Análisis estadios con datos recodificados

Estefanía Gómez Guil

2023-06-02

Table of Contents

## 1.Importación de datos

params <- list(  
 estadio1="Est1",  
 estadio4="Est4",  
 directorio= "/Users/estefaniagomezguil/Desktop/Estadios Recod"  
)  
ruta\_est1<- file.path(params$directorio, "1")  
ruta\_est4<- file.path(params$directorio, "4")  
  
files\_est1 <- list.files(path = ruta\_est1, pattern = ".txt", full.names = TRUE)  
files\_est4 <- list.files(path = ruta\_est4, pattern = ".txt", full.names = TRUE)  
Grupo <- c (rep(params$estadio1, length(files\_est1)), rep(params$estadio4, length(files\_est4)))  
files <- c(files\_est1, files\_est4)  
print(files)

## [1] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2701.txt"  
## [2] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2858.txt"  
## [3] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2894.txt"  
## [4] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2907.txt"  
## [5] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2912.txt"  
## [6] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2708.txt"  
## [7] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2730.txt"  
## [8] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2736.txt"  
## [9] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2801.txt"  
## [10] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2823.txt"  
## [11] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2824.txt"  
## [12] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2891.txt"  
## [13] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2893.txt"  
## [14] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2911.txt"  
## [15] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2934.txt"  
## [16] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2948.txt"

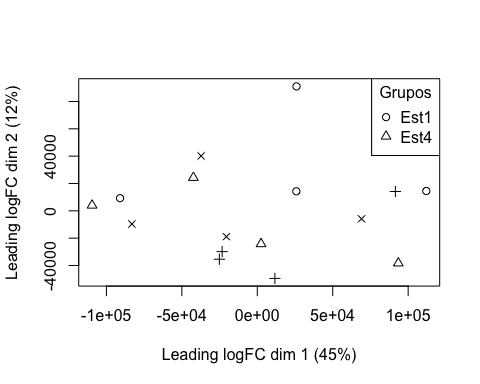
targets = read.maimages(file=files, source="agilent", green.only=TRUE)

## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2701.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2858.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2894.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2907.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2912.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2708.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2730.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2736.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2801.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2823.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2824.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2891.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2893.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2911.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2934.txt   
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2948.txt

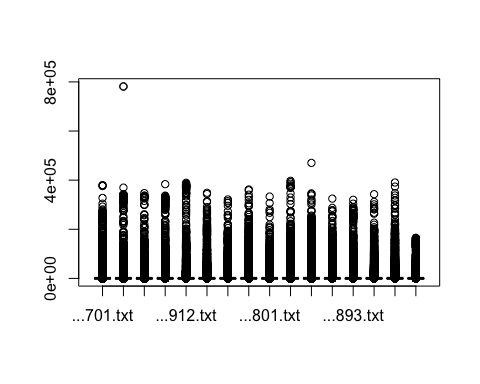
targets$targets$class <- Grupo  
files2<- str\_remove\_all(string=files, pattern= "US85003608\_253949442215")  
files2<- str\_remove\_all(string=files2, patter= "\_S01\_GE1\_107\_Sep09")

## 2.Quality Control

limma::plotMA3by2(targets, status = targets$genes$ControlType, prefix = "MA", device = "pdf")  
# Obtener los grupos únicos y sus correspondientes colores  
grupos\_unicos <- unique(Grupo)  
colores <- 1:4  
Grupo2<-colores  
plotMDS(targets, top = 100, gene.selection = "pairwise", pch=colores, type="p", plot = TRUE)  
  
# Agregar la leyenda  
legend("topright", legend = grupos\_unicos, pch = colores, title = "Grupos")  
legend("topright", legend = grupos\_unicos, pch = colores, title = "Grupos")



boxplot(targets$E, col = "royalblue", names = str\_trunc(targets$targets$FileName, width = 10, side = "left"))



## 3. Preprocesado

### 3.1 Background

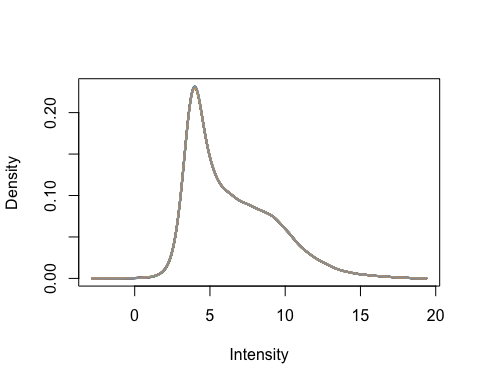
BK <- backgroundCorrect(targets, method ="normexp", printer=targets$printer, verbose=TRUE)

