

## IST 444

### Assignment 2: Learning to use NCBI

#### Step 1

**1. Start at the main NCBI page. Use all databases on the NCBI home page. To retrieve a large amount of returns, use “retinol binding protein” as your search term. How many proteins did you find in the Entrez Page?**

Using all the databases available on NCBI, Retinol Binding Protein was searched for. There were approximately 545 Proteins found.

**2. Now try “retinol binding protein 4” on All Databases search. How many proteins do you find in the Entrez Page?**

Retinol binding protein 4 was then searched for. It yielded 193 proteins.

**3. What about “RBP4” on all the databases search? To make it even more specific, let’s add “RBP4 homo sapiens.” How many proteins do you find using the all databases search?**

RBP4 was then searched for. This yielded 64 proteins. RBP4 homo sapiens yielded 15 proteins after being searched for.

**4. If you consider in question 3 you are looking for the full length rbp4 Homo sapiens with accession number NP\_006735. What about searching using this tool do you think still gets other returns when you type in “rbp4 homo sapiens?”.**

The other proteins come up in the search because they all pertain to homo sapiens and some aspect of retinol binding proteins.

**5. What is the full name of this gene’s protein product?**

The gene’s protein product is retinol-binding protein 4, plasma precursor [homo sapiens].

**6. Give a brief description of what the protein does. If you quote a record, give me the link you used.**

This protein is responsible for carrying retinol (vitamin A alcohol) in the blood. Retinol is moved from the liver to the tangential tissues.

**7. How many amino acids are in this protein?**

This protein has 201 amino acids.

**8. Are there functional domains described for this protein? You will find this in the conserved domain database. This is either in RefSeq or can be linked from domains through the record. List them:**

There are 21 domains for "Rbp 4 homo sapiens"

- a. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=71614>
- b. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=71601>
- c. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70908>
- d. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70876>
- e. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70455>
- f. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70420>
- g. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70236>
- h. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=70230>
- i. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=69903>
- j. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=69361>
- k. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=69208>
- l. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=68443>
- m. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=68258>
- n. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=68171>
- o. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=67152>
- p. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=66499>
- q. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=79595>
- r. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=79590>
- s. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=58014>
- t. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=29089>
- u. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=68884>

**9. How many amino acids are in the sig peptide? What is a signal peptide? How many are in the mat peptide? What is a mature peptide? What does CDS stand for and how many nucleotides are in the CDS?**

There are 18 amino acids for the sig\_peptide. A signal peptide is a peptide that is approximately 3-60 peptides long and it directs the post-translational modification of a peptide. A mature peptide is a protein that has been altered by proteolysis.

There are 182 amino acids for the mat\_peptide.

There are 201 nucleotides in the CDS. CDS stands for coding sequence.

**10. Can you find any PubMed references for this gene? Give me the link(s).**

Articles from PubMed include:

- a. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=17618858](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=17618858)
- b. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=17405846](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=17405846)
- c. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=17337499](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=17337499)
- d. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=17299074](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=17299074)
- e. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=17259477](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=17259477)
- f. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=1623143](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=1623143)
- g. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=2217163](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=2217163)
- h. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=574085](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=574085)
- i. [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list\\_uids=571335](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&list_uids=571335)

**11. What does it mean when the record states that it has been “curated by NCBI staff?”**

When it is stated that the article has been “curated by NCBI staff” this means that it has been edited from the original submissions from the various different scientists. Instead of having redundant information, the curators at NCBI condensed everything into a simplified entity.

**12. Does this gene have any homologs in other species? Give me some links to some of them. You can find this through blink or homologue database:**

This gene does exist in other species. It exists in 8 other species. They are:

- a. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=4908&ordinalpos=1&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=4908&ordinalpos=1&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)

- b. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=101049&ordinalpos=2&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=101049&ordinalpos=2&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- c. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=100570&ordinalpos=3&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=100570&ordinalpos=3&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- d. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=99569&ordinalpos=4&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=99569&ordinalpos=4&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- e. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=95712&ordinalpos=6&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=95712&ordinalpos=6&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- f. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=95712&ordinalpos=6&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=95712&ordinalpos=6&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- g. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=94706&ordinalpos=7&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=94706&ordinalpos=7&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)
- h. [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=39391&ordinalpos=8&itool=EntrezSystem2.PEntrez.Homologene.Homologene\\_ResultsPanel.Homologene\\_RVDocSum](http://www.ncbi.nlm.nih.gov/sites/entrez?Db=homologene&Cmd=ShowDetailView&TermToSearch=39391&ordinalpos=8&itool=EntrezSystem2.PEntrez.Homologene.Homologene_ResultsPanel.Homologene_RVDocSum)

**13. Click on the link associated with domains under RefSeq. What is a conserved domain of this protein called? What is its function?**

It is associated with many biological processes such as immune response, cancer cell interactions, prostaglandin synthesis, and more.

**14. What chromosome is this gene on? What chromosome arm is it on? How many nucleotides are listed in this entire chromosome? You will find this information in Entrez GENE database or in the mapviewer links to the right of the page.**

It is on chromosome 10, the q arm. There are 135Mbp located on the chromosome.

**15. Click on mapviewer. What is the accession number of the genomic contig for RBP4? How many nucleotides does it contain?**

The accession number for the contig is NM\_006744. It contains 606 nucleotides.

**16. Click on the annotation links labeled sv mv ev mm in mapviewer. What are each of these links abbreviations for?**

sv: sequence viewer  
mv: map viewer  
ev: evidence viewer  
mm: model maker

**17. Does this gene contain introns? If so, how many and where are the splice junctions? Which link did you use to discover this?**

There are several, including looking for the gene name in the genomic contig sequence, or looking in the whole chromosome sequence.

**18. Click on the OMIM link. What is a biological consequence of a mutation in this protein for humans?**

A biological consequence of a mutation in this retinol binding protein is a deficiency of the retinol binding to the protein.

**19. Can you find your gene in SwissProt database? Give me the accession number in Swiss Prot.**

The gene, FADD can be found in the SwissProt database. Its accession number is Q13158.

**20. What is the advantage of Swiss Prot vs. NCBI?**

The advantage of using Swiss Prot versus NCBI is because although it is not as big as the other networks, it is increasingly more curated than the others. This means that more scientists have gone through the genomes and have spent more time validating the results given to those who search.

## **Step 2**

**1. Go to Entrez and search in the genome database. Type in "Hiv-1." How many results did you get? We are interested in NC\_001802.**

There are 7 hits in the genomic database.

**2. Click on this Hiv-1 accession number. How many different coding regions do you find?**

There are 9 protein coding regions.

**3. How many open reading frames (ORF)'s are there?**

There are 122 open reading frames

**4. Where would we go to find the reverse transcriptase sequence of this genome?**

a. Unigene

b. OMIM

c. PubMed

**d. Swiss Prot**

**5. Find the reverse transcriptase accession number P03369. How many amino acids are there?**

There are 1437 amino acids.