# GSE4619

*Natalia*

*14 May 2019*

```r
set.seed(42)

# 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:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind,
##     colMeans, colnames, colSums, dirname, do.call, duplicated,
##     eval, evalq, Filter, Find, get, grep, grepl, intersect,
##     is.unsorted, lapply, lengths, Map, mapply, match, mget, order,
##     paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,
##     Reduce, rowMeans, rownames, rowSums, 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)
```

```r
library(limma)
```

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

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

```
#library(org.Mm.eg.db)
library(org.Hs.eg.db)
```

```
## Loading required package: AnnotationDbi

## Loading required package: stats4

## Loading required package: IRanges

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

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

##
## Attaching package: 'IRanges'

## The following object is masked from 'package:grDevices':
##
##     windows

##
```

```
# for collapseBy:

source("C://Users//Natalia//Desktop//ITMO//SystemBiology//RNAseq_analysis//RNAseq_analysis//dataset#3//


#Gene expression analysis of bone marrow mesenchymal stromal cells
#from myelodysplastic syndrome (MDS) patients and normal controls:

es <- getGEO("GSE4619", AnnotGPL = TRUE, parseCharacteristics = FALSE)[[1]]
```

```
## Found 1 file(s)

## GSE4619_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:

## C:\Users\Public\Documents\iSkysoft\CreatorTemp\RtmpcDfoZ3/GPL570.annot.gz

## Warning: 62 parsing failures.
##   row             col           expected    actual         file
## 54614 Platform_SPOTID 1/0/T/F/TRUE/FALSE --Control literal data
## 54615 Platform_SPOTID 1/0/T/F/TRUE/FALSE --Control literal data
## 54616 Platform_SPOTID 1/0/T/F/TRUE/FALSE --Control literal data
## 54617 Platform_SPOTID 1/0/T/F/TRUE/FALSE --Control literal data
## 54618 Platform_SPOTID 1/0/T/F/TRUE/FALSE --Control literal data
## ..... ............... .................. ......... ............
## See problems(...) for more details.
```

```r
str(experimentData(es))
```

```
## Formal class 'MIAME' [package "Biobase"] with 13 slots
##   ..@ name             : chr "Andrea,,Pellagatti"
##   ..@ lab              : chr ""
##   ..@ contact          : chr "andreapellagatti@yahoo.co.uk"
##   ..@ title            : chr "Gene expression profiling of CD34+ cells from MDS patients and normal c
##   ..@ abstract         : chr "In order to gain insight into the poorly understood pathophysiology of
##   ..@ url              : chr "https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4619"
##   ..@ pubMedIds        : chr "16527891"
##   ..@ samples          : list()
##   ..@ hybridizations   : list()
##   ..@ normControls     : list()
##   ..@ preprocessing    : list()
##   ..@ other            :List of 25
##   .. ..$ contact_address        : chr "John Radcliffe Hospital"
##   .. ..$ contact_city           : chr "Oxford"
##   .. ..$ contact_country        : chr "United Kingdom"
##   .. ..$ contact_department     : chr "NDCLS, RDM"
##   .. ..$ contact_email          : chr "andreapellagatti@yahoo.co.uk"
##   .. ..$ contact_institute      : chr "University of Oxford"
##   .. ..$ contact_laboratory     : chr "LLR Molecular Haematology Unit"
##   .. ..$ contact_name           : chr "Andrea,,Pellagatti"
##   .. ..$ contact_phone          : chr "00441865222911"
##   .. ..$ contact_zip/postal_code: chr "OX3 9DU"
##   .. ..$ contributor            : chr "Andrea,,Pellagatti\nJames,S,Wainscoat\nJacqueline,,Boultwood"
##   .. ..$ geo_accession          : chr "GSE4619"
##   .. ..$ last_update_date       : chr "Mar 25 2019"
##   .. ..$ overall_design         : chr "55 patients with MDS and 11 healthy controls were included in
##   .. ..$ platform_id            : chr "GPL570"
##   .. ..$ platform_taxid         : chr "9606"
##   .. ..$ pubmed_id              : chr "16527891"
##   .. ..$ relation               : chr "BioProject: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA9532
##   .. ..$ sample_id              : chr "GSM103343 GSM103344 GSM103345 GSM103346 GSM103347 GSM103348 G
##   .. ..$ sample_taxid           : chr "9606"
##   .. ..$ status                 : chr "Public on Jul 01 2006"
##   .. ..$ submission_date        : chr "Apr 05 2006"
##   .. ..$ summary                : chr "In order to gain insight into the poorly understood pathophysi
##   .. ..$ title                  : chr "Gene expression profiling of CD34+ cells from MDS patients and
##   .. ..$ type                   : chr "Expression profiling by array"
##   ..@ .__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
```

```r
str(pData(es))
```

```
## 'data.frame':    66 obs. of  35 variables:
##  $ title                  : Factor w/ 66 levels "MDS1","MDS10",..: 56 59 60 1 12 23 34 45 52 53 ...
##  $ geo_accession          : chr  "GSM103343" "GSM103344" "GSM103345" "GSM103346" ...
##  $ status                 : Factor w/ 1 level "Public on Jul 01 2006": 1 1 1 1 1 1 1 1 1 1 ...
##  $ submission_date        : Factor w/ 1 level "Apr 05 2006": 1 1 1 1 1 1 1 1 1 1 ...
##  $ last_update_date       : Factor w/ 1 level "Aug 22 2006": 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/ 3 levels "CD34+ cells isolated from the bone marrow of a normal
##  $ organism_ch1           : Factor w/ 1 level "Homo sapiens": 1 1 1 1 1 1 1 1 1 1 ...
##  $ characteristics_ch1    : Factor w/ 5 levels "Subtype: normal",..: 1 1 1 3 2 5 2 3 5 3 ...
##  $ characteristics_ch1.1  : Factor w/ 3 levels "Karyotype: del(5q)",..: 2 2 2 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 "Trizol extraction of total RNA was performed accordin
##  $ label_ch1              : Factor w/ 1 level "biotin": 1 1 1 1 1 1 1 1 1 1 ...
##  $ label_protocol_ch1     : Factor w/ 1 level "Biotinylated cRNA were prepared according to the Two-
##  $ taxid_ch1              : Factor w/ 1 level "9606": 1 1 1 1 1 1 1 1 1 1 ...
##  $ hyb_protocol           : Factor w/ 3 levels "Following fragmentation 10 micrograms of cRNA were hy
##  $ scan_protocol          : Factor w/ 1 level "GeneChips were scanned using a GeneChip Scanner 3000"
##  $ description            : Factor w/ 1 level "Gene expression data from CD34+ cells isolated from bo
##  $ data_processing        : Factor w/ 1 level "Affymetrix CEL files were used for Robust MultiChip An
##  $ platform_id            : Factor w/ 1 level "GPL570": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_name           : Factor w/ 1 level "Andrea,,Pellagatti": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_email          : Factor w/ 1 level "andreapellagatti@yahoo.co.uk": 1 1 1 1 1 1 1 1 1 1 ..
##  $ contact_phone          : Factor w/ 1 level "00441865222911": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_laboratory     : Factor w/ 1 level "LLR Molecular Haematology Unit": 1 1 1 1 1 1 1 1 1 1
##  $ contact_department     : Factor w/ 1 level "NDCLS, RDM": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_institute      : Factor w/ 1 level "University of Oxford": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_address        : Factor w/ 1 level "John Radcliffe Hospital": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_city           : Factor w/ 1 level "Oxford": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_zip/postal_code: Factor w/ 1 level "OX3 9DU": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact_country        : Factor w/ 1 level "United Kingdom": 1 1 1 1 1 1 1 1 1 1 ...
##  $ supplementary_file     : Factor w/ 1 level "NONE": 1 1 1 1 1 1 1 1 1 1 ...
##  $ data_row_count         : Factor w/ 1 level "54675": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Karyotype:ch1          : chr  "normal" "normal" "normal" "del(5q)" ...
##  $ Subtype:ch1            : chr  "normal" "normal" "normal" "RAEB1" ...
```

```r
head(fData(es))
```

```
##                  ID
## 1007_s_at 1007_s_at
## 1053_at     1053_at
## 117_at       117_at
## 121_at       121_at
## 1255_g_at 1255_g_at
## 1294_at     1294_at
##                                                             Gene title
## 1007_s_at microRNA 4640///discoidin domain receptor tyrosine kinase 1
## 1053_at                                replication factor C subunit 2
## 117_at                   heat shock protein family A (Hsp70) member 6
## 121_at                                                   paired box 8
## 1255_g_at                              guanylate cyclase activator 1A
## 1294_at   microRNA 5193///ubiquitin like modifier activating enzyme 7
##              Gene symbol          Gene ID UniGene title UniGene symbol
## 1007_s_at MIR4640///DDR1  100616237///780
## 1053_at             RFC2             5982
## 117_at             HSPA6             3310
## 121_at              PAX8             7849
## 1255_g_at         GUCA1A             2978
## 1294_at   MIR5193///UBA7 100847079///7318
##           UniGene ID
## 1007_s_at
## 1053_at
## 117_at
## 121_at
## 1255_g_at
## 1294_at
##                                                                                  Nucleotide Title
## 1007_s_at                                 Human receptor tyrosine kinase DDR gene, complete cds
## 1053_at                    Human replication factor C, 40-kDa subunit (A1) mRNA, complete cds
## 117_at                                                 Human heat-shock protein HSP70B' gene
## 121_at                                                                   H.sapiens Pax8 mRNA
## 1255_g_at Homo sapiens guanylate cyclase activating protein (GCAP) gene exons 1-4, complete cds
## 1294_at          Homo sapiens ubiquitin-activating enzyme E1 related protein mRNA, complete cds
##                GI GenBank Accession Platform_CLONEID Platform_ORF
## 1007_s_at 1753221            U48705               NA           NA
## 1053_at   1590810            M87338               NA           NA
## 117_at      35221            X51757               NA           NA
## 121_at      38425            X69699               NA           NA
## 1255_g_at  623404            L36861               NA           NA
## 1294_at    520832            L13852               NA           NA
##           Platform_SPOTID Chromosome location
## 1007_s_at              NA              6p21.3
## 1053_at                NA             7q11.23
## 117_at                 NA                1q23
## 121_at                 NA                2q13
## 1255_g_at              NA              6p21.1
## 1294_at                NA                3p21
##
## 1007_s_at                        Chromosome 6, NC_000006.12 (30890883..30890972)///Chromosome 6, NC_
## 1053_at                                                              Chromosome 7, NC_000007.14 (7
## 117_at                                                                         Chromosome 1, NC_00
## 121_at                                                             Chromosome 2, NC_000002.12 (113
## 1255_g_at                                                                        Chromosome 6, NC_
## 1294_at   Chromosome 3, NC_000003.12 (49806137..49806245, complement)///Chromosome 3, NC_000003.12 (4
##
## 1007_s_at
## 1053_at
## 117_at
## 121_at    DNA binding///DNA binding///RNA polymerase II core promoter proximal region sequence-speci
## 1255_g_at
## 1294_at
##
## 1007_s_at
## 1053_at
## 117_at
## 121_at    anatomical structure morphogenesis///branching involved in ureteric bud morphogenesis///cel
## 1255_g_at
## 1294_at
##
## 1007_s_at basolateral plasma membrane///extracellular exosome///extracellular space///integral compon
## 1053_at                                                                               Ctf18 RFC-li
## 117_at                                      colocalizes_with COP9 signalosome///blood microparticl
## 121_at
## 1255_g_at                                                                         photoreceptor di
## 1294_at
##
## 1007_s_at                           GO:0005524///GO:0005518///GO:0005518///GO:0046872///GO:0005515//
## 1053_at                         GO:0005524///contributes_to GO:0003689///GO:0019899///GO:0005515///c
## 117_at                                          GO:0005524///GO:0042623///GO:0019899///GO:0031072//
## 121_at    GO:0003677///GO:0003677///GO:0000978///GO:0000979///GO:0005515///GO:0004996///GO:0003700//
## 1255_g_at                                                                               GO:0005509//
## 1294_at                                         GO:0005524///GO:0019782///GO:0005515///GO:0004839//
##
## 1007_s_at
## 1053_at
## 117_at
## 121_at    GO:0009653///GO:0001658///GO:0071371///GO:0007417///GO:0042472///GO:0001822///GO:0003337//
## 1255_g_at
## 1294_at
##                                                                                        GO:Component
## 1007_s_at                  GO:0016323///GO:0070062///GO:0005615///GO:0005887///GO:0005886///GO:00432
## 1053_at                                                           GO:0031390///GO:0005663///GO:00056
## 117_at    colocalizes_with GO:0008180///GO:0072562///GO:0005814///GO:0005737///GO:0005829///GO:00700
## 121_at                                                            GO:0005654///GO:0005654///GO:00056
## 1255_g_at                                                         GO:0097381///GO:0001917///GO:00058
## 1294_at                                              GO:0005829///GO:0005829///GO:0005654///GO:00056
```

```
es$`Subtype:ch1`
```

```
##  [1] "normal" "normal" "normal" "RAEB1"  "RA"     "RARS"   "RA"    
##  [8] "RAEB1"  "RARS"   "RAEB1"  "RAEB1"  "RA"     "RA"     "RARS"  
## [15] "RA"     "RA"     "RAEB1"  "RAEB1"  "RA"     "RAEB2"  "RA"    
## [22] "normal" "normal" "normal" "RA"     "RARS"   "normal" "normal"
## [29] "RAEB2"  "RARS"   "RAEB2"  "RAEB2"  "RARS"   "RARS"   "RAEB2" 
## [36] "RAEB2"  "RARS"   "RAEB2"  "RA"     "RA"     "RAEB1"  "RA"    
## [43] "RARS"   "RAEB2"  "RARS"   "normal" "normal" "normal" "RA"    
## [50] "RARS"   "RARS"   "RA"     "RARS"   "RARS"   "RAEB2"  "RARS"  
## [57] "RA"     "RAEB1"  "RA"     "RARS"   "RARS"   "RA"     "RARS"  
## [64] "RAEB1"  "RARS"   "RA"
```

```
#The condition is the "Subtype:ch1" in this dataset:

es$condition <- gsub("\\+", "_", es$`Subtype:ch1`)
es$condition
```

```
##  [1] "normal" "normal" "normal" "RAEB1"  "RA"     "RARS"   "RA"    
##  [8] "RAEB1"  "RARS"   "RAEB1"  "RAEB1"  "RA"     "RA"     "RARS"  
## [15] "RA"     "RA"     "RAEB1"  "RAEB1"  "RA"     "RAEB2"  "RA"    
## [22] "normal" "normal" "normal" "RA"     "RARS"   "normal" "normal"
## [29] "RAEB2"  "RARS"   "RAEB2"  "RAEB2"  "RARS"   "RARS"   "RAEB2" 
## [36] "RAEB2"  "RARS"   "RAEB2"  "RA"     "RA"     "RAEB1"  "RA"    
## [43] "RARS"   "RAEB2"  "RARS"   "normal" "normal" "normal" "RA"    
## [50] "RARS"   "RARS"   "RA"     "RARS"   "RARS"   "RAEB2"  "RARS"  
## [57] "RA"     "RAEB1"  "RA"     "RARS"   "RARS"   "RA"     "RARS"  
## [64] "RAEB1"  "RARS"   "RA"
```

```
#Then we collapse the dataset with gene ID as in phantasus:

es <- collapseBy(es, fData(es)$`Gene symbol`, FUN=median)
es <- es[!grepl("///", rownames(es)), ]
es <- es[rownames(es) != "", ]
```

```
# there is a lot of garbage there.
# Annotate the symbols with human database entries:

fData(es) <- data.frame(row.names = rownames(es))
fData(es)$entrez <- row.names(fData(es))

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

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

