R Notebook

Load the libraries

library(clusterProfiler)

##

## clusterProfiler v4.12.0 For help: https://yulab-smu.top/biomedical-knowledge-mining-book/  
##   
## If you use clusterProfiler in published research, please cite:  
## T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021, 2(3):100141

##   
## Attaching package: 'clusterProfiler'

## The following object is masked from 'package:stats':  
##   
## filter

library(org.Hs.eg.db)

## Loading required package: AnnotationDbi

## Loading required package: stats4

## Loading required package: BiocGenerics

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
## get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
## Position, rank, rbind, Reduce, rownames, sapply, setdiff, table,  
## tapply, union, unique, unsplit, which.max, which.min

## Loading required package: Biobase

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: IRanges

## Loading required package: S4Vectors

##   
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:clusterProfiler':  
##   
## rename

## The following object is masked from 'package:utils':  
##   
## findMatches

## The following objects are masked from 'package:base':  
##   
## expand.grid, I, unname

##   
## Attaching package: 'IRanges'

## The following object is masked from 'package:clusterProfiler':  
##   
## slice

##   
## Attaching package: 'AnnotationDbi'

## The following object is masked from 'package:clusterProfiler':  
##   
## select

##

library(edgeR)

## Loading required package: limma

##   
## Attaching package: 'limma'

## The following object is masked from 'package:BiocGenerics':  
##   
## plotMA

counts <- read.csv("Group1.csv", row.names = 1)

Set the factors

treatment\_pattern <- (rep(c("Untreated", "Treatment"), times = 9))  
treatment <- factor(c(treatment\_pattern))  
  
donor\_pattern <- c(rep("M24", times = 6), rep("M31", times = 6), rep("M32", times = 6))  
donors <- factor(c(donor\_pattern))

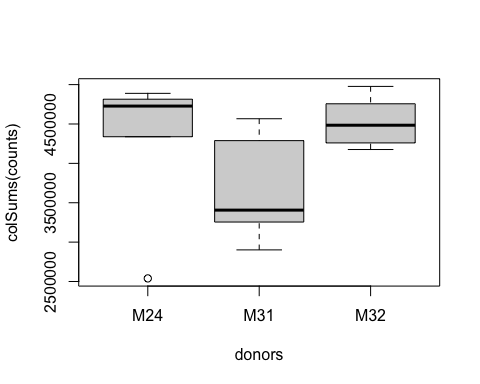
Cbind the factors

groups <- data.frame(sample = colnames(counts), treatment, donors)  
  
cbind(groups,colnames(counts))

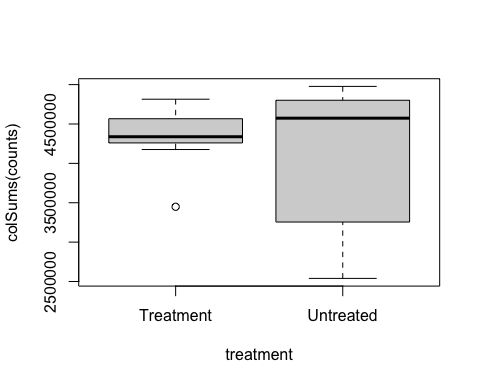
## sample treatment donors colnames(counts)  
## 1 CF\_M24\_EV1 Untreated M24 CF\_M24\_EV1  
## 2 CF\_M24\_EVPA1 Treatment M24 CF\_M24\_EVPA1  
## 3 CF\_M24\_EV2 Untreated M24 CF\_M24\_EV2  
## 4 CF\_M24\_EVPA2 Treatment M24 CF\_M24\_EVPA2  
## 5 CF\_M24\_EV3 Untreated M24 CF\_M24\_EV3  
## 6 CF\_M24\_EVPA3 Treatment M24 CF\_M24\_EVPA3  
## 7 CF\_M31\_EV1 Untreated M31 CF\_M31\_EV1  
## 8 CF\_M31\_EVPA1 Treatment M31 CF\_M31\_EVPA1  
## 9 CF\_M31\_EV2 Untreated M31 CF\_M31\_EV2  
## 10 CF\_M31\_EVPA2 Treatment M31 CF\_M31\_EVPA2  
## 11 CF\_M31\_EV3 Untreated M31 CF\_M31\_EV3  
## 12 CF\_M31\_EVPA3 Treatment M31 CF\_M31\_EVPA3  
## 13 CF\_M32\_EV1 Untreated M32 CF\_M32\_EV1  
## 14 CF\_M32\_EVPA1 Treatment M32 CF\_M32\_EVPA1  
## 15 CF\_M32\_EV2 Untreated M32 CF\_M32\_EV2  
## 16 CF\_M32\_EVPA2 Treatment M32 CF\_M32\_EVPA2  
## 17 CF\_M32\_EV3 Untreated M32 CF\_M32\_EV3  
## 18 CF\_M32\_EVPA3 Treatment M32 CF\_M32\_EVPA3

