RNA_seq_GSE28166

Sedreh

5/7/2019

```
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
```

```
# BiocManager::install('GEOquery')
# read the dataset into R
library(GEOquery)
```

```
## Loading required package: Biobase
```

```
## 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 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,
##     which.max, which.min
```

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

```
## Setting options('download.file.method.GEOquery'='auto')
```

```
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```
library(limma)
```

```
##
## Attaching package: 'limma'
```

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

```
# library for mouse annotation
library(org.Mm.eg.db)
```

```
## Loading required package: AnnotationDbi
```

```
## Loading required package: stats4
```

```
## Loading required package: IRanges
```

```
## 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
```

```
##
## Attaching package: 'IRanges'
```

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

```
##
## Attaching package: 'AnnotationDbi'
```

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

```
##
```

```
# for collapseBy and other functions
source("/home/sedreh/Documents/rnaseq/functions.r")
### load the dataset here
res <- getGEO("GSE28166", AnnotGPL = TRUE)[[1]]
```

```
## Found 1 file(s)
```

```
## GSE28166_series_matrix.txt.gz
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   ID_REF = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
## File stored at:
```

```
## /tmp/RtmpW9UIEv/GPL6480.annot.gz
```

```
# GEOquery is working, this is a list of files, I can see all the information
# to access individual list I need to use this format res$data@data
# for example, res@experimentData@title will give us details about the experiment
res@experimentData@title
```

```
## [1] ""
```

```
# this is mouse dataset
res@experimentData@abstract
```

```
## [1] ""
```

```
#  HPAI H5N1 pathogenesis is paper's pathway of consideration
```

```
# every GEO data has these internal identifiers: pData is phenotypeData, fData is featureData
str(experimentData(res))
```

```
## Formal class 'MIAME' [package "Biobase"] with 13 slots
##   ..@ name             : chr ""
##   ..@ lab              : chr ""
##   ..@ contact          : chr ""
##   ..@ title            : chr ""
##   ..@ abstract         : chr ""
##   ..@ url              : chr ""
##   ..@ pubMedIds        : chr ""
##   ..@ samples          : list()
##   ..@ hybridizations   : list()
##   ..@ normControls     : list()
##   ..@ preprocessing    : list()
##   ..@ other            : list()
##   ..@ .__classVersion__:Formal class 'Versions' [package "Biobase"] with 1 slot
##   .. .. ..@ .Data:List of 2
##   .. .. .. ..$ : int [1:3] 1 0 0
##   .. .. .. ..$ : int [1:3] 1 1 0
```

```
str(pData(res))
```

```
## 'data.frame':    36 obs. of  40 variables:
##  $ title                 : Factor w/ 36 levels "Mock_0H_1","Mock_0H_2",..: 1 2 3 13 14 15 16 17 18 4 ...
##  $ geo_accession         : chr  "GSM697564" "GSM697565" "GSM697566" "GSM697567" ...
##  $ status                : Factor w/ 1 level "Public on Sep 02 2011": 1 1 1 1 1 1 1 1 1 1 ...
##  $ submission_date       : Factor w/ 1 level "Mar 24 2011": 1 1 1 1 1 1 1 1 1 1 ...
##  $ last_update_date      : Factor w/ 1 level "Sep 02 2011": 1 1 1 1 1 1 1 1 1 1 ...
##  $ type                  : Factor w/ 1 level "RNA": 1 1 1 1 1 1 1 1 1 1 ...
##  $ channel_count         : Factor w/ 1 level "1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ source_name_ch1       : Factor w/ 12 levels "calu3, mock, 0H",..: 1 1 1 5 5 5 6 6 6 2 ...
##  $ organism_ch1          : Factor w/ 1 level "Homo sapiens": 1 1 1 1 1 1 1 1 1 1 ...
##  $ characteristics_ch1   : Factor w/ 1 level "cell line: Calu-3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ characteristics_ch1.1 : Factor w/ 1 level "cell type: lung adenocarcinoma": 1 1 1 1 1 1 1 1 1 1 ...
##  $ characteristics_ch1.2 : Factor w/ 2 levels "infection: mock",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ characteristics_ch1.3 : Factor w/ 6 levels "infection duration: 0h",..: 1 1 1 5 5 5 6 6 6 2 ...
##  $ treatment_protocol_ch1 : Factor w/ 1 level "For RNA isolation, Calu-3 cells were seeded in 6-well plates
(1 x 10^6 cells/well) two days prior to infection."| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ growth_protocol_ch1   : Factor w/ 1 level "Calu-3 cells, a human lung adenocarcinoma cell line, were ki
ndly provided by Dr. Raymond Pickles (University of"| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ molecule_ch1          : Factor w/ 1 level "total RNA": 1 1 1 1 1 1 1 1 1 1 ...
##  $ extract_protocol_ch1  : Factor w/ 1 level "At 0, 3, 7, 12, 18 and 24 hours post-infection (hpi), tripli
cate wells of mock-infected and VN1203-infected Cal"| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ label_ch1             : Factor w/ 1 level "Cy3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ label_protocol_ch1    : Factor w/ 1 level "The Agilent One-Color Microarray-Based Gene Expression Analy
sis Protocol was followed for all processing steps,"| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ taxid_ch1             : Factor w/ 1 level "9606": 1 1 1 1 1 1 1 1 1 1 ...
##  $ hyb_protocol          : Factor w/ 1 level "The Agilent One-Color Microarray-Based Gene Expression Analy
sis Protocol was followed for all processing steps,"| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ scan_protocol         : Factor w/ 1 level "Dry slides were scanned on an Agilent DNA microarray scanner
(Model G2505B) using the XDR setting.": 1 1 1 1 1 1 1 1 1 1 ...
##  $ description           : Factor w/ 36 levels "251485048465_1_1",..: 5 35 21 15 1 14 9 24 34 25 ...
##  $ description.1         : Factor w/ 12 levels "Mock host response 0H.",..: 1 1 1 5 5 5 6 6 6 2 ...
##  $ data_processing       : Factor w/ 1 level "Raw images were analyzed using the Agilent Feature Extractio
n software (version 9.5.3.1) and the GE1-v5_95_Feb0"| __truncated__: 1 1 1 1 1 1 1 1 1 1 ...
##  $ platform_id           : Factor w/ 1 level "GPL6480": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_name          : Factor w/ 1 level "Armand,,Bankhead III": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_department    : Factor w/ 1 level "Depatartment of Medical Informatics and Clinical Epidemiolog
y": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_institute     : Factor w/ 1 level "Oregon Health and Science University": 1 1 1 1 1 1 1 1 1 1
...
##  $ contact_address       : Factor w/ 1 level "3181 SW Sam Jackson Park Rd.": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_city          : Factor w/ 1 level "Portland": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_state         : Factor w/ 1 level "OR": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_zip/postal_code: Factor w/ 1 level "97080": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_country       : Factor w/ 1 level "USA": 1 1 1 1 1 1 1 1 1 1 ...
##  $ supplementary_file    : Factor w/ 36 levels "ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM697nnn/GSM69756
4/suppl/GSM697564.txt.gz",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ data_row_count        : Factor w/ 1 level "41000": 1 1 1 1 1 1 1 1 1 1 ...
##  $ cell line:ch1         : chr  "Calu-3" "Calu-3" "Calu-3" "Calu-3" ...
##  $ cell type:ch1         : chr  "lung adenocarcinoma" "lung adenocarcinoma" "lung adenocarcinoma" "lung ad
enocarcinoma" ...
##  $ infection duration:ch1 : chr  "0h" "0h" "0h" "3h" ...
##  $ infection:ch1         : chr  "mock" "mock" "mock" "mock" ...
```

```
head(fData(res))
```

```
##                        ID
## A_23_P100001 A_23_P100001
## A_23_P100011 A_23_P100011
## A_23_P100022 A_23_P100022
## A_23_P100056 A_23_P100056
## A_23_P100074 A_23_P100074
## A_23_P100092 A_23_P100092
##                                                 Gene title Gene symbol
## A_23_P100001      family with sequence similarity 174 member B     FAM174B
## A_23_P100011 adaptor related protein complex 3 sigma 2 subunit       AP3S2
## A_23_P100022                  synaptic vesicle glycoprotein 2B        SV2B
## A_23_P100056      RNA binding protein with multiple splicing 2      RBPMS2
## A_23_P100074        apoptosis and caspase activation inhibitor        AVEN
## A_23_P100092         zinc finger and SCAN domain containing 29     ZSCAN29
##              Gene ID UniGene title UniGene symbol UniGene ID
## A_23_P100001  400451
## A_23_P100011   10239
## A_23_P100022    9899
## A_23_P100056  348093
## A_23_P100074   57099
## A_23_P100092  146050
##                                                                                 Nucleotide T
itle
## A_23_P100001                           Homo sapiens family with sequence similarity 174 member B (FAM174B),
mRNA
## A_23_P100011 Homo sapiens adaptor related protein complex 3 sigma 2 subunit (AP3S2), transcript variant 1,
mRNA
## A_23_P100022                  Homo sapiens synaptic vesicle glycoprotein 2B (SV2B), transcript variant 1,
mRNA
## A_23_P100056                           Homo sapiens RNA binding protein with multiple splicing 2 (RBPMS2),
mRNA
