

## RNA\_seq

```
library(limma)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(DT)
library(GenomicRanges)

## Loading required package: stats4

## Loading required package: BiocGenerics

## Loading required package: parallel

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:dplyr':
##
##     combine, intersect, setdiff, union

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

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

```

## The following objects are masked from 'package:base':
##
##   anyDuplicated, 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, sort, table, tapply,
##   union, unique, unsplit, which.max, which.min

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:dplyr':
##
##   first, rename

## The following object is masked from 'package:base':
##
##   expand.grid

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':
##
##   collapse, desc, slice

## Loading required package: GenomeInfoDb

library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:S4Vectors':
##
##   expand

library(pheatmap)
library(DESeq2)

## Loading required package: SummarizedExperiment

## Loading required package: MatrixGenerics

## Loading required package: matrixStats

```

```

## 
## Attaching package: 'matrixStats'

## The following object is masked from 'package:dplyr':
## 
##     count

## 
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':
## 
##     colAlls, colAnyNAs, colAnyNs, colAvgsPerRowSet, colCollapse,
##     colCounts, colCummmaxs, colCummins, colCumprods, colCumsums,
##     colDiffss, colIQRDiffss, colIQRs, colLogSumExps, colMadDiffss,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffss, colSds,
##     colSums2, colTabulates, colVarDiffss, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnyNs, rowAvgsPerColSet,
##     rowCollapse, rowCounts, rowCummmaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffss, rowIQRDiffss, rowIQRs, rowLogSumExps,
##     rowMadDiffss, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffss, rowSds, rowSums2, rowTabulates, rowVarDiffss, rowVars,
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars

## 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")'.

## 
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
## 
##     rowMedians

## The following objects are masked from 'package:matrixStats':
## 
##     anyMissing, rowMedians

library(BSgenome.Mmusculus.UCSC.mm10)

## Loading required package: BSgenome

```

```

## Loading required package: Biostrings

## Loading required package: XVector

##
## Attaching package: 'Biostrings'

## The following object is masked from 'package:base':
## 
##     strsplit

## Loading required package: rtracklayer

library(tracktables)
library(Rsubread)
library(clusterProfiler)

## 

## clusterProfiler v3.18.1 For help: https://guangchuangyu.github.io/software/clusterProfiler
## 
## If you use clusterProfiler in published research, please cite:
## Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He. clusterProfiler: an R package for comparing bio

##
## Attaching package: 'clusterProfiler'

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

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

## The following object is masked from 'package:S4Vectors':
## 
##     rename

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

library(enrichplot)
library(ChIPseeker)

## ChIPseeker v1.26.2 For help: https://guangchuangyu.github.io/software/ChIPseeker
## 
## If you use ChIPseeker in published research, please cite:
## Guangchuang Yu, Li-Gen Wang, Qing-Yu He. ChIPseeker: an R/Bioconductor package for ChIP peak annotat

```

```

library(org.Mm.eg.db)

## Loading required package: AnnotationDbi

##
## Attaching package: 'AnnotationDbi'

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

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

## 

library(TxDb.Mmusculus.UCSC.mm10.knownGene)

## Loading required package: GenomicFeatures

library(gplots)

##
## Attaching package: 'gplots'

## The following object is masked from 'package:rtracklayer':
##      space

## The following object is masked from 'package:IRanges':
##      space

## The following object is masked from 'package:S4Vectors':
##      space

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

library(ggplot2)

setwd("~/Documents/GitHub/RNA_n_ATAC_hematopoiesis")
countData=read.csv('complete.csv', header=TRUE, as.is = TRUE, row.names = 1)
head(countData)

```

```

##          SEDM_2P_1 SEDM_2P_2 SEDM_2P_3 SEEM_2P_1 SEEM_2P_2 SEEM_2P_3
## a            1        1        1        1        7        2
## a.1          0        0        0        0        0        0
## A030001D20Rik    20       33       30       21       19       23
## A030003K21Rik     0        0        0        0        0        0
## A030005K14Rik     0        0        0        0        0        0
## A030005L19Rik    13       10       12       0        6        3
##          SEEM_4P_1 SEEM_4P_2 SEEM_4P_3 PERKi_1 PERKi_2 PERKi_3 T_1 T_2 T_3
## a            1        3        4        8        1        6        1        1        1
## a.1          0        0        0        0        0        0        0        0        0
## A030001D20Rik    27       28       17       33       0       11       38       48       31
## A030003K21Rik     0        0        0        0        0        0        0        0        0
## A030005K14Rik     0        0        0        0        0        0        0        0        0
## A030005L19Rik    7        1        3       23       14       16       16       10       18
##          P1_dP_redoA_1 P2_dP_redoA_1 P3_dP_redoA_1
## a            1        4        4
## a.1          0        0        0
## A030001D20Rik    40       22       36
## A030003K21Rik     0        0        0
## A030005K14Rik     0        0        0
## A030005L19Rik     0        0        0

#countData$gene_id=unlist(lapply(strsplit(countData$gene_id, '/', fixed = TRUE), '[' , 2))

#countData=na.omit(countData)

rownames(countData) = make.names(countData[,1],unique=TRUE)
countData=countData[,-1]

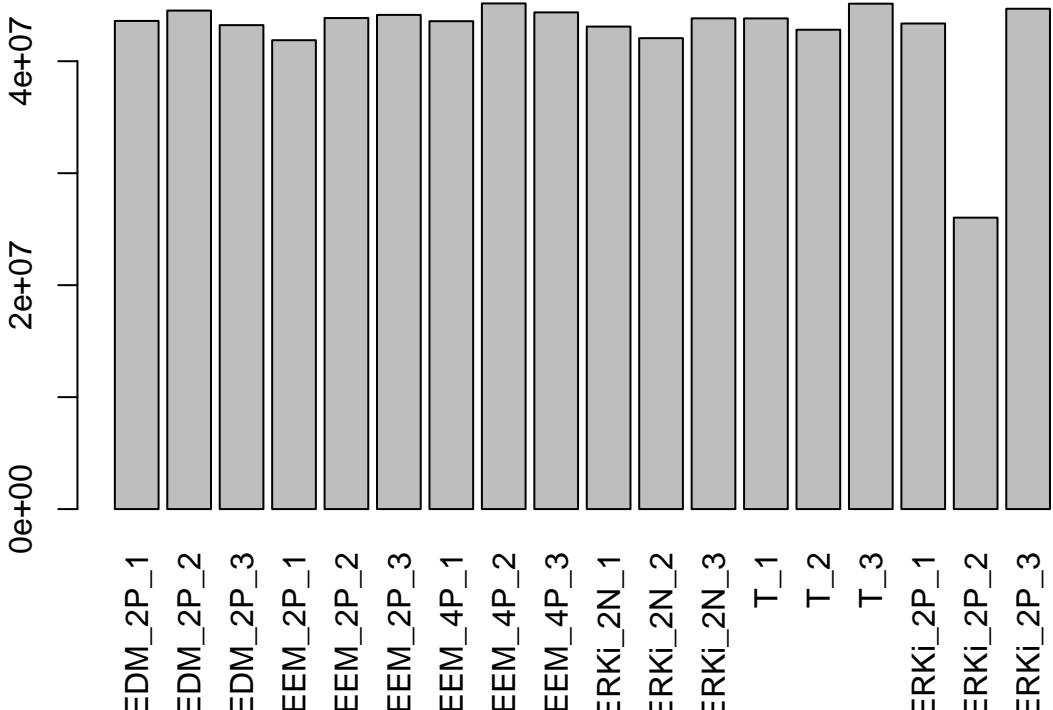
columns=c("SEDM_2P_1","SEDM_2P_2","SEDM_2P_3","SEEM_2P_1","SEEM_2P_2","SEEM_2P_3","SEEM_4P_1","SEEM_4P_2")

colnames(countData)=columns
colSums(countData)

