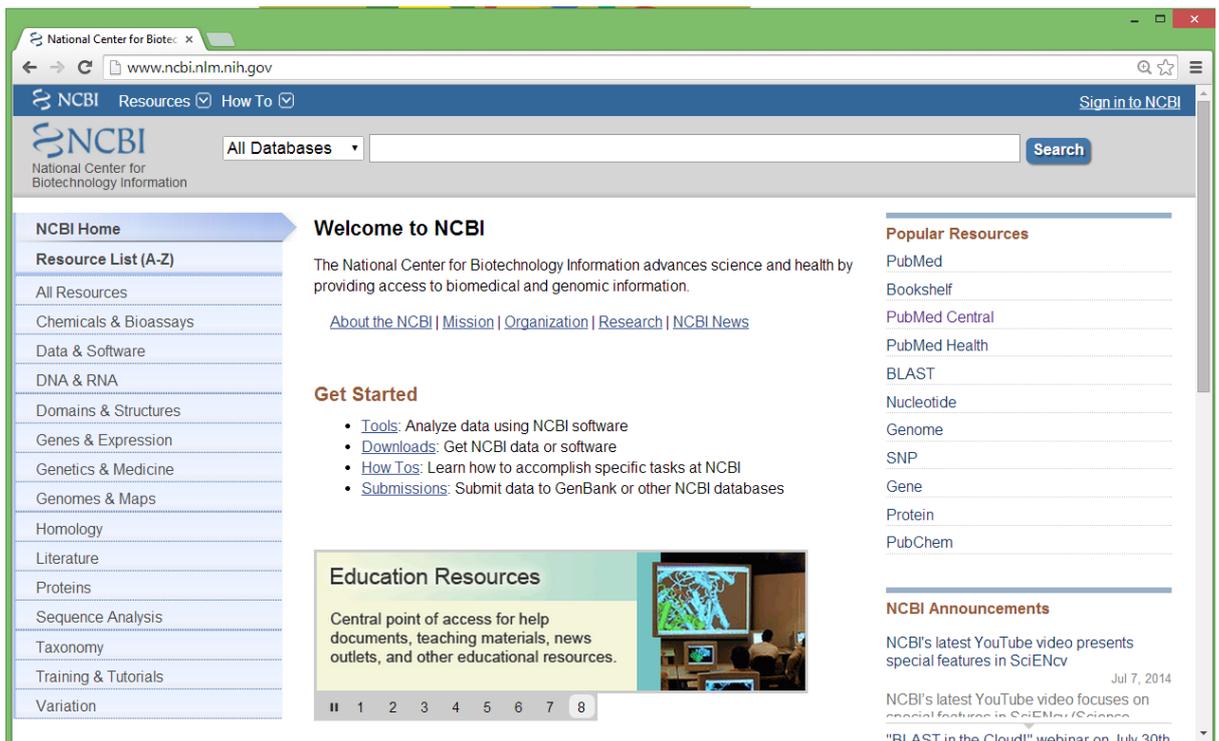


Figure 1: NCBI web page

The primary portal for accessing data at NCBI is called GQuery. When you enter any search request through **All Databases**, it will go to GQuery page which shows like Figure 2

