BIOS 497/897-1

HW5: differential expression analysis with edgeR

Due on Feb. 24th, 11:59PM

1. Two data files are included in this package. One file, “RNA\_seq\_data.txt”, is data file for the numbers of raw reads in each gene from a maize experiment (six columns: three for control and three for treatment), and the other is the list of corresponding gene IDs. They are all plain text files.

2. Installing edgeR

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

> biocLite("edgeR")

3. Using edgeR

Please modify the following code to make it work.

> library(edgeR)

> library(stats) #edgeR needs this lib

> setwd("???") #<== set the directory

> set.seed(???) #<= give the random number seed here.

> y <- as.matrix(read.table(“???”)) #<= give file name here

> tags <- as.matrix(read.table("tags.txt"))

> rownames(y)=tags

> lib.sizes <- c(???,???,???,???,???,???) #calculate the total number of reads

> d<-DGEList(counts=y, group=factor(c(”Ctr",”Ctr",”Ctr",”Tr",”Tr”,”Tr”)), remove.zeros = TRUE)

> d<-estimateCommonDisp(d)

> ms<-exactTest(d)

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

Questions:

1. What do the following two R statements work for?

> tags <- as.matrix(read.table("tags.txt"))

> rownames(y)=tags

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?

4. To plot a volcano plot

To submit your homework, please submit answers to all questions, source code, results of edgeR (it is better to have a screen shot).