BIOS 426/826

HW5: Using edgeR and DESeq to identify differentially expressed genes

Due on Oct 12th, 11:59PM

3. Learn using edgeR

Please modify the following codes to make it work.

The R code for using edgeR

> library(edgeR)

> library(stats)

> setwd("/Users/chizhang/homework/HW5")

> y <- as.matrix(read.table("RNA\_seq\_data.txt", header=TRUE, row.names="Gene"))

> g mygroups =factor(c("c","c","c","t","t","t"))

> mylibsizes <- colSums(y)

> 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=5, adjust.method="fdr")

Questions:

* 1. How do you calculate the total number of reads? (Hints: you may use R command to calculate.)

 > mylibsizes <- colSums(y)

* 1. What are the top five genes that have the largest Fold Change in absolute values? What are their Fold Change and P-values?

 > result=topTags(ms, n=5, adjust.method="fdr", sort.by="logFC")

* 1. How can you save your results into a text file? Can you use MS Excel to open and display this file?

> write.csv(result$table, file="your\_file\_name.csv")

4. Plot a volcano plot from edgeR’s results

>plot(result$table$logFC, -log2(result$table$Pvalue), xlab="log(FC)", ylab="-log(P-value)")

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 questions:

The R code for using DESeq

> library(DESeq)

> setwd("/Users/chizhang/homework/HW5")

> y <- as.matrix(read.table("RNA\_seq\_data.txt", header=TRUE, row.names="Gene"))

> groups=c("c", "T")

> d<-newCountDataSet(y, groups)

> d<-estimateSizeFactors(d)

> d <- estimateDispersions (d)

> res <- nbinomTest (d, "c", "T")

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

 > sizeFactors(d)

5.2 Can you display the dispersion of this experiment?

 > plotDispEsts(d) # to visualize dispersion

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

 > head( res[order(res$pval),], n=10)

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

 > write.csv(res, file="your\_file\_name.csv")

6. Plot a volcano plot from DESeq’s results

>plot(res$log2FoldChange, -log2(res$pval), xlab="log(FC)", ylab="-log(P-value)")