Final Project - Introduction to Bioinformatics

אלמוג

- Students' names and IDs: Ofer Nissim – 312367576

    Tom Smolin – 313552739

- Your disease of choice: HIV vs COVID-19

- The biological question you want to answer: We would like to explore the biological crosstalk between HIV and COVID-19. i.e., we want to examine whether one condition may affect the patient's prognosis, due to the second condition (e.g. finding bio-markers which are present in COVID-19 and HIV patients and which may indicate on worse COVID-19 prognosis among HIV patients)

- How can gene expression data help? By performing the gene expression analysis we'll be able to identify the differentially expressed genes in each condition. After doing that, by intersecting the results we'll have the common differentially expressed genes. At this point we'll have hopefully enough information to identify some bio-markers which may predict when the prognosis might be worse for patients with both conditions.

- What other type of analysis are you planning to use and how can it help? In the first analysis for finding the DEG, we plan to use methods such as "Deseq2" and "limma" (and use visualizations such as volcano plots and heatmaps). Afterwards, within the first analysis we'll also use the "clusterProfiler" package and perform "GSEA" as we want to find hallmarks pathways as well (which will help identify bio-markers).

    In the second analysis we'll have survival analysis (TCGA). But, as we don't have the mortality feature, and we do have the "severity" feature and the amount of time from symptoms onset, we'll instead label patients who reached the ICU as 1 (=imitating the "dead patients" - similarly to what was done in the tutorial) and the rest as 0 (hence performing ICU risk analysis).

- A table of the datasets (at least five) you collected with the following information:

Accession ID	Location	Data type	Count Matrix	Groups
GSE171110	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171110	Bulk RNA-Seq	Raw Count Matrix	44 COVID-19 subjects and 10 healthy controls
GSE152418	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152418	Bulk RNA-Seq	Raw Count Matrix	17 COVID-19 subjects and 17 healthy controls
GSE179850	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179850	Bulk RNA-Seq	Raw Count Matrix	31 COVID-19 subjects and 16 healthy controls
GSE37250	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37250	Bulk RNA-Seq	Raw Count Matrix	274 HIV+ subjects and 263 HIV- subjects
PRJEB34025	https://www.ebi.ac.uk/gxa/experiments/E-MTAB-8249/Experiment%20Design	Bulk RNA-Seq	Raw Count Matrix	22 HIV+ subjects and 15 HIV- subjects