

# STAT718/BIOL703 Homework 6

**DUE: November 8, 2024**

**Total Points: 100**

## Part 1 DE Analysis

This homework is a continuation of Homework 5. Use the data downloaded from Homework 5 and continue the analysis. The `targets.txt` file describes all 18 FASTQ files and experimental conditions: three treatment conditions: Mock, Avr, and Vir and three time points: 1h, 6h, and 12h.

```
url <- "https://people.stat.sc.edu/hoyen/STAT718/Data/targets.txt"
sampleData <- read.delim(file = url, comment = "#")
```

Use the following design (design= ~ treat + time) and perform the following tasks:

- Identify DE genes between treatment conditions: Mock and Vir. (20 points)
- Perform Wald test as well as shrunken foldchange. Are the results similar? (20 points)
- Use Adjust p-value=0.05 as the cutoff and identify significant DE genes. (10 points)
- Create volcano plots using the unshrunken and shrunken estimates and the corresponding p values. Label significant DE genes labeled in different color than non-DE genes. (15 points)

## Part 2 Heatmap and Clustering

Use only the read counts from the DE genes and perform the following tasks:

- Generate a heatmap which clusters both samples and genes. (20 points)
- Use hierarchical clustering to cluster the sample and present a dendrogram. (15 points)

## Reference:

1. Howard BE, Hu Q, Babaoglu AC, Chandra M, Borghi M, Tan X, He L, Winter-Sederoff H, Gassmann W, Veronese P, Heber S. High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. PLoS One. 2013 Oct 1;8(10):e74183. doi: 10.1371/journal.pone.0074183. PMID: 24098335; PMCID: PMC3788074.