Technische Universität München   Prof. Burkhard Rost
Lehrstuhl XII: Bioinformatik   Tobias Hamp
                                                      Esmeralda Vicedo

Exercise 'Protein Prediction II'
Winter Term 2011-2012

Sheet 1

General information

- Our course homepage, containing lecture slides and exercise sheets:
  http://rostlab.org/cms/teaching/teaching-overview/hs-ws-11-12/
- Time and place: **Friday, 13:00 – 14:30**, Seminar room 'John von Neumann' (MI 00.11.038)
- Grading („Schein"):
    - **40% Exercise** (theory/programming),
      60% Final exam (lecture/exercise content)
    - T.b.d.: minimum of exercise points required for exam attendance
- Contact:
  hampt@rostlab.org, vicedo@rostlab.org
- Send an email (one per group) to **the two of us** including the paths only (no files as attachments!) to your results (answers, program code, figures ‚...) until **November 4, 10:00 am**. Scripts should be executable for us so that we can reproduce your results.
  Everything has to be readable by us, so please check the permissions of your directories/files.

Exercise 1:   The Gene Ontology (3 Points)

The Gene Ontology plays an important role on this and upcoming exercise sheets. Therefore, a good understanding of its ideas and organization is essential not here, but throughout the rest of the course. Use the following tasks to make yourself familiar.

  a) What is the Gene Ontology (www.geneontology.org)? Explain its major principles and organization.
  b) Explain the relationships between the following three databases:
       - Gene Ontology
       - UniProt
       - SwissProt
  c) What is a Gene Ontology Evidence Code? We want to predict protein function in-silico. Do we need to consider the Evidence Code in this context? Explain.

Exercise 2:   Myosin X, hemoglobin and pyruvat kinase (5 Points)

The proteins above will give you good hands-on experience with the GO from the perspective of single annotations.

Go to UniProt (www.uniprot.org/uniprot) and explore the protein records for human myosin X, hemoglobin (any organism) and pyruvat kinase (any organism)

a) What do these proteins do?
b) List all the Gene Ontology terms of human myosin X that you find there.
c) Is there any experimental evidence for the annotations of b)?
d) Find all the direct descendents for all the GO terms of b). Are these descendent also part of the terms of the protein? Explain your answer.
e) Explore the GO annotations of hemoglobin and pyruvat kinase directly on the GO website. When you come across it, put special emphasis on the understanding of the "Inferred Tree View" and "Graph View". What do you think: Could the Gene Ontology successfully annotate the two proteins? Explain.

Exercise 3:   First programming steps (3 Points)

Now, we switch the perspective to the whole protein universe. Use the following task to familiarize yourself with handling functional annotations on a large scale.

a) Write a parser which extracts all Gene Ontology terms for each protein in SwissProt (/mnt/opt/data/pp2_11_12/sprot.dat).

Exercise 4:   Critical Assessment of Function Annotations (CAFA) (3 Points)

The CAFA experiment (biofunctionprediction.org) was carried out for the first time this year. Hopefully, next year we will again be part of it.

a) What is CAFA?
b) Who is CAFA?
c) How was it carried out?

Give a rough overview.