

# CTS505 Project 1

Tobias Jakobi (tjakobi@arizona.edu)

October 26, 2023

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](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`.

Please send your R scripts & plots as attachment via email to [tjakobi@arizona.edu](mailto:tjakobi@arizona.edu).

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