```
#To normalize the data:

es.qnorm <- es
summary(exprs(es.qnorm))
```

```
##    GSM103343          GSM103344          GSM103345      
##  Min.   :    6.093   Min.   :    5.963   Min.   :    5.909  
##  1st Qu.:   24.358   1st Qu.:   24.880   1st Qu.:   23.543  
##  Median :   62.805   Median :   62.265   Median :   61.490  
##  Mean   :  293.274   Mean   :  293.084   Mean   :  288.843  
##  3rd Qu.:  209.600   3rd Qu.:  207.000   3rd Qu.:  214.250  
##  Max.   :23508.000   Max.   :24523.000   Max.   :23459.000  
##    GSM103346          GSM103347          GSM103348      
##  Min.   :    6.244   Min.   :    5.887   Min.   :    6.231  
##  1st Qu.:   25.580   1st Qu.:   21.560   1st Qu.:   26.600  
##  Median :   63.015   Median :   62.810   Median :   67.240  
##  Mean   :  293.601   Mean   :  330.469   Mean   :  301.105  
##  3rd Qu.:  199.300   3rd Qu.:  267.425   3rd Qu.:  215.700  
##  Max.   :22777.000   Max.   :21180.000   Max.   :23156.000  
##    GSM103349          GSM103350          GSM103351      
##  Min.   :    6.153   Min.   :    6.053   Min.   :    6.00  
##  1st Qu.:   23.880   1st Qu.:   24.350   1st Qu.:   24.66  
##  Median :   61.425   Median :   62.560   Median :   60.98  
##  Mean   :  297.064   Mean   :  309.206   Mean   :  280.40  
##  3rd Qu.:  212.200   3rd Qu.:  222.500   3rd Qu.:  197.05  
##  Max.   :23038.000   Max.   :22948.000   Max.   :23793.00  
##    GSM103352          GSM103353          GSM103354      
##  Min.   :    5.924   Min.   :    6.119   Min.   :    6.124  
##  1st Qu.:   25.935   1st Qu.:   26.370   1st Qu.:   23.680  
##  Median :   62.100   Median :   65.975   Median :   64.990  
##  Mean   :  252.690   Mean   :  309.566   Mean   :  344.386  
##  3rd Qu.:  170.900   3rd Qu.:  208.200   3rd Qu.:  253.725  
##  Max.   :21990.000   Max.   :22325.000   Max.   :23010.000  
##    GSM103355          GSM103356          GSM103357      
##  Min.   :    6.129   Min.   :    5.681   Min.   :    6.323  
##  1st Qu.:   25.130   1st Qu.:   25.587   1st Qu.:   25.520  
##  Median :   64.090   Median :   63.625   Median :   60.880  
##  Mean   :  289.867   Mean   :  278.352   Mean   :  279.908  
##  3rd Qu.:  211.025   3rd Qu.:  197.700   3rd Qu.:  188.725  
##  Max.   :22974.000   Max.   :23730.000   Max.   :22476.000  
##    GSM103358          GSM103359          GSM103360      
##  Min.   :    6.148   Min.   :    6.237   Min.   :    5.986  
##  1st Qu.:   25.030   1st Qu.:   25.207   1st Qu.:   24.040
##  Median :   61.805   Median :   62.625   Median :   60.580
##  Mean   :  273.013   Mean   :  280.845   Mean   :  288.298
##  3rd Qu.:  191.225   3rd Qu.:  206.600   3rd Qu.:  206.000
##  Max.   :22774.000   Max.   :23699.000   Max.   :23327.000
##    GSM103361           GSM103362           GSM103363
##  Min.   :    5.904   Min.   :    6.096   Min.   :    6.059
##  1st Qu.:   23.017   1st Qu.:   24.610   1st Qu.:   24.157
##  Median :   58.430   Median :   62.050   Median :   62.400
##  Mean   :  278.634   Mean   :  288.875   Mean   :  302.779
##  3rd Qu.:  196.800   3rd Qu.:  202.800   3rd Qu.:  216.125
##  Max.   :21694.000   Max.   :23551.000   Max.   :23488.000
##    GSM103364           GSM103365           GSM103366
##  Min.   :    5.839   Min.   :    6.117   Min.   :    6.065
##  1st Qu.:   24.530   1st Qu.:   24.730   1st Qu.:   24.777
##  Median :   61.900   Median :   62.275   Median :   64.015
##  Mean   :  287.743   Mean   :  274.043   Mean   :  268.433
##  3rd Qu.:  207.800   3rd Qu.:  201.600   3rd Qu.:  199.225
##  Max.   :23290.000   Max.   :23103.000   Max.   :23480.000
##    GSM103367           GSM103368           GSM103369
##  Min.   :    6.056   Min.   :    6.049   Min.   :    5.752
##  1st Qu.:   23.240   1st Qu.:   22.360   1st Qu.:   21.460
##  Median :   64.840   Median :   63.800   Median :   57.815
##  Mean   :  312.107   Mean   :  322.509   Mean   :  316.241
##  3rd Qu.:  243.300   3rd Qu.:  247.700   3rd Qu.:  234.200
##  Max.   :24174.000   Max.   :23389.000   Max.   :23562.000
##    GSM103370           GSM103371          GSM103372
##  Min.   :    6.065   Min.   :    5.94   Min.   :    5.968
##  1st Qu.:   20.740   1st Qu.:   20.68   1st Qu.:   22.350
##  Median :   61.995   Median :   59.42   Median :   63.785
##  Mean   :  335.004   Mean   :  342.37   Mean   :  324.729
##  3rd Qu.:  266.675   3rd Qu.:  262.10   3rd Qu.:  251.600
##  Max.   :24608.000   Max.   :25175.00   Max.   :23713.000
##    GSM103373           GSM103374           GSM103375
##  Min.   :    5.882   Min.   :    5.964   Min.   :    6.222
##  1st Qu.:   20.927   1st Qu.:   22.330   1st Qu.:   23.740
##  Median :   60.030   Median :   65.815   Median :   65.035
##  Mean   :  321.317   Mean   :  325.225   Mean   :  310.180
##  3rd Qu.:  251.525   3rd Qu.:  248.700   3rd Qu.:  235.100
##  Max.   :24799.000   Max.   :23805.000   Max.   :24219.000
##    GSM103376           GSM103377           GSM103378
##  Min.   :    6.294   Min.   :    5.853   Min.   :    5.464
##  1st Qu.:   23.047   1st Qu.:   22.137   1st Qu.:   19.730
##  Median :   65.270   Median :   64.950   Median :   54.905
##  Mean   :  336.690   Mean   :  317.077   Mean   :  325.924
##  3rd Qu.:  253.900   3rd Qu.:  253.000   3rd Qu.:  254.700
##  Max.   :24388.000   Max.   :25083.000   Max.   :23800.000
##    GSM103379           GSM103380           GSM103381
##  Min.   :    6.184   Min.   :    5.395   Min.   :    6.093
##  1st Qu.:   23.308   1st Qu.:   22.040   1st Qu.:   20.738
##  Median :   64.210   Median :   63.800   Median :   60.970
##  Mean   :  298.470   Mean   :  317.211   Mean   :  335.684
##  3rd Qu.:  233.225   3rd Qu.:  245.725   3rd Qu.:  266.200
##  Max.   :23902.000   Max.   :24228.000   Max.   :24446.000
##    GSM103382           GSM103383          GSM103384
##  Min.   :    5.902   Min.   :    6.01   Min.   :    6.121
##  1st Qu.:   20.170   1st Qu.:   21.50   1st Qu.:   20.220
##  Median :   61.195   Median :   66.19   Median :   61.435
##  Mean   :  337.915   Mean   :  315.76   Mean   :  335.311
##  3rd Qu.:  279.675   3rd Qu.:  260.10   3rd Qu.:  276.700
##  Max.   :24228.000   Max.   :22386.00   Max.   :23914.000
##    GSM103385           GSM103386           GSM103387
##  Min.   :    6.129   Min.   :    5.903   Min.   :    5.978
##  1st Qu.:   23.070   1st Qu.:   21.460   1st Qu.:   22.040
##  Median :   65.610   Median :   61.005   Median :   64.975
##  Mean   :  296.021   Mean   :  303.422   Mean   :  317.437
##  3rd Qu.:  234.125   3rd Qu.:  248.300   3rd Qu.:  258.825
##  Max.   :23880.000   Max.   :23899.000   Max.   :24948.000
##    GSM103388          GSM103389           GSM103390
##  Min.   :    5.82   Min.   :    6.109   Min.   :    5.887
##  1st Qu.:   20.70   1st Qu.:   20.760   1st Qu.:   22.098
##  Median :   61.69   Median :   61.805   Median :   63.730
##  Mean   :  336.51   Mean   :  340.728   Mean   :  309.979
##  3rd Qu.:  273.20   3rd Qu.:  269.225   3rd Qu.:  237.050
##  Max.   :25299.00   Max.   :24740.000   Max.   :24041.000
##    GSM103391           GSM103392           GSM103393
##  Min.   :    5.869   Min.   :    6.202   Min.   :    6.048
##  1st Qu.:   21.070   1st Qu.:   22.707   1st Qu.:   22.010
##  Median :   64.675   Median :   64.355   Median :   63.110
##  Mean   :  350.067   Mean   :  319.218   Mean   :  321.909
##  3rd Qu.:  280.850   3rd Qu.:  254.800   3rd Qu.:  251.600
##  Max.   :23620.000   Max.   :24631.000   Max.   :22315.000
##    GSM103394           GSM103395           GSM103396
##  Min.   :    5.906   Min.   :    6.022   Min.   :    5.754
##  1st Qu.:   21.290   1st Qu.:   24.977   1st Qu.:   22.410
##  Median :   65.945   Median :   67.525   Median :   63.355
##  Mean   :  325.781   Mean   :  320.109   Mean   :  309.978
##  3rd Qu.:  267.725   3rd Qu.:  231.200   3rd Qu.:  236.700
##  Max.   :24740.000   Max.   :21839.000   Max.   :22668.000
##    GSM103397           GSM103398           GSM103399
##  Min.   :    5.932   Min.   :    5.994   Min.   :    5.828
##  1st Qu.:   22.468   1st Qu.:   22.008   1st Qu.:   23.295
##  Median :   62.970   Median :   63.140   Median :   64.770
##  Mean   :  313.914   Mean   :  318.539   Mean   :  308.555
##  3rd Qu.:  246.700   3rd Qu.:  252.825   3rd Qu.:  232.800
##  Max.   :23508.000   Max.   :23858.000   Max.   :23250.000
##    GSM103400          GSM103401           GSM103402
##  Min.   :    5.72   Min.   :    5.527   Min.   :    5.89
##  1st Qu.:   20.46   1st Qu.:   23.570   1st Qu.:   21.66
##  Median :   60.26   Median :   64.640   Median :   64.84
##  Mean   :  322.13   Mean   :  305.877   Mean   :  330.31
##  3rd Qu.:  260.00   3rd Qu.:  216.800   3rd Qu.:  269.90
##  Max.   :22987.00   Max.   :21894.000   Max.   :23615.00
##    GSM103403           GSM103404           GSM103405
##  Min.   :    5.832   Min.   :    6.097   Min.   :    5.875
##  1st Qu.:   21.740   1st Qu.:   21.810   1st Qu.:   21.880
##  Median :   65.510   Median :   63.055   Median :   62.225
##  Mean   :  335.372   Mean   :  307.677   Mean   :  309.069
##  3rd Qu.:  275.125   3rd Qu.:  246.600   3rd Qu.:  245.100
##  Max.   :23594.000   Max.   :24601.000   Max.   :23551.000
##    GSM103406           GSM103407           GSM103408
##  Min.   :    6.006   Min.   :    6.025   Min.   :    5.673
##  1st Qu.:   21.808   1st Qu.:   22.500   1st Qu.:   20.277
##  Median :   59.755   Median :   63.895   Median :   60.405
##  Mean   :  309.458   Mean   :  319.534   Mean   :  323.487
##  3rd Qu.:  235.400   3rd Qu.:  248.800   3rd Qu.:  272.475
##  Max.   :23805.000   Max.   :23778.000   Max.   :23409.000
```

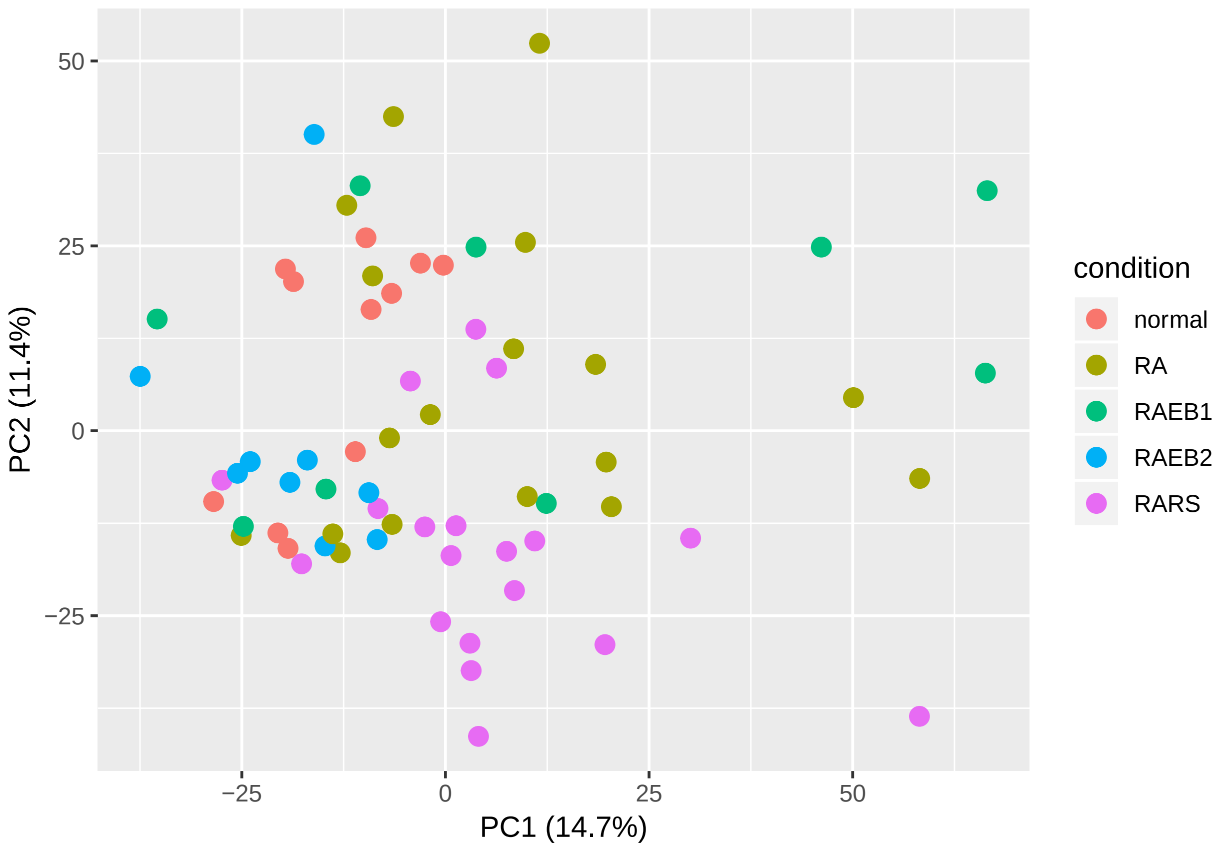
```
exprs(es.qnorm) <- normalizeBetweenArrays(log2(exprs(es.qnorm)+1), method="quantile")
summary(exprs(es.qnorm))
```