## Array 1 corrected  
## Array 2 corrected  
## Array 3 corrected  
## Array 4 corrected  
## Array 5 corrected  
## Array 6 corrected  
## Array 7 corrected  
## Array 8 corrected  
## Array 9 corrected  
## Array 10 corrected  
## Array 11 corrected  
## Array 12 corrected  
## Array 13 corrected  
## Array 14 corrected  
## Array 15 corrected  
## Array 16 corrected

### 3.2 Normalización

MA.p<-normalizeBetweenArrays(BK, method = 'quantile')  
transposed <- t(MA.p$E)  
df<-data.frame(value = transposed, group = as.character(Grupo))  
plotDensities(MA.p, legend = FALSE, type="1")

## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded  
  
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...) :  
## extra argument 'type' will be disregarded



### 3.3 Anotación

MA.p$genes$EntrezID<- mapIds(HsAgilentDesign026652.db, MA.p$genes$ProbeName, keytype = "PROBEID", column = "ENTREZID")

## 'select()' returned 1:1 mapping between keys and columns

MA.p$genes$Symbol <- mapIds(HsAgilentDesign026652.db, MA.p$genes$ProbeName, keytype = "PROBEID", column = "SYMBOL")

## 'select()' returned 1:1 mapping between keys and columns

map=getGEO("GPL20844")

### 3.3 Filtrado

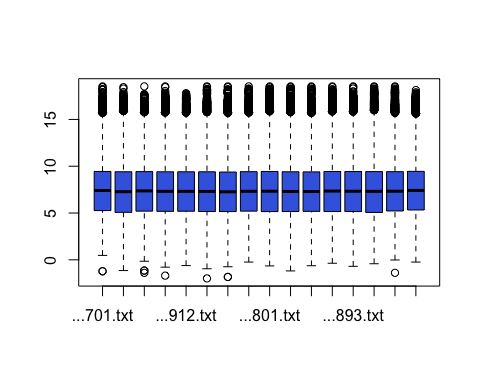
control\_1 =MA.p$genes$ControlType == 1  
control\_b= MA.p$genes$ControlType == -1

NoSym = is.na(MA.p$genes$Symbol)  
NoEnID= is.na(MA.p$genes$EntrezID)  
table(NoEnID)

## NoEnID  
## FALSE TRUE   
## 36037 26939

yfilt = MA.p[!control\_1& !control\_b & !NoSym & !NoEnID,]

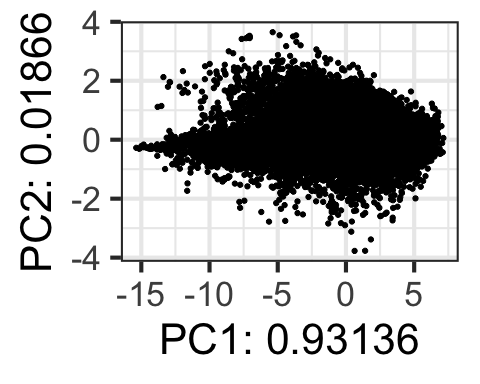
boxplot(yfilt$E, col = "royalblue", names = str\_trunc(targets$targets$FileName, width = 10, side = "left"))



pcomp=prcomp(yfilt$E, scale. = TRUE)  
eso=summary(pcomp)  
eso

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7  
## Standard deviation 3.8603 0.54638 0.37118 0.32929 0.29281 0.28341 0.25835  
## Proportion of Variance 0.9314 0.01866 0.00861 0.00678 0.00536 0.00502 0.00417  
## Cumulative Proportion 0.9314 0.95002 0.95863 0.96541 0.97076 0.97579 0.97996  
## PC8 PC9 PC10 PC11 PC12 PC13 PC14  
## Standard deviation 0.25461 0.22436 0.20594 0.19374 0.17947 0.16337 0.1598  
## Proportion of Variance 0.00405 0.00315 0.00265 0.00235 0.00201 0.00167 0.0016  
## Cumulative Proportion 0.98401 0.98715 0.98981 0.99215 0.99416 0.99583 0.9974  
## PC15 PC16  
## Standard deviation 0.14591 0.14095  
## Proportion of Variance 0.00133 0.00124  
## Cumulative Proportion 0.99876 1.00000

com=as.data.frame(pcomp$x)  
ggplot(com, aes(PC1, PC2)) + geom\_point()+theme\_bw(base\_size=32)+xlab(paste("PC1:",eso$importance[2,1])) + ylab(paste("PC2:",eso$importance[2,2]))+theme(legend.position="top")



