Description of the project work for bioinformatic practicals

3-6 pages long manuscript of 3 authors

Send: 3 January 2020 as a **pdf** attachment of an email to ELTEbioinfo@gmail.com. (If you send it earlier, we will correct it earlier.)

Topic: Description, comparison and summariation of certain properties of the 3 different proteins the authors analyzed during the semester.

Description: Write down the details of how you solved the below mentioned tasks and insert the results. Comment why you did certain steps and what is your opinion about the results. The sent document should be able to explain all above mentioned thing itself.

Authors: There should be three authors who have different proteins. (If necessary you can choose other proteins from this list: RASK, ERK1, JAK1, IGF1R, GSK3B, AXIN1, SMAD2, NOTCH1.) Write the name of the authors to the document.

Tasks:

1. Describe the proteins based on the basic descriptions you find in the Uniprot database. What are the unique and common properties of your proteins based on the Gene Ontology descriptions? (You can find the Gene Ontology categories of your proteins at Uniprot or at Ensembl Biomart.)

2. How well are your proteins represented by know structures? Compare your proteins! For this, check, which part of your protein has a known structure (using the Uniprot or the PDB resource). Select the protein that has *the least structural coverage* (in terms of residue). Write down which protein did you choose and *why*. Create a homology model for a region that does not have a known structure. You can use the HHPred method for this. For this, you can use the HHPred method as described in the tutorial of the practical. Write down the code of the template structure, the sequence identity between the selected protein and the template, and the region that is covered by your model. Do you trust this model? Describe why or why not. Download the PDB file of the model, then load it into jmol. Create an image using the ribbon representation colored by secondary structure. Attach an image of your model to the document.

3. Draw a majority rule (not extended) consensus tree from the maximum parsimony trees of the (enshortened) mRNA sequences of your proteins.[1] (Species: human, chimp, dog, cow, mouse and opossum.) You can use the trees you calculated during the practicals or calculate new ones. Insert the original maximum parsimony trees and the consensus tree to the document as figures. Use readable font size on the figures. Show the consensus values on the consensus tree (how many (or %) maximum parsimony trees supported the branches). Discuss the the similarities and differences of trees.

4. Find two expression profiles on the GEO database that used HG-U133_Plus_2 Affimetrix platform (*platform ID* or *reporter*: GPL570). Describe these investigations, what treatment groups they used. Analyze these groups by GEO2R (differential expression analysis). Write down the adjusted *p*-values and logFCs of your genes. Are they significantly differentially expressed between the groups? Insert figures of the expression profiles of your genes.

5. Download the first neighbors of your proteins from two public databases using the

---

1 Since you cannot use UGENE for this you can do the followings: Export all trees from UGENE to Newick format. Copy these trees to a single file. Upload the file to http://bioinfo.nhri.org.tw/cgi-bin/emboss/fconsense. Calculate the consensus, download the outtreefile. Read it into the Figtree software and (after formatting) export the tree as a figure.

<u>Cytoscape</u> software. Merge the networks. Discard all non-human proteins and all subgraphs that don't connect to the greatest component (graph). Insert a figure in which the 3 proteins of yours are visually distinguishable from the others and from each other. Make their names readable (only for those three). Make the first neighbors and the common first neighbors of your proteins visually distinguishable too. Write a figure description for this.

<div align="right">21.11.2019</div>