CTS505 Project 1

Tobias Jakobi (tjakobi@arizona.edu)

December 9, 2025

The solutions for the tasks listed below should be saved in separate `.R` files, e.g. `task1.R`, `task2.R`. Please include loaded libraries, loading of files, and all other commands in the files.

Task 1.

(a) Download a list of gene IDs with expression values and false discovery rate (FDR) columns from `https://links.jakobilab.org/cts505_genes`.

(b) Identify the organism used for the experiment based on the provided gene IDs for the next tasks.

Task 2.

(a) Read the provided input file into R with an appropriate function.

(b) Use the `biomaRt` package to retrieve gene names and Uniprot IDs for the provided gene IDs using the identified organism.

(c) Print the top 10 highest expressed genes based on the logFC value into a new CSV file, `top10.csv`. Include the gene names and Uniprot IDs.

Task 3.

(a) Use the `ggplot2` package to create an X-Y graph that plots `log10(FDR value)` on the Y axis and the logFC value on the X axis.

(b) Set a title and descriptive X and Y axis labels and save the plot to PDF file named `xy_plot.pdf`.

Task 4.

- Download the list of gene counts stored in the following file
  `https://links.jakobilab.org/p2_counts`
- Identify the organism used for the experiment based on the provided gene IDs for the next tasks.

(a) Establish a reasonable grouping of the provided samples based on the sample names.

(b) Read the provided input file into R with an appropriate function.

(c) Use the `biomaRt` package or any other suitable ID conversion package to retrieve gene names for gene IDs.

Task 5.

(a) Develop a suitable design formula as discussed in the class to perform differential gene expression analysis between your chosen groups.

(b) Use the `edgeR` workflow to build an `DGElist` object.

(c) Using `edgeR`, perform filtering and normalization of the imported count data.

(d) Lastly, perform differential gene expression analysis and note the number of up- and down regulated genes.

Task 6.

(a) Save a result table with differentially expressed genes filtered by p-value and FDR (0.05) and sorted by logFC.

Please send your R scripts & plots as attachment via email to `tjakobi@arizona.edu`.

In case of questions please do not hesitate to contact me.