

PHS 675 Special Topics: Analysis of RNA-Sequencing Data

Homework 1: Differential Expression

For this homework, submit a Word file with a narrative, along with an R script file. This homework is due by 11:59pm on Thursday, September 3rd.

The narrative should contain a brief description of what was done, and the results, including any tables and graphs. If it helps, you can organize this as a "methods" section and a "results" section.

The R script should run as submitted without any modification and should generate the results and figures that are included in the narrative.

Try to avoid formatting this in a question and answer format. The narrative, while answering the instructions below, should stand alone and should make sense without reference to these instructions or the questions.

1. Download the count table from

https://denvirlab.marshall.edu/PHS675-2026/data/M_hf_20wk_th_vs_b6_counts.csv

This count table contains raw read counts from RNA-sequencing from the pancreas of 20-week old male mice fed a high-fat diet. There are two different strains in this data set: seven of the mice are C57BL/6 mice (the reference strain) and seven are TALLYHO mice (the experimental strain; these mice are insulin resistant at birth). The sample names for the C57BL/6 mice contain "B6"; the sample names for the TALLYHO mice contain "TH". The sample names in full denote the sex, strain, diet, age, id, and tissue of the sample.

2. Use the DESeq2 package to create a DESeqDataSet which we will use to compare gene expression by the mouse strain. You will need to create a "column data" table and specify an experimental design that specifies you are making comparisons by strain.
3. Perform the differential expression analysis.
 - a. Create an MA-plot and a PCA plot. Use these to determine if the experiment looks like it is producing data that would generally be expected.
 - b. Determine how many genes are upregulated (higher expression in TALLYHO mice than C57BL/6 mice) and downregulated (lower expression in TALLYHO mice than C57BL/6 mice) at a false discovery rate (adjusted p-value) of less than 0.1.
 - c. Create a volcano plot with all the genes which have more-than-minimal expression (i.e. those for which the adjusted p-value is not denoted as "NA"). Use colors to indicate the genes which have adjusted p-value less than 0.1 and a fold change of at least 2 (either up- or down-regulated)

- d. Create a table with the 20 differentially expressed genes (adjusted p-value less than 0.1) with the largest fold changes. Show the log2 fold change, adjusted p-value, and gene symbol in the table.
- e. Create a heatmap with the genes from the table. Use the colors in the heatmap to show the expression values after the variance-stabilizing transformation provided by DESeq2.