##   SEDM_2P_1   SEDM_2P_2   SEDM_2P_3   SEEM_2P_1   SEEM_2P_2   SEEM_2P_3   SEEM_4P_1
## 43599396  44523440  43223051  41877778  43855707  44137369  43575802
##   SEEM_4P_2   SEEM_4P_3 PERKi_2N_1 PERKi_2N_2 PERKi_2N_3      T_1      T_2
## 45162148  44363909  43094604  42060873  43829962  43823034  42812832
##      T_3 PERKi_2P_1 PERKi_2P_2 PERKi_2P_3
## 45140511  43369677  26029564  44687702

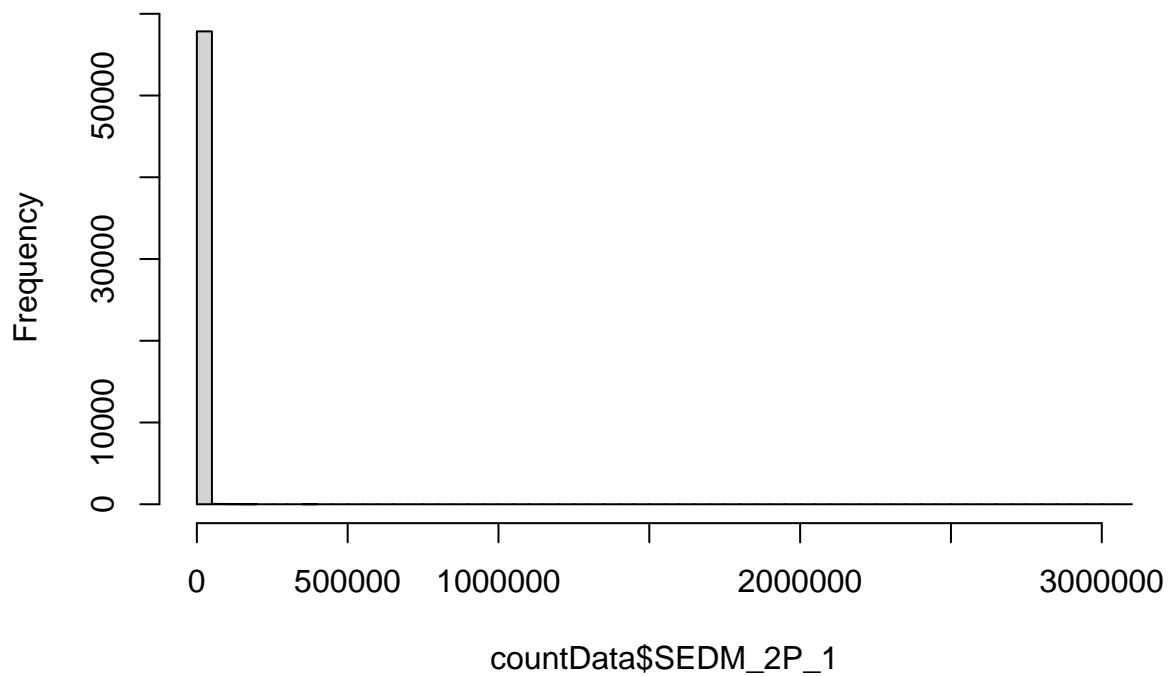
barplot(colSums(countData),las=3)

```



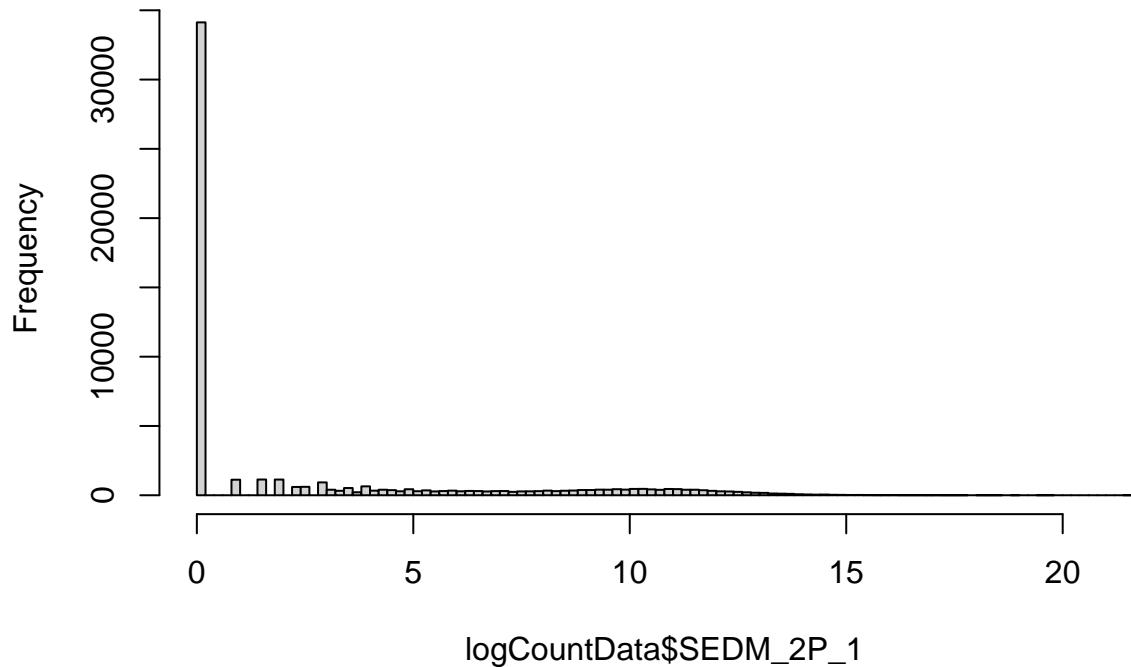
```
hist(countData$SEDM_2P_1, br=100)
```

**Histogram of countData\$SEDM\_2P\_1**

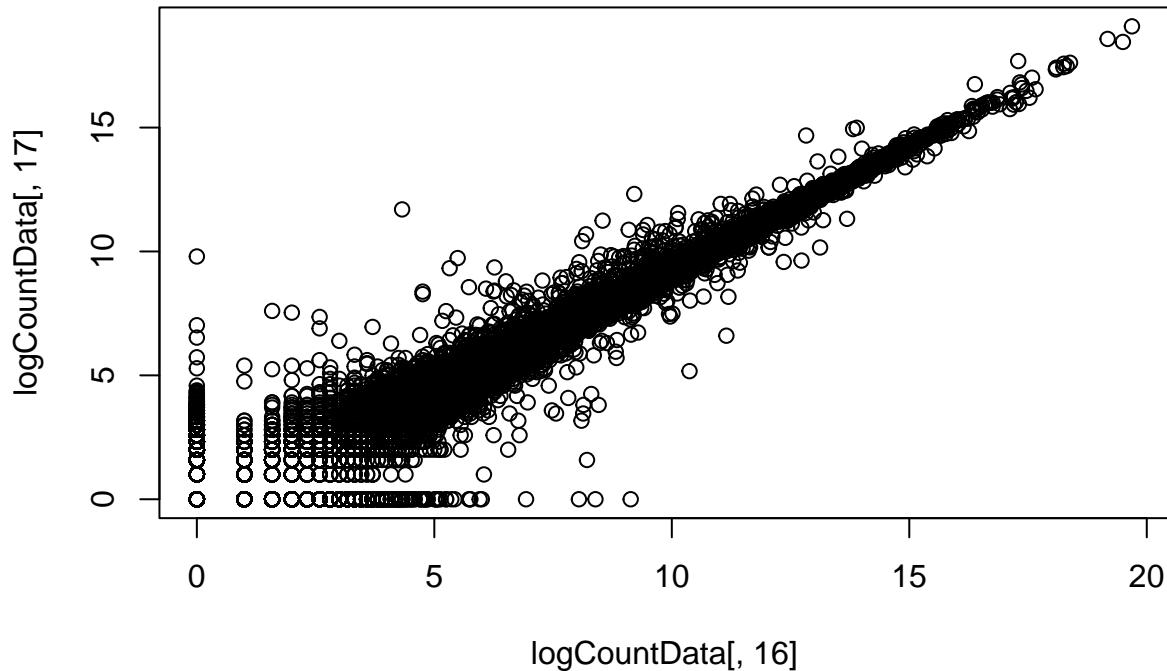


