

Tutorial Assignment 2: *C. aerofaciens* DESeq2 analysis

Due: 10am (before class), 10/8

For this homework assignment, we will be using the same dataset and notebook as we used for the in-class tutorial. We will investigate differential gene expression in *C. aerofaciens*.

- Substitute "Cint.normalized" with "Caero.normalized" in the 3rd code block of the in-class tutorial to use it for the homework.
- For this homework, upload both a PDF of answers as well as a PDF of your completed .ipynb.

Questions:

1. How many genes are significantly upregulated in fructoselysine vs. glucose+lysine after FDR correction? (10 points)
2. How many of these are anticipated to be false positives? (10 points)
3. Describe the difference before and after log-fold change shrinkage in the MA plot. (10 points)
4. In PCA, which group is the most separated? Along which axis? (10 points)
5. In the heatmap, hierarchically cluster by columns in addition to rows by adjusting the "Colv" command. Does each sample type cluster with its replicates? Based on the observed sample clustering, which treatment group is separate from the other three? (20 points)

[Hint: add the following variable to read the entire x axis : margins=c(10,5)]
6. Which gene is changing most significantly (lowest adjusted p-value)? Which treatment is it most highly expressed in? (10 points)
7. Calculate and report the number of significantly changing genes between glucose vs. glucose+fructoselysine treatments. (20 points)