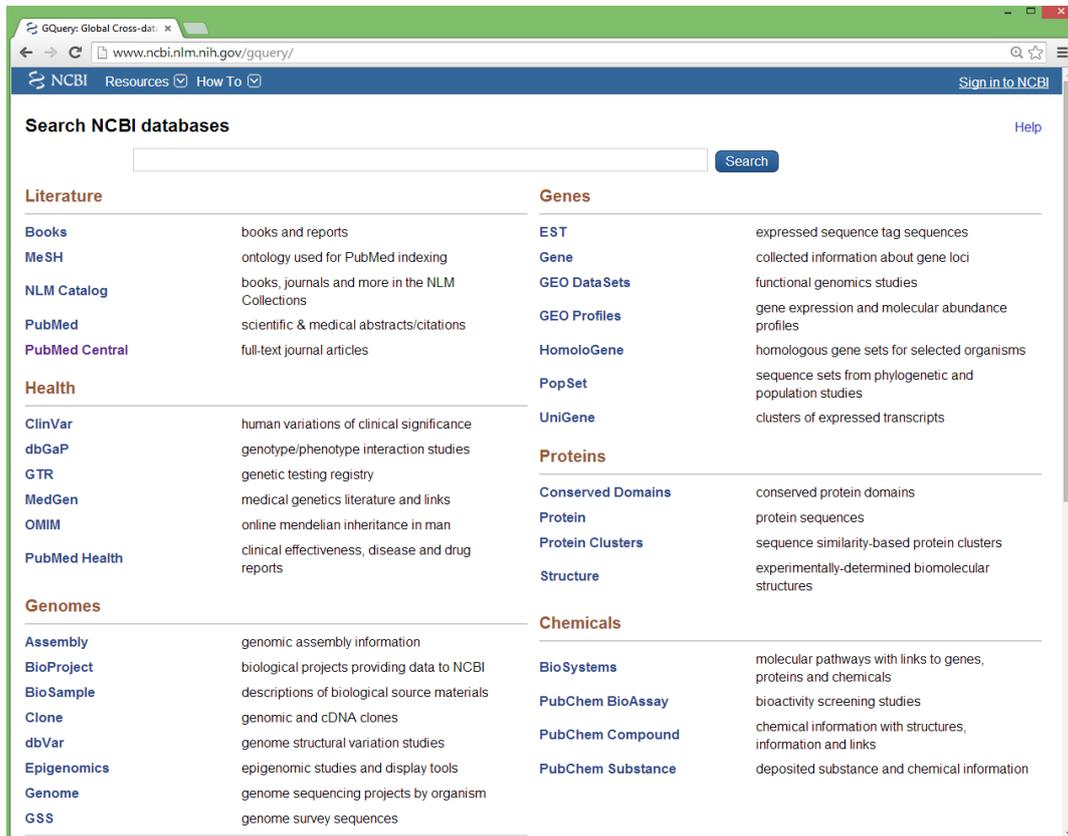


Figure 2: GQuery page of NCBI

Now try searching a keyword – Hemoglobin in GQuery page. Now you will end up with large number of results right? We need to specify which entry you would like to retrieve. This will take you to the exact result, you would like to retrieve.

Now go back to GQuery page. Type - **human leptin** and hit enter. This will take you to the GQuery page with results. Now click on Gene (See Figure 2). This will take you to the page as in Figure 3

