BIOS 426/826

HW5: Using edgeR and DESeq to identify differentially expressed genes

Due on Oct 12th, 11:59PM

Name:

1. Please download data file, “RNA\_seq\_data.txt”, from our class website. The data file has the numbers of raw reads in each gene from a maize experiment (six columns: three as controls and three for the treatment). It is a plain text file and you can open it with any text editor.

2. Install edgeR

> source("http://bioconductor.org/biocLite.R")

> biocLite("edgeR")

3. Learn using edgeR

Please modify the following codes to make it work.

> library(edgeR)

> library(stats) #edgeR needs this lib

> setwd("???") #<== set the work directory for data & output files

> y <- as.matrix(read.table(???)) #<= change “???” to correct parameters.

> mylibsizes <- ??? #calculate the total number of reads

> mygroups <- ??? #how do you group your samples?

> d<-DGEList(counts=y, lib.size= mylibsizes, group=mygroups, remove.zeros = TRUE)

> d<-estimateCommonDisp(d)

> d= estimateTagwiseDisp(d)

> ms<-exactTest(d)

> result=topTags(ms, n=???, adjust.method= "fdr") #how many genes do you want to output

More Questions:

* 1. How do you calculate the total number of reads? (Hints: you may use R command to calculate.)
  2. What are the top five genes that have the largest Fold Change in absolute values? What are their Fold Change and P-values?
  3. How can you save your results into a text file? Can you use MS Excel to open and display this file?

4. Plot a volcano plot from edgeR’s results

5. Learn using DESeq

Install DESeq first. Please use the same data set as the input, and make a script for DESeq to analyze the RNA-seq data, like Question 3 for edgeR. And answer the following additional quesionts:

5.1 Can you find the size factors for all six samples?

5.2 Can you display the dispersion of this experiment?

5.3 Can you show me the top 10 differentially expressed genes ranked by their p-values?

5.4 Can you save your all results in a text file? Can you open this file with MS Excel?

6. Plot a volcano plot from DESeq’s results

To submit your homework, please submit answers to all questions, source code with screen snapshots for each step, results of edgeR and DESeq in both and MS Excel format, and figures as image files formats, such as PDF.