**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, for example,

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

biocLite("limma")

#To install R packages, for example,

install.packages("gplots")

**##set work directory**

setwd("F:/STUDY/SwapnaJoshi\_quanSeq/DEA\_voom")

**##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)

**##normalize counts**

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

**##design matrix**

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

lane<-as.factor(targets$Lane)

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

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

colnames(design)[1:8]<-c( "HC\_F\_N", "IBS\_F\_M", "IBS\_F\_D", "HC\_M\_N", "IBS\_M\_D", "IBS\_M\_M", "IBS\_F\_C", "IBS\_M\_C")

my.contrasts<- makeContrasts( HC.M.N\_vs\_HC.F.N = HC\_M\_N - HC\_F\_N,

IBS.M.C\_vs\_IBS.F.C = IBS\_M\_C - IBS\_F\_C,

IBS.M.D\_vs\_IBS.F.D = IBS\_M\_D - IBS\_F\_D,

IBS.M.M\_vs\_IBS.F.M = IBS\_M\_M - IBS\_F\_M,

IBS.M\_vs\_IBS.F = (IBS\_M\_M+ IBS\_M\_D+ IBS\_M\_C)/3-(IBS\_F\_M+ IBS\_F\_D+ IBS\_F\_C)/3,

IBS.M\_vs\_HC.M=(IBS\_M\_M+ IBS\_M\_D+ IBS\_M\_C)/3 - HC\_M\_N,

IBS.F\_vs\_HC.F=(IBS\_F\_M+ IBS\_F\_D+ IBS\_F\_C)/3 - HC\_F\_N,

IBS\_vs\_HC=( IBS\_M\_M+ IBS\_M\_D+ IBS\_M\_C+ IBS\_F\_M+ IBS\_F\_D+ IBS\_F\_C)/6 - (HC\_M\_N + HC\_F\_N)/2,

IBS.M.C\_vs\_HC.M.N = IBS\_M\_C - HC\_M\_N,

IBS.F.C\_vs\_HC.F.N = IBS\_F\_C - HC\_F\_N,

IBS.M.D\_vs\_HC.M.N = IBS\_M\_D - HC\_M\_N,

IBS.F.D\_vs\_HC.F.N = IBS\_F\_D - HC\_F\_N,

IBS.M.M\_vs\_HC.M.N = IBS\_M\_M - HC\_M\_N,

IBS.F.M\_vs\_HC.F.N = IBS\_F\_M - HC\_F\_N,

levels=design)

**##model fittiing**

vm <- voom(yy,design,plot=TRUE, normalize="quantile")

vfit<- lmFit(vm, design)

fit2.anova<- contrasts.fit(vfit, my.contrasts)

# Empirical Bayes

fitb<- eBayes(fit2.anova)

#selecting the statistical cutoff

decide <- matrix(c("fdr",0.05, "fdr",0.1,"none",0.001,"none",0.005, "none", 0.01),nrow=5,ncol=2,byr=T)

# initialize:

mysum <- as.list(1:nrow(decide))

mynum <- 0

maxmax <- 0

for (test in 1:nrow(decide)){

results<-decideTests(fitb, adjust.method=decide[test,1],p=as.numeric(decide[test,2]))

summary(results) -> mysum[[test]]

mynum[test] <-length(which(apply(results,1,function(x)any(x,na.rm=T))))

maxmax <- max(c(maxmax, as.vector(mysum[[test]][c(1,3),])))

}

pdf("5\_threshold\_selection.pdf", width=13,height=6)

par(mfrow=c(1,nrow(decide)))

for (test in 1:nrow(decide))

{

as.numeric(as.vector(mysum[[test]][3,]))->plotMe1

as.numeric(as.vector(mysum[[test]][1,]))->plotMe2

maxData = max(plotMe1)

maxData2 = max(plotMe2)

barplot(plotMe1,horiz=T,col="red",xlim=c(-maxmax,maxmax),

main=paste("Gene Changes \np<",decide[test,2], ", " , decide[test,1],

" (" ,mynum[test] ,")",sep=""))->yy

barplot(-plotMe2,horiz=T,col="green",add=T)->yy

xx<-vector("integer",ncol(mysum[[test]]))

text(xx,yy,colnames(mysum[[test]]))

text((plotMe1+10)\*0 + .9\*maxData,yy+0.1,format(plotMe1,digits=3))

text((-plotMe2-10)\*0 - .9\*maxData2,yy+0.1,format(plotMe2,digits=3))