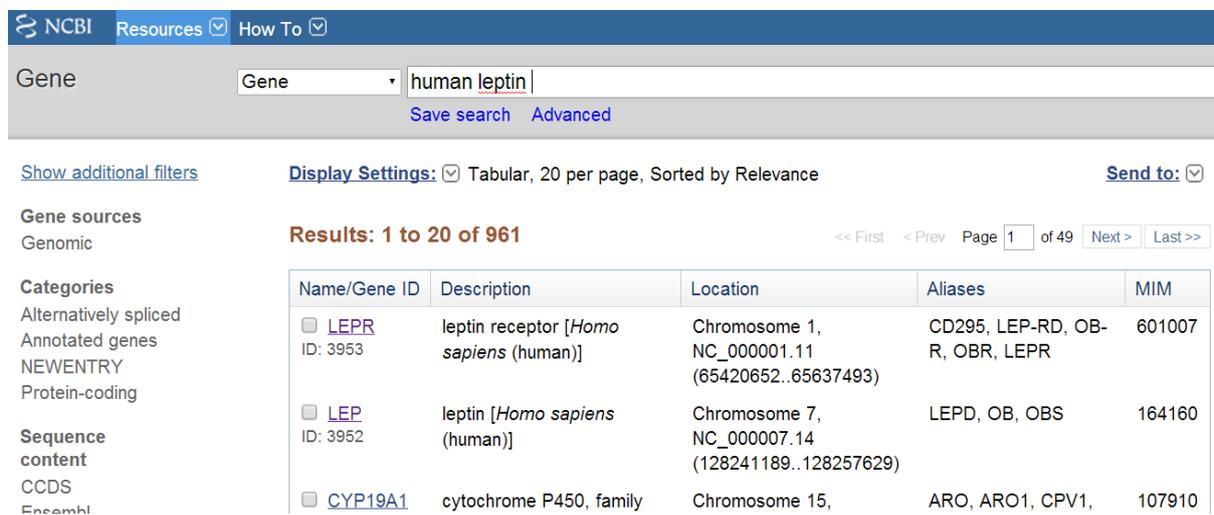


Figure 3: Page for Gene results of the query- human leptin

### ***Do you Know?***

Leptin is a hormone made by fat cells which regulates the amount of fat stored in the body. It does this by adjusting both the sensation of hunger, and adjusting energy expenditures. Hunger is inhibited (satiety) when the amount of fat stored reaches a certain level. Leptin is then secreted and circulates through the body, eventually activating leptin receptors in the arcuate nucleus of the hypothalamus

Now click on LEP, which is the official symbol for leptin. Also note that in the records, we can see that it is of human. We can see the gene ID as 3952. On clicking on LEP, it will take you to the detailed record of Leptin.