```
logCountData=log2(1+countData)
hist(logCountData$SEDM_2P_1,br=100)
```

## Histogram of logCountData\$SEDM\_2P\_1



```
plot(logCountData[,16],logCountData[,17])
```



```
treatment=as.factor(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6))
inh=c(1,1,1,1,1,1,1,2,2,2,1,1,1,2,2,2)
colData=as.data.frame(cbind(colnames(countData), inh, treatment))
```

```
dds=DESeqDataSetFromMatrix(countData = countData, colData = colData, design=~inh)
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```
dds=DESeq(dds)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
## fitting model and testing
```

```

## -- replacing outliers and refitting for 336 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)

## estimating dispersions

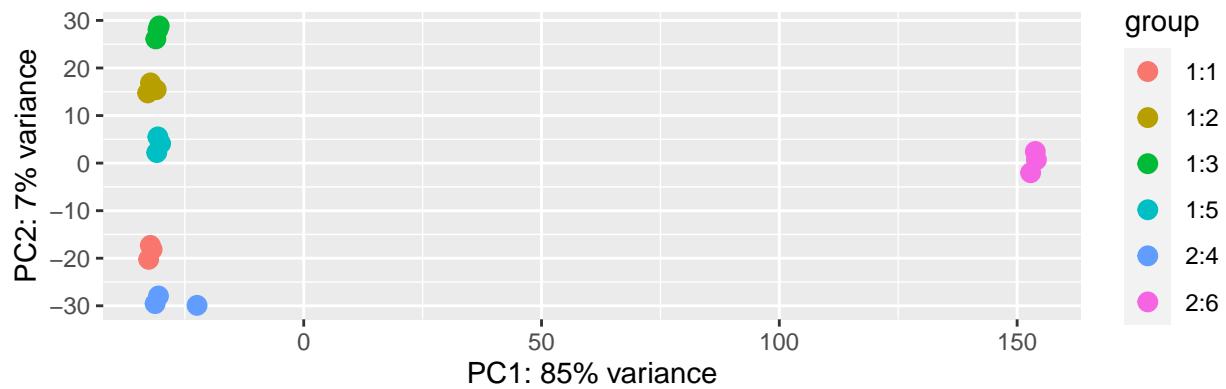
## fitting model and testing

dds= dds[rowSums(counts(dds))>5,]
nrow(dds)

## [1] 29367

library(ggplot2)
rld=rlog(dds)
plotPCA(rld, intgroup= c("inh","treatment"))

```



#courtesy of [https://github.com/ACSoupir/Bioinformatics\\_YouTube/blob/master/Visualizing%20Counts/README.md](https://github.com/ACSoupir/Bioinformatics_YouTube/blob/master/Visualizing%20Counts/README.md)

```

detectGroups <- function (x){ # x are col names
  tem <- gsub("[0-9]*$", "", x) # Remove all numbers from end
  #tem = gsub("_Rep/_rep/_REP","",tem)
  tem <- gsub("_$","",tem); # remove "_" from end

```

```

tem <- gsub("_Rep$","",tem); # remove "_Rep" from end
tem <- gsub("_rep$","",tem); # remove "_rep" from end
tem <- gsub("_REP$","",tem) # remove "_REP" from end
return( tem )
}

detectGroups(colnames(countData))

## [1] "SEDM_2P"  "SEDM_2P"  "SEDM_2P"  "SEEM_2P"  "SEEM_2P"  "SEEM_2P"
## [7] "SEEM_4P"  "SEEM_4P"  "SEEM_4P"  "PERKi_2N" "PERKi_2N" "PERKi_2N"
## [13] "T"         "T"        "T"        "PERKi_2P" "PERKi_2P" "PERKi_2P"

dist2 <- function(x, ...) # distance function = 1-PCC (Pearson's correlation coefficient)
  as.dist(1-cor(t(x), method="pearson"))

library(gplots)

hclust2 <- function(x, method="average", ...) # average linkage in hierarchical clustering
  hclust(x, method=method, ...)

n=10000 # number of top genes by standard deviation

x = assay(rld)
if(n>dim(x)[1]) n = dim(x)[1] # max as data

x = x[order(apply(x,1,sd),decreasing=TRUE),] # sort genes by standard deviation

x = x[1:n,] # only keep the n genes

# this will cutoff very large values, which could skew the color
x=as.matrix(x[1:n,])-apply(x[1:n,],1,mean)
cutoff = median(unlist(x)) + 4*sd (unlist(x))
x[x>cutoff] <- cutoff
cutoff = median(unlist(x)) - 4*sd (unlist(x))
x[x< cutoff] <- cutoff