## 4. Expressión Diferencial

### 4.1 Modelización i Regresión

design <- model.matrix(~0+Grupo)  
colnames(design) = c(params$estadio1, params$estadio4)  
design

## Est1 Est4  
## 1 1 0  
## 2 1 0  
## 3 1 0  
## 4 1 0  
## 5 1 0  
## 6 0 1  
## 7 0 1  
## 8 0 1  
## 9 0 1  
## 10 0 1  
## 11 0 1  
## 12 0 1  
## 13 0 1  
## 14 0 1  
## 15 0 1  
## 16 0 1  
## attr(,"assign")  
## [1] 1 1  
## attr(,"contrasts")  
## attr(,"contrasts")$Grupo  
## [1] "contr.treatment"

fit = lmFit(yfilt, design)

parametre1 = params$estadio1  
parametre2 = params$estadio4  
  
contraste = makeContrasts(Est1-Est4, levels = design)  
fit2 = contrasts.fit(fit, contraste)  
fit2 = eBayes(fit2)

### Resultados

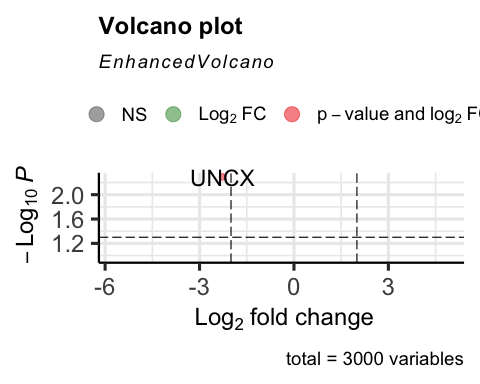
topTable(fit2, adjust.method = "bonferroni", n=30)