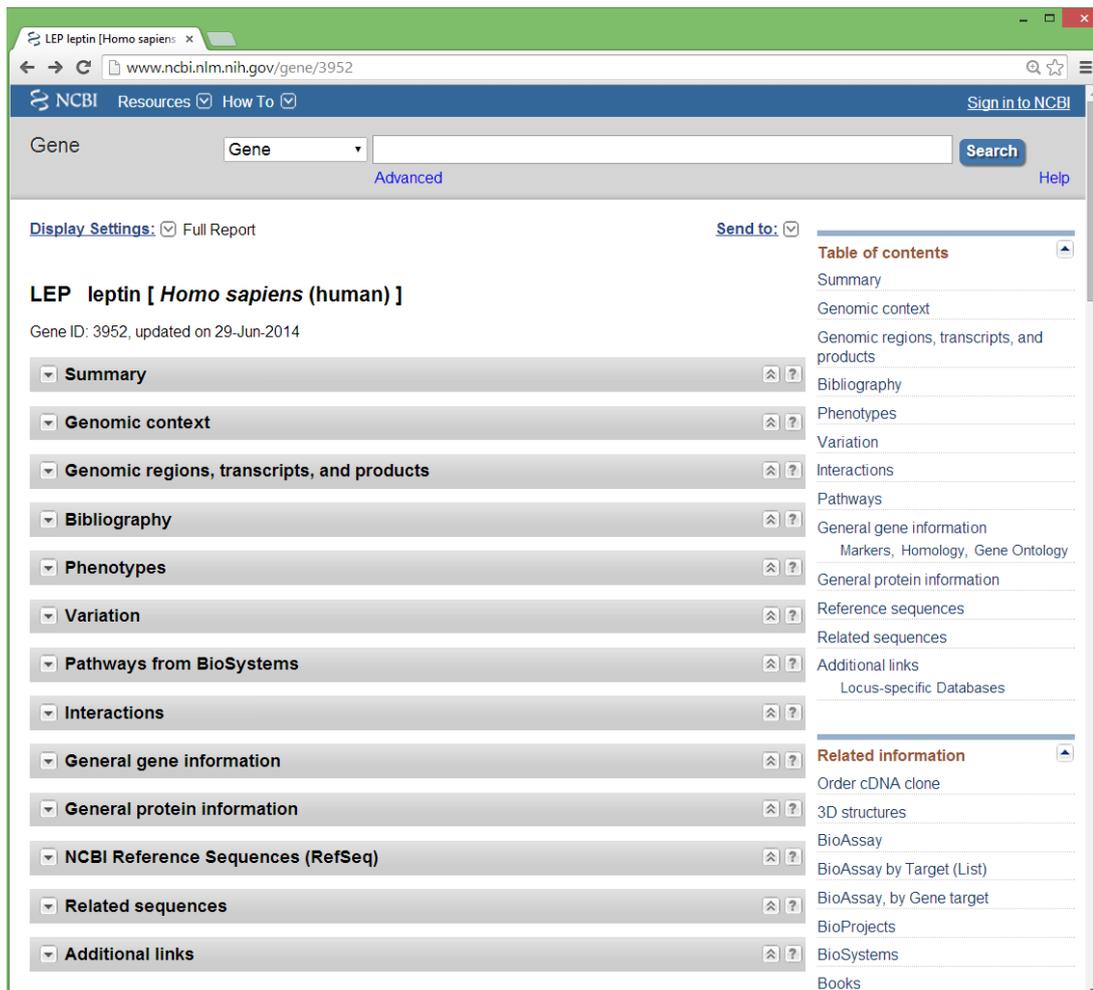


Figure 4: Result page (gene – human leptin) of NCBI

In the result page you can see the following divisions- Summary, Genomic context, Genomic regions, transcripts, and products, Bibliography, Phenotypes, Variation, Pathways from BioSystems, Interactions, General gene information, General protein information, NCBI Reference Sequences (RefSeq), Related sequences and some additional links.

LEP leptin [ *Homo sapiens* (human) ]  
 Gene ID: 3952, updated on 29-Jun-2014

**Summary**

**Official Symbol** LEP provided by [HGNC](#)  
**Official Full Name** leptin provided by [HGNC](#)  
**Primary source** [HGNC:6553](#)  
**See related** [Ensembl:ENSG00000174697](#); [HPRD:01249](#); [MIM:164160](#); [Vega:OTTHUMG00000157564](#)  
**Gene type** protein coding  
**RefSeq status** REVIEWED  
**Organism** [Homo sapiens](#)  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo  
**Also known as** OB; OBS; LEPD  
**Summary** This gene encodes a protein that is secreted by white adipocytes, and which plays a major role in the regulation of body weight. This protein, which acts through the leptin receptor, functions as part of a signaling pathway that can inhibit food intake and/or regulate energy expenditure to maintain constancy of the adipose mass. This protein also has several endocrine functions, and is involved in the regulation of immune and inflammatory responses, hematopoiesis, angiogenesis and wound healing. Mutations in this gene and/or its regulatory regions cause severe obesity, and morbid obesity with hypogonadism. This gene has also been linked to type 2 diabetes mellitus development. [provided by RefSeq, Jul 2008]

**Genomic context**

**Genomic regions, transcripts, and products**

**Table of Contents**  
 Summary  
 Genomic  
 Genomic products  
 Bibliography  
 Phenotypic  
 Variation  
 Interactions  
 Pathways  
 General gene  
 Marker  
 General protein  
 Reference  
 Related sequences  
 Additional  
 Locus

**Related information**  
 Order cDNA  
 3D structure  
 BioAssay  
 BioAssay  
 BioAssay  
 BioProject  
 BioSystem  
 Books

Figure 5: Summary tab of human leptin

Summary contains - Official Symbol, Official Full Name, Gene type, RefSeq status, Organism, Lineage and Summary. After that you can find The Genomic Regions, Transcripts, and Products section which is a graphical window to retrieve protein products etc.

When you need to download gene sequence/mRNA sequence you may go to the NCBI Reference Sequences (RefSeq) tab (see figure 4) and click on the download link,

You can access refseq information in three file formats- (1) GenBank, (2) FASTA, (3) Sequence Viewer (Graphics)

