Pasilla Analysis in EdgeR

Duncan Brian

Last modified: 30 July, 2015

Load edgeR and set working directory

library(edgeR)  
setwd("/home/participant/Course\_Materials")

Load the count file

datafile = system.file( "extdata/pasilla\_gene\_counts.tsv", package="pasilla" )  
  
datafile

## [1] "/home/participant/R/x86\_64-pc-linux-gnu-library/3.2/pasilla/extdata/pasilla\_gene\_counts.tsv"

## [1] "/home/participant/R/x86\_64-pc-linux-gnu-library/3.2/pasilla/extdata/pasilla\_gene\_counts.tsv"

Read the count table and print out

pasillaCountTable = read.table(datafile, header=TRUE, row.names=1)  
  
head(pasillaCountTable)

## untreated1 untreated2 untreated3 untreated4 treated1 treated2  
## FBgn0000003 0 0 0 0 0 0  
## FBgn0000008 92 161 76 70 140 88  
## FBgn0000014 5 1 0 0 4 0  
## FBgn0000015 0 2 1 2 1 0  
## FBgn0000017 4664 8714 3564 3150 6205 3072  
## FBgn0000018 583 761 245 310 722 299  
## treated3  
## FBgn0000003 1  
## FBgn0000008 70  
## FBgn0000014 0  
## FBgn0000015 0  
## FBgn0000017 3334  
## FBgn0000018 308

Construct the design and setup the conditions. Report how many samples were in the treated and untreated groups

pasillaDesign = data.frame(  
 row.names = colnames( pasillaCountTable ),  
 condition = c( "untreated", "untreated", "untreated",  
 "untreated", "treated", "treated", "treated" ),  
 libType = c( "single-end", "single-end", "paired-end",  
 "paired-end", "single-end", "paired-end", "paired-end" ) )  
  
pasillaDesign

## condition libType  
## untreated1 untreated single-end  
## untreated2 untreated single-end  
## untreated3 untreated paired-end  
## untreated4 untreated paired-end  
## treated1 treated single-end  
## treated2 treated paired-end  
## treated3 treated paired-end

Normalise

pairedSamples = pasillaDesign$libType == "paired-end"  
countTable = pasillaCountTable[ , pairedSamples ]  
condition = pasillaDesign$condition[ pairedSamples ]

Do the differential expression test and print the table of top hits

Use the results of the previous code-chunk to comment on how many up and down regulated genes were found at a p-value cut-off of 0.05. [Embed your answers in this section of text quoting the p-value cut-off and number of genes]

Produce the smear plot, but hide the code