## Row Col ControlType ProbeName SystematicName EntrezID Symbol  
## 14027 86 87 0 A\_33\_P3411388 NM\_001080461 340260 UNCX  
## 46463 284 51 0 A\_33\_P3358457 NM\_005270 2736 GLI2  
## 58569 358 21 0 A\_23\_P123704 NR\_024032 58483 LINC00474  
## 45947 281 27 0 A\_33\_P3292064 BC062656 254240 BPIFC  
## 48397 296 17 0 A\_33\_P3381478 NM\_001105659 127255 LRRIQ3  
## 35434 217 10 0 A\_33\_P3348352 NR\_015407 339535 LINC01139  
## 57904 354 12 0 A\_23\_P112296 NM\_000787 1621 DBH  
## 36631 224 59 0 A\_33\_P3296427 NR\_015407 339535 LINC01139  
## 46435 284 23 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 6403 40 7 0 A\_23\_P120428 NM\_014477 27296 TP53TG5  
## 13838 85 62 0 A\_33\_P3388183 AK126237 4057 LTF  
## 13906 85 130 0 A\_33\_P3382867 NR\_027420 389834 LOC389834  
## 42053 257 69 0 A\_33\_P3356990 NM\_130464 23117 NPIPB3  
## 7120 44 68 0 A\_33\_P3546363 BQ950045 400128 TUSC8  
## 50931 311 91 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 51722 316 62 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 49376 302 12 0 A\_24\_P366652 NM\_005353 3681 ITGAD  
## 47072 288 4 0 A\_33\_P3397545 NM\_001033017 613209 DEFB135  
## 13611 83 163 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 12032 74 60 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 31774 194 122 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 20025 123 17 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 36707 224 135 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 43645 267 21 0 A\_23\_P66593 NM\_002055 2670 GFAP  
## 62794 383 146 0 A\_23\_P35782 NM\_024771 79829 NAA40  
## 11591 71 111 0 A\_33\_P3267185 NM\_001001323 490 ATP2B1  
## 21404 131 84 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 26345 161 105 0 A\_23\_P1102 NM\_001100 58 ACTA1  
## 60401 369 49 0 A\_33\_P3275226 ENST00000414750 101927156 LINC01934  
## 60921 372 77 0 A\_33\_P3217495 NM\_003041 6524 SLC5A2  
## logFC AveExpr t P.Value adj.P.Val B  
## 14027 -2.2720234 13.209101 -8.443014 1.413578e-07 0.005094109 4.4868391  
## 46463 -1.2069692 5.267545 -6.484043 4.909597e-06 0.176927144 2.5530295  
## 58569 -1.2854009 3.643822 -6.073209 1.103130e-05 0.397534883 2.0635939  
## 45947 -1.3244308 3.307873 -5.629304 2.712331e-05 0.977442620 1.5000412  
## 48397 -1.8589308 3.580196 -5.609381 2.825746e-05 1.000000000 1.4739049  
## 35434 -1.3502628 3.902014 -5.465142 3.806826e-05 1.000000000 1.2825423  
## 57904 -0.9751315 4.096276 -5.329178 5.053085e-05 1.000000000 1.0987463  
## 36631 -0.9799991 3.770903 -5.269513 5.725650e-05 1.000000000 1.0170574  
## 46435 3.2923483 7.156777 5.243505 6.046939e-05 1.000000000 0.9812534  
## 6403 -1.0694685 4.305319 -5.236029 6.142664e-05 1.000000000 0.9709407  
## 13838 0.9245064 4.514476 5.222460 6.320404e-05 1.000000000 0.9521970  
## 13906 -1.5415469 3.930114 -5.149474 7.371099e-05 1.000000000 0.8508280  
## 42053 -1.2769467 4.106202 -5.137577 7.558622e-05 1.000000000 0.8342178  
## 7120 -1.7213817 12.313771 -5.108727 8.033933e-05 1.000000000 0.7938369  
## 50931 3.1631896 7.386266 4.966034 1.087638e-04 1.000000000 0.5920462  
## 51722 3.1561055 7.382460 4.944645 1.138366e-04 1.000000000 0.5615061  
## 49376 -1.0980037 5.163349 -4.926253 1.183916e-04 1.000000000 0.5351863  
## 47072 -1.2091881 3.490359 -4.918647 1.203293e-04 1.000000000 0.5242857  
## 13611 3.0010779 7.160533 4.912428 1.219379e-04 1.000000000 0.5153645  
## 12032 3.0831708 7.337670 4.908267 1.230263e-04 1.000000000 0.5093934  
## 31774 2.9856592 7.348353 4.894459 1.267100e-04 1.000000000 0.4895555  
## 20025 3.0869178 7.125093 4.847597 1.400726e-04 1.000000000 0.4220021  
## 36707 3.1956223 7.221084 4.813673 1.506362e-04 1.000000000 0.3728815  
## 43645 -1.0144588 3.425805 -4.800023 1.551129e-04 1.000000000 0.3530649  
## 62794 -0.8808876 6.220432 -4.743723 1.750590e-04 1.000000000 0.2710219  
## 11591 0.7490237 5.726711 4.704939 1.903034e-04 1.000000000 0.2142186  
## 21404 2.9290395 7.639089 4.675732 2.026709e-04 1.000000000 0.1712908  
## 26345 3.0460645 7.585673 4.652421 2.131267e-04 1.000000000 0.1369374  
## 60401 -1.2492521 3.980746 -4.644361 2.168685e-04 1.000000000 0.1250400  
## 60921 -0.8253286 6.267019 -4.631065 2.231878e-04 1.000000000 0.1053922

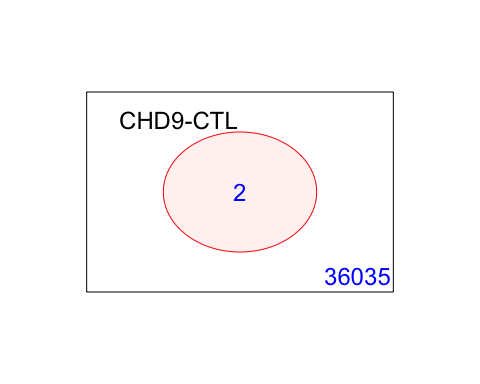
tT=topTable(fit2, adjust.method = "bonferroni", n=3000)  
  
results = decideTests(fit2, p.value = 0.1)  
summary(decideTests(fit2), p.value = 0.1)

## Est1 - Est4  
## Down 1  
## NotSig 36036  
## Up 0

EnhancedVolcano(tT, lab = tT$Symbol, x = "logFC", y = "adj.P.Val", pCutoff = 0.05, FCcutoff=2, labSize = 6.0, ylim=0.95)



vennDiagram(results, include=c("both"), circle.col = "red", counts.col = "blue", names = c("CHD9-CTL"))



## 5. Gene Enrichment Analysis

### Gene Ontology

g = goana(fit2, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05)  
topGO(g, n=20)