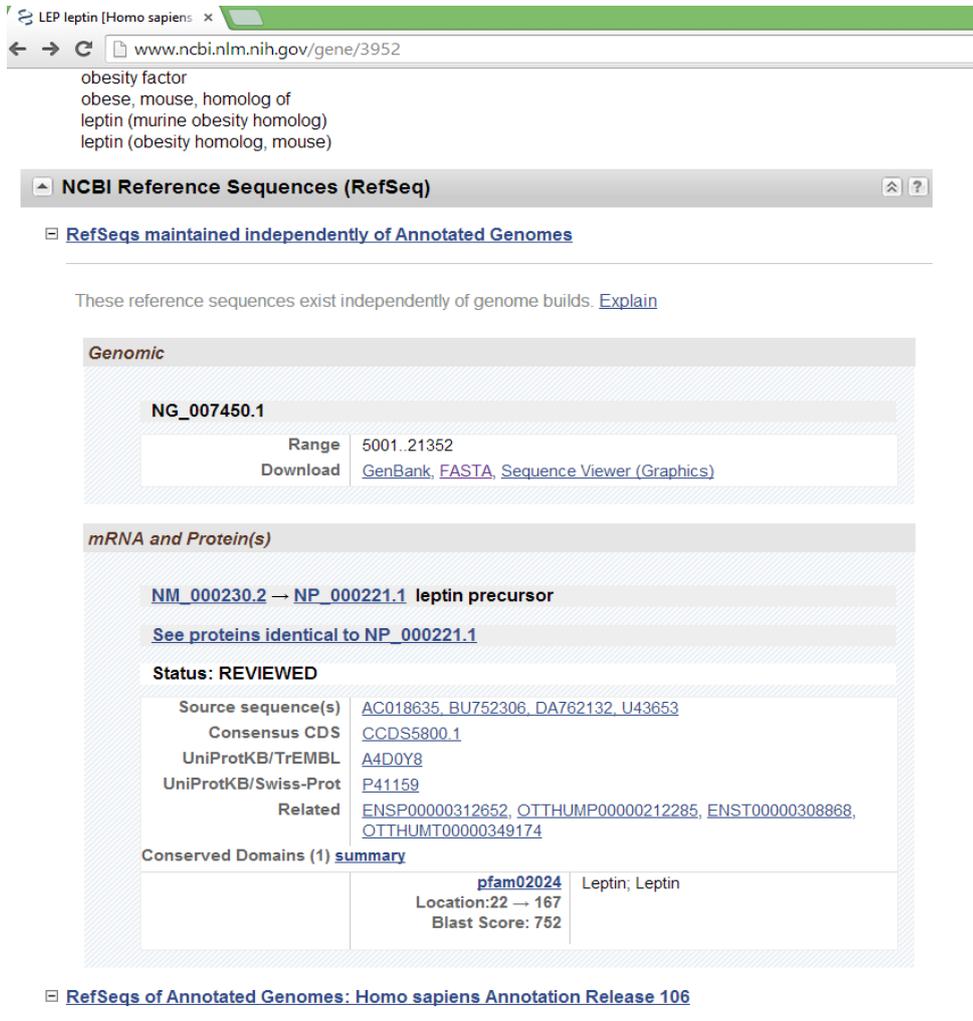


Figure 6: NCBI Reference Sequences (RefSeq) tab of human leptin

Fasta is the simplest file format which is the most commonly used format for tools and software. It contain a ">" symbol, an Accession number, (GenInfo Identifier – gi), and small description about the sequence (see figure 7)



Figure 7: FASTA file record of human leptin gene

Accession number is a unique identifier for a particular sequence record. An accession number is assigned to a specific record and stays with that record forever. Version numbers follow the Accession number and indicate the revision history of that entry starting with 1 and increasing with each revision.

RNA and protein products that are generated by the eukaryotic genome annotation pipeline, use accession prefixes XM\_, XR\_, and XP\_.

RNA and protein products that are mainly derived from GenBank cDNA and EST data and are supported by the RefSeq eukaryotic curation group, use accession prefixes NM\_, NR\_, and NP\_.

Category	Description
NC	Complete genomic molecules
NG	Incomplete genomic region
NM	mRNA
NR	ncRNA
NP	Protein
XM	predicted mRNA model
XR	predicted ncRNA model
XP	predicted Protein model

**Homo sapiens leptin (LEP), RefSeqGene on chromosome 7**  
 NCBI Reference Sequence: NG\_007450.1  
[FASTA](#) [Graphics](#)

Go to: [v]