## A_23_P100074                              Homo sapiens apoptosis and caspase activation inhibitor (AVEN),
mRNA
## A_23_P100092                            Homo sapiens zinc finger and SCAN domain containing 29 (ZSCAN29),
mRNA
##                      GI GenBank Accession Platform_CLONEID Platform_ORF
## A_23_P100001  150170693         NM_207446              NA           NA
## A_23_P100011  189409107         NM_005829              NA           NA
## A_23_P100022 1019366977         NM_014848              NA           NA
## A_23_P100056   34915989         NM_194272              NA           NA
## A_23_P100074   56699476         NM_020371              NA           NA
## A_23_P100092  109715824         NM_152455              NA           NA
##              Platform_SPOTID Chromosome location
## A_23_P100001    A_23_P100001             15q26.1
## A_23_P100011    A_23_P100011             15q26.1
## A_23_P100022    A_23_P100022             15q26.1
## A_23_P100056    A_23_P100056            15q22.31
## A_23_P100074    A_23_P100074             15q13.1
## A_23_P100092    A_23_P100092             15q15.3
##                                                 Chromosome annotation
## A_23_P100001 Chromosome 15, NC_000015.10 (92617447..92734219, complement)
## A_23_P100011 Chromosome 15, NC_000015.10 (89830599..89894385, complement)
## A_23_P100022             Chromosome 15, NC_000015.10 (91099588..91301309)
## A_23_P100056 Chromosome 15, NC_000015.10 (64739894..64775571, complement)
## A_23_P100074 Chromosome 15, NC_000015.10 (33858602..34074877, complement)
## A_23_P100092 Chromosome 15, NC_000015.10 (43358172..43371032, complement)
##
GO:Function
## A_23_P100001
## A_23_P100011                                                                          protein tr
ansporter activity
## A_23_P100022                                                              protein binding///transmembrane tr
ansporter activity
## A_23_P100056                           mRNA binding///nucleotide binding///protein binding///protein homodim
erization activity
## A_23_P100074
protein binding
## A_23_P100092 DNA binding///RNA polymerase II transcription factor activity, sequence-specific DNA bindin
g///metal ion binding
##
GO:Process
## A_23_P100001
## A_23_P100011
```

```
anterograde axonal transport///anterograde synaptic vesicle transport///intracellular protein transport
## A_23_P100022
neurotransmitter transport///transmembrane transport
## A_23_P100056 embryonic digestive tract morphogenesis///negative regulation of BMP signaling pathway///negat
ive regulation of smooth muscle cell differentiation///negative regulation of smooth muscle cell differentiati
on///positive regulation of smooth muscle cell proliferation
## A_23_P100074
apoptotic process///negative regulation of apoptotic process
## A_23_P100092
regulation of transcription from RNA polymerase II promoter///transcription, DNA-templated
##
GO:Component
## A_23_P100001
integral component of membrane
## A_23_P100011                                  AP-3 adaptor complex///Golgi apparatus///axon cytoplasm///cyto
plasmic vesicle membrane///intracellular membrane-bounded organelle
## A_23_P100022 acrosomal vesicle///cell junction///integral component of membrane///membrane///plasma membran
e///synaptic vesicle///synaptic vesicle///synaptic vesicle membrane
## A_23_P100056
cytoplasm
## A_23_P100074
endomembrane system///intracellular///membrane
## A_23_P100092
nucleus
##                                             GO:Function ID
## A_23_P100001
## A_23_P100011                                      GO:0008565
## A_23_P100022                             GO:0005515///GO:0022857
## A_23_P100056 GO:0003729///GO:0000166///GO:0005515///GO:0042803
## A_23_P100074                                      GO:0005515
## A_23_P100092              GO:0003677///GO:0000981///GO:0046872
##                                                          GO:Process ID
## A_23_P100001
## A_23_P100011                             GO:0008089///GO:0048490///GO:0006886
## A_23_P100022                                      GO:0006836///GO:0055085
## A_23_P100056 GO:0048557///GO:0030514///GO:0051151///GO:0051151///GO:0048661
## A_23_P100074                                      GO:0006915///GO:0043066
## A_23_P100092                                      GO:0006357///GO:0006351
##                                                                                                  GO:Compo
nent ID
## A_23_P100001                                                                                          GO:
0016021
## A_23_P100011                                      GO:0030123///GO:0005794///GO:1904115///GO:0030659///GO:
0043231
## A_23_P100022 GO:0001669///GO:0030054///GO:0016021///GO:0016020///GO:0005886///GO:0008021///GO:0008021///GO:
0030672
## A_23_P100056                                                                                          GO:
0005737
## A_23_P100074                                                                GO:0012505///GO:0005622///GO:
0016020
## A_23_P100092                                                                                          GO:
0005634
##                                                    Platform_SEQUENCE
## A_23_P100001 ATCTCATGGAAAAGCTGGATTCCTCTGCCTTACGCAGAAACACCCGGGCTCCATCTGCCA
## A_23_P100011 TCAAGTATTGGCCTGACATAGAGTCCTTAAGACAAGCAAAGACAAGCAAGGCAAGCACGT
## A_23_P100022 ATGTCGGCTGTGGAGGGTTAAAGGGATGAGGCTTTCCTTTGTTTAGCAAATCTGTTCACA
## A_23_P100056 CCCTGTCAGATAAGTTTAATGTTTAGTTTGAGGCATGAAGAAGAAAAGGGTTTCCATTCT
## A_23_P100074 GACCAGCCAGTTTACAAGCATGTCTCAAGCTAGTGTGTTCCATTATGCTCACAGCAGTAA
## A_23_P100092 AGAGAAACCCTATGGGTGTCATGACTGTGGTAAGTGCTTCAGTAAAAGCTCTGCCCTTAA
```

```
# this will help us in identifying condition, we need to modify the data according to condition
res$`infection:ch1`
```

```
##  [1] "mock"                   "mock"
##  [3] "mock"                   "mock"
##  [5] "mock"                   "mock"
##  [7] "mock"                   "mock"
##  [9] "mock"                   "mock"
## [11] "mock"                   "mock"
## [13] "mock"                   "mock"
## [15] "mock"                   "mock"
## [17] "mock"                   "mock"
## [19] "VN1203 influenza virus" "VN1203 influenza virus"
## [21] "VN1203 influenza virus" "VN1203 influenza virus"
## [23] "VN1203 influenza virus" "VN1203 influenza virus"
## [25] "VN1203 influenza virus" "VN1203 influenza virus"
## [27] "VN1203 influenza virus" "VN1203 influenza virus"
## [29] "VN1203 influenza virus" "VN1203 influenza virus"
## [31] "VN1203 influenza virus" "VN1203 influenza virus"
## [33] "VN1203 influenza virus" "VN1203 influenza virus"
## [35] "VN1203 influenza virus" "VN1203 influenza virus"
```

```
# so we can see that infection versus normal is our condition, let's store this to a new column
# of condition
# here with gsub, we are just cleaning the data. each entry begins with the symbols \\+, _
# we have to CLEAN these symbols
res$condition <- gsub("\\+", "_", res$`infection:ch1`)
```

```
res$condition
```

```
##  [1] "mock"                   "mock"
##  [3] "mock"                   "mock"
##  [5] "mock"                   "mock"
##  [7] "mock"                   "mock"
##  [9] "mock"                   "mock"
## [11] "mock"                   "mock"
## [13] "mock"                   "mock"
## [15] "mock"                   "mock"
## [17] "mock"                   "mock"
## [19] "VN1203 influenza virus" "VN1203 influenza virus"
## [21] "VN1203 influenza virus" "VN1203 influenza virus"
## [23] "VN1203 influenza virus" "VN1203 influenza virus"
## [25] "VN1203 influenza virus" "VN1203 influenza virus"
## [27] "VN1203 influenza virus" "VN1203 influenza virus"
## [29] "VN1203 influenza virus" "VN1203 influenza virus"
## [31] "VN1203 influenza virus" "VN1203 influenza virus"
## [33] "VN1203 influenza virus" "VN1203 influenza virus"
## [35] "VN1203 influenza virus" "VN1203 influenza virus"
```

```
# our conditions are Mock versus Infection, as we can see there is white spaces in the name of contition! we t
ried to clean(we just made a vector with names without spaces)

res$condition <- c("mock","mock","mock","mock","mock","mock","mock","mock","mock","mock","mock","mock","mock"
,"mock","mock","mock","mock","mock","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus"
,"VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_i
nfluenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_v
irus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1203_influenza_virus","VN1
203_influenza_virus","VN1203_influenza_virus")
```

```
# Now we collapse the dataset with genesymbols, similar to what we did in phantasus
res <- collapseBy(res, fData(res)$`Gene symbol`, FUN=median)
res <- res[!grepl("///", rownames(res)), ]
res <- res[rownames(res) != "", ]

# We can see that expressionset size has been reduced from 44 Mb to 33.5 Mb
```

```
# let's annotate the symbols with the mouse database entries

fData(res) <- data.frame(row.names = rownames(res))

fData(res)$entrez <- row.names(fData(res))

fData(res)$symbol <- mapIds(org.Mm.eg.db, keys=fData(res)$entrez, keytype = "SYMBOL",
                            column="ENTREZID" )
```