```
##    GSM103343        GSM103344        GSM103345        GSM103346
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.882   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103347        GSM103348        GSM103349        GSM103350
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103351        GSM103352        GSM103353        GSM103354
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.882   3rd Qu.: 7.882   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103355        GSM103356        GSM103357        GSM103358
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.882
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103359        GSM103360        GSM103361        GSM103362
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.575
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.882   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103363        GSM103364        GSM103365        GSM103366
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.882
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103367        GSM103368        GSM103369        GSM103370
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103371        GSM103372        GSM103373        GSM103374
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.816
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.575   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103375        GSM103376        GSM103377        GSM103378
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.882   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103379        GSM103380        GSM103381        GSM103382
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.575   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103383        GSM103384        GSM103385        GSM103386
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103387        GSM103388        GSM103389        GSM103390
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.573   1st Qu.: 4.574   1st Qu.: 4.575
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.882   3rd Qu.: 7.882
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103391        GSM103392        GSM103393        GSM103394
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.575
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103395        GSM103396        GSM103397        GSM103398
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.575
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.882
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103399        GSM103400        GSM103401        GSM103402
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.575   1st Qu.: 4.575   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.882   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103403        GSM103404        GSM103405        GSM103406
##  Min.   : 2.802   Min.   : 2.802   Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.575   1st Qu.: 4.574   1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997   Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344   Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524   Max.   :14.524   Max.   :14.524
##    GSM103407        GSM103408
##  Min.   : 2.802   Min.   : 2.802
##  1st Qu.: 4.574   1st Qu.: 4.574
##  Median : 5.997   Median : 5.997
##  Mean   : 6.344   Mean   : 6.344
##  3rd Qu.: 7.883   3rd Qu.: 7.883
##  Max.   :14.524   Max.   :14.524
```

```r
#To get get first 12000 entries:

es.qnorm.top12K <- es.qnorm
es.qnorm.top12K <- es.qnorm.top12K[head(order(apply(exprs(es.qnorm.top12K), 1, mean),
                                            decreasing = TRUE), 12000), ]
```

```r
#Have a look at the data  - make pca plot:

pcaPlot(es.qnorm.top12K,1,2) + aes(color = condition)
```

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



```
#To make a design matrix that will be used to make a model for given data:

es.design <- model.matrix(~0+condition, data=pData(es.qnorm.top12K))
es.design
```

```
##           conditionnormal conditionRA conditionRAEB1 conditionRAEB2
## GSM103343               1           0              0              0
## GSM103344               1           0              0              0
## GSM103345               1           0              0              0
## GSM103346               0           0              1              0
## GSM103347               0           1              0              0
## GSM103348               0           0              0              0
## GSM103349               0           1              0              0
## GSM103350               0           0              1              0
## GSM103351               0           0              0              0
## GSM103352               0           0              1              0
## GSM103353               0           0              1              0
## GSM103354               0           1              0              0
## GSM103355               0           1              0              0
## GSM103356               0           0              0              0
## GSM103357               0           1              0              0
## GSM103358               0           1              0              0
## GSM103359               0           0              1              0
## GSM103360               0           0              1              0
## GSM103361               0           1              0              0
## GSM103362               0           0              0              1
## GSM103363               0           1              0              0
## GSM103364               1           0              0              0
## GSM103365               1           0              0              0
## GSM103366               1           0              0              0
## GSM103367               0           1              0              0
## GSM103368               0           0              0              0
## GSM103369               1           0              0              0
## GSM103370               1           0              0              0
## GSM103371               0           0              0              1
## GSM103372               0           0              0              0
## GSM103373               0           0              0              1
## GSM103374               0           0              0              1
## GSM103375               0           0              0              0
## GSM103376               0           0              0              0
## GSM103377               0           0              0              1
## GSM103378               0           0              0              1
## GSM103379               0           0              0              0
## GSM103380               0           0              0              1
## GSM103381               0           1              0              0
## GSM103382               0           1              0              0
## GSM103383               0           0              1              0
## GSM103384               0           1              0              0
## GSM103385               0           0              0              0
## GSM103386               0           0              0              1
## GSM103387               0           0              0              0
## GSM103388               1           0              0              0
## GSM103389               1           0              0              0
## GSM103390               1           0              0              0
## GSM103391               0           1              0              0
## GSM103392               0           0              0              0
## GSM103393               0           0              0              0
## GSM103394               0           1              0              0
## GSM103395               0           0              0              0
## GSM103396               0           0              0              0
## GSM103397               0           0              0              1
## GSM103398               0           0              0              0
## GSM103399               0           1              0              0
## GSM103400               0           0              1              0
## GSM103401               0           1              0              0
## GSM103402               0           0              0              0
## GSM103403               0           0              0              0
## GSM103404               0           1              0              0
## GSM103405               0           0              0              0
## GSM103406               0           0              1              0
## GSM103407               0           0              0              0
## GSM103408               0           1              0              0
##           conditionRARS
## GSM103343             0
## GSM103344             0
## GSM103345             0
## GSM103346             0
## GSM103347             0
## GSM103348             1
## GSM103349             0
## GSM103350             0
## GSM103351             1
## GSM103352             0
## GSM103353             0
## GSM103354             0
## GSM103355             0
## GSM103356             1
## GSM103357             0
## GSM103358             0
## GSM103359             0
## GSM103360             0
## GSM103361             0
## GSM103362             0
## GSM103363             0
## GSM103364             0
## GSM103365             0
## GSM103366             0
## GSM103367             0
## GSM103368             1
## GSM103369             0
## GSM103370             0
## GSM103371             0
## GSM103372             1
## GSM103373             0
## GSM103374             0
## GSM103375             1
## GSM103376             1
## GSM103377             0
## GSM103378             0
## GSM103379             1
## GSM103380             0
## GSM103381             0
## GSM103382             0
## GSM103383             0
## GSM103384             0
## GSM103385             1
## GSM103386             0
## GSM103387             1
## GSM103388             0
## GSM103389             0
## GSM103390             0
## GSM103391             0
## GSM103392             1
## GSM103393             1
## GSM103394             0
## GSM103395             1
## GSM103396             1
## GSM103397             0
## GSM103398             1
## GSM103399             0
## GSM103400             0
## GSM103401             0
## GSM103402             1
## GSM103403             1
## GSM103404             0
## GSM103405             1
## GSM103406             0
## GSM103407             1
## GSM103408             0
## attr(,"assign")
## [1] 1 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$condition
## [1] "contr.treatment"
```