LOCUS NG\_007450 16352 bp DNA linear PRI 04-MAY-2014  
 DEFINITION Homo sapiens leptin (LEP), RefSeqGene on chromosome 7.  
 ACCESSION NG\_007450 REGION: 5001..21352  
 VERSION NG\_007450.1 GI:169808406  
 KEYWORDS RefSeq; RefSeqGene.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.  
 COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from [AC018662.3](#) and [AC018635.10](#). This sequence is a reference standard in the [RefSeqGene](#) project.  
 Summary: This gene encodes a protein that is secreted by white adipocytes, and which plays a major role in the regulation of body weight. This protein, which acts through the leptin receptor, functions as part of a signaling pathway that can inhibit food intake and/or regulate energy expenditure to maintain constancy of the adipose mass. This protein also has several endocrine functions, and is involved in the regulation of immune and inflammatory responses, hematopoiesis, angiogenesis and wound healing. Mutations in this gene and/or its regulatory regions cause severe obesity, and morbid obesity with hypogonadism. This gene has also been linked to type 2 diabetes mellitus development. [provided by RefSeq, Jul 2008].  
 PRIMARY REFSEQ\_SPAN PRIMARY\_IDENTIFIER PRIMARY\_SPAN COMP  
 1-5476 AC018662.3 168370-173845  
 5477-23351 AC018635.10 20142-38016  
 FEATURES Location/Qualifiers  
 source 1..16352  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q31.3"  
 gene 1..16352  
 /gene="LEP"  
 /gene\_synonym="LEPD; OB; OBS"  
 /note="leptin"  
 /db\_xref="GeneID:3952"  
 /db\_xref="HGNC:HGNC:6553"  
 /db\_xref="MIM:164160"  
 mRNA join(1..29,10714..10885,13127..16352)  
 /gene="LEP"  
 /gene\_synonym="LEPD; OB; OBS"  
 /product="leptin"  
 /transcript\_id="NM\_000230.2"  
 /db\_xref="GI:169790920"  
 /db\_xref="GeneID:3952"

Figure 8: Genbank file record of human leptin gene in GenPept format.

You can also view the result in graphical format by clicking graphics link. You will get a page as in Figure 9.