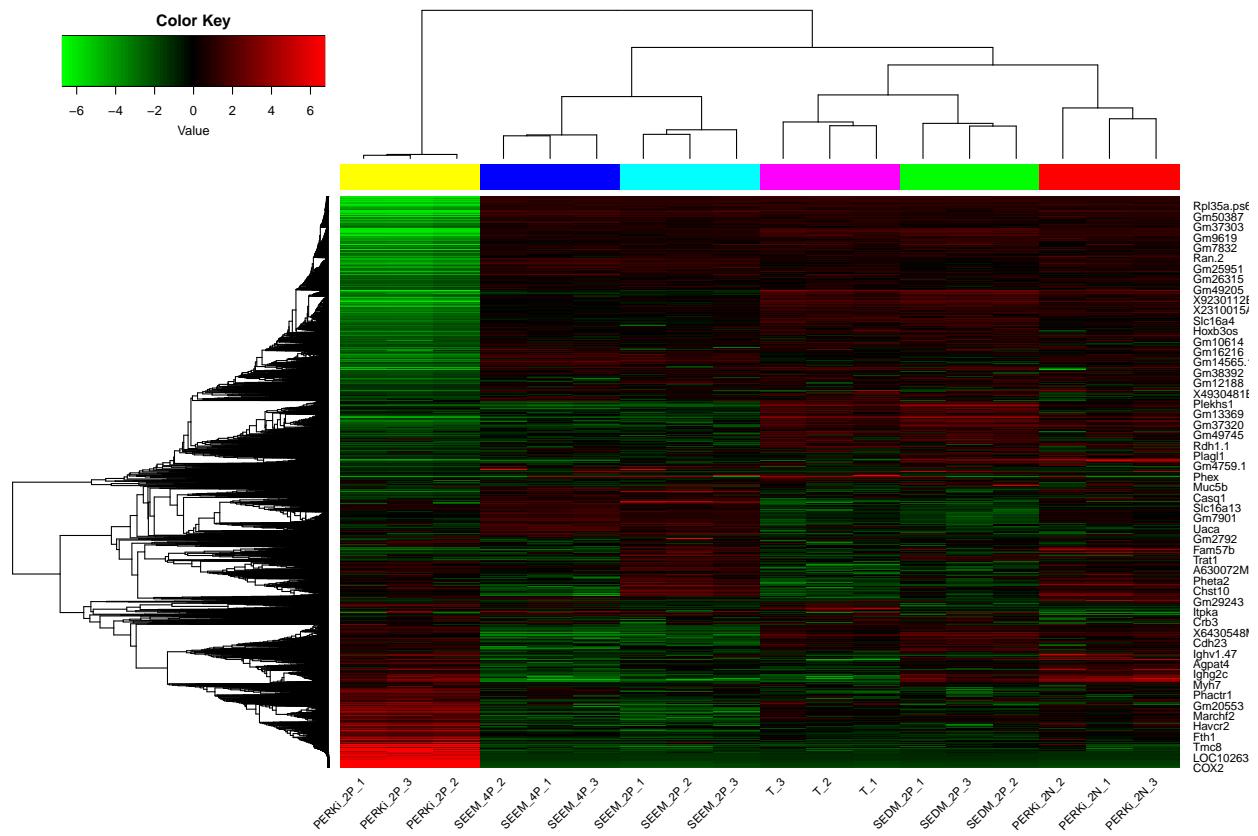
groups = detectGroups(colnames(x) )
groups.colors = rainbow(length(unique(groups)) )

lmat = rbind(c(5,4),c(0,1),c(3,2))
lwid = c(1.5,4)
lhei = c(1,.2,4)

heatmap.2(x, distfun = dist2,hclustfun=hclust2,
  col=greenred(75), density.info="none", trace="none", scale="none", keysize=.5
  ,key=T, symkey=F
  ,ColSideColors=groups.colors[ as.factor(groups)]
  ,cexRow=1
  ,srtCol=45
  ,cexCol=1. # size of font for sample names

```

```
, lmat = lmat, lwid = lwid, lhei = lhei
)
```



```
resultsNames(dds)
```

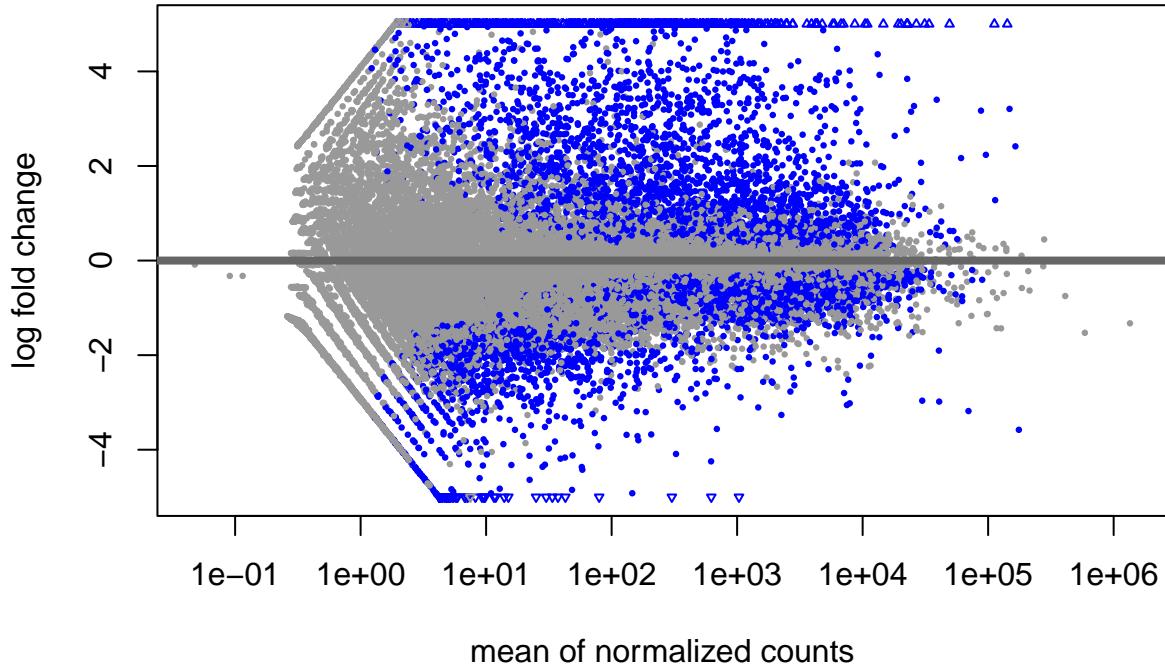
```
## [1] "Intercept" "inh_2_vs_1"
```