```
#we have 5 conditions:
im <- data.frame(es.design)
colnames(im) <- c("conditionnormal", "conditionRA", "conditionRAEB1",
                  "conditionRAEB2", "conditionRARS")

rm(es.design)
es.design <- as.matrix(im)

#On the base of this matrix, we fit our data:
fit <- lmFit(es.qnorm.top12K, es.design)

#Also we make bayisian model for the data called fit2:
#NB! we need to choose contrast names which specify the sample groups to compare!
# we need to specify the condion of interest and level to compare:

fit2 <- contrasts.fit(fit, makeContrasts(conditionnormal, conditionRA,
                                         conditionRAEB1, conditionRAEB2,
                                         conditionRARS,  levels=es.design))

fit2 <- eBayes(fit2)

#To do Bonferonni-hochback correction:
de <- topTable(fit2, adjust.method="BH", number=Inf)
head(de)
```

```
##        entrez symbol conditionnormal conditionRA conditionRAEB1
## EEF1A1 EEF1A1   1915        14.52369    14.50776       14.47589
## RPL37A RPL37A   6168        14.26571    14.24020       14.23359
## RPL41   RPL41   6171        14.38027    14.35944       14.39197
## RPL9     RPL9   6133        13.99103    13.94559       13.92852
## RPS12   RPS12   6206        14.00190    13.95918       13.95078
## RPL39   RPL39   6170        13.90554    13.89138       13.88007
##        conditionRAEB2 conditionRARS  AveExpr        F       P.Value
## EEF1A1       14.52369      14.52369 14.51283 337686.4 4.379843e-141
## RPL37A       14.23754      14.24334 14.24409 284434.3 1.111277e-138
## RPL41        14.31769      14.36116 14.36215 257673.5 2.692449e-137
## RPL9         13.99796      13.93178 13.95400 236430.6 4.320667e-136
## RPS12        14.03068      13.97745 13.98016 228916.3 1.224747e-135
## RPL39        13.86507      13.80120 13.86265 225073.3 2.114668e-135
##            adj.P.Val
## EEF1A1 5.255812e-137
## RPL37A 6.667661e-135
## RPL41  1.076980e-133
## RPL9   1.296200e-132
## RPS12  2.939392e-132
## RPL39  4.229337e-132
```

```
# Here, we have a matrix that contains the enriched genes,
#we take the top genes and submit to database (msigdbr) to get the enriched pathways.
#We first target the hallmark pathways, which are well studied and
#then we target all the pathways. We try to find out what special pathways
#are involved in our normal versus condition.
#This will further give us insight into the comparision.
```

```
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
```

```
de <- as.data.table(de, keep.rownames=TRUE)
de[entrez == "EEF1A1"]
```

```
##        rn entrez symbol conditionnormal conditionRA conditionRAEB1
## 1: EEF1A1 EEF1A1   1915        14.52369    14.50776       14.47589
##    conditionRAEB2 conditionRARS  AveExpr        F       P.Value
## 1:       14.52369      14.52369 14.51283 337686.4 4.379843e-141
##        adj.P.Val
## 1: 5.255812e-137
```

```
#BioConductor: install fgsea:

library(fgsea)
```

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

```
library(tibble)
library(Rcpp)
```

```
# To make a new matrix de2 which will store information about pathways:

de2 <- data.frame(de$entrez, de$P.Value)
colnames(de2) <- c('ENTREZ', 'stat')
```

```
# To get the rank of genes from top differentially expressed to non significant:

ranks <- deframe(de2)
head(ranks, 20)
```

```
##        EEF1A1        RPL37A         RPL41          RPL9         RPS12
## 4.379843e-141 1.111277e-138 2.692449e-137 4.320667e-136 1.224747e-135
##         RPL39        RPL10A         HUWE1         RPS4X         RPLP0
## 2.114668e-135 1.183251e-134 3.521782e-134 3.896013e-133 7.069068e-133
##         RPS15         RPS23         RPS16          RPL7         RPS29
## 9.536100e-133 2.603295e-132 3.160321e-132 3.330305e-132 5.161301e-132
##         RPS10         RPL31         RPL14         RPLP1         RPL30
## 6.154875e-132 3.801197e-131 8.153319e-130 2.680015e-129 6.516108e-129
```

```r
# Load the pathways into a named list:

library(msigdbr)
```

```
## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

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

## The following objects are masked from 'package:IRanges':
##
##     collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':
##
##     first, intersect, rename, setdiff, setequal, union