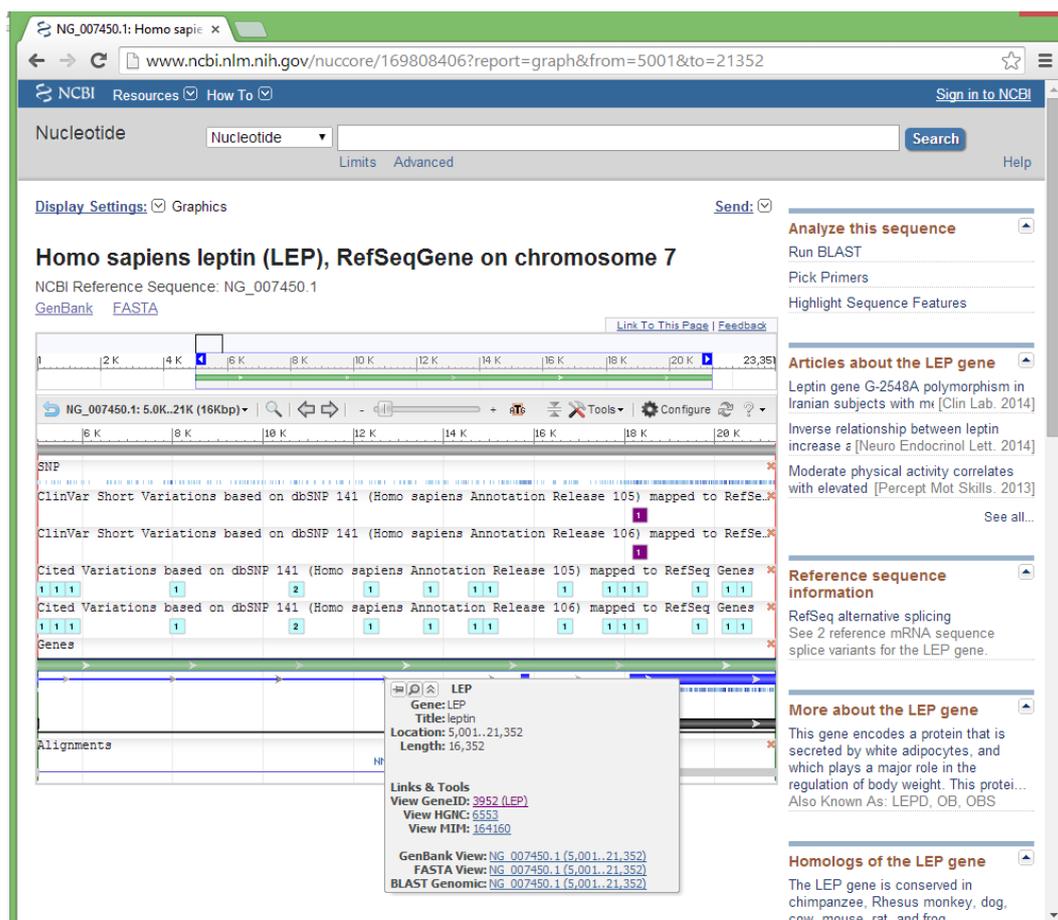


Figure 9: Graphical file record of human leptin gene

If you right click on the (green) band in the graphical window, you can download sequence records directly in the Fasta format.

The sequence can be accessed by another method also. If you know the location of a particular gene in the chromosome, you can access via map viewer.

Let us do and exercise with a protein well known to even a layman: haemoglobin.

Let us find out in human being's genome, where exactly the genes for hemoglobin reside. Select map viewer (from the submenu maps and markers take map viewer) from NCBI home page or use direct URL: <http://www.ncbi.nlm.nih.gov/mapview/>

Click on the Annotation Release 106.

