Introduction to Bioinformatics (236523)

HW 4 – Spring 2021

**General instructions:**

• Deadline: 01/07/21 23:59.

• Submission in pairs only.

• The submission is via the course website.

**Detailed instruction for the HW:**

The answers to the questions 1-2 should be submitted as a .pdf file. All the relevant figures should be placed together with the text, except for the PDF figure from UCSC that should be attached separately. The answers to the question 3 should be submitted within the HW4\_TCGA.Rmd file that is provided together with the HW. You should submit al the files in a .ZIP format named according to next format: <HW#>\_<ID1>\_<ID2>.zip .

**Question 1:**

Combine 4 datasets from TCGA studies on cBioPortal:

Breast Invasive Carcinoma

Cervical Squamous Cell Carcinoma

Ovarian Serous Cystadenocarcinoma

Uterine Corpus Endometrial Carcinoma

Choose the appropriate source for the datasets, regarding the fact that you combine them to perform further comparison analysis.

1.1. Compare the 4 cancers by patients’ race. Deselect the samples with “NA” race. Go to clinical data and choose `cancer type` among attributes. Can you claim that Native Hawaiian or Other Pacific Islander race group is not predisposed to specific types of cancer analyzed here? Explain the result that you see for this this race group.

Add the 100% stacked barplot generated at this analysis to your answer. The plots can be easily downloaded from the page, find a way to do it.

1.2. Compare between overall survival statuses of the patients in the 4 cancers by tumor type. Which tumor type has the lowest median of overall survival time ?

1.3. Which gene is the gene with the highest frequency of copy number alterations in this combined study? What is its functional role in the cell?

1.4. Query the combined studies for the gene from 1.3. Perform a comparison/survival analysis for the gene versus unaltered group.

Relying on these results what can you say regarding progression free survival in the three cancers of patients with the altered gene versus unaltered group? Attach the progression-free survival plot here.

1.5. Browse each of the cancers separately and attach its progression free plot of altered gene vs unaltered group here. Relying on the results that you got, what can you say regarding the meaning of the plot you’ve got at 1.4 ?

1.6. What is the most abundant genetic alteration in this gene for all the three cancers analyzed? Attach a plot that shows it.

**Question 2:**

2.1 Look for let-7g in internet. What is let-7g? Search for relevant literature and shortly describe its functions and the disease/s it is known to be dysregulated at.

2.2. Search for let-7g in UCSC genome browser, hg38.

2.2.1. What gene is overlapping let-7g ?

2.2.3. Show conservation of the area at +- 1,000 bp from both sides of let-7g. In your conservation track you should show primates, mammals, and fish only. What can you say regarding the conservation of let-7g versus conservation of the area adjacent +-1000 bps to it ? Find a way to export all the browser view to PDF at (it can be found at UCSC FAQs) and attach the exported file to your HW submission.

2.3 Query the mir-34a, TP53 and AGO2 at the TCGA PanCancer studies in cBioPortal.

2.3.1. Which among mir-34a, TP53 and AGO2 tend to co-occur or to be mutually exclusive?

2.3.2. Search in the literature and find a study/s that support the tendency of the most significant pair. Cite at least one study.

2.3.3. Can you find in literature any connection between let-7 family and this pair? Cite if yes.

**Question 3:**

Go over the explanations in HW4\_TCGA.Rmd file and answer the questions in the file.