

RNA Seq Plot Easy Output

Jiarun Liu (A17384789) | Sicheng Jing (A16910209) | Zhijun Qian (A16909815)

Our tool will be based on a Python script with imports like NumPy, Panda, and Matplot. The tool will also utilize some R packages to handle complicated data analysis and return a corresponding file for the script to keep analyzing. Our tool will stand out for its user-friendliness, offering a single, convenient command line with a plethora of options to cater to our users' diverse data analysis needs. Its functions will encompass data quality checks, differential expression analysis(based on DESeq2), and data visualization. Notably, our tool will provide users with the flexibility to customize their data visualization based on their desired p-value or/and fold of change.

Under the worst-case scenario, our tool will undertake steps, including data quality checks, differential expression analysis(based on DESeq2), and data visualization. Therefore, we will compare the runtime of our tool against the runtime of DESeq2, tximport, and matplotlib combination.

We will use the gene expression results from replicates of mice fed by either a chow diet or a high-fat diet:

- Chow_Rep[1-3].genes.results & HFD_Rep[1-3].genes.results

A reference file (GRCm38):

- GRCm38.75.gene_names

All the above files come from lab4.

The second dataset is sourced from the study titled "**Global Transcriptome Analysis of Brown Adipose Tissue of Diet-Induced Obese Mice**," published with the DOI:

10.3390/ijms19041095. The data is accessible through the GEO accession number

GSE112740 and includes the following samples:

- GSM3082780: HFD BAT 1
- GSM3082781: HFD BAT 2
- GSM3082782: HFD BAT 3
- GSM3082783: LFD BAT 1
- GSM3082784: LFD BAT 2
- GSM3082785: LFD BAT 3

HFD : High fat diet, LFD: Low fat diet.