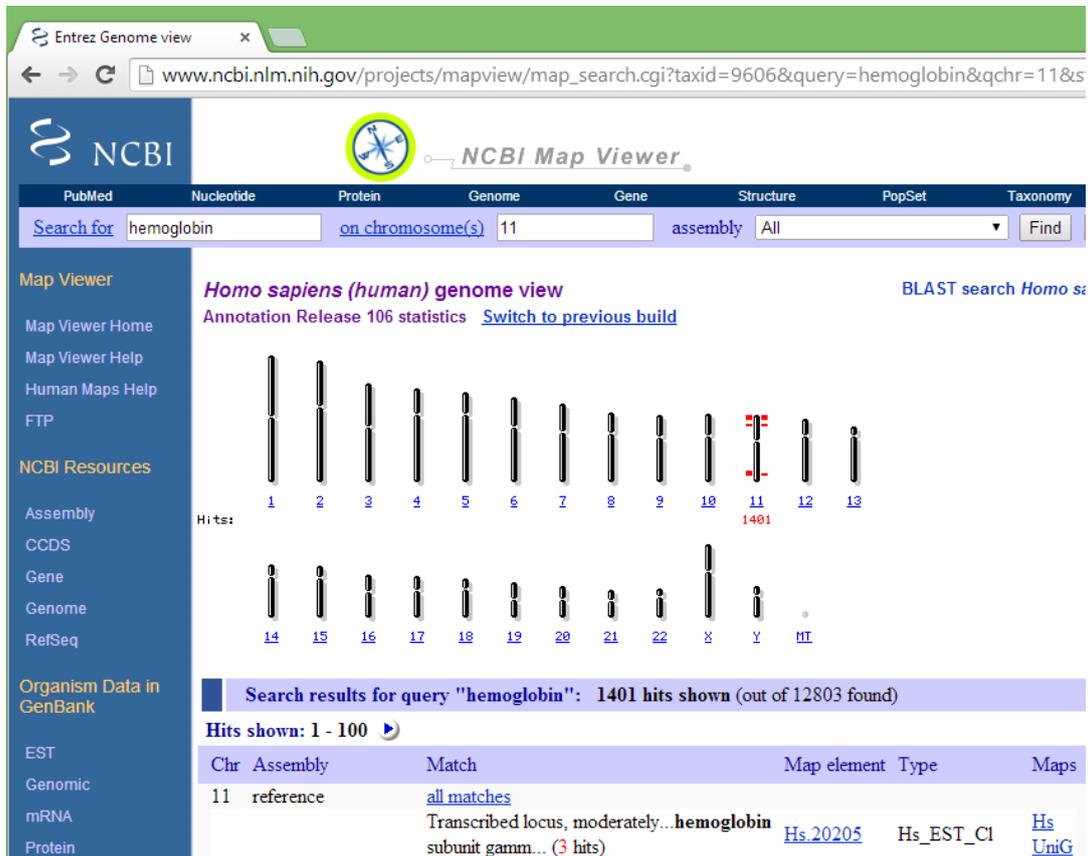


Figure 11: NCBI Map viewer page

See that all chromosomes (the 22 autosomes, the sex chromosomes along with the mitochondria) are represented by symbols. Search for **hemoglobin** in **all**.

In the result page, find out the chromosomes with the hemoglobin genes in them. The hits are prolific in chromosome No 11. Go down and have a look at the document and click at **all matches of reference assembly on chromosome 11** in hemoglobin hits.

In the new page, on the Genes map, click at **HBB-symbol HBB** to get the Entrez gene document on HBB- Take your own time to go through the document.

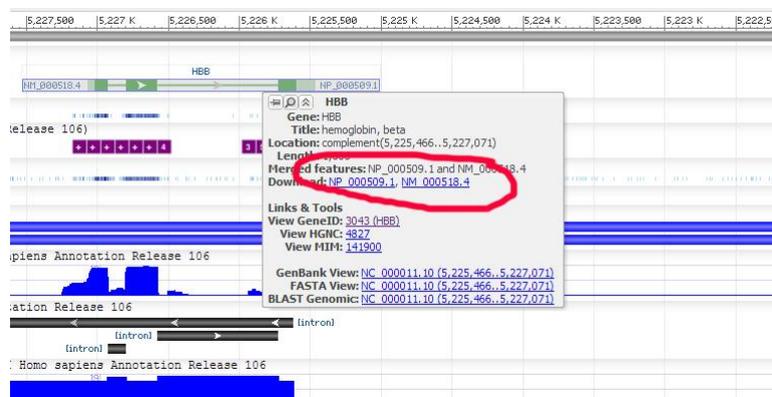


Figure 10: Downloading file record of a gene by using graphics page

Go back to GQuery page on HBB. Quite a lot of literature is available on hemoglobin. Click on pubmed links (**related articles in pubmed**) and then click on anyone paper to view the abstract.

Now it is time to get the actual sequences of hemoglobin genes and the protein. Let us go back to GQuery page of HBB and this time select **refseqs**.

Click on format FASTA and then save the FASTA file as plain text file –cut and paste and save as hbbdna.txt using notepad.

Now let us see how the protein structure can be downloaded from Molecular Modeling Database (MMDB)

For this go to GQuery page and type human haemoglobin and select protein- structure link. This will take you to the page from where you can select protein structure of your interest and download the structure.

The screenshot displays the NCBI MMDB protein structure page for Human Obesity Protein, Leptin. The page includes the following information:

- Protein Name:** Human Obesity Protein, Leptin
- Citation:** Crystal structure of the obese protein leptin-e100. Zhang F, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, Dimarchi RD, Furman TC, Hale JE, Hsiung HM, Schonher BE, Smith DP, Zhang XY, Wery JP, Schevitz RW. *Nature* (1997) 387 p.206
- Biological Unit:** monomeric; determined by author
- Structure Summary:** MMDb ID: 55396, PDB ID: 1AX8, PDB Deposition Date: 1997/10/31, Updated in MMDB: 05/2012, Experimental Method: X-Ray Diffraction, Resolution: 2.4 Å, Source Organism: Homo sapiens
- View or Save 3D Structure:** File Format: Cn3D, Display As: 3D structure, Data Set: Single 3D structure. Buttons: View structure, Download Cn3D.
- Notice:** In order to view this biological unit properly, please upgrade to Cn3D 4.3.
- Molecules and interactions:** A table with columns for Label, Count, Molecule, and Interactions. It shows "Protein and interactions (1 molecule)".

Figure 11: Molecular Modeling Database (MMDB)