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

```
# let's normalize this data

res.qnorm <- res

summary(exprs(res.qnorm))
```

```
##    GSM697564          GSM697565           GSM697566
##  Min.   : 0.03565   Min.   : 0.007331   Min.   : 0.03565
##  1st Qu.: 3.64027   1st Qu.: 3.631503   1st Qu.: 3.65700
##  Median : 7.48323   Median : 7.454318   Median : 7.48433
##  Mean   : 7.23275   Mean   : 7.220964   Mean   : 7.24057
##  3rd Qu.:10.14789   3rd Qu.:10.112703   3rd Qu.:10.15102
##  Max.   :18.34811   Max.   :18.348106   Max.   :18.34811
##    GSM697567          GSM697568           GSM697569
##  Min.   : 0.0233   Min.   : 0.007331   Min.   : 0.01697
##  1st Qu.: 3.6378   1st Qu.: 3.625616   1st Qu.: 3.66566
##  Median : 7.4235   Median : 7.459471   Median : 7.41386
##  Mean   : 7.2116   Mean   : 7.229724   Mean   : 7.21759
##  3rd Qu.:10.1211   3rd Qu.:10.139448   3rd Qu.:10.10947
##  Max.   :18.3481   Max.   :18.348106   Max.   :18.34811
##    GSM697570           GSM697571          GSM697572
##  Min.   : 0.007331   Min.   : 0.02879   Min.   : 0.0233
##  1st Qu.: 3.656028   1st Qu.: 3.66078   1st Qu.: 3.6757
##  Median : 7.455088   Median : 7.48520   Median : 7.4618
##  Mean   : 7.235898   Mean   : 7.25657   Mean   : 7.2417
##  3rd Qu.:10.129304   3rd Qu.:10.15556   3rd Qu.:10.1131
##  Max.   :18.348106   Max.   :18.34811   Max.   :18.3481
##    GSM697573          GSM697574         GSM697575
##  Min.   : 0.02879   Min.   : 0.0233   Min.   : 0.05121
##  1st Qu.: 3.60308   1st Qu.: 3.6704   1st Qu.: 3.66195
##  Median : 7.48614   Median : 7.4747   Median : 7.49215
##  Mean   : 7.24124   Mean   : 7.2371   Mean   : 7.24580
##  3rd Qu.:10.15392   3rd Qu.:10.1425   3rd Qu.:10.17747
##  Max.   :18.26469   Max.   :18.3481   Max.   :18.34811
##    GSM697576          GSM697577         GSM697578         GSM697579
##  Min.   : 0.01697   Min.   : 0.0233   Min.   : 0.0233   Min.   : 0.08434
##  1st Qu.: 3.61087   1st Qu.: 3.6285   1st Qu.: 3.6654   1st Qu.: 3.73452
##  Median : 7.47417   Median : 7.4962   Median : 7.4705   Median : 7.54964
##  Mean   : 7.23536   Mean   : 7.2392   Mean   : 7.2479   Mean   : 7.27571
##  3rd Qu.:10.13945   3rd Qu.:10.1461   3rd Qu.:10.1593   3rd Qu.:10.18674
##  Max.   :18.34811   Max.   :18.2333   Max.   :18.3481   Max.   :18.23333
##    GSM697580           GSM697581           GSM697582
##  Min.   : 0.007331   Min.   : 0.007331   Min.   : 0.007331
##  1st Qu.: 3.656292   1st Qu.: 3.694247   1st Qu.: 3.642581
##  Median : 7.539701   Median : 7.545358   Median : 7.457009
##  Mean   : 7.251238   Mean   : 7.265125   Mean   : 7.224317
##  3rd Qu.:10.161803   3rd Qu.:10.160712   3rd Qu.:10.127042
##  Max.   :18.104417   Max.   :18.297660   Max.   :18.297660
##    GSM697583          GSM697584           GSM697585
##  Min.   : 0.03565   Min.   : 0.007331   Min.   : 0.06702
##  1st Qu.: 3.61846   1st Qu.: 3.649248   1st Qu.: 3.61913
##  Median : 7.46085   Median : 7.450258   Median : 7.42460
##  Mean   : 7.22663   Mean   : 7.221045   Mean   : 7.21118
##  3rd Qu.:10.13625   3rd Qu.:10.116507   3rd Qu.:10.09863
##  Max.   :18.34811   Max.   :18.348106   Max.   :18.34811
##    GSM697586          GSM697587          GSM697588
##  Min.   : 0.02879   Min.   : 0.01697   Min.   : 0.03565
##  1st Qu.: 3.64729   1st Qu.: 3.65977   1st Qu.: 3.73344
##  Median : 7.43622   Median : 7.42150   Median : 7.37744
##  Mean   : 7.22061   Mean   : 7.20490   Mean   : 7.21066
##  3rd Qu.:10.11871   3rd Qu.:10.09924   3rd Qu.:10.07747
##  Max.   :18.34811   Max.   :18.34811   Max.   :18.29766
##    GSM697589          GSM697590          GSM697591
##  Min.   : 0.01697   Min.   : 0.03565   Min.   : 0.01697
##  1st Qu.: 3.78029   1st Qu.: 3.69254   1st Qu.: 3.74005
##  Median : 7.38426   Median : 7.37466   Median : 7.24044
##  Mean   : 7.21240   Mean   : 7.20245   Mean   : 7.13731
##  3rd Qu.:10.06635   3rd Qu.:10.08596   3rd Qu.: 9.96806
##  Max.   :18.34811   Max.   :18.34811   Max.   :18.34811
##    GSM697592          GSM697593          GSM697594
##  Min.   : 0.01697   Min.   : 0.02879   Min.   : 0.01697
##  1st Qu.: 3.81396   1st Qu.: 3.79480   1st Qu.: 3.60007
##  Median : 7.28819   Median : 7.26864   Median : 7.11797
##  Mean   : 7.17099   Mean   : 7.15159   Mean   : 7.05523
##  3rd Qu.:10.00746   3rd Qu.: 9.98313   3rd Qu.: 9.90169
##  Max.   :18.29766   Max.   :18.34811   Max.   :18.34811
##    GSM697595           GSM697596          GSM697597
##  Min.   : 0.007331   Min.   : 0.02879   Min.   : 0.0233
##  1st Qu.: 3.769260   1st Qu.: 3.76377   1st Qu.: 3.7418
```

```
##  Median : 7.189588   Median : 7.17627   Median : 7.1448
##  Mean   : 7.118579   Mean   : 7.09740   Mean   : 7.0773
##  3rd Qu.: 9.937001   3rd Qu.: 9.91234   3rd Qu.: 9.8920
##  Max.   :18.264694   Max.   :18.34811   Max.   :18.1044
##    GSM697598         GSM697599
##  Min.   : 0.0233   Min.   : 0.2458
##  1st Qu.: 3.7399   1st Qu.: 3.7945
##  Median : 7.2307   Median : 7.1886
##  Mean   : 7.1128   Mean   : 7.0844
##  3rd Qu.: 9.9280   3rd Qu.: 9.8855
##  Max.   :18.2977   Max.   :18.2977
```

```
exprs(res.qnorm) <- normalizeBetweenArrays(log2(exprs(res.qnorm)+1), method="quantile")
summary(exprs(res.qnorm))
```

```
##    GSM697564         GSM697565         GSM697566         GSM697567
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697568         GSM697569         GSM697570         GSM697571
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697572         GSM697573         GSM697574         GSM697575
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697576         GSM697577         GSM697578         GSM697579
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697580         GSM697581         GSM697582         GSM697583
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697584         GSM697585         GSM697586         GSM697587
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697588         GSM697589         GSM697590         GSM697591
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697592         GSM697593         GSM697594         GSM697595
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
##    GSM697596         GSM697597         GSM697598         GSM697599
##  Min.   :0.04288   Min.   :0.04288   Min.   :0.04288   Min.   :0.04288
##  1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722   1st Qu.:2.22722
##  Median :3.06967   Median :3.06967   Median :3.06967   Median :3.06967
##  Mean   :2.84366   Mean   :2.84366   Mean   :2.84366   Mean   :2.84366
##  3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009   3rd Qu.:3.47009
##  Max.   :4.27165   Max.   :4.27165   Max.   :4.27165   Max.   :4.27165
```

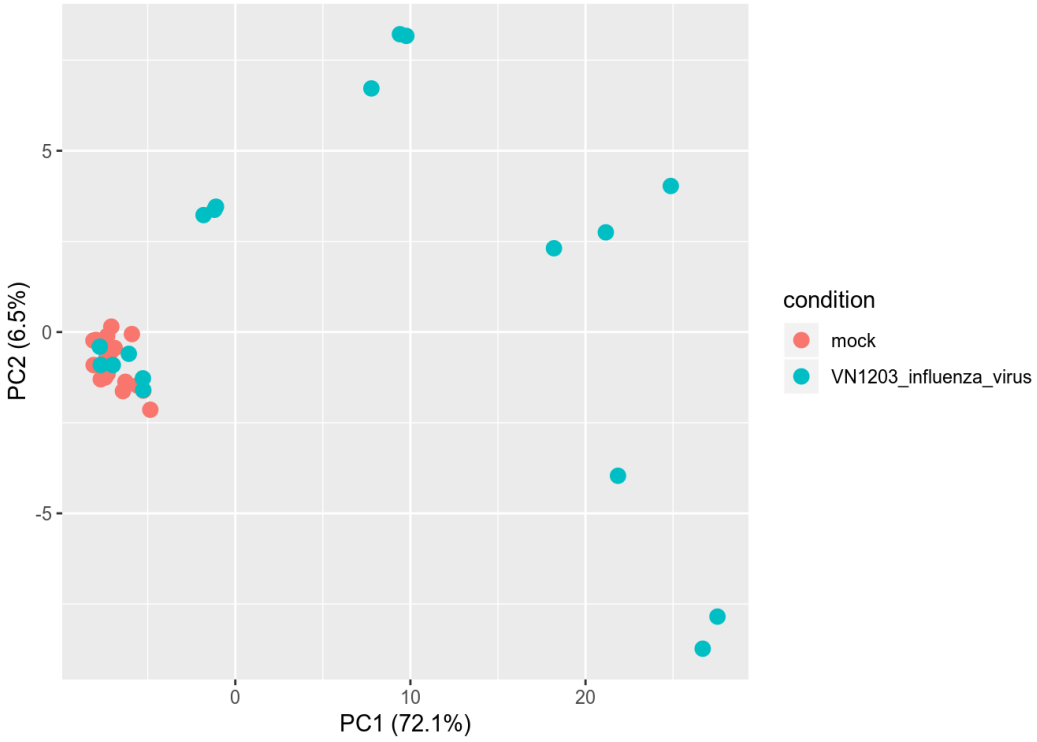
```
res.qnorm.top12K <- res.qnorm
# let's get top 12000 entries
res.qnorm.top12K <- res.qnorm.top12K[head(order(apply(exprs(res.qnorm.top12K), 1, mean),
                                                decreasing = TRUE), 12000), ]
```

```
# Now let's look at the dataset
#pdf('pca_dataset1.pdf')

#also we can make PCA plot from our dataset
pcaPlot(res.qnorm.top12K, 1, 2) + aes(color=condition)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```




