**Giovanni Coppola Laboratory, UCLA**

**This analysis is performed with software R version 3.4.2**

<https://www.r-project.org/>

**##load packages**

library("gplots")

library("heatmap.plus")

library("MASS")

library(edgeR)

library(DESeq)

library(limma)

library("RColorBrewer")

#To install bioconductor packages

source("https://bioconductor.org/biocLite.R")

biocLite("limma")

#To install R packages

install.packages("gplots")

**##set work directory**

setwd("F:/STUDY/2017-9162\_SwapnaJoshi\_quanSeq/DEA\_edgeR")

**##read in count**

rawCount<-read.csv("metaReadCount\_subset.csv")

dim(rawCount)

rownames(rawCount)<- rawCount$gene\_id

annot<- rawCount[,1:5]

cc<- rawCount[,6:138]

**##read in phenodata**

targets<-read.csv("targets.csv")

labls<-paste(targets$SampleName, targets$Dx, targets$Gender, targets$BH\_Colon\_Exam, sep="\_")

colnames(cc)<-labls

**##create DGEList**

y<- DGEList(counts=cc, genes=annot)

**# filter out low expressed genes**

keep <- rowSums(cpm(y)>1) >= 36

y <- y[keep, keep.lib.sizes=FALSE]

dim(y)

**## plot density with logCPM**

dx<-ifelse(targets$Dx =="HC","orange","cyan")

col<-dx

pdf("1\_density\_0.5cpm.pdf",width=14,height=7)

nsamples <- ncol(y)

par(mfrow=c(1,2))

lcpm0 <- cpm(y0, log=TRUE)

plot(density(lcpm0[,1]), col=col[1], lwd=2, ylim=c(0,2), las=2,

main="", xlab="")

title(main=" Raw data", xlab="Log-cpm")

abline(v=0, lty=3)

for (i in 2:nsamples){

den <- density(lcpm0[,i])

lines(den$x, den$y, col=col[i], lwd=2)

}

legend(x="topright",legend=unique(targets$Dx),col=unique(dx),fill= unique(dx), cex=1)

lcpm <- cpm(y, log=TRUE)

plot(density(lcpm[,1]), col=col[1], lwd=2, ylim=c(0,0.5), las=2,

main="", xlab="")

title(main=" Filtered data", xlab="Log-cpm")

abline(v=0, lty=3)

for (i in 2:nsamples){

den <- density(lcpm[,i])

lines(den$x, den$y, col=col[i], lwd=2)

}

legend(x="topright",legend=unique(targets$Dx),col=unique(dx),fill= unique(dx), cex=1)dev.off()

**##boxplot of expression with logCPM**

x2<-y

x2$samples$norm.factors <- 1

yy <- calcNormFactors(y, method="TMM")

pdf("2\_boxplot.pdf",width=15,height=8)

par(mfrow=c(1,2))

par(cex.axis=0.7)

lcpm <- cpm(x2, log=TRUE)

boxplot(lcpm, las=2, col= dx, main="")

title(main="Filtered and Unnormalized data",ylab="Log-cpm")

legend(x="topright",legend=unique(targets$Dx),col=unique(dx),fill= unique(dx), cex=1)

lcpmn <- cpm(yy, log=TRUE)

par(cex.axis=0.7)

boxplot(lcpmn, las=2, col= dx, main="")

title(main=" Filtered and Normalized data",ylab="Log-cpm")

legend(x="topright",legend=unique(targets$Dx),col=unique(dx),fill= unique(dx), cex=1)

dev.off()

**##design matrix**

TS<- targets$Dx

lane<-as.factor(targets$Lane)

TS <- factor(TS, levels=unique(TS))

design <- model.matrix(~0+TS+lane)

my.contrasts<- makeContrasts( IBS\_vs\_HC=TSIBS-TSHC, levels=design)

**##mean-variance trend**

pdf("2\_voom\_qn.pdf", height=10, width=10)

v <- voom(y,design,plot=TRUE, normalize="quantile")

dev.off()

**##normalize counts**

yy <- calcNormFactors(y, method="TMM")

**##estimate common dispersion**

yy <- estimateGLMCommonDisp(yy,design, verbose=TRUE)

**##estimate trend dispersion**

yy <- estimateGLMTrendedDisp(yy,design)

**##estimate tagwise dispersion**

yy <- estimateGLMTagwiseDisp(yy, design)

**##fit model**

fit <- glmFit(yy,design)

**##contrast**

##contrast 1

lrt1 <- glmLRT(fit, contrast=my.contrasts[,"IBS\_vs\_HC"])

lrt1$table <- cbind(lrt1$table, FDR=p.adjust(lrt1$table$PValue,method="BH"))

names(lrt1$table) <- paste(names(lrt1$table),"IBS\_vs\_HC", sep="\_")

lrt1.out<-lrt1$table

#

lrtTot <- as.data.frame(cbind(y$genes,lrt1.out, yy$counts))

head(lrtTot)

dim(lrtTot)

**# Ratio for heatmap**

all.samples<-as.data.frame(lcpmn)

#1. Controls,

hc<-all.samples[,targets$Dx== "HC" ]

hcM<-rowMeans(hc)

#2. Exp,

ibs<-all.samples[,targets$Dx== "IBS" ]

#3. Ratios

ibshc<- ibs - hcM

colnames(ibshc)<-paste(colnames(ibshc), "\_vs\_HC" ,sep="")

**##arrange output**

colnames(lcpmn)<-labls

colnames(lcpmn)<-paste(colnames(lcpmn), ".logCPM", sep="")

#exporting all the data

ratio.exp <- as.data.frame(cbind(yy$genes, ibshc , yy$counts, lcpmn ))

