EC Innovative Training Network MATURE-NK

Bioinformatics Course

Exercises Day 1:

Sequence databases + retrieval:

NCBI

1) Go to the NCBI webpage https://www.ncbi.nlm.nih.gov
2) Enter the name of an NK receptor or trans-membrane signaling protein of your interest into the search field, e.g. DAP10.
3) In how many of the categories do you get hits?
4) How many hits do you get in Homologene and Taxonomy? Look at them. Do they all look expected.
5) How many SRA hits are there. What are those? Are they experiments related to the gene?
6) How many OMIM hits are there. What are those? Are they related to the gene?
7) Go to the NCBI Gene database.
8) How many hits do you have? What kinds of species do the reflect (on the first page)?
9) Choose one of the entries from the first page. Note down the species, protein name, genomic location, etc.
10) Is the genome in a finalized state?
11) Find the link to the sequence and to the Genbank entry. Are you able to obtain the DNA and protein sequence? Store it in FASTA format (or create a file in FASTA format) and store the sequence for later.
12) Visualize the gene in a Genome Browser/Viewer.
13) Explore other parts of the NCBI hits and other like of the Gene entry. Is there anything useful for you?
14) Enter the species name "Nitrososphaera" into the NCBI Taxonomy. How many sequences of different types do exist? How many genomes are there.

ENSEMBL

15) Go to the Ensembl webpages http://ensembl.org/ and http://ensemblgenomes.org/ - do you see differences? What do the contain? Try to figure out how many genomes the contain.
16) Enter the species name "Nitrososphaera". Do you get any hits.
17) Enter the name of your protein of interest in there. Do you get hits here?
18) Try to restrict the search to one or a few species you are interested in. How many hits do you get here.
19) Go to BioMart and obtain a table with the Gene name, product/protein name, genomic position, TSS and at least two accession numbers from different databases.