1. Visualize your library sizes.

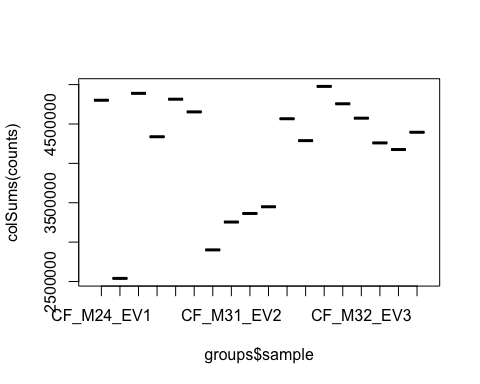
boxplot(colSums(counts) ~ donors)



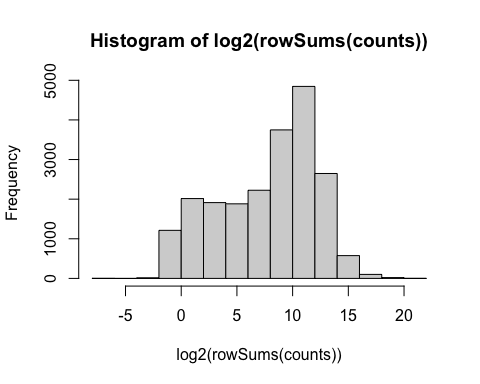
boxplot(colSums(counts) ~ treatment)



boxplot(colSums(counts) ~ groups$sample)



hist(log2(rowSums(counts)))



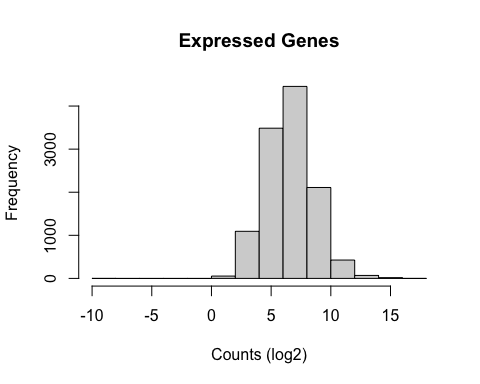
##3. Determine how many genes had zero counts and remove these genes (rows) from the data set. What does your count distribution look like before and after this?##

MinVals <- apply(counts, 1, min)  
sum(MinVals == 0)

## [1] 24868

Exp <- counts[MinVals > 0, ]

ExpLog2 <- log2(Exp)  
hist(rowMeans(ExpLog2), xlab = "Counts (log2)", main = "Expressed Genes")

 ##4. Make a boxplot of log2 raw counts after removing all genes that had zero counts.##

Plot “raw” (our log2 transformed) sample counts

boxplot(ExpLog2, ylab = "log2 counts", main = "Raw RNA-Seq Counts", las = 2)