```
# dev.off()
```

```
# we can see that there are spaces in the names, we need to convert these names
#res.design <- gsub(" ", "_", names(res.design))
#res.design$condition <- c("conditionmock", "conditionVN1203_influenza _virus")
```

```
# Now we make a design matrix that will be used to make a model for the given data
res.design <- model.matrix(~0+condition, data=pData(res.qnorm.top12K))
# based on this matrix we fit our data
fit <- lmFit(res.qnorm.top12K, res.design)

# we will also make a bayisian model for the data called fit2
# this is the tricky part, because we need to choose contrast names
# we call mock as 'conditionmock' and infection as 'conditionvn1203_influenza_virus'

fit2 <- contrasts.fit(fit, makeContrasts(conditionmock-conditionVN1203_influenza_virus,
                                         levels=res.design))

# View(res.design)
fit2 <- eBayes(fit2)

# now let's do a bonferroni-hochback correction
de <- topTable(fit2, adjust.method="BH", number=Inf)
head(de)
```

```
##              entrez symbol      logFC  AveExpr         t      P.Value
## IFI30         IFI30   <NA>  0.1156056 3.564850  8.316578 6.831861e-10
## LOC338620 LOC338620   <NA> -0.1179840 3.439782 -8.201655 9.519608e-10
## BORCS6       BORCS6   <NA>  0.1773008 3.492265  8.006566 1.677763e-09
## NR2F1-AS1 NR2F1-AS1   <NA> -0.1691660 3.380849 -7.917554 2.175934e-09
## IDI2-AS1   IDI2-AS1   <NA> -0.2425770 3.397217 -7.829545 2.816195e-09
## CCDC58       CCDC58   <NA> -0.1085168 3.414474 -7.726773 3.810072e-09
##              adj.P.Val        B
## IFI30     5.711765e-06 12.55917
## LOC338620 5.711765e-06 12.24208
## BORCS6    5.772264e-06 11.69996
## NR2F1-AS1 5.772264e-06 11.45104
## IDI2-AS1  5.772264e-06 11.20399
## CCDC58    5.772264e-06 10.91434
```

```
res.design
```

```
##           conditionmock conditionVN1203_influenza_virus
## GSM697564             1                               0
## GSM697565             1                               0
## GSM697566             1                               0
## GSM697567             1                               0
## GSM697568             1                               0
## GSM697569             1                               0
## GSM697570             1                               0
## GSM697571             1                               0
## GSM697572             1                               0
## GSM697573             1                               0
## GSM697574             1                               0
## GSM697575             1                               0
## GSM697576             1                               0
## GSM697577             1                               0
## GSM697578             1                               0
## GSM697579             1                               0
## GSM697580             1                               0
## GSM697581             1                               0
## GSM697582             0                               1
## GSM697583             0                               1
## GSM697584             0                               1
## GSM697585             0                               1
## GSM697586             0                               1
## GSM697587             0                               1
## GSM697588             0                               1
## GSM697589             0                               1
## GSM697590             0                               1
## GSM697591             0                               1
## GSM697592             0                               1
## GSM697593             0                               1
## GSM697594             0                               1
## GSM697595             0                               1
## GSM697596             0                               1
## GSM697597             0                               1
## GSM697598             0                               1
## GSM697599             0                               1
## attr(,"assign")
## [1] 1 1
## attr(,"contrasts")
## attr(,"contrasts")$condition
## [1] "contr.treatment"
```

```
# now we can use the big matrix de, to select top differentially expressed genes using p-values
# we can also make pca's, heatmaps etc. But most importantly, we can do pathway analysis
####
# FGSEA
####

library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following object is masked from 'package:IRanges':
##
##     shift
```

```
## The following objects are masked from 'package:S4Vectors':
##
##     first, second
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
de <- as.data.table(de, keep.rownames=TRUE)
de[entrez == "REST"]
```

```
##      rn entrez symbol      logFC  AveExpr         t      P.Value
## 1: REST   REST   <NA> -0.1520649 3.376595 -4.522153 6.433924e-05
##       adj.P.Val        B
## 1: 0.0003463754 1.566821
```

```
# we can see that de matrix stores information about the gene expression
```

```
# Let's make a new matrix de2 which will store information about pathways
de2 <- data.frame(de$entrez, de$t)
colnames(de2) <- c('ENTREZ', 'stat')

# BiocManager::install('fgsea')
library(fgsea)
```

```
## Loading required package: Rcpp
```

```
library(tibble)

# let's get the rank of genes from top differentially expressed to non significant
ranks <- deframe(de2)
head(ranks, 20)
```

```
##     IFI30 LOC338620    BORCS6 NR2F1-AS1  IDI2-AS1    CCDC58     IP6K2
##  8.316578 -8.201655  8.006566 -7.917554 -7.829545 -7.726773  7.726724
##     PDAP1     BUD31    ANP32E      NAT9      LRR1    TXNL4A    GOLT1B
## -7.723396 -7.643699 -7.571418  7.563921 -7.557678 -7.459139 -7.433903
##      ERI3    EIF1AY      SSR1     AHSA2     MKNK1      DENR
## -7.412102 -7.384894 -7.381185  7.336755  7.306017 -7.296162
```

```
# Load the pathways into a named list
# BiocManager::install('msigdbr')
library(msigdbr)

m_df <- msigdbr(species = "Mus musculus")
# View(m_df)
pathways <- split(m_df$human_gene_symbol, m_df$gs_name)
head(pathways)
```

