June W2

Awais Choudhry

2017-6-16

Over the course of this month i focused heavily on my RHEP project hence i was writing less code for my honours project and conducting less investigations

## 16//6/2017

## Investigating the microarray data of Quille

# Unable to generate script analyzing differential expression.  
# Invalid input: at least two groups of samples should be selected.  
  
################################################################  
# Boxplot for selected GEO samples  
library(Biobase)  
library(GEOquery)  
  
# load series and platform data from GEO  
  
gset <- getGEO("GSE12609", GSEMatrix =TRUE, getGPL=FALSE)  
if (length(gset) > 1) idx <- grep("GPL1261", attr(gset, "names")) else idx <- 1  
gset <- gset[[idx]]

library(limma)  
  
gset <- getGEO("GSE12609", GSEMatrix =TRUE, getGPL=FALSE)  
if (length(gset) > 1) idx <- grep("GPL1261", attr(gset, "names")) else idx <- 1  
gset <- gset[[idx]]  
  
   
 f <- rbind(c(rep("control", 4), rep("knockout", 4)))  
 f <- as.factor(f)  
 design <- model.matrix(~0+f)  
   
   
 colnames(design) <- c("control", "knockout")  
 rownames(design)<-c("GSM315884", "GSM315886", "GSM315888", "GSM315891",  
 "GSM315890", "GSM315889", "GSM315887", "GSM315885")  
   
  
   
 fit<- lmFit(gset, design)  
   
 names(fit)  
   
 cont.matrix<- makeContrasts(KO\_vs\_control = knockout - control, levels = colnames(design))  
 fit2 <- contrasts.fit(fit, cont.matrix)  
 fit2 <- eBayes(fit2)  
 colnames(fit2)  
 topTable(fit2,coef="KO\_vs\_control")  
 topTable(fit2,coef="KO\_vs\_control",adjust="fdr", number = nrow(fit2))  
   
   
 ## MDS plots and stuff  
plotMDS(exprs(gset), main="MDS plot")  
  
# fit2<-rownames\_to\_column(as.data.frame(fit2))  
# Glimma::glMDPlot(fit2, coef="KO\_vs\_control",  
# counts = exprs(gset), transform=TRUE)

Examining the MDS plot we can see 0 clustering of our samples. I am worried i did something wrong hence why i do not see any clustering. Therefore, i am going to leave this for the moment.

## Examining Fulp et als data.

getGEOSuppFiles("GSE12609", makeDirectory = TRUE, baseDir = setwd("~/DataFiles"))  
  
## Load packages  
library(affy) # Affymetrix pre-processing  
library(limma) # two-color pre-processing; differential  
 # expression  
   
## import "phenotype" data, describing the experimental design  
phenoData <- gset@phenoData  
  
## RMA normalization  
eset <- justRMA(phenoData = phenoData)  
  
## differential expression  
   
 f <- rbind(c(rep("control", 4), rep("knockout", 4)))  
 f <- as.factor(f)  
 design <- model.matrix(~0+f)  
   
   
 colnames(design) <- c("control", "knockout")  
 rownames(design)<-c("GSM315884", "GSM315886", "GSM315888", "GSM315891",  
 "GSM315890", "GSM315889", "GSM315887", "GSM315885")  
fit <- lmFit(eset, design) # fit each probeset to model  
 cont.matrix<- makeContrasts(KO\_vs\_control = knockout - control, levels = colnames(design))  
 fit2 <- contrasts.fit(fit, cont.matrix)  
 efit <- eBayes(fit2) # empirical Bayes adjustment  
 topTable(efit, coef=1, adjust="fdr", number = 4500) %>% View() # table of differentially expressed probesets

With this dataset again, we see no-differenital expression.

## Analysisizng Fulpps data using the limma package

## Load packages  
library(affy) # Affymetrix pre-processing  
library(limma) # two-color pre-processing; differential  
 # expression  
   
## import "phenotype" data, describing the experimental design  
phenoData <- gset@phenoData  
  
## RMA normalization  
celfiles <- system.file("extdata", package="arrays")  
eset <- justRMA(phenoData=phenoData,  
 celfile.path="~/GSE12609/")  
  
## Warning: replacing previous import 'AnnotationDbi::tail' by 'utils::tail' when  
## loading 'hgfocuscdf'  
  
## Warning: replacing previous import 'AnnotationDbi::head' by 'utils::head' when  
## loading 'hgfocuscdf'  
  
##   
  
## differential expression  
combn <- factor(paste(pData(phenoData)[,1],  
 pData(phenoData)[,2], sep = "\_"))  
design <- model.matrix(~combn) # describe model to be fit  
  
c  
  
fit <- lmFit(eset, design) # fit each probeset to model  
efit <- eBayes(fit) # empirical Bayes adjustment  
topTable(efit, coef=2) # table of differentially expressed probesets

Again, we do not see any changes in gene expression in Arx mutations.

I am now pretty confident that i am missing something as i am following the limma pipeline in from the R package however, i am not see any results for any of these 3 datasets. I will have to ask Jimmy/Steve as to whats happening. I believe either i am missing a normalization step somehwere or this data is not particulary good. Likely the first.

## Trying to download the raw cell files and use them to read in the data.

library(affy)  
celfileNames = list.celfiles('~/GSE12609/')  
brainBatch=ReadAffy(filenames=celfileNames,celfile.path='~/GSE12609/',compress=FALSE)