```
res=results(dds)
summary(res)
```

```
##
## out of 29358 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)      : 3143, 11%
## LFC < 0 (down)    : 3516, 12%
## outliers [1]       : 249, 0.85%
## low counts [2]     : 4490, 15%
## (mean count < 1)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
res=results(dds, lfcThreshold = 0.01)
```

```
DESeq2::plotMA(res,ylim=c(-5,5))
```



```
library(dplyr)
res1=as.data.frame(res)
head(res1)
```

```
##          baseMean log2FoldChange      lfcSE       stat      pvalue
## a        2.640186     1.20800170  0.6242685  1.919048804 0.054978158
## A030001D20Rik 27.148364    -0.10656893 0.4818935 -0.200394738 0.841171875
## A030005L19Rik  8.599040    -0.01394906 0.9223589 -0.004281483 0.996583881
## A130006I12Rik  1.306081    -1.25849561 1.3279849 -0.940142941 0.347144245
## A130010J15Rik 269.164900     0.02832540 0.1791695  0.102279677 0.918534686
## A130012E19Rik  3.945066    -2.69399004 0.9113581 -2.945044435 0.003229083
##          padj
## a        0.16902093
## A030001D20Rik 0.91410585
## A030005L19Rik 1.00000000
## A130006I12Rik 0.54143849
## A130010J15Rik 0.96410127
## A130012E19Rik 0.02120132
```

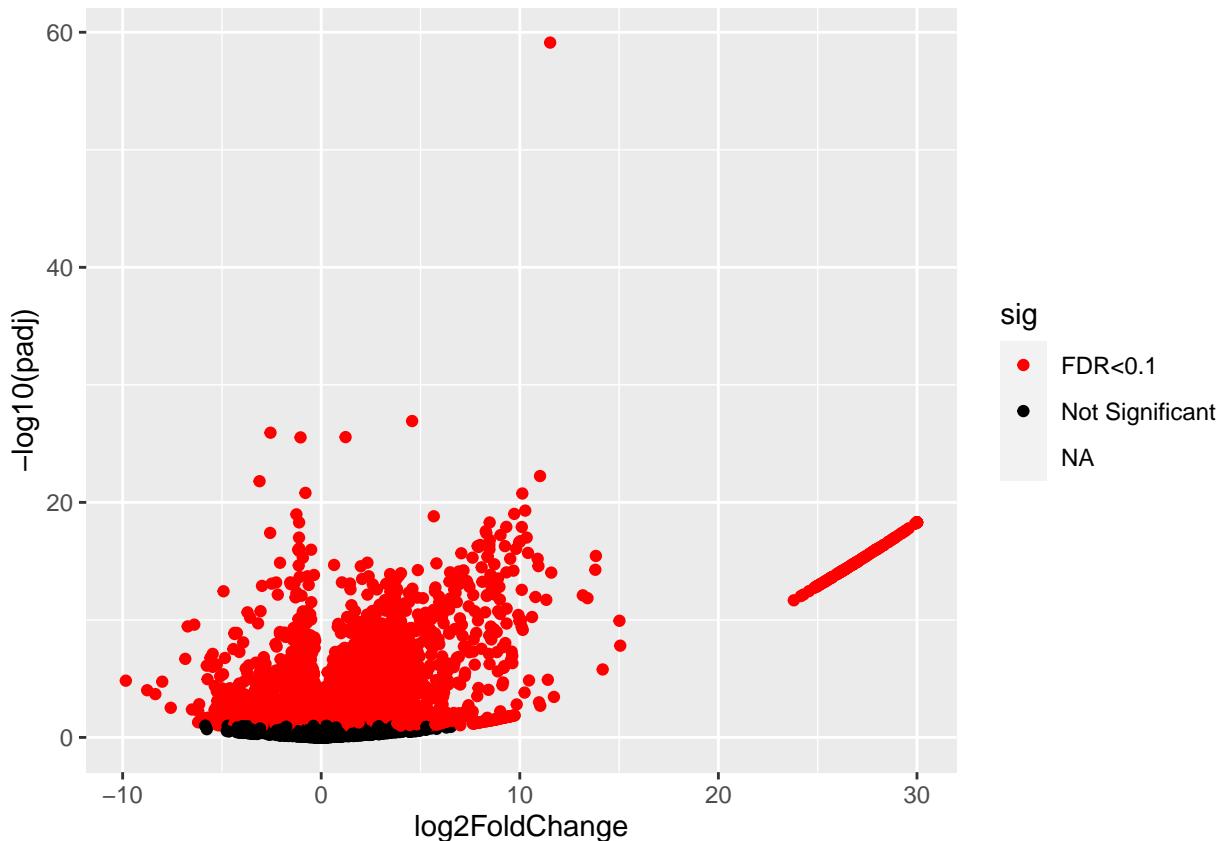
```
res1=mutate(res1,sig=ifelse(res1$padj<0.1, "FDR<0.1","Not Significant"))
res1[which(abs(res1$log2FoldChange)<1.0),"Significant"] = "Not Significant"
```

```

library(ggplot2)
ggplot(res1, aes(log2FoldChange, -log10(padj))) + geom_point(aes(col=sig)) + scale_color_manual(values=)

