

Computational Biology (BT401)

NCBI and EBI InLab

Part 1: (NCBI)

Go to the National Center for Biotechnology Information (NCBI) site to solve problem number 1, 2 and 3. You may use the entrez search on NCBI, the Genome viewer and any other tools or services.

- Note: If you are asked to copy a sequence in FASTA format it should contain the whole content of the file with the sequence name and comments that precede the sequence.

Example:

```
>NP_000509.1 hemoglobin subunit beta [Homo sapiens]
MVHLTPEEKSAVTALWGKVVNDEVGGGEALGRLLVVPWTQRRFFESFGDLSTPDVAMGNPKVKAHGKKVLG
AFSDGLAHLDDLKGTFTSLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
```

Problem No.1: Access to Sequence Data and Related information

1. Using the accession number NM_005368.2, answer the following set of questions.
 - A. What kind of sequences is it (DNA , RNA or protein) ?
 - B. What is the chromosome number that hosting this gene?

Screen shot for this sequence from [Genome Data Viewer](#)

- C. What is the source of this gene?
- D. How many base pairs this gene has?

E. Download the sequence to your desktop in FASTA format.

Paste the sequence in FASTA format

2. Download the **transcript** for neutrophil gelatinase-associated lipocalin precursor [Homo sapiens] (NP_005555.2)

Paste the sequence in FASTA format

3. You are asked to search for *human complement component 8 gamma (C8G)* on NCBI.
- A. What is the Gene ID for this gene? -----
 - B. Is it protein coding gene? -----
 - C. If it is protein coding gene what is the accession number for this protein? ----

Problem No. 2: Basic Local Alignment Search Tool (BLAST)

1. **Pairwise sequence alignment** : Apply pairwise sequence alignment between [caspase-2 isoform 1 preproprotein\(NP_116764.2\)](#) and [caspase-2 isoform 2 \(NP_001215.1\)](#) sequences.

A. Show is the result of the alignment .

Paste a print screen for the alignment description

B. Print out the allignemnt between the two sequences.

Paste a print screen for the alignment result

2. **BLAST Search:** You are asked to search for [caspase-2 isoform 1 preproprotein \[Homo sapiens\] NP_116764.2](#) protein against human [refseq_protein](#) database using BLAST tool on NCBI with the default parameter values. Answer the following set of questions:
- A. Print out your searchdescription using print screen button.

- B. How many hits the search returns? -----
- C. Print out your search graphical result here.

- D. What is the protein sequence that is most homology/similar for our query excluding the sequence with it self. -----

- E. What is the identity percentage between the query sequence and caspase-9 isoform 1 precursor [Mus musculus]? -----
- F. View the pairwise alignment between these two sequences and print the screen here.

Problem No. 3:

1. Find the similarity/identity between **neutrophil gelatinase-associated lipocalin precursor [Homo sapiens] (NP_005555.2)** and **human complement component 8 gamma (C8G) protein**?
2. View the pairwise alignment between these two sequences and print the screen here.

3. Print out the Distance tree between your query and search results.

Part 2: (EBI)

Problem No. 4:

Using the Clustal MSA tool on The European Bioinformatics Institute (EBI) :

1. Build a multiple sequence alignment for the following sequences
[beta_globin](#), [myoglobin](#), [neuroglobin](#), [soybean_globin](#), and [rice_globin](#).
2. Print out the multiple sequence alignment of the five globins.

3. print out the guid tree for the following MSA between