}

dev.off()

#select the chosen paramaters

chosen.adjust<-"fdr"

chosen.p<-0.05

current.contrast<-"contrast"

results<-decideTests(fitb,adjust.method=chosen.adjust,p=as.numeric(chosen.p))

summary(results)

summary(results)->mysum05

mysum05\_no<-length(which(apply(results,1,function(x)any(x,na.rm=T))))

maxmax<-max(as.vector(mysum05[c(1,3),]))

pdf("5\_selected\_contrast\_fdr0.05.pdf", width=10,height=10)

as.numeric(as.vector(mysum05[3,]))->plotMe1

as.numeric(as.vector(mysum05[1,]))->plotMe2

maxData = max(plotMe1)

maxData2 = max(plotMe2)

barplot(plotMe1,horiz=T,col="red",xlim=c(-maxmax,maxmax), main=paste("Gene Changes, fdr p<0.05 (",mysum05\_no,")",sep=""))->yy

barplot(-plotMe2,horiz=T,col="green",add=T)->yy

xx<-vector("integer",ncol(mysum05))

text(xx,yy,colnames(mysum05))

text((plotMe1+10)\*0 + .9\*maxmax,yy,format(plotMe1,digits=3))

text((-plotMe2-10)\*0 - .9\*maxmax,yy,format(plotMe2,digits=3))

dev.off()

##write out fit object

write.fit(fitb,file="dummy5.xls",digits=30,adjust=chosen.adjust,results=results)

treat.de<-read.delim(file="dummy5.xls",head=T)

dim(treat.de)

#output for contrasts

myNames<-names(treat.de)

res.col<- which(regexpr("Res.",myNames)>0)

anovalist<- which(apply(treat.de[,res.col],1,function(x)any(x,na.rm=T)))

length(anovalist)

treat.de.anova<-treat.de[anovalist,]

dim(treat.de.anova )

##significant gene list

fitsel.ratio<-merge(treat.de.anova, ratio.exp, by.x="Genes.gene\_id", by.y="gene\_id",all.x=T)

myNames <-names(fitsel.ratio)

#selects the relevant columns for output

res.col<- which(regexpr("Res.",myNames)>0)

coefs.col <- which(regexpr("Coef.",myNames)>0)

ts.col<- coefs.col+length(coefs.col)

pvals.col <- which(regexpr("p.value.",myNames)>0)

fitsel.ratio2<-cbind(fitsel.ratio[,c(1, 75:78)],

fitsel.ratio[,coefs.col],

fitsel.ratio[,pvals.col],

F=fitsel.ratio$F,

F.p.value=fitsel.ratio$F.p.value,

fitsel.ratio[,res.col],

fitsel.ratio[,ts.col],

AverageExpr= fitsel.ratio$A,

fitsel.ratio[,c(85:860)])

dim(fitsel.ratio2)

fitsel.ratio2<-fitsel.ratio2[order(fitsel.ratio2$F,decreasing=T),]

write.csv(fitsel.ratio2,file="significant\_geneList\_FDR0.05.csv")

#complete list of genes

fitsel.ratioAll<-merge(treat.de,ratio.exp, by.x="Genes.gene\_id", by.y="gene\_id")

myNames <-names(fitsel.ratio)

#selects the relevant columns for output

res.col<- which(regexpr("Res.",myNames)>0)

coefs.col <- which(regexpr("Coef.",myNames)>0)

ts.col<- coefs.col+length(coefs.col)

pvals.col <- which(regexpr("p.value.",myNames)>0)

fitsel.ratioN<-cbind(fitsel.ratioAll[,c(1, 75:78)],

fitsel.ratioAll[,coefs.col],

fitsel.ratioAll[,pvals.col],

F=fitsel.ratioAll$F,

F.p.value=fitsel.ratioAll$F.p.value,

fitsel.ratioAll[,res.col],

fitsel.ratioAll[,ts.col],

AverageExpr= fitsel.ratioAll$A,

fitsel.ratioAll[,c(85:860)])

fitsel.ratioN<-fitsel.ratioN[order(fitsel.ratioN$F,decreasing=T),]

write.csv(fitsel.ratioN,file="complete\_geneList.csv")