## Warning: Removed 4739 rows containing missing values (geom_point).

```

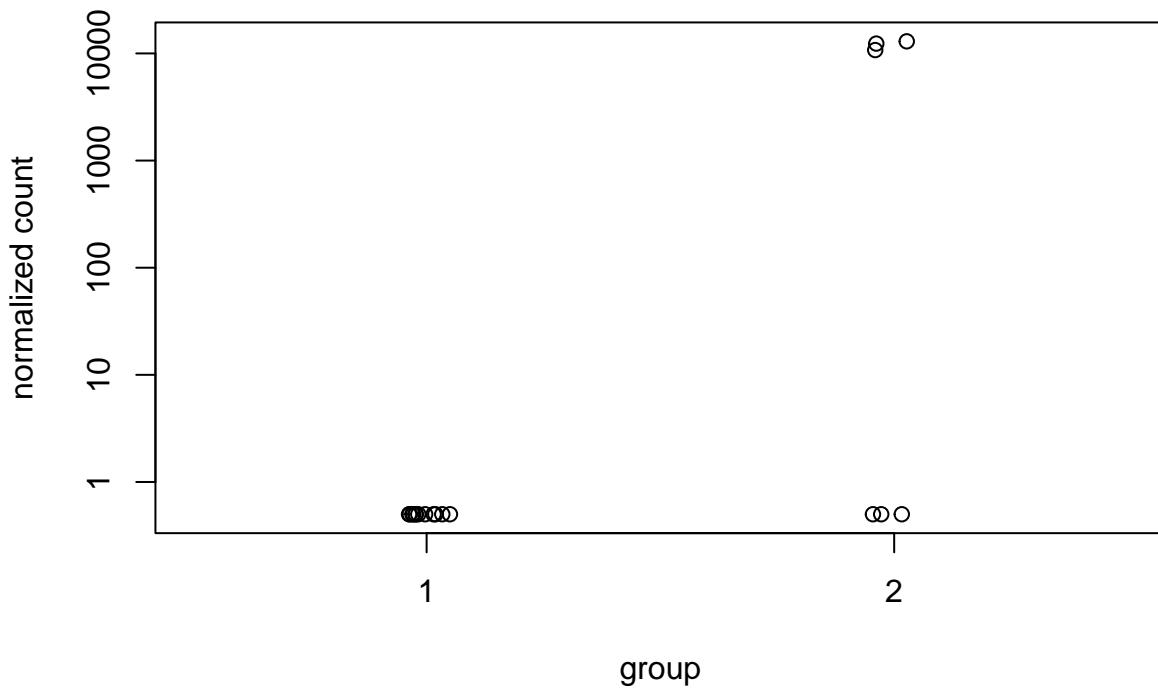


```

res=res[order(abs(res$log2FoldChange), decreasing=TRUE),]
topGene=rownames(res)[1]
plotCounts(dds, gene=topGene, intgroup = c("inh"))

```

## Abcc5.1



```
library(AnnotationDbi)
library(org.Mm.eg.db)

res$SYMBOL=gsub("\\..*", "", row.names(res))
res$ENTREZID= mapIds(org.Mm.eg.db, key=res$SYMBOL, column="ENTREZID", keytype = "SYMBOL", multiVals = "")

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

#write.csv(res, file='treatments-0.1_lfc-1_results.csv')
```

GO!

```
library(GO.db)
library(GOstats)

## Loading required package: Category

## Loading required package: Matrix

##
## Attaching package: 'Matrix'
```

```

## The following objects are masked from 'package:tidyR':
##
##     expand, pack, unpack

## The following object is masked from 'package:S4Vectors':
##
##     expand

## Loading required package: graph

##
## Attaching package: 'graph'

## The following object is masked from 'package:Biostrings':
##
##     complement

##
## Attaching package: 'GOstats'

## The following object is masked from 'package:AnnotationDbi':
##
##     makeGOGraph

```

```

res_05=as.data.frame(subset(res, padj<0.05))
sig_lfc= 4

```

```

selectGenesUp= unique(res_05[res_05$log2FoldChange>sig_lfc, "ENTREZID"])
selectGenesDown= unique(res_05[res_05$log2FoldChange>-(sig_lfc), "ENTREZID"])

```

```

universeGenes= unique(res_05$ENTREZID)
cutOff= 0.01

```

```

UpParams=new("GOHyperGParams", geneIds=selectGenesUp, universeGeneIds=universeGenes, annotation=org.Mm.)

```

```

DownParams=new("GOHyperGParams", geneIds=selectGenesDown, universeGeneIds=universeGenes, annotation=org.

```

```

upBP=hyperGTest(UpParams)
summary(upBP)[1:10,]

```

	GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size
## 1	GO:0006119	9.115216e-07	10.325212	2.5875434	12	20
## 2	GO:0098662	1.420142e-04	2.306359	16.6896552	32	129
## 3	GO:0046034	1.688105e-04	3.351117	6.7276129	17	52
## 4	GO:0006091	1.886448e-04	2.851226	9.3151564	21	72
## 5	GO:0042773	2.076579e-04	7.803121	1.9406576	8	15
## 6	GO:1904064	2.634837e-04	4.332739	4.0106923	12	31
## 7	GO:0015988	2.771600e-04	Inf	0.5175087	4	4
## 8	GO:0015990	2.771600e-04	Inf	0.5175087	4	4

```

## 9 GO:0034765 3.060005e-04 2.522530 11.6439455    24   90
## 10 GO:0098660 3.157285e-04 2.154092 18.1128041   33  140
##
##                                Term
## 1                      oxidative phosphorylation
## 2          inorganic cation transmembrane transport
## 3                  ATP metabolic process
## 4      generation of precursor metabolites and energy
## 5          ATP synthesis coupled electron transport
## 6      positive regulation of cation transmembrane transport
## 7  energy coupled proton transmembrane transport, against electrochemical gradient
## 8          electron transport coupled proton transport
## 9      regulation of ion transmembrane transport
## 10         inorganic ion transmembrane transport

```

```

downBP=hyperGTest(DownParams)
summary(downBP)[1:10,]

```

```

##           GOBPID      Pvalue OddsRatio ExpCount Count Size
## 1  GO:0043170 0.0005679058  3.639448 1985.3312 1995 2003
## 2  GO:0044237 0.0008918533  3.296063 2174.6464 2184 2194
## 3  GO:0009058 0.0011083564  5.116184 1246.9030 1255 1258
## 4  GO:0010467 0.0014204443  4.963680 1222.1235 1230 1233
## 5  GO:0010033 0.0014393955      Inf 664.0898  670  670
## 6  GO:1901576 0.0014629746  4.945586 1219.1500 1227 1230
## 7  GO:0044249 0.0017106035  4.849820 1203.2911 1211 1214
## 8  GO:0051641 0.0019243684      Inf 637.3280  643  643
## 9  GO:0009059 0.0020442276  5.969555 1023.8877 1031 1033
## 10 GO:0034645 0.0023431591  5.857860 1010.0112 1017 1019
##
##                                Term
## 1          macromolecule metabolic process
## 2          cellular metabolic process
## 3          biosynthetic process
## 4          gene expression
## 5          response to organic substance
## 6          organic substance biosynthetic process
## 7          cellular biosynthetic process
## 8          cellular localization
## 9          macromolecule biosynthetic process
## 10         cellular macromolecule biosynthetic process

```

```

UpParams=new("GOHyperGParams", geneIds=selectGenesUp, universeGeneIds=universeGenes, annotation=org.Mm.

```

```

DownParams=new("GOHyperGParams", geneIds=selectGenesDown, universeGeneIds=universeGenes, annotation=org.

```

```

upCC=hyperGTest(UpParams)
summary(upCC)[1:10,]

```

```

##           GOCCID      Pvalue OddsRatio ExpCount Count Size
## 1  GO:0070469 1.975174e-05  6.362845  3.225594    12   25
## 2  GO:0098803 1.089392e-04  6.244069  2.709499    10   21
## 3  GO:0005746 4.141240e-04  5.607955  2.580475     9   20

```

```

## 4 GO:0005747 1.165944e-03 8.188820 1.419261 6 11
## 5 GO:0030964 1.165944e-03 8.188820 1.419261 6 11
## 6 GO:0045271 1.165944e-03 8.188820 1.419261 6 11
## 7 GO:0098800 1.263309e-03 4.032912 3.483641 10 27
## 8 GO:1990204 2.774838e-03 3.849609 3.225594 9 25
## 9 GO:0005743 3.308621e-03 2.112717 12.128232 22 94
## 10 GO:0099056 7.104746e-03 3.980463 2.451451 7 19
##
##                                     Term
## 1                               respirasome
## 2             respiratory chain complex
## 3           mitochondrial respirasome
## 4   mitochondrial respiratory chain complex I
## 5           NADH dehydrogenase complex
## 6           respiratory chain complex I
## 7   inner mitochondrial membrane protein complex
## 8           oxidoreductase complex
## 9           mitochondrial inner membrane
## 10      integral component of presynaptic membrane

```