## The following object is masked from 'package:Biobase':
##
##     combine

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

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

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

```r
m_df <- msigdbr(species = "Homo sapiens")

# View(m_df):
pathways <- split(m_df$human_gene_symbol, m_df$gs_name)
head(pathways)
```

```
## $AAACCAC_MIR140
##   [1] "ABCC4"     "ACTN4"     "ACVR1"     "ADAM9"     "ADAMTS5"
##   [6] "AGER"      "ANK2"      "API5"      "BACH1"     "BAZ2B"
##  [11] "BCL11A"    "BCL2L2"    "BCL9"      "C15orf29"  "C1orf21"
##  [16] "C3orf58"   "C7orf60"   "CACNA1C"   "CEBPA"     "CHD4"
##  [21] "CIT"       "COL23A1"   "CSK"       "CSNK1G3"   "CTCF"
##  [26] "CUL3"      "DAZL"      "DBNDD2"    "DCUN1D4"   "DDX3X"
##  [31] "DDX3Y"     "DHX57"     "DPP4"      "DSCAM"     "DTNA"
##  [36] "E2F3"      "EHD1"      "EPHB1"     "ERC2"      "ETV3"
##  [41] "EYA2"      "FAM123A"   "FAM175B"   "FAM178A"   "GABARAP"
##  [46] "GALNTL1"   "GDF6"      "GIT1"      "GYS1"      "HDAC4"
##  [51] "HNRNPH3"   "HSPA13"    "IGFBP5"    "KCND2"     "KIAA1370"
##  [56] "LOC440742" "LOXL3"     "LRRC4"     "LRRC8E"    "MAP3K8"
##  [61] "MDGA2"     "MEX3C"     "MGAT1"     "MMD"       "NAV3"
##  [66] "NKIRAS2"   "NR3C1"     "NUTF2"     "OGT"       "OSTM1"
##  [71] "PDGFRA"    "PFN1"      "PHF20L1"   "PHYHIP"    "PITX2"
##  [76] "PPP1CC"    "PRIMA1"    "R3HDM1"    "REEP1"     "RNF19A"
##  [81] "RTKN2"     "SENP1"     "SIAH1"     "SLC25A13"  "SLC38A2"
##  [86] "SLC41A2"   "SLMAP"     "SNX2"      "SOX4"      "SRR"
##  [91] "STAG1"     "STRADB"    "SYT6"      "TAF9B"     "TBX3"
##  [96] "TP53INP2"  "TSHZ1"     "TSPAN2"    "TSSK2"     "TTYH2"
## [101] "UBASH3B"   "USP6"      "VEGFA"     "WHSC1L1"   "WNT1"
## [106] "YES1"      "ZBED4"     "ZBTB10"    "ZNF182"    "ZNF608"
## [111] "ZNF654"
##
## $AAAGACA_MIR511
##   [1] "ABCG8"     "ACE"       "ADAMTSL3"  "ADGRF5"    "ADSS"
##   [6] "AGBL3"     "ALCAM"     "ANKZF1"    "AQP6"      "ARHGEF17"
##  [11] "ATL2"      "ATP2B2"    "ATRX"      "BCL11A"    "BTG1"
##  [16] "BUB3"      "BZRAP1"    "C11orf51"  "C18orf34"  "C1orf21"
##  [21] "C1QL2"     "C21orf59"  "C2orf71"   "C5orf41"   "C6orf106"
##  [26] "C7orf23"   "C7orf42"   "CALM1"     "CAMK2N1"   "CAMTA1"
##  [31] "CAPRIN1"   "CCND1"     "CCNT2"     "CDH2"      "CDK14"
##  [36] "CDK19"     "CELF1"     "CELF6"     "CEP350"    "CLK2"
##  [41] "CLTC"      "CNOT4"     "CORIN"     "CREM"      "CRIM1"
##  [46] "DCTN4"     "DDX3X"     "DDX3Y"     "DEDD"      "DNAJB12"
##  [51] "DNAJC13"   "DSC1"      "DUSP6"     "DYRK1B"    "E2F3"
##  [56] "EDEM3"     "EFR3A"     "EIF2C1"    "EIF2C2"    "EIF2C4"
##  [61] "ELAVL3"    "EMILIN2"   "EML4"      "ENPP1"     "ENPP4"
##  [66] "EPHA4"     "ESRRG"     "EYA1"      "EYA4"      "FAM117A"
##  [71] "FAM60A"    "FGF13"     "FIP1L1"    "FMR1"      "FN1"
##  [76] "FNDC1"     "FNDC5"     "FOXK2"     "FOXN3"     "GAD2"
##  [81] "GEMIN2"    "GFAP"      "GJA1"      "GLRA2"     "GPR116"
##  [86] "HAS2"      "HCN4"      "HLF"       "HLTF"      "HOXA13"
##  [91] "IGF2BP1"   "IGF2BP3"   "KCNE1"     "KCNMA1"    "KHDRBS2"
##  [96] "KIAA1429"  "KLF9"      "KLHL18"    "KLHL24"    "LATS1"
## [101] "LINC00483" "LMCD1"     "LPP"       "LRCH4"     "LUC7L3"
## [106] "MAP3K2"    "MAP4K4"    "MAPK1IP1L" "MBD2"      "MBD6"
## [111] "MDGA2"     "METAP2"    "MIB1"      "MINK1"     "MRPL21"
## [116] "MSTN"      "MTAP"      "MYCBP"     "MYO19"     "NACC1"
## [121] "NEUROD6"   "NHLH2"     "NLK"       "NR4A2"     "NRXN3"
## [126] "NTRK2"     "NXPH1"     "ONECUT2"   "PAX8"      "PCDH10"
## [131] "PCDH17"    "PELI1"     "PHLPP1"    "PIK3R3"    "PMEPA1"
## [136] "POGK"      "POU4F2"    "PPARGC1A"  "PRELP"     "PRPF4B"
## [141] "PSMA1"     "PSMD10"    "QKI"       "RAB22A"    "RAB2A"
## [146] "RBM15B"    "RBM26"     "RECK"      "REV3L"     "RGL1"
## [151] "RHOJ"      "RHOT1"     "RNF19A"    "ROBO2"     "RPS6KB1"
## [156] "RPS6KL1"   "SATB2"     "SCN4B"     "SEMA3F"    "SEMA6D"
## [161] "SEPP1"     "SLC22A17"  "SLC25A26"  "SLC6A6"    "SLITRK1"
## [166] "SMARCE1"   "SOCS2"     "SORCS3"    "SOST"      "SOX12"
## [171] "SPTBN4"    "SPTLC2"    "SRGAP3"    "SS18"      "ST18"
## [176] "SYT11"     "T"         "TAF5"      "THOC5"     "TIAL1"
## [181] "TMEM196"   "TNRC6A"    "TNRC6B"    "TOB1"      "TRAPPC3"
## [186] "TRAPPC8"   "TRIM2"     "TRIM24"    "TXNL1"     "UBE2H"
## [191] "VANGL2"    "VAV3"      "VKORC1L1"  "VMP1"      "WNT16"
## [196] "YTHDF2"    "YY1"       "ZADH2"     "ZCCHC24"   "ZDHHC21"
## [201] "ZNF319"    "ZNF654"    "ZNF706"
##
## $AAAGGAT_MIR501
##   [1] "ACACA"    "ACADSB"   "ADCYAP1"  "ADIPOR2"  "ALS2"     "AMMECR1"
##   [7] "APOLD1"   "ATP6V1H"  "BCL6"     "BCLAF1"   "C8orf82"  "CA6"
##  [13] "CACHD1"   "CAMTA1"   "CCDC140"  "CD164"    "CELF2"    "CELSR2"
##  [19] "CHODL"    "CLK1"     "CLK2"     "CTDSP1"   "CTDSPL2"  "CUL1"
##  [25] "CUX2"     "DCX"      "DNAJB12"  "ELAVL4"   "ERRFI1"   "FAM179B"
##  [31] "GIF"      "GRAMD4"   "GRB10"    "H2AFX"    "HAS2"     "HES5"
##  [37] "HOXB8"    "JUN"      "KCND2"    "KCNRG"    "KIAA2022" "KIF1C"
##  [43] "KIF2A"    "KLHL14"   "KRR1"     "LARP1"    "LEPROTL1" "LPGAT1"
##  [49] "LPIN1"    "LRRC1"    "MAP2K1"   "MAP3K8"   "MCU"      "MEF2C"
##  [55] "MYB"      "MYCL1"    "MYLK"     "NFASC"    "NFIL3"    "NFIX"
##  [61] "NPR3"     "NR2F2"    "NR4A3"    "PCDH19"   "PDK1"     "PHC1"
##  [67] "PHF16"    "PHF6"     "PIK3AP1"  "PITX2"    "PLP1"     "PLXNB1"
##  [73] "PNN"      "PPP1CB"   "PPP2R5E"  "PPP6R3"   "PRKCE"    "PURA"
##  [79] "QKI"      "RAB22A"   "RABGEF1"  "RASL10B"  "RCN1"     "RDX"
##  [85] "RET"      "RGL1"     "RNF11"    "ROBO2"    "RPGRIP1L" "RSBN1"
##  [91] "SATB2"    "SCN3A"    "SENP3"    "SEPHS1"   "SGPP1"    "SLC25A3"
##  [97] "SLC35B3"  "SLITRK5"  "SMC1A"    "SMEK1"    "SNAP29"   "SOX11"
## [103] "SOX4"     "SPOPL"    "SRR"      "SRSF2"    "SYNC"     "SYNJ1"
## [109] "SYT7"     "TAF5L"    "TAPT1"    "TNNI2"    "TOMM70A"  "TRIM39"
## [115] "UBAP1"    "UBE2Q1"   "UBE4B"    "USP12"    "VDAC2"    "WDFY3"
## [121] "WIPF2"    "WT1-AS"   "ZC3H7A"   "ZIC4"     "ZMYM5"    "ZNF238"
##
## $AAAGGGA_MIR204_MIR211
##   [1] "ADAMTS9"  "ADCY6"    "AKAP1"    "ALPL"     "ANGPT1"   "ANKRD13A"
##   [7] "ANXA11"   "AP1S1"    "AP1S3"    "AP2A2"    "AP3M1"    "APH1A"
##  [13] "ARAP2"    "ARCN1"    "ARGLU1"   "ARHGAP29" "ARL8B"    "ATF2"
##  [19] "ATP2B1"   "AUP1"     "BAZ2A"    "BCL11B"   "BCL2"     "BCL9"
##  [25] "BCL9L"    "BRD4"     "BRPF3"    "BUD31"    "C16orf72" "C17orf48"
##  [31] "C1orf144" "C21orf63" "CAPRIN1"  "CCNT2"    "CCPG1"    "CDC25B"
##  [37] "CDC42"    "CDH2"     "CELSR3"   "CHD5"     "CHN2"     "CHP"
##  [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"    "EIF2C4"
##  [67] "ELAVL3"   "ELF2"     "ELL2"     "ELMOD3"   "ELOVL6"   "EPHA7"
##  [73] "EPHB6"    "ESR1"     "ESRRG"    "EZR"      "FAM117B"  "FAM120C"
##  [79] "FAM122B"  "FAM160A2" "FAM175B"  "FARP1"    "FBN2"     "FBXW7"
##  [85] "FJX1"     "FNIP1"    "FRAS1"    "FREM1"    "FRY"      "GABRB3"
##  [91] "GAPVD1"   "GGA2"     "GLIS3"    "GPM6A"    "GRM1"     "HIC2"
##  [97] "HMGA2"    "HOOK3"    "HOXC8"    "HS2ST1"   "IGF2R"    "ING4"
## [103] "ITPR1"    "JPH3"     "KCNA3"    "KCTD1"    "KDM2A"    "KHDRBS1"
## [109] "KHDRBS3"  "KITLG"    "KLF12"    "KLHL13"   "LATS1"    "LRRC8D"
## [115] "MALL"     "MAML3"    "MAP1LC3B" "MAP3K3"   "MBNL1"    "MED13L"
## [121] "METAP1"   "MIR600HG" "MLL"      "MLLT3"    "MMGT1"    "MON2"
## [127] "MRPL35"   "MRPL52"   "MYO10"    "NAA15"    "NBEA"     "NCOA7"
## [133] "NEUROG1"  "NOVA1"    "NPTX1"    "NR3C1"    "NR4A2"    "NRBF2"
## [139] "NTRK2"    "P4HB"     "PCDH9"    "PHF13"    "PID1"     "PLAG1"
## [145] "POU3F2"   "PPARGC1A" "PPP3R1"   "PRDM2"    "PRPF38B"  "PRRX1"
## [151] "RAB10"    "RAB14"    "RAB1A"    "RAP2C"    "REEP1"    "RERE"
## [157] "RHOBTB3"  "RHOT1"    "RICTOR"   "RPS6KA3"  "RPS6KA5"  "RPS6KC1"
## [163] "RSPO3"    "RTKN2"    "RUNX2"    "SATB2"    "SCRT2"    "SEC24D"
## [169] "SEC61A2"  "SERINC3"  "SETD8"    "SF3B1"    "SGCZ"     "SGIP1"
## [175] "SHC1"     "SIN3A"    "SIRT1"    "SLC17A7"  "SLC22A2"  "SLC37A3"
## [181] "SLITRK4"  "SLTM"     "SMOC1"    "SOCS6"    "SOX11"    "SOX4"
## [187] "SPOP"     "SPRED1"   "SPRYD7"   "SSRP1"    "ST7"      "STXBP5"
## [193] "SUMO2"    "SUMO4"    "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"    "ZFYVE20"  "ZNF282"
## [223] "ZNF335"   "ZNF423"
##
## $AAANWWTGC_UNKNOWN
##   [1] "ACTB"      "ADHFE1"    "AFF4"      "ANK2"      "ANK3"
##   [6] "APP"       "ASPA"      "ATOH7"     "ATP1B1"    "ATP2B4"
##  [11] "ATXN7L1"   "BCL11A"    "BCL6"      "BNC2"      "C11orf87"
##  [16] "C17orf85"  "CACNA1D"   "CACNG3"    "CALM1"     "CD14"
##  [21] "CDC42EP3"  "CDC42EP5"  "CDH13"     "CDK2AP1"   "CEPT1"
##  [26] "CHD2"      "CITED2"    "CNTFR"     "DAB1"      "DCAF11"
##  [31] "DCHS2"     "DDIT3"     "DIS3L"     "DLG2"      "DLGAP4"
##  [36] "DMD"       "DNAJB5"    "DPYSL5"    "DRD3"      "DSCAM"
##  [41] "DSEL"      "DSTN"      "DTX3L"     "DUSP1"     "DYNC1I2"
##  [46] "EBF1"      "EFNA5"     "EGFLAM"    "EIF4EBP2"  "ELAVL4"
##  [51] "ELF4"      "EPHA7"     "EPHB2"     "ESR1"      "FBXW7"
##  [56] "FGF7"      "FGFR2"     "FLJ45983"  "FN1"       "FOXN3"
##  [61] "FOXP1"     "FOXP2"     "FTHL17"    "FZD7"      "GANAB"
##  [66] "GATA3"     "GLRA2"     "GPC3"      "GPC6"      "GPR21"
##  [71] "GPRIN3"    "GRHL3"     "GRIN2B"    "GTF2E2"    "HEPACAM"
##  [76] "HHEX"      "HOXA2"     "HOXA3"     "HOXB2"     "HOXB6"
##  [81] "HOXC4"     "IGF2BP1"   "INHBA"     "ITM2C"     "KANK1"
##  [86] "KCNJ13"    "KLF12"     "KLF14"     "KRTAP8-1"  "LEAP2"
##  [91] "LECT1"     "LIPG"      "LOC148872" "LOX"       "LOXL4"
##  [96] "LRRC3B"    "LRRN1"     "LSAMP"     "LUC7L3"    "MAML3"
## [101] "MAN2A2"    "MAP3K4"    "MAPK3"     "MBNL1"     "MEF2C"
## [106] "MEIS1"     "MGLL"      "MID1"      "MLLT6"     "MMP3"
## [111] "MPZL3"     "MRPL24"    "MRPS18B"   "MYCL1"     "MYH2"
## [116] "MYLK"      "NEK6"      "NEUROG1"   "NFE2L2"    "NNAT"
## [121] "NR2F2"     "NRAS"      "NTN1"      "NTRK3"     "OLFM1"
## [126] "OLIG2"     "OMG"       "OTX2"      "PATZ1"     "PAX1"
## [131] "PAX6"      "PCSK1"     "PCTP"      "PDGFRB"    "PHF15"
## [136] "PHOX2B"    "PHTF1"     "PIK3R3"    "POU2F1"    "POU4F1"
## [141] "PPARGC1A"  "PPFIA2"    "PPP1R10"   "PPP2R2A"   "PPP3CC"
## [146] "PRDM16"    "PRIMA1"    "PRKRIR"    "PRPF4B"    "RAB10"
## [151] "RBMX"      "RORA"      "RRS1"      "RSPO2"     "S100PBP"
## [156] "SALL3"     "SAMD12"    "SATB2"     "SEMA6C"    "SESN2"
## [161] "SFRP2"     "SGCD"      "SHC3"      "SIX5"      "SKIL"