**#output complete gene list**

Complete<-merge(lrtTot,ratio.exp,by.x="gene\_id",by.y="gene\_id")

write.csv(Complete, file= "Complete\_geneList.csv")

**##significant list**

Sig\_fdr0.05<- Complete[Complete$FDR\_IBS\_vs\_HC <0.05,]

dim(Sig\_fdr0.05)

write.csv(Sig\_fdr0.05, file= "Significant\_geneList\_FDR0.05.csv")

Sig\_fdr0.1<- Complete[ Complete$FDR\_IBS\_vs\_HC <0.1,]

dim(Sig\_fdr0.1)

write.csv(Sig\_fdr0.1, file= "Significant\_geneList\_FDR0.1.csv")

Sig\_p0.005<- Complete[Complete$PValue\_IBS\_vs\_HC <0.005,]

dim(Sig\_p0.005)

write.csv(Sig\_p0.005, file= "Significant\_geneList\_pLess0.005.csv")

**###plot heapmap with significant genes**

toplot<- Sig\_p0.005 [,11:107]

rownames(toplot)<- Sig\_p0.005$gene\_name

pdf("5\_heatmap\_p0.005.pdf", height=10, width=10)

heatmap.2(as.matrix(toplot), col=rev(redgreen(68)), main="IBS vs. HC, p<0.005", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.5, cexRow=0.7, keysize=1, )

dev.off()

**##p-value distribution**

pdf("2-pVal\_histogram.pdf", height=8,width=10)

hist(as.matrix(Complete$PValue\_IBS\_vs\_HC), breaks=100, col="lightblue", main="PValue\_IBS\_vs\_HC ", xlab="")

dev.off()

pdf("3\_FDR\_histogram.pdf", height=8,width=10)

hist(as.matrix(Complete$FDR\_IBS\_vs\_HC), breaks=100, col="lightblue", main="PValue\_IBS\_vs\_HC ", xlab="")

dev.off()

**# barplot for number of sig genes**

pdf("1\_contrastAnalysis\_p0.005.pdf", height=10, width=12)

par(mar=c(3,3.5,2.5,3.5))

SigGene05 <- Sig\_p0.005

padj=0.005

FDR<- Sig\_p0.005 [,7]

logFC<- Sig\_p0.005[,6]

lfc <- Sig\_p0.005 [,c(6:7)]

head(lfc)

dim(lfc)

ups <- NA

downs <- NA

for(i in 1:1)

{

downs <- c(downs, -length(which(lfc[which(lfc[i+1] <= 0.005),i] < 0)))

ups <- c(ups, length(which(lfc[which(lfc[i+1] <= 0.005),i]>0)))

}

print(ups)

print(downs)

mx <- max(ups[is.na(ups)==FALSE])

mn <- min(downs[is.na(downs)==FALSE])

bp1 <- barplot(downs,horiz=TRUE,xlim=c(mn,mx), col="green", )

bp2 <- barplot(ups, horiz=TRUE,xlim=c(mn,mx), col="red",add=TRUE,axes=TRUE)

axis(2, at=bp1[2:length(bp1)],tick=FALSE,labels=downs[is.na(downs)==FALSE],las=1)

axis(4,at=bp2[2:length(bp2)],tick=FALSE,labels=ups[is.na(ups)==FALSE], las=1)

#axis(4,at=bp2[2:length(bp2)],tick=FALSE,labels=ups[is.na(ups)==FALSE], las=1,line=-2)

labs <- NA

fullNames <- "IBS vs. HC"

for(tis in fullNames){

labs <- c(labs, tis)

}

labs <- labs[is.na(labs)==FALSE]

text(x=0,y=bp2[2:length(bp2)], labels=labs)

title(main=paste("Gene Changes, p <",padj, "(",dim(SigGene05)[[1]], ")", sep=" "))

dev.off()

**#R sessionInfo()**

R version 3.4.2 (2017-09-28)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 7 x64 (build 7601) Service Pack 1

Matrix products: default

locale:

[1] LC\_COLLATE=English\_United States.1252 LC\_CTYPE=English\_United States.1252

[3] LC\_MONETARY=English\_United States.1252 LC\_NUMERIC=C

[5] LC\_TIME=English\_United States.1252

attached base packages:

[1] parallel stats graphics grDevices utils datasets methods base

other attached packages:

[1] RColorBrewer\_1.1-2 DESeq\_1.28.0 lattice\_0.20-35 locfit\_1.5-9.1

[5] Biobase\_2.36.2 BiocGenerics\_0.22.1 edgeR\_3.18.1 limma\_3.32.10

[9] MASS\_7.3-47 heatmap.plus\_1.3 gplots\_3.0.1

loaded via a namespace (and not attached):

[1] Rcpp\_0.12.14 compiler\_3.4.2 pillar\_1.0.1

[4] bitops\_1.0-6 digest\_0.6.13 bit\_1.1-12

[7] annotate\_1.54.0 RSQLite\_2.0 memoise\_1.1.0

[10] tibble\_1.4.1 rlang\_0.1.6 Matrix\_1.2-11

[13] DBI\_0.7 genefilter\_1.58.1 S4Vectors\_0.14.7

[16] gtools\_3.5.0 caTools\_1.17.1 IRanges\_2.10.5

[19] stats4\_3.4.2 bit64\_0.9-7 grid\_3.4.2

[22] AnnotationDbi\_1.38.2 XML\_3.98-1.9 survival\_2.41-3

[25] gdata\_2.18.0 geneplotter\_1.54.0 blob\_1.1.0

[28] splines\_3.4.2 xtable\_1.8-2 KernSmooth\_2.23-15

[31] RCurl\_1.95-4.10