```

downCC=hyperGTest(DownParams)
summary(downCC)[1:10,]

```

	GOCCID	Pvalue	OddsRatio	ExpCount	Count	Size
## 1	GO:0005622	0.0001398901	3.815809	3000.037	3010	3028
## 2	GO:0043231	0.0001482556	3.742063	2360.993	2372	2383
## 3	GO:0043227	0.0004205359	3.331104	2479.885	2490	2503
## 4	GO:0043226	0.0005891630	3.227413	2735.503	2745	2761
## 5	GO:0043229	0.0009884805	3.059226	2695.872	2705	2721
## 6	GO:0005634	0.0010010018	3.715497	1622.873	1632	1638
## 7	GO:0031981	0.0016971521	6.071038	1002.654	1010	1012
## 8	GO:0031974	0.0037162174	4.340202	1078.943	1086	1089
## 9	GO:0043233	0.0037162174	4.340202	1078.943	1086	1089
## 10	GO:0070013	0.0037162174	4.340202	1078.943	1086	1089

```

UpParams=new("GOHyperGParams", geneIds=selectGenesUp, universeGeneIds=universeGenes, annotation=org.Mm...)

```

```

DownParams=new("GOHyperGParams", geneIds=selectGenesDown, universeGeneIds=universeGenes, annotation=org...)

```

```

upMF=hyperGTest(UpParams)
summary(upMF)[1:10,]

```

```

##          GOMFID      Pvalue OddsRatio   ExpCount Count Size
## 1  GO:0003954 2.613535e-05 41.560345  0.8928087    6    7
## 2  GO:0016651 9.832323e-05  9.261183  1.7856174    8   14
## 3  GO:0008137 1.778444e-04 34.559140  0.7652646    5    6
## 4  GO:0016655 1.778444e-04 34.559140  0.7652646    5    6
## 5  GO:0050136 1.778444e-04 34.559140  0.7652646    5    6
## 6  GO:0016491 1.875742e-03 1.961490 17.6010855   30  138
## 7  GO:0097677 2.063278e-03       Inf  0.3826323    3    3
## 8  GO:0022836 4.598007e-03  2.661411  5.9945726   13   47
## 9  GO:0015077 5.758726e-03  2.375874  7.5251018   15   59
## 10 GO:0015035 7.468035e-03 20.646681  0.5101764    3    4
##                                         Term
## 1                               NADH dehydrogenase activity
## 2                               oxidoreductase activity, acting on NAD(P)H
## 3                               NADH dehydrogenase (ubiquinone) activity
## 4  oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor
## 5                               NADH dehydrogenase (quinone) activity
## 6                               oxidoreductase activity
## 7                               STAT family protein binding
## 8                               gated channel activity
## 9  monovalent inorganic cation transmembrane transporter activity
## 10                          protein disulfide oxidoreductase activity

```

```

downMF=hyperGTest(DownParams)
summary(downMF)[1:10,]

```

```

##          GOMFID      Pvalue OddsRatio   ExpCount Count Size
## 1  GO:0097159 0.001298260 4.969210 1151.9539 1160 1163
## 2  GO:1901363 0.001517234 4.875582 1137.0963 1145 1148
## 3  GO:0005515 0.007712512 2.458464 2043.4057 2051 2063
## 4  GO:0043168 0.009972830 7.256649 636.8928  642  643
## NA      <NA>        NA        NA        NA        NA        NA
## NA.1    <NA>        NA        NA        NA        NA        NA
## NA.2    <NA>        NA        NA        NA        NA        NA
## NA.3    <NA>        NA        NA        NA        NA        NA
## NA.4    <NA>        NA        NA        NA        NA        NA
## NA.5    <NA>        NA        NA        NA        NA        NA
##                                         Term
## 1  organic cyclic compound binding
## 2  heterocyclic compound binding
## 3  protein binding
## 4  anion binding
## NA      <NA>
## NA.1    <NA>
## NA.2    <NA>
## NA.3    <NA>
## NA.4    <NA>
## NA.5    <NA>

```

```

library(pathview)

```

```

##

```

```

## ##### #####
## Pathview is an open source software package distributed under GNU General
## Public License version 3 (GPLv3). Details of GPLv3 is available at
## http://www.gnu.org/licenses/gpl-3.0.html. Particullary, users are required to
## formally cite the original Pathview paper (not just mention it) in publications
## or products. For details, do citation("pathview") within R.
##
## The pathview downloads and uses KEGG data. Non-academic uses may require a KEGG
## license agreement (details at http://www.kegg.jp/kegg/legal.html).
## #####
library(gage)
library(gageData)

foldChanges<- res$log2FoldChange
names(foldChanges)=res$ENTREZID

data("go.sets.mm")
data("go.subs.mm")

gobpsets= go.sets.mm[go.subs.mm$BP]
gobpres=gage(foldChanges, gsets= gobpsets, same.dir=TRUE)
head(gobpres$greater)

##                                     p.geomean
## GO:0006954 inflammatory response      1.957082e-08
## GO:0051240 positive regulation of multicellular organismal process 1.638978e-07
## GO:0051050 positive regulation of transport          1.920971e-07
## GO:0032101 regulation of response to external stimulus 2.482292e-07
## GO:0006935 chemotaxis                  2.639040e-07
## GO:0042330 taxis                      2.946461e-07
##                                     stat.mean
## GO:0006954 inflammatory response      5.558304
## GO:0051240 positive regulation of multicellular organismal process 5.146690
## GO:0051050 positive regulation of transport          5.114171
## GO:0032101 regulation of response to external stimulus 5.078830
## GO:0006935 chemotaxis                  5.068210
## GO:0042330 taxis                      5.046169
##                                     p.val
## GO:0006954 inflammatory response      1.957082e-08
## GO:0051240 positive regulation of multicellular organismal process 1.638978e-07
## GO:0051050 positive regulation of transport          1.920971e-07
## GO:0032101 regulation of response to external stimulus 2.482292e-07
## GO:0006935 chemotaxis                  2.639040e-07
## GO:0042330 taxis                      2.946461e-07
##                                     q.val
## GO:0006954 inflammatory response      7.726558e-05
## GO:0051240 positive regulation of multicellular organismal process 1.938771e-04
## GO:0051050 positive regulation of transport          1.938771e-04
## GO:0032101 regulation of response to external stimulus 1.938771e-04
## GO:0006935 chemotaxis                  1.938771e-04
## GO:0042330 taxis                      1.938771e-04

```