## Term  
## GO:0035726 common myeloid progenitor cell proliferation  
## GO:0021889 olfactory bulb interneuron differentiation  
## GO:0001502 cartilage condensation  
## GO:0021516 dorsal spinal cord development  
## GO:0098743 cell aggregation  
## GO:0021772 olfactory bulb development  
## GO:0021988 olfactory lobe development  
## GO:0021510 spinal cord development  
## GO:0051216 cartilage development  
## GO:0048705 skeletal system morphogenesis  
## GO:0021537 telencephalon development  
## GO:0061448 connective tissue development  
## GO:0030900 forebrain development  
## GO:0007389 pattern specification process  
## GO:0001501 skeletal system development  
## GO:0007420 brain development  
## GO:0060322 head development  
## GO:0009887 animal organ morphogenesis  
## GO:0007417 central nervous system development  
## GO:0000981 DNA-binding transcription factor activity, RNA polymerase II-specific  
## Ont N Up Down P.Up P.Down  
## GO:0035726 BP 8 0 1 1 0.0003843198  
## GO:0021889 BP 11 0 1 1 0.0005284397  
## GO:0001502 BP 20 0 1 1 0.0009607994  
## GO:0021516 BP 20 0 1 1 0.0009607994  
## GO:0098743 BP 24 0 1 1 0.0011529593  
## GO:0021772 BP 33 0 1 1 0.0015853190  
## GO:0021988 BP 35 0 1 1 0.0016813989  
## GO:0021510 BP 102 0 1 1 0.0049000769  
## GO:0051216 BP 199 0 1 1 0.0095599539  
## GO:0048705 BP 226 0 1 1 0.0108570331  
## GO:0021537 BP 266 0 1 1 0.0127786318  
## GO:0061448 BP 270 0 1 1 0.0129707917  
## GO:0030900 BP 393 0 1 1 0.0188797079  
## GO:0007389 BP 464 0 1 1 0.0222905457  
## GO:0001501 BP 531 0 1 1 0.0255092237  
## GO:0007420 BP 770 0 1 1 0.0369907763  
## GO:0060322 BP 818 0 1 1 0.0392966949  
## GO:0009887 BP 1021 0 1 1 0.0490488086  
## GO:0007417 BP 1056 0 1 1 0.0507302075  
## GO:0000981 MF 1136 0 1 1 0.0545734051

### KEGG

k = kegga(fit2, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05)  
topKEGG(k, number = 20)

## Pathway N Up Down  
## hsa00780 Biotin metabolism 3 0 0  
## hsa00290 Valine, leucine and isoleucine biosynthesis 4 0 0  
## hsa03267 Virion - Adenovirus 4 0 0  
## hsa03264 Virion - Flavivirus 4 0 0  
## hsa03265 Virion - Lyssavirus 4 0 0  
## hsa00232 Caffeine metabolism 5 0 0  
## hsa00524 Neomycin, kanamycin and gentamicin biosynthesis 5 0 0  
## hsa03260 Virion - Human immunodeficiency virus 5 0 0  
## hsa00470 D-Amino acid metabolism 6 0 0  
## hsa00400 Phenylalanine, tyrosine and tryptophan biosynthesis 6 0 0  
## hsa00440 Phosphonate and phosphinate metabolism 6 0 0  
## hsa00750 Vitamin B6 metabolism 6 0 0  
## hsa00740 Riboflavin metabolism 8 0 0  
## hsa04122 Sulfur relay system 8 0 0  
## hsa03266 Virion - Herpesvirus 9 0 0  
## hsa00920 Sulfur metabolism 10 0 0  
## hsa00130 Ubiquinone and other terpenoid-quinone biosynthesis 11 0 0  
## hsa03450 Non-homologous end-joining 13 0 0  
## hsa00533 Glycosaminoglycan biosynthesis - keratan sulfate 14 0 0  
## hsa00603 Glycosphingolipid biosynthesis - globo and isoglobo series 14 0 0  
## P.Up P.Down  
## hsa00780 1 1  
## hsa00290 1 1  
## hsa03267 1 1  
## hsa03264 1 1  
## hsa03265 1 1  
## hsa00232 1 1  
## hsa00524 1 1  
## hsa03260 1 1  
## hsa00470 1 1  
## hsa00400 1 1  
## hsa00440 1 1  
## hsa00750 1 1  
## hsa00740 1 1  
## hsa04122 1 1  
## hsa03266 1 1  
## hsa00920 1 1  
## hsa00130 1 1  
## hsa03450 1 1  
## hsa00533 1 1  
## hsa00603 1 1