```
## $AAACCAC_MIR140
##   [1] "ABCC4"    "ABRAXAS2" "ACTN4"    "ACVR1"    "ADAM9"    "ADAMTS5"
##   [7] "AGER"     "AMER2"    "ANK2"     "API5"     "BACH1"    "BAZ2B"
##  [13] "BCL11A"   "BCL2L2"   "BCL9"     "BMT2"     "C1orf21"  "CACNA1C"
##  [19] "CEBPA"    "CHD4"     "CIT"      "COL23A1"  "CSK"      "CSNK1G3"
##  [25] "CTCF"     "CUL3"     "DAZL"     "DBNDD2"   "DCUN1D4"  "DDX3X"
##  [31] "DDX3Y"    "DHX57"    "DIPK2A"   "DPP4"     "DSCAM"    "DTNA"
##  [37] "E2F3"     "EHD1"     "EPHB1"    "ERC2"     "ETV3"     "EYA2"
##  [43] "FAM214A"  "GABARAP"  "GALNT16"  "GDF6"     "GIT1"     "GYS1"
##  [49] "HDAC4"    "HNRNPH3"  "HSPA13"   "IGFBP5"   "KATNBL1"  "KCND2"
##  [55] "LOXL3"    "LRRC4"    "LRRC8E"   "MAP3K8"   "MDGA2"    "MEX3C"
##  [61] "MGAT1"    "MMD"      "NAV3"     "NKIRAS2"  "NR3C1"    "NSD3"
##  [67] "NUTF2"    "OGT"      "OSTM1"    "PDGFRA"   "PFN1"     "PHF20L1"
##  [73] "PHYHIP"   "PITX2"    "PPP1CC"   "PRIMA1"   "R3HDM1"   "REEP1"
##  [79] "RNF19A"   "RTKN2"    "SENP1"    "SIAH1"    "SLC25A13" "SLC38A2"
##  [85] "SLC41A2"  "SLF2"     "SLMAP"    "SNX2"     "SOX4"     "SRR"
##  [91] "STAG1"    "STRADB"   "SYT6"     "TAF9B"    "TBX3"     "TP53INP2"
##  [97] "TSHZ1"    "TSPAN2"   "TSSK2"    "TTYH2"    "UBASH3B"  "USP6"
## [103] "VEGFA"    "WNT1"     "YES1"     "ZBED4"    "ZBTB10"   "ZNF182"
## [109] "ZNF608"   "ZNF654"
##
## $AAAGACA_MIR511
##   [1] "ABCG8"     "ACE"       "ADAMTSL3"  "ADGRF5"    "ADSS"
##   [6] "AGBL3"     "AGO1"      "AGO2"      "AGO4"      "ALCAM"
##  [11] "ANAPC15"   "ANKRD40CL" "ANKZF1"    "AQP6"      "ARHGEF17"
##  [16] "ATL2"      "ATP2B2"    "ATRX"      "BCL11A"    "BTG1"
##  [21] "BUB3"      "C1orf21"   "C1QL2"     "C6orf106"  "CALM1"
##  [26] "CAMK2N1"   "CAMTA1"    "CAPRIN1"   "CCDC178"   "CCND1"
##  [31] "CCNT2"     "CDH2"      "CDK14"     "CDK19"     "CELF1"
##  [36] "CELF6"     "CEP350"    "CFAP298"   "CLK2"      "CLTC"
##  [41] "CNOT4"     "CORIN"     "CREBRF"    "CREM"      "CRIM1"
##  [46] "DCTN4"     "DDX3X"     "DDX3Y"     "DEDD"      "DNAJB12"
##  [51] "DNAJC13"   "DSC1"      "DUSP6"     "DYRK1B"    "E2F3"
##  [56] "EDEM3"     "EFR3A"     "ELAVL3"    "EMILIN2"   "EML4"
##  [61] "ENPP1"     "ENPP4"     "EPHA4"     "ESRRG"     "EYA1"
##  [66] "EYA4"      "FAM117A"   "FGF13"     "FIP1L1"    "FMR1"
##  [71] "FN1"       "FNDC1"     "FNDC5"     "FOXK2"     "FOXN3"
##  [76] "GAD2"      "GEMIN2"    "GFAP"      "GJA1"      "GLRA2"
##  [81] "HAS2"      "HCN4"      "HLF"       "HLTF"      "HOXA13"
##  [86] "IGF2BP1"   "IGF2BP3"   "KCNE1"     "KCNMA1"    "KHDRBS2"
##  [91] "KLF9"      "KLHL18"    "KLHL24"    "LATS1"     "LMCD1"
##  [96] "LPP"       "LRCH4"     "LUC7L3"    "MAP3K2"    "MAP4K4"
## [101] "MAPK1IP1L" "MBD2"      "MBD6"      "MDGA2"     "METAP2"
## [106] "MIB1"      "MINK1"     "MRPL21"    "MSTN"      "MTAP"
## [111] "MYCBP"     "MYO19"     "NACC1"     "NEUROD6"   "NHLH2"
## [116] "NLK"       "NR4A2"     "NRXN3"     "NTRK2"     "NXPH1"
## [121] "ONECUT2"   "PAX8"      "PCARE"     "PCDH10"    "PCDH17"
## [126] "PELI1"     "PHLPP1"    "PIK3R3"    "PMEPA1"    "POGK"
## [131] "POU4F2"    "PPARGC1A"  "PRELP"     "PRPF4B"    "PSMA1"
## [136] "PSMD10"    "QKI"       "RAB22A"    "RAB2A"     "RBM15B"
## [141] "RBM26"     "RECK"      "REV3L"     "RGL1"      "RHOJ"
## [146] "RHOT1"     "RNF19A"    "ROBO2"     "RPS6KB1"   "RPS6KL1"
## [151] "SATB2"     "SCN4B"     "SELENOP"   "SEMA3F"    "SEMA6D"
## [156] "SINHCAF"   "SLC22A17"  "SLC25A26"  "SLC6A6"    "SLITRK1"
## [161] "SMARCE1"   "SOCS2"     "SORCS3"    "SOST"      "SOX12"
## [166] "SPTBN4"    "SPTLC2"    "SRGAP3"    "SS18"      "ST18"
## [171] "SYT11"     "TAF5"      "TBXT"      "THOC5"     "TIAL1"
## [176] "TMEM196"   "TMEM243"   "TMEM248"   "TNRC6A"    "TNRC6B"
## [181] "TOB1"      "TRAPPC3"   "TRAPPC8"   "TRIM2"     "TRIM24"
## [186] "TSPOAP1"   "TXNL1"     "UBE2H"     "VANGL2"    "VAV3"
## [191] "VIRMA"     "VKORC1L1"  "VMP1"      "WNT16"     "YTHDF2"
## [196] "YY1"       "ZADH2"     "ZCCHC24"   "ZDHHC21"   "ZNF319"
## [201] "ZNF654"    "ZNF706"
##
## $AAAGGAT_MIR501
##   [1] "ACACA"    "ACADSB"   "ADCYAP1"  "ADIPOR2"  "ALS2"     "AMMECR1"
##   [7] "APOLD1"   "ATP6V1H"  "BCL6"     "BCLAF1"   "C8orf82"  "CA6"
##  [13] "CACHD1"   "CAMTA1"   "CD164"    "CELF2"    "CELSR2"   "CHODL"
##  [19] "CLK1"     "CLK2"     "CTDSP1"   "CTDSPL2"  "CUL1"     "CUX2"
##  [25] "DCX"      "DNAJB12"  "ELAVL4"   "ERRFI1"   "GIF"      "GRAMD4"
##  [31] "GRB10"    "H2AFX"    "HAS2"     "HES5"     "HOXB8"    "JADE3"
##  [37] "JUN"      "KCND2"    "KCNRG"    "KIF1C"    "KIF2A"    "KLHL14"
##  [43] "KRR1"     "LARP1"    "LEPROTL1" "LPGAT1"   "LPIN1"    "LRRC1"
```