## [166] "SKP2"      "SLMAP"     "SNCAIP"    "SNX25"     "SORT1"
## [171] "SOX13"     "SOX4"      "SOX5"      "SPAG9"     "SPARCL1"
## [176] "SSBP3"     "STEAP2"    "TBC1D8B"   "TFAP4"     "TFDP2"
## [181] "TGIF1"     "THBS2"     "TLE4"      "TLK1"      "TLX3"
## [186] "TRAM1"     "TRPM3"     "TSC22D4"   "ZFPM1"     "ZHX3"
## [191] "ZNF462"    "ZNF827"    "ZW10"
##
## $AAAYRNCTG_UNKNOWN
##   [1] "ABT1"      "ACVR1"     "ADAM12"    "ADD3"      "AGGF1"
##   [6] "ANKRD12"   "ANKRD28"   "AP4S1"     "APBB2"     "APOBR"
##  [11] "AQP2"      "ARHGAP44"  "ARID1A"    "ARID4A"    "ARPC2"
##  [16] "ARSG"      "ARX"       "ASB4"      "ASPH"      "ATOH8"
##  [21] "ATP1A2"    "ATP5L"     "ATPIF1"    "AXDND1"    "B4GALT6"
##  [26] "BAI3"      "BAMBI"     "BCL2L1"    "BCL9"      "BMPR1B"
##  [31] "BMX"       "BRSK2"     "BTBD3"     "BUB3"      "C11orf84"
##  [36] "C11orf92"  "C12orf65"  "C13orf30"  "C14orf1"   "C15orf26"
##  [41] "C17orf28"  "C20orf197" "C3orf19"   "C6orf138"  "CA3"
##  [46] "CACNA2D3"  "CACNB2"    "CAPN1"     "CAPZA1"    "CASQ2"
##  [51] "CBX2"      "CCNJ"      "CCNY"      "CDC23"     "CDH2"
##  [56] "CER1"      "CHRM1"     "CITED2"    "CLDN5"     "CLTC"
##  [61] "CMKLR1"    "CNTLN"     "CNTN1"     "COCH"      "COL12A1"
##  [66] "COL1A2"    "COL4A5"    "COL4A6"    "COLEC10"   "CRAT"
##  [71] "CRH"       "CRKL"      "CRYGD"     "CRYGS"     "CSNK1A1"
##  [76] "CSRNP3"    "CSTF3"     "CYBRD1"    "DAAM1"     "DBNDD2"
##  [81] "DCAKD"     "DDAH2"     "DDX4"      "DEF6"      "DENND4A"
##  [86] "DGKB"      "DHH"       "DHRS4"     "DHRS4L2"   "DIDO1"
##  [91] "DMD"       "DMRT1"     "DNAJA2"    "DNAJB3"    "DNAJB4"
##  [96] "DSCAML1"   "DUSP4"     "DYNC1I1"   "DYRK1A"    "EDA"
## [101] "EFNA1"     "EGFLAM"    "EIF5"      "EMX2"      "EPC1"
## [106] "EPHA7"     "ERBB4"     "ERRFI1"    "ESRP2"     "ESRRB"
## [111] "ESRRG"     "EYA1"      "FAM49A"    "FAM83F"    "FCER1A"
## [116] "FGD4"      "FGF10"     "FGF12"     "FGFR1"     "FGFR1OP2"
## [121] "FIZ1"      "FKRP"      "FMNL3"     "FNDC9"     "FOXA1"
## [126] "FOXG1"     "FOXO4"     "FOXP2"     "FSIP2"     "FST"
## [131] "GABRA3"    "GDNF"      "GFI1"      "GGNBP2"    "GJB4"
## [136] "GLDN"      "GNAQ"      "GPR85"     "GPRC5D"    "GRIN2B"
## [141] "H3F3A"     "HDAC8"     "HESX1"     "HEXIM2"    "HGF"
## [146] "HIC2"      "HIP1R"     "HN1"       "HOXA10"    "HOXA5"
## [151] "HOXB8"     "HPSE2"     "HSD3B7"    "ICAM4"     "ID1"
## [156] "IGF1"      "IL1RAPL1"  "INHBC"     "IP6K2"     "ITGA10"
## [161] "ITGA8"     "JPH1"      "KANK2"     "KCNIP2"    "KCNK5"
## [166] "KCNN3"     "KCNQ1DN"   "KIAA0182"  "KITLG"     "KLF5"
## [171] "KLHDC10"   "KLHL20"    "KLHL3"     "LARS2"     "LENG9"
## [176] "LHFP"      "LHX9"      "LMO7"      "LOC151534" "LRP5"
## [181] "LRRC4"     "LRRN4CL"   "LTBP1"     "MAML1"     "MANF"
## [186] "MAP2"      "MAP3K5"    "MAP6"      "MEIS1"     "MGAT1"
## [191] "MGAT4A"    "MID1"      "MLL"       "MOAP1"     "MPP6"
## [196] "MPPED2"    "MRPL13"    "MTA2"      "MTBP"      "MYF6"
## [201] "MYH1"      "MYH10"     "MYO18A"    "NAGLU"     "NAPB"
## [206] "NAV2"      "NAV3"      "NCDN"      "NDNF"      "NDST4"
## [211] "NDUFS4"    "NEK1"      "NEK2"      "NFATC4"    "NFYB"
## [216] "NMI"       "NMT1"      "NR2F1"     "NRG1"      "NTRK2"
## [221] "NUP54"     "NXPH4"     "OMA1"      "OMG"       "OR2L13"
## [226] "OTX2"      "PACRG"     "PAPD5"     "PARK2"     "PART1"
## [231] "PCDH17"    "PCDH18"    "PCF11"     "PCYT1B"    "PDGFB"
## [236] "PDGFRA"    "PDLIM2"    "PDS5B"     "PDZRN4"    "PFN2"
## [241] "PHC2"      "PHEX"      "PHF1"      "PHF15"     "PHF6"
## [246] "PHOX2B"    "PLAGL2"    "PLEC"      "PLEKHM1"   "PLP2"
## [251] "PMCH"      "PMCHL1"    "PODXL2"    "POFUT1"    "POU2AF1"
## [256] "POU4F1"    "PPAP2B"    "PPP1R9B"   "PPP2R3A"   "PPP2R4"
## [261] "PPP2R5E"   "PPP3CA"    "PRELP"     "PRKCG"     "PRKCQ"
## [266] "PROK2"     "PTH1R"     "PXN"       "R3HDM1"    "RAB30"
## [271] "RAB5B"     "RAB5C"     "RAPGEF4"   "RBMS3"     "RGS17"
## [276] "RNF146"    "ROBO4"     "ROR1"      "RPLP0"     "RTN1"
## [281] "RUFY3"     "S1PR2"     "SCN3B"     "SCN5A"     "SCN8A"
## [286] "SCOC"      "SDCBP"     "SEMA6D"    "SEPT7"     "SESN3"
## [291] "SGCD"      "SH2D6"     "SHC3"      "SHCBP1L"   "SIPA1"
## [296] "SIRPA"     "SLC26A6"   "SLC4A1"    "SLC6A1"    "SMARCA2"
## [301] "SNX9"      "SORBS2"    "SOX12"     "SOX21"     "SOX30"
## [306] "SOX5"      "SPOCK2"    "SPTLC2"    "SRGAP2"    "SRSF8"
## [311] "SSBP2"     "ST7L"      "STAC3"     "STAG1"     "STAG2"
## [316] "STC2"      "STRN3"     "STRN4"     "TAS1R2"    "TEF"
## [321] "TFAP4"     "TFDP2"     "TM2D3"     "TMEM182"   "TMEM27"
## [326] "TMEM69"    "TMSB4X"    "TMSB4XP1"  "TMSL3"     "TMSL6"
## [331] "TNFAIP8"   "TNS1"      "TNXB"      "TP53INP2"  "TRDN"
## [336] "TREML1"    "TRIM28"    "TRIM68"    "TRIM8"     "TRIML1"
## [341] "TRPS1"     "TSC22D3"   "TSPAN7"    "TSPY26P"   "TSSK3"
## [346] "TTC17"     "TUSC2"     "UBE2W"     "UBXN10"    "USP1"
## [351] "VDR"       "VIP"       "VKORC1L1"  "VWA5A"     "WBP1"
## [356] "WNT2B"     "WT1"       "WT1-AS"    "XRCC1"     "ZADH2"
## [361] "ZBTB11"    "ZFP91"     "ZFPM2"     "ZIC1"      "ZIC4"
## [366] "ZMAT3"     "ZNF238"    "ZNF296"    "ZNF503"    "ZNF521"
## [371] "ZNF524"    "ZNF654"    "ZNF687"    "ZNF710"
```