##p-value plot

pVal <- fitsel.ratioN [ , grep("p.value", colnames(fitsel.ratioN))]

pVal<- pVal[,1:14]

FDR <- fitsel.ratioN [ , grep("p.value.adj", colnames(fitsel.ratioN))]

pdf("2-pVal\_histogram.pdf", height=8,width=12)

par(mfrow=c(4,4))

for(i in 1:ncol(pVal)){

hist(as.matrix(pVal[i]), breaks=100, col=rainbow(ncol(pVal))[i], main=names(pVal[i]))}

dev.off()

#

pdf("3\_FDR\_histogram.pdf", height=8,width=12)

par(mfrow=c(4,4))

for(i in 1:ncol(FDR)){

hist(as.matrix(FDR[i]), breaks=100, col=rainbow(ncol(FDR))[i], main=names(FDR[i]))}

dev.off()

#heatmap

toplot<-fitsel.ratio2[,79:455]

rownames(toplot)<-fitsel.ratio2$gene\_name

toplotf5a<- toplot [,1:16]

toplotf5b<- toplot [,17:51]

toplotf5c<- toplot [,52:86]

toplotf5d<- toplot [,87:183]

toplotf5e<- toplot [,184:280]

toplot5g<- toplot [,281:311]

toplot5h<- toplot [,312:346]

toplot5j<- toplot [,347:377]

pdf("7\_anova\_heatmap\_fdr0.05.pdf", height=10,width=10)

heatmap.2(as.matrix(toplotf5a), col=rev(redgreen(68)), main="HC, Male vs. Femal, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.7, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5b), col=rev(redgreen(68)), main="IBS, Male vs. Female within each Exam, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), ColSideColors=colu1, cexCol=0.5, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5c), col=rev(redgreen(68)), main="IBS, Male vs. Female, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.7, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5d), col=rev(redgreen(68)), main="IBS vs. HC, within each Gender, fdr p<0.05",trace="none", breaks=breaks, margins=c(6,5), ColSideColors=colu2, cexCol=0.5, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5e), col=rev(redgreen(68)), main="IBS vs. HC, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.4, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5g), col=rev(redgreen(68)), main="IBS ExamC vs. HC Female ExamN, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), ColSideColors=colu3, cexCol=0.7, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5h), col=rev(redgreen(68)), main="IBS ExamD vs. HC ExamN, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.7, ColSideColors=colu4, cexRow=0.7, keysize=1, labRow = "")

heatmap.2(as.matrix(toplotf5j), col=rev(redgreen(68)), main="IBS ExamM vs. HC ExamN, fdr p<0.05", trace="none", breaks=breaks, margins=c(6,5), cexCol=0.7, ColSideColors=colu5, cexRow=0.7, keysize=1, labRow = "")

dev.off()

#venn

sumV <- apply(summary(results),2,function(x) x[1]+x[3])

v <- paste(names(sumV), " (",sumV, ")",sep="")

pdf("6\_venn\_fdr0.05.pdf", width=20,height=10)

par(mfrow=c(1,2))

vennDiagram(results[,c(1,5)],names=c(v[1],v[5]),main="FDR p<0.05", include=c("up","down"),counts.col=c(2,3), cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(1,5)],names=c(v[1],v[5]),main="FDR p<0.05", cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(2,3,4)],names=c(v[2],v[3],v[4]),main="FDR p<0.05", include=c("up","down"),counts.col=c(2,3), cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(2,3,4)],names=c(v[2],v[3],v[4]),main="FDR p<0.05", cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(6,7,8)],names=c(v[6],v[7],v[8]),main="FDR p<0.05", include=c("up","down"),counts.col=c(2,3), cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(6,7,8)],names=c(v[6],v[7],v[8]),main="FDR p<0.05", cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(9,11,13)],names=c(v[9],v[11],v[13]),main="FDR p<0.05", include=c("up","down"),counts.col=c(2,3), cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(9,11,13)],names=c(v[9],v[11],v[13]),main="FDR p<0.05", cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(10,12,14)],names=c(v[10],v[12],v[14]),main="FDR p<0.05", include=c("up","down"),counts.col=c(2,3), cex=1.2, circle.col=c("magenta","blue","cyan"))

vennDiagram(results[,c(10,12,14)],names=c(v[10],v[12],v[14]),main="FDR p<0.05", cex=1.2, circle.col=c("magenta","blue","cyan"))

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