```
##  [49] "MAP2K1"   "MAP3K8"   "MCU"      "MEF2C"    "MYB"      "MYCL"
##  [55] "MYLK"     "NEXMIF"   "NFASC"    "NFIL3"    "NFIX"     "NPR3"
##  [61] "NR2F2"    "NR4A3"    "PCDH19"   "PDK1"     "PHC1"     "PHF6"
##  [67] "PIK3AP1"  "PITX2"    "PLP1"     "PLXNB1"   "PNN"      "PPP1CB"
##  [73] "PPP2R5E"  "PPP4R3A"  "PPP6R3"   "PRKCE"    "PURA"     "QKI"
##  [79] "RAB22A"   "RABGEF1"  "RASL10B"  "RCN1"     "RDX"      "RET"
##  [85] "RGL1"     "RNF11"    "ROBO2"    "RPGRIP1L" "RSBN1"    "SATB2"
##  [91] "SCN3A"    "SENP3"    "SEPHS1"   "SGPP1"    "SLC25A3"  "SLC35B3"
##  [97] "SLITRK5"  "SMC1A"    "SNAP29"   "SOX11"    "SOX4"     "SPOPL"
## [103] "SRR"      "SRSF2"    "SYNC"     "SYNJ1"    "SYT7"     "TAF5L"
## [109] "TAPT1"    "TNNI2"    "TOGARAM1" "TOMM70"   "TRIM39"   "UBAP1"
## [115] "UBE2Q1"   "UBE4B"    "USP12"    "VDAC2"    "WDFY3"    "WIPF2"
## [121] "WT1-AS"   "ZBTB18"   "ZC3H7A"   "ZIC4"     "ZMYM5"
##
## $AAAGGGA_MIR204_MIR211
##   [1] "ABRAXAS2" "ADAMTS9"  "ADCY6"    "ADPRM"    "AGO4"     "AKAP1"
##   [7] "ALPL"     "ANGPT1"   "ANKRD13A" "ANXA11"   "AP1S1"    "AP1S3"
##  [13] "AP2A2"    "AP3M1"    "APH1A"    "ARAP2"    "ARCN1"    "ARGLU1"
##  [19] "ARHGAP29" "ARL8B"    "ATF2"     "ATP2B1"   "AUP1"     "BAZ2A"
##  [25] "BCL11B"   "BCL2"     "BCL9"     "BCL9L"    "BRD4"     "BRPF3"
##  [31] "BUD31"    "C16orf72" "CAPRIN1"  "CCNT2"    "CCPG1"    "CDC25B"
##  [37] "CDC42"    "CDH2"     "CELSR3"   "CHD5"     "CHN2"     "CHP1"
##  [43] "CLIP1"    "CORO1C"   "COX5A"    "CPD"      "CPNE8"    "CREB5"
##  [49] "CRKL"     "CTDNEP1"  "DAG1"     "DCAF5"    "DCUN1D3"  "DENND5A"
##  [55] "DHH"      "DLG5"     "DMTF1"    "DNAJC13"  "DNM2"     "DTX1"
##  [61] "DVL3"     "DYRK1A"   "EDEM1"    "EEF1E1"   "EFNB3"    "ELAVL3"
##  [67] "ELF2"     "ELL2"     "ELMOD3"   "ELOVL6"   "EPHA7"    "EPHB6"
##  [73] "ESR1"     "ESRRG"    "EVA1C"    "EZR"      "FAM117B"  "FAM120C"
##  [79] "FAM122B"  "FAM160A2" "FARP1"    "FBN2"     "FBXW7"    "FJX1"
##  [85] "FNIP1"    "FRAS1"    "FREM1"    "FRY"      "GABRB3"   "GAPVD1"
##  [91] "GGA2"     "GLIS3"    "GPM6A"    "GRM1"     "HIC2"     "HMGA2"
##  [97] "HOOK3"    "HOXC8"    "HS2ST1"   "IGF2R"    "ING4"     "ITPR1"
## [103] "JPH3"     "KCNA3"    "KCTD1"    "KDM2A"    "KHDRBS1"  "KHDRBS3"
## [109] "KITLG"    "KLF12"    "KLHL13"   "KMT2A"    "KMT5A"    "LATS1"
## [115] "LRRC8D"   "MALL"     "MAML3"    "MAP1LC3B" "MAP3K3"   "MBNL1"
## [121] "MED13L"   "METAP1"   "MLLT3"    "MMGT1"    "MON2"     "MRPL35"
## [127] "MRPL52"   "MYO10"    "NAA15"    "NBEA"     "NCOA7"    "NEUROG1"
## [133] "NOVA1"    "NPTX1"    "NR3C1"    "NR4A2"    "NRBF2"    "NTRK2"
## [139] "P4HB"     "PCDH9"    "PHF13"    "PID1"     "PLAG1"    "POU3F2"
## [145] "PPARGC1A" "PPP3R1"   "PRDM2"    "PRPF38B"  "PRRX1"    "RAB10"
## [151] "RAB14"    "RAB1A"    "RAP2C"    "RBSN"     "REEP1"    "RERE"
## [157] "RHOBTB3"  "RHOT1"    "RICTOR"   "RPS6KA3"  "RPS6KA5"  "RPS6KC1"
## [163] "RSPO3"    "RTKN2"    "RUNX2"    "SATB2"    "SCRT2"    "SEC24D"
## [169] "SEC61A2"  "SERINC3"  "SF3B1"    "SGCZ"     "SGIP1"    "SHC1"
## [175] "SIN3A"    "SIRT1"    "SLC17A7"  "SLC22A2"  "SLC37A3"  "SLITRK4"
## [181] "SLTM"     "SMOC1"    "SOCS6"    "SOX11"    "SOX4"     "SPOP"
## [187] "SPRED1"   "SPRYD7"   "SSRP1"    "ST7"      "STXBP5"   "SUMO2"
## [193] "SUMO4"    "SZRD1"    "TAF5"     "TCF12"    "TCF7L1"   "TGFBR2"
## [199] "TMEM30A"  "TMOD3"    "TNRC6B"   "TP53INP1" "TRIAP1"   "TRIP12"
## [205] "TRPC5"    "TTYH1"    "UBE2R2"   "UHRF2"    "USP6"     "WEE1"
## [211] "WNT4"     "WSB1"     "XRN1"     "YTHDF3"   "YWHAG"    "ZCCHC14"
## [217] "ZCCHC24"  "ZDHHC17"  "ZFC3H1"   "ZFP91"    "ZNF282"   "ZNF335"
## [223] "ZNF423"
##
## $AAANWWTGC_UNKNOWN
##   [1] "ACTB"     "ADHFE1"   "AFF4"     "ANK2"     "ANK3"     "APP"
##   [7] "ASPA"     "ATOH7"    "ATP1B1"   "ATP2B4"   "ATXN7L1"  "BCL11A"
##  [13] "BCL6"     "BNC2"     "C11orf87" "CACNA1D"  "CACNG3"   "CALM1"
##  [19] "CD14"     "CDC42EP3" "CDC42EP5" "CDH13"    "CDK2AP1"  "CEPT1"
##  [25] "CHD2"     "CITED2"   "CNMD"     "CNTFR"    "DAB1"     "DCAF11"
##  [31] "DCHS2"    "DDIT3"    "DIS3L"    "DLG2"     "DLGAP4"   "DMD"
##  [37] "DNAJB5"   "DPYSL5"   "DRD3"     "DSCAM"    "DSEL"     "DSTN"
##  [43] "DTX3L"    "DUSP1"    "DYNC1I2"  "EBF1"     "EFNA5"    "EGFLAM"
##  [49] "EIF4EBP2" "ELAVL4"   "ELF4"     "EPHA7"    "EPHB2"    "ESR1"
##  [55] "FBXW7"    "FGF7"     "FGFR2"    "FN1"      "FOXN3"    "FOXP1"
##  [61] "FOXP2"    "FTHL17"   "FZD7"     "GANAB"    "GATA3"    "GLRA2"
##  [67] "GPC3"     "GPC6"     "GPR21"    "GPRIN3"   "GRHL3"    "GRIN2B"
##  [73] "GTF2E2"   "HEPACAM"  "HHEX"     "HOXA2"    "HOXA3"    "HOXB2"
##  [79] "HOXB6"    "HOXC4"    "IGF2BP1"  "INHBA"    "ITM2C"    "JADE2"
##  [85] "KANK1"    "KCNJ13"   "KLF12"    "KLF14"    "KRTAP8-1" "LEAP2"
##  [91] "LIPG"     "LOX"      "LOXL4"    "LRRC3B"   "LRRN1"    "LSAMP"
##  [97] "LUC7L3"   "MAML3"    "MAN2A2"   "MAP3K4"   "MAPK3"    "MBNL1"
## [103] "MEF2C"    "MEIS1"    "MGLL"     "MID1"     "MLLT6"    "MMP3"
```

```
## [109] "MPZL3"    "MRPL24"   "MRPS18B"  "MYCL"     "MYH2"     "MYLK"
## [115] "NCBP3"    "NEK6"     "NEUROG1"  "NFE2L2"   "NNAT"     "NR2F2"
## [121] "NRAS"     "NTN1"     "NTRK3"    "OLFM1"    "OLIG2"    "OMG"
## [127] "OTX2"     "PATZ1"    "PAX1"     "PAX6"     "PCSK1"    "PCTP"
## [133] "PDGFRB"   "PHOX2B"   "PHTF1"    "PIK3R3"   "POU2F1"   "POU4F1"
## [139] "PPARGC1A" "PPFIA2"   "PPP1R10"  "PPP2R2A"  "PPP3CC"   "PRDM16"
## [145] "PRIMA1"   "PRPF4B"   "RAB10"    "RBMX"     "RORA"     "RRS1"
## [151] "RSPO2"    "S100PBP"  "SALL3"    "SAMD12"   "SATB2"    "SEMA6C"
## [157] "SESN2"    "SFRP2"    "SGCD"     "SHC3"     "SIX5"     "SKIL"
## [163] "SKP2"     "SLMAP"    "SNCAIP"   "SNX25"    "SORT1"    "SOX13"
## [169] "SOX4"     "SOX5"     "SPAG9"    "SPARCL1"  "SSBP3"    "STEAP2"
## [175] "TBC1D8B"  "TFAP4"    "TFDP2"    "TGIF1"    "THAP12"   "THBS2"
## [181] "TLE4"     "TLK1"     "TLX3"     "TRAM1"    "TRPM3"    "TSC22D4"
## [187] "ZFPM1"    "ZHX3"     "ZNF462"   "ZNF827"   "ZW10"
##
## $AAAYRNCTG_UNKNOWN
##   [1] "ABT1"     "ACVR1"    "ADAM12"   "ADD3"     "ADGRB3"   "AGGF1"
##   [7] "ANKRD12"  "ANKRD28"  "AP4S1"    "APBB2"    "APOBR"    "AQP2"
##  [13] "ARHGAP44" "ARID1A"   "ARID4A"   "ARPC2"    "ARSG"     "ARX"
##  [19] "ASB4"     "ASPH"     "ATOH8"    "ATP1A2"   "ATP5IF1"  "ATP5MG"
##  [25] "AXDND1"   "B4GALT6"  "BAMBI"    "BCL2L1"   "BCL9"     "BMPR1B"
##  [31] "BMX"      "BRSK2"    "BTBD3"    "BUB3"     "C12orf65" "CA3"
##  [37] "CACNA2D3" "CACNB2"   "CAPN1"    "CAPZA1"   "CASQ2"    "CBX2"
##  [43] "CCDC174"  "CCNJ"     "CCNY"     "CDC23"    "CDH2"     "CER1"
##  [49] "CFAP161"  "CHRM1"    "CITED2"   "CLDN5"    "CLTC"     "CLTRN"
##  [55] "CMKLR1"   "CNTLN"    "CNTN1"    "COCH"     "COL12A1"  "COL1A2"
##  [61] "COL4A5"   "COL4A6"   "COLEC10"  "CRAT"     "CRH"      "CRKL"
##  [67] "CRYGD"    "CRYGS"    "CSNK1A1"  "CSRNP3"   "CSTF3"    "CYBRD1"
##  [73] "DAAM1"    "DBNDD2"   "DCAKD"    "DDAH2"    "DDX4"     "DEF6"
##  [79] "DENND4A"  "DGKB"     "DHH"      "DHRS4"    "DHRS4L2"  "DIDO1"
##  [85] "DMD"      "DMRT1"    "DNAJA2"   "DNAJB3"   "DNAJB4"   "DSCAML1"
##  [91] "DUSP4"    "DYNC1I1"  "DYRK1A"   "EDA"      "EFNA1"    "EGFLAM"
##  [97] "EIF5"     "EMX2"     "EPC1"     "EPHA7"    "ERBB4"    "ERG28"
## [103] "ERRFI1"   "ESRP2"    "ESRRB"    "ESRRG"    "EYA1"     "FAM216B"
## [109] "FAM49A"   "FAM83F"   "FCER1A"   "FGD4"     "FGF10"    "FGF12"
## [115] "FGFR1"    "FGFR1OP2" "FIZ1"     "FKRP"     "FMNL3"    "FNDC9"
## [121] "FOXA1"    "FOXG1"    "FOXO4"    "FOXP2"    "FSIP2"    "FST"
## [127] "GABRA3"   "GDNF"     "GFI1"     "GGNBP2"   "GJB4"     "GLDN"
## [133] "GNAQ"     "GPR85"    "GPRC5D"   "GRIN2B"   "GSE1"     "H3F3A"
## [139] "HDAC8"    "HESX1"    "HEXIM2"   "HGF"      "HIC2"     "HID1"
## [145] "HIP1R"    "HOXA10"   "HOXA5"    "HOXB8"    "HPSE2"    "HSD3B7"
## [151] "ICAM4"    "ID1"      "IGF1"     "IL1RAPL1" "INHBC"    "IP6K2"
## [157] "ITGA10"   "ITGA8"    "JADE2"    "JPH1"     "JPT1"     "KANK2"
## [163] "KCNIP2"   "KCNK5"    "KCNN3"    "KITLG"    "KLF5"     "KLHDC10"
## [169] "KLHL20"   "KLHL3"    "KMT2A"    "LARS2"    "LENG9"    "LHFPL6"
## [175] "LHX9"     "LMO7"     "LRP5"     "LRRC4"    "LRRN4CL"  "LTBP1"
## [181] "MAML1"    "MANF"     "MAP2"     "MAP3K5"   "MAP6"     "MEIS1"
## [187] "MGAT1"    "MGAT4A"   "MID1"     "MOAP1"    "MPP6"     "MPPED2"
## [193] "MRPL13"   "MTA2"     "MTBP"     "MYF6"     "MYH1"     "MYH10"
## [199] "MYO18A"   "NAGLU"    "NAPB"     "NAV2"     "NAV3"     "NCDN"
## [205] "NDNF"     "NDST4"    "NDUFS4"   "NEK1"     "NEK2"     "NFATC4"
## [211] "NFYB"     "NMI"      "NMT1"     "NR2F1"    "NRG1"     "NTRK2"
## [217] "NUP54"    "NXPH4"    "OMA1"     "OMG"      "OR2L13"   "OTX2"
## [223] "PACRG"    "PCDH17"   "PCDH18"   "PCF11"    "PCYT1B"   "PDGFB"
## [229] "PDGFRA"   "PDLIM2"   "PDS5B"    "PDZRN4"   "PFN2"     "PHC2"
## [235] "PHEX"     "PHF1"     "PHF6"     "PHOX2B"   "PLAGL2"   "PLEC"
## [241] "PLEKHM1"  "PLP2"     "PLPP3"    "PMCH"     "PODXL2"   "POFUT1"
## [247] "POU2AF1"  "POU4F1"   "PPP1R9B"  "PPP2R3A"  "PPP2R5E"  "PPP3CA"
## [253] "PRELP"    "PRKCG"    "PRKCQ"    "PRKN"     "PROK2"    "PTCHD4"
## [259] "PTH1R"    "PTPA"     "PXN"      "R3HDM1"   "RAB30"    "RAB5B"
## [265] "RAB5C"    "RAPGEF4"  "RBMS3"    "RGS17"    "RNF146"   "ROBO4"
## [271] "ROR1"     "RPLP0"    "RTN1"     "RUFY3"    "S1PR2"    "SCN3B"
## [277] "SCN5A"    "SCN8A"    "SCOC"     "SDCBP"    "SEMA6D"   "SEPT7"
## [283] "SESN3"    "SGCD"     "SH2D6"    "SHC3"     "SHCBP1L"  "SIPA1"
## [289] "SIRPA"    "SLC26A6"  "SLC4A1"   "SLC6A1"   "SMARCA2"  "SNX9"
## [295] "SORBS2"   "SOX12"    "SOX21"    "SOX30"    "SOX5"     "SPINDOC"
## [301] "SPOCK2"   "SPTLC2"   "SRGAP2"   "SRSF8"    "SSBP2"    "ST7L"
## [307] "STAC3"    "STAG1"    "STAG2"    "STC2"     "STRN3"    "STRN4"
## [313] "TAS1R2"   "TEF"      "TENT4B"   "TFAP4"    "TFDP2"    "TM2D3"
## [319] "TMEM182"  "TMEM69"   "TMSB4X"   "TNFAIP8"  "TNS1"     "TNXB"
## [325] "TP53INP2" "TRDN"     "TREML1"   "TRIM28"   "TRIM68"   "TRIM8"
## [331] "TRIML1"   "TRPS1"    "TSC22D3"  "TSPAN7"   "TSPY26P"  "TSSK3"
## [337] "TTC17"    "TUSC2"    "UBE2W"    "UBXN10"   "USP1"     "VDR"
```