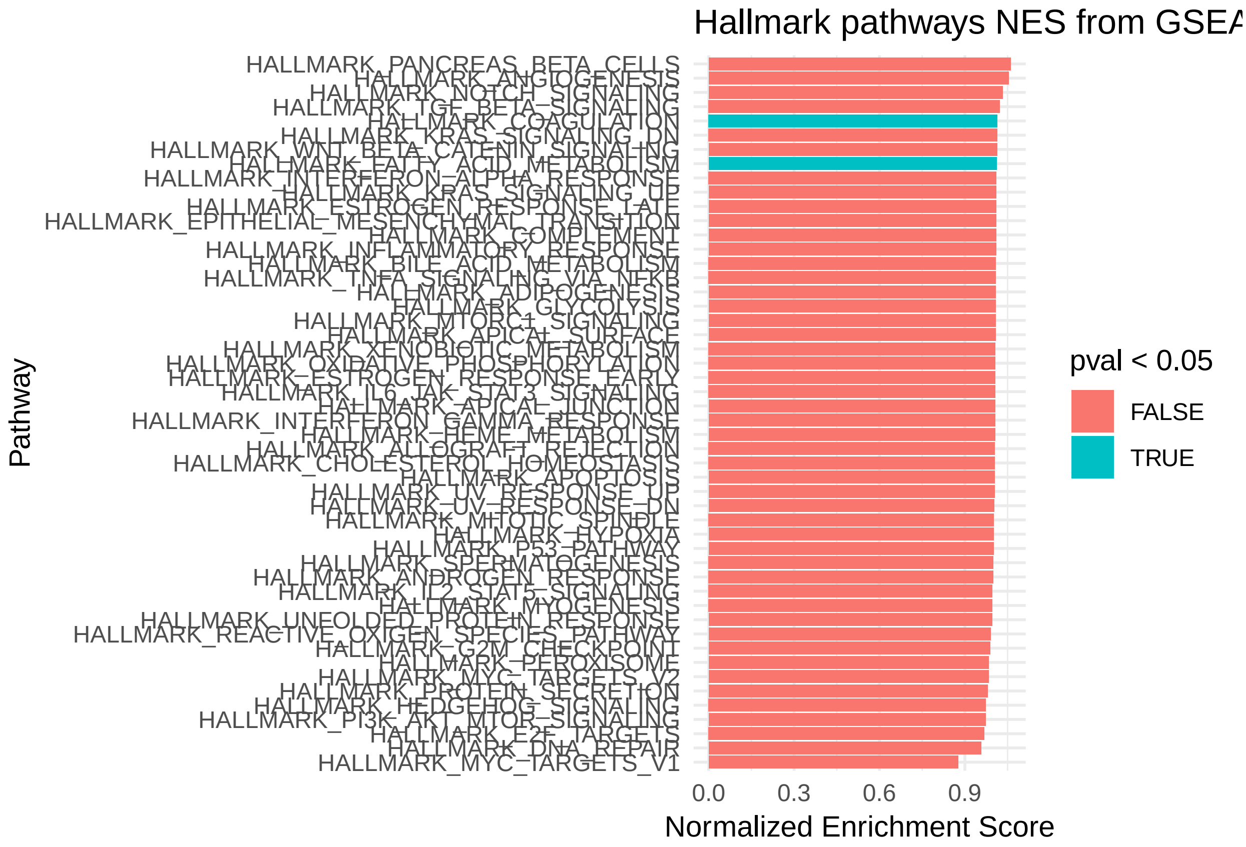
```r
# filter the list to include only hallmark pathways:

library(dplyr)
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)
```

```r
# 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" "AIM1"    "AKAP7"   "APP"
```

```
# To run the fgsea algorithm on hallmark.pathways:

fgseaEs <- fgsea(pathways=pathways.hallmark, stats=ranks, nperm=1000)
fgseaEsTidy <- fgseaEs %>%
  as_tibble() %>%
  arrange(desc(NES))  #ggploting for halmark pathways
```

```
# ggplot for hallmark pathways:
library(ggplot2)

 #pdf("fgseaEsTidy.pdf", width = 10, height = 20)

ggplot(fgseaEsTidy, 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 plotted all the significant patways in the hallmark pathways as 'turquoise'
# We can see that:
    # HALLMARK_ANGIOGENESIS, HALLMARK_ADIPOGENESIS, HALLMARK_ANDROGEN_RESPONSE etc.
# pathway are activated!
# Let's look at all pathways involving the following genes that
#were mentioned in the initial work (paper):
```

```
# FIT1, IFITM1, hIAN2,  DLK1, RAB27B,   PTHR2,  GNPDA1, COCH,   FLJ22457,
#XK,    TMEM46, AKAP12, ARPP-21,    PAX5,   GPR,    LOC90925,   TCF8,   VPREB3,
#EBF,   MME, FECH, ALAS2,   ALAD,   HMBS,   UROD,   GATA1,  CA2,    EPO-R,  CGI-69, TRAP1,  TIMM10.

# We are going to search the entire pathway list for any pathway
#that contains these genes, this can be done by subsetting and
#appending to a new dataframe of pathways.
```

```
# To make a list of all pathways fgseares.all:

fgseaEs.all <- fgsea(pathways=pathways, stats=ranks, nperm=1000)

item <- data.frame('FIT1',  'IFITM1',   'hIAN2',    'DLK1', 'RAB27B',
                   'PTHR2', 'GNPDA1',   'COCH', 'FLJ22457', 'XK',
                   'TMEM46',    'AKAP12',   'ARPP-21',  'PAX5', 'GPR',
                   'LOC90925',  'TCF8', 'VPREB3',   'EBF',  'MME',
                   'FECH', 'ALAS2', 'ALAD', 'HMBS', 'UROD', 'GATA1',
                   'CA2',   'EPO-R',    'CGI-69', 'TRAP1',  'TIMM10')

item<- t(item)
rownames(item) <- NULL

entry <- function(){

  x<- for (i in item){
    print(de[entrez == i])

  }

  return(x)

}
```

```
# searching for the genes in pathway and appending the rownumbers
#sink('numbers.csv')

options(