```

##                                         set.size
## GO:0006954 inflammatory response          349
## GO:0051240 positive regulation of multicellular organismal process 461
## GO:0051050 positive regulation of transport      461
## GO:0032101 regulation of response to external stimulus    327
## GO:0006935 chemotaxis                      320
## GO:0042330 taxis                          321
##                                         exp1
## GO:0006954 inflammatory response          1.957082e-08
## GO:0051240 positive regulation of multicellular organismal process 1.638978e-07
## GO:0051050 positive regulation of transport      1.920971e-07
## GO:0032101 regulation of response to external stimulus    2.482292e-07
## GO:0006935 chemotaxis                      2.639040e-07
## GO:0042330 taxis                          2.946461e-07

data("kegg.sets.mm")
data("sigmet.idx.mm")
kegg.sets.mm = kegg.sets.mm[sigmet.idx.mm]

keggres= gage(exprs = foldChanges, gsets= kegg.sets.mm, same.dir = TRUE)

head(keggres$greater)

##                                         p.geomean stat.mean      p.val
## mmu04380 Osteoclast differentiation     0.001280799 3.057634 0.001280799
## mmu04810 Regulation of actin cytoskeleton 0.001281518 3.038649 0.001281518
## mmu00190 Oxidative phosphorylation      0.001313525 3.059661 0.001313525
## mmu04510 Focal adhesion                 0.002368070 2.843336 0.002368070
## mmu04020 Calcium signaling pathway       0.002489652 2.830976 0.002489652
## mmu04260 Cardiac muscle contraction      0.002703403 2.853849 0.002703403
##                                         q.val set.size      exp1
## mmu04380 Osteoclast differentiation     0.05578634      114 0.001280799
## mmu04810 Regulation of actin cytoskeleton 0.05578634      190 0.001281518
## mmu00190 Oxidative phosphorylation      0.05578634      128 0.001313525
## mmu04510 Focal adhesion                 0.05578634      183 0.002368070
## mmu04020 Calcium signaling pathway       0.05578634      144 0.002489652
## mmu04260 Cardiac muscle contraction      0.05578634      71 0.002703403

head(keggres$less)

##                                         p.geomean stat.mean      p.val      q.val
## mmu03040 Spliceosome                  0.05132618 -1.6418132 0.05132618 0.9987192
## mmu03030 DNA replication              0.08350215 -1.4107175 0.08350215 0.9987192
## mmu03440 Homologous recombination     0.09171664 -1.3653022 0.09171664 0.9987192
## mmu03430 Mismatch repair              0.13868921 -1.1142382 0.13868921 0.9987192
## mmu03013 RNA transport                0.16078925 -0.9933107 0.16078925 0.9987192
## mmu03015 mRNA surveillance pathway    0.16819399 -0.9653503 0.16819399 0.9987192
##                                         set.size      exp1
## mmu03040 Spliceosome                  130 0.05132618
## mmu03030 DNA replication              35 0.08350215
## mmu03440 Homologous recombination     27 0.09171664
## mmu03430 Mismatch repair              22 0.13868921

```

```

## mmu03013 RNA transport          148 0.16078925
## mmu03015 mRNA surveillance pathway      84 0.16819399

keggrespathways=data.frame(id = rownames(keggres$greater), keggres$greater) %>%
  tibble::as_tibble() %>%
  filter(row_number() <= 20) %>%
  .$id %>
  as.character()
keggrespathways

## [1] "mmu04380 Osteoclast differentiation"
## [2] "mmu04810 Regulation of actin cytoskeleton"
## [3] "mmu00190 Oxidative phosphorylation"
## [4] "mmu04510 Focal adhesion"
## [5] "mmu04020 Calcium signaling pathway"
## [6] "mmu04260 Cardiac muscle contraction"
## [7] "mmu04640 Hematopoietic cell lineage"
## [8] "mmu04062 Chemokine signaling pathway"
## [9] "mmu04010 MAPK signaling pathway"
## [10] "mmu04620 Toll-like receptor signaling pathway"
## [11] "mmu04142 Lysosome"
## [12] "mmu04610 Complement and coagulation cascades"
## [13] "mmu04970 Salivary secretion"
## [14] "mmu04512 ECM-receptor interaction"
## [15] "mmu04621 NOD-like receptor signaling pathway"
## [16] "mmu04270 Vascular smooth muscle contraction"
## [17] "mmu04630 Jak-STAT signaling pathway"
## [18] "mmu04912 GnRH signaling pathway"
## [19] "mmu04910 Insulin signaling pathway"
## [20] "mmu04144 Endocytosis"

keggres_ids= substr(keggrespathways, start=1, stop=8)

tmp=sapply(keggres_ids, function(pid) pathview(gene.data= foldChanges, pathway.id = pid, species="mmu"))

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04380.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04810.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

```

```
## Info: Writing image file mmu00190.pathview.png

## Info: some node width is different from others, and hence adjusted!
## Info: some node width is different from others, and hence adjusted!

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04510.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04020.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04260.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04640.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04062.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04010.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04620.pathview.png

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

```

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04142.pathview.png

## Info: some node width is different from others, and hence adjusted!

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04610.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04970.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04512.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04621.pathview.png

## Info: some node width is different from others, and hence adjusted!

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04270.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04630.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

```

```

## Info: Writing image file mmu04912.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04910.pathview.png

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

## Info: Working in directory /home/juanfraitu/Documents/GitHub/RNA_n_ATAC_hematopoiesis/RNA_ATAC

## Info: Writing image file mmu04144.pathview.png

## Info: some node width is different from others, and hence adjusted!

head(tmp)

##          mmu04380      mmu04810      mmu00190      mmu04510
## plot.data.gene data.frame,10 data.frame,10 data.frame,10 data.frame,10
## plot.data.cpd  data.frame,10 data.frame,10 data.frame,10 data.frame,10
##          mmu04020      mmu04260      mmu04640      mmu04062
## plot.data.gene data.frame,10 data.frame,10 data.frame,10 data.frame,10
## plot.data.cpd  data.frame,10 data.frame,10 NULL        data.frame,10
##          mmu04010      mmu04620      mmu04142      mmu04610
## plot.data.gene data.frame,10 data.frame,10 data.frame,10 data.frame,10
## plot.data.cpd  data.frame,10 NULL        data.frame,10 NULL
##          mmu04970      mmu04512      mmu04621      mmu04270
## plot.data.gene data.frame,10 data.frame,10 data.frame,10 data.frame,10
## plot.data.cpd  data.frame,10 data.frame,10 data.frame,10 data.frame,10
##          mmu04630      mmu04912      mmu04910      mmu04144
## plot.data.gene data.frame,10 data.frame,10 data.frame,10 data.frame,10
## plot.data.cpd  NULL        data.frame,10 data.frame,10 data.frame,10

```