```
## [343] "VIP"      "VKORC1L1" "VWA5A"    "WBP1"     "WNT2B"    "WT1"
## [349] "WT1-AS"   "XRCC1"    "ZADH2"    "ZBTB11"   "ZBTB18"   "ZFP91"
## [355] "ZFPM2"    "ZIC1"     "ZIC4"     "ZMAT3"    "ZNF296"   "ZNF503"
## [361] "ZNF521"   "ZNF524"   "ZNF654"   "ZNF687"   "ZNF710"
```

```
# filter the list to include only hallmark pathways

library(data.table)

pathways.hallmark <- m_df[m_df$gs_name %like% "HALLMARK_", ]
pathways.hallmark <- split(pathways.hallmark$human_gene_symbol, pathways.hallmark$gs_name)

# Show the first few pathways, and within those, show only the first few genes.
pathways.hallmark %>%
  head() %>%
  lapply(head)
```

```
## $HALLMARK_ADIPOGENESIS
## [1] "ABCA1" "ABCB8" "ACAA2" "ACADL" "ACADM" "ACADS"
##
## $HALLMARK_ALLOGRAFT_REJECTION
## [1] "AARS"   "ABCE1"  "ABI1"   "ACHE"   "ACVR2A" "AKT1"
##
## $HALLMARK_ANDROGEN_RESPONSE
## [1] "ABCC4"   "ABHD2"   "ACSL3"   "ACTN1"   "ADAMTS1" "ADRM1"
##
## $HALLMARK_ANGIOGENESIS
## [1] "APOH"   "APP"    "CCND2"  "COL3A1" "COL5A2" "CXCL6"
##
## $HALLMARK_APICAL_JUNCTION
## [1] "ACTA1" "ACTB"  "ACTC1" "ACTG1" "ACTG2" "ACTN1"
##
## $HALLMARK_APICAL_SURFACE
## [1] "ADAM10"   "ADIPOR2"  "AFAP1L2"  "AKAP7"    "APP"      "ATP6V0A4"
```

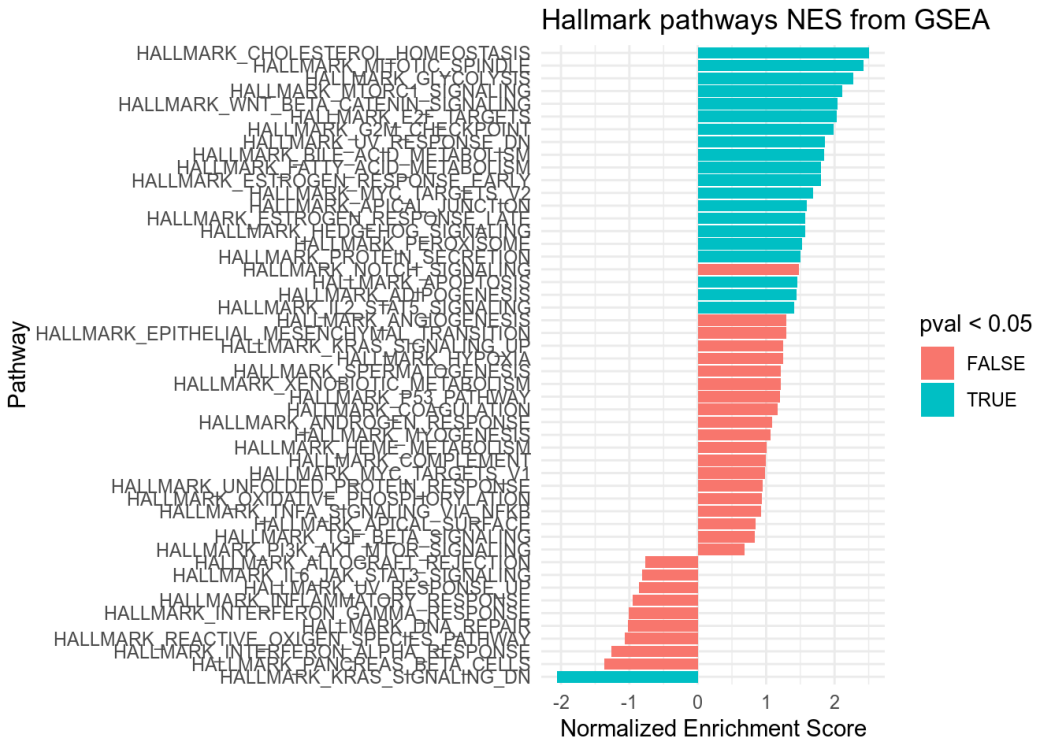
```
# running the fgsea algorithm on hallmark.pathways

fgseaRes <- fgsea(pathways=pathways.hallmark, stats=ranks, nperm=1000)

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))
```

```
# ggplotting for hallmark pathways
library(ggplot2)
#pdf("fgseaResTidy.pdf", width = 10, height = 10)

ggplot(fgseaResTidy, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=pval<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```




```
#dev.off()
```

```
# We have just plotted all the significant patways in the hallmark pathways as 'blue'
# We can see that:
    # HALLMARK_APOPTOSIS (cell death)
    # HALLMARK_GLYCOLYSIS
    # HALLMARK_IL2_STAT5_SIGNALING (interleukin)
# et cetera pathways are activated!

# Let's look at all viral pathways

pathways.viral <- m_df[m_df$gs_name %like% "VIRAL_", ]
pathways.viral <- split(pathways.viral$human_gene_symbol, pathways.viral$gs_name)

# Show the first few pathways, and within those, show only the first few genes.
pathways.hallmark %>%
  head() %>%
  lapply(head)
```

```
## $HALLMARK_ADIPOGENESIS
## [1] "ABCA1" "ABCB8" "ACAA2" "ACADL" "ACADM" "ACADS"
##
## $HALLMARK_ALLOGRAFT_REJECTION
## [1] "AARS"   "ABCE1"  "ABI1"   "ACHE"   "ACVR2A" "AKT1"
##
## $HALLMARK_ANDROGEN_RESPONSE
## [1] "ABCC4"   "ABHD2"   "ACSL3"   "ACTN1"   "ADAMTS1" "ADRM1"
##
## $HALLMARK_ANGIOGENESIS
## [1] "APOH"   "APP"    "CCND2"  "COL3A1" "COL5A2" "CXCL6"
##
## $HALLMARK_APICAL_JUNCTION
## [1] "ACTA1" "ACTB"  "ACTC1" "ACTG1" "ACTG2" "ACTN1"
##
## $HALLMARK_APICAL_SURFACE
## [1] "ADAM10"   "ADIPOR2"  "AFAP1L2"  "AKAP7"    "APP"      "ATP6V0A4"
```

```
# running the fgsea algorithm on viral pathways

fgseaRes_viral <- fgsea(pathways=pathways.viral, stats=ranks, nperm=1000)

fgseaResTidy_viral <- fgseaRes_viral %>%
  as_tibble() %>%
  arrange(desc(NES))

# Let's look at the plot

# ggplotting for hallmark pathways
library(ggplot2)

ggplot(fgseaResTidy_viral, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=pval<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Viral pathways NES from GSEA") +
  theme_minimal()
```


```
# install.packages('DT')
library(DT)
# Show in a nice table for all pathways
fgseaResTidy %>%
  dplyr::select(-leadingEdge, -ES, -nMoreExtreme) %>%
  arrange(padj) %>%
  DT::datatable()
```

Show 10 entries Search:

	pathway	pval	padj	NES	size
1	HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.00207900207900208	0.0107991360691145	2.50604303128592	66
2	HALLMARK_MITOTIC_SPINDLE	0.00215982721382289	0.0107991360691145	2.42747838761065	181
3	HALLMARK_GLYCOLYSIS	0.00209205020920502	0.0107991360691145	2.27600624101148	171
4	HALLMARK_MTORC1_SIGNALING	0.00214592274678112	0.0107991360691145	2.11173603644745	188
5	HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.00191938579654511	0.0107991360691145	2.04291622246835	26
6	HALLMARK_E2F_TARGETS	0.00206185567010309	0.0107991360691145	2.02782060303682	193

	pathway	pval	padj	NES	size
7	HALLMARK_G2M_CHECKPOINT	0.00212765957446809	0.0107991360691145	1.99131261489342	189
8	HALLMARK_UV_RESPONSE_DN	0.00213675213675214	0.0107991360691145	1.85636790239389	109
9	HALLMARK_FATTY_ACID_METABOLISM	0.00204081632653061	0.0107991360691145	1.80509775283643	126
10	HALLMARK_KRAS_SIGNALING_DN	0.00191938579654511	0.0107991360691145	-2.05955042365381	77

Showing 1 to 10 of 50 entries

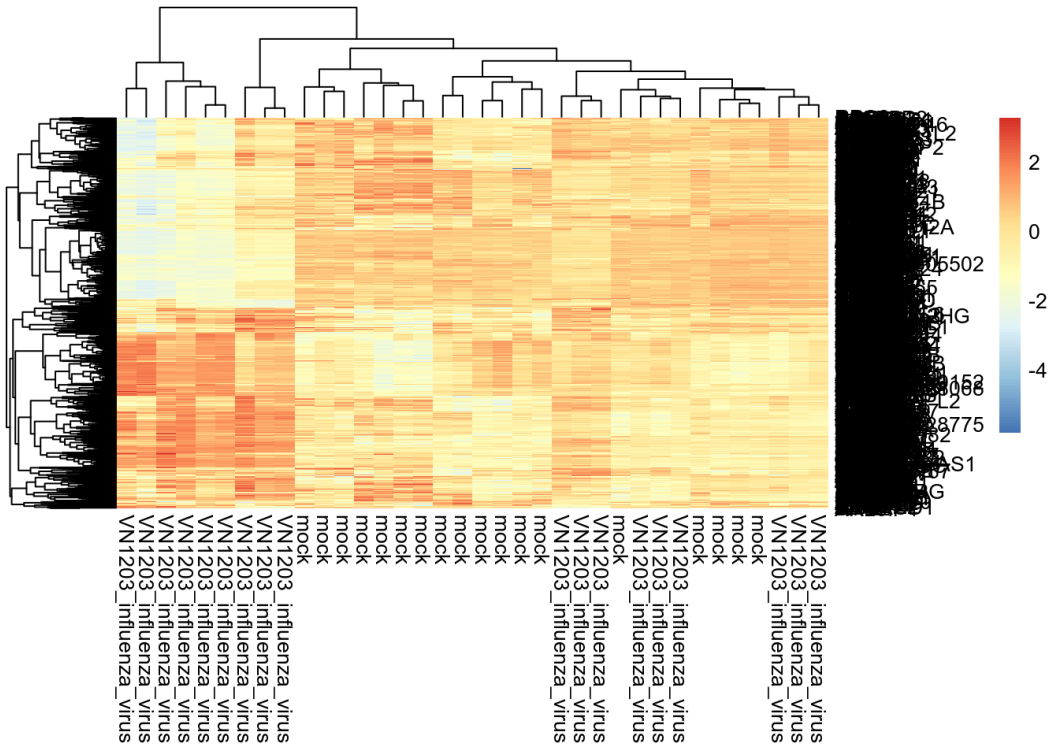
Previous 1 2 3 4 5 Next

```
# heatmap
library(pheatmap)
#scale rows
xt<-t(as.matrix(res.qnorm.top12K)) # this is a matrix of normalised 12k genes
xts<-scale(xt)
xtst<-t(xts)
xtst <- na.omit(xtst)
colnames(xtst) <- res$condition

#only grab top 1000 by p-value
h<-head(xtst, n = 1000L)

#set layout options - adjust if labels get cut off
# pdf("heatmap.pdf",width=10, height=100)

#draw heatmap allowing larger margins and adjusting row label font size
pheatmap(h)
```



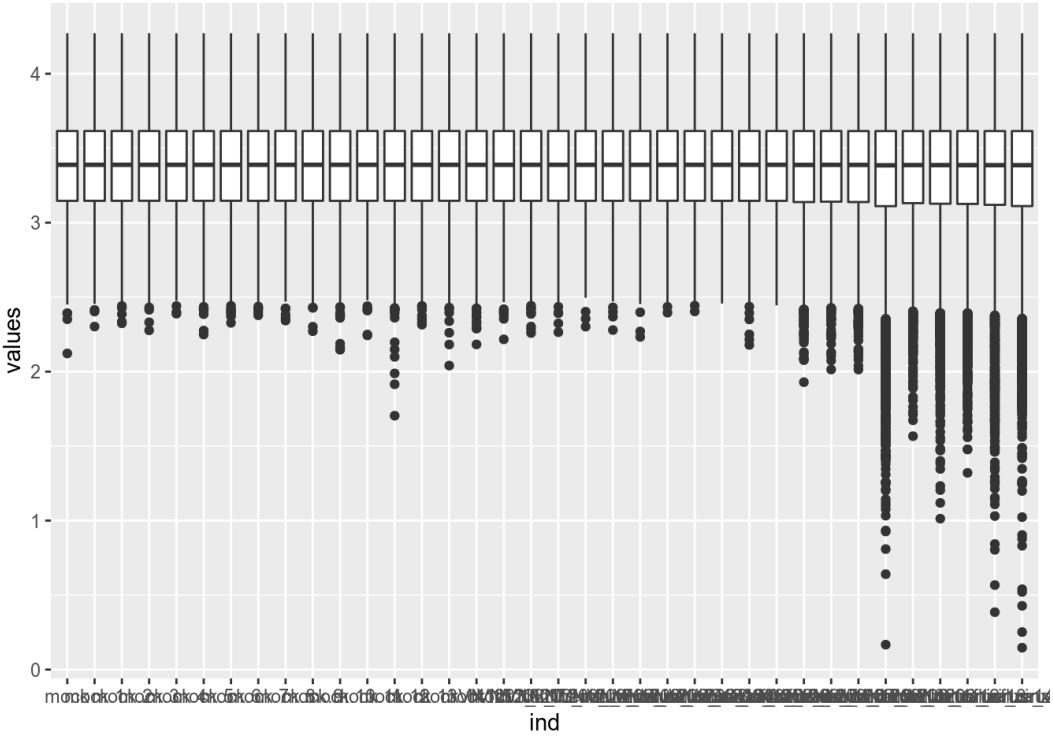

```
#output plot to file
# dev.off()
```

```
# let's make a boxplot of the data

# install.packages('devtools')
library(devtools)
# devtools::install_github("sinhrks/ggfortify")
library(ggfortify)

# pdf('box_dataset1.pdf', width = 50)

gt <- t(xt) # taking xt from the heatmap and transposing it
colnames(gt)<- res$condition # now giving it labels from condition
ggplot(stack(data.frame(gt)), aes(x = ind, y = values)) +
  geom_boxplot()
```




```
# dev.off()

#
# we can see that various pathways associated with viral infection, consistent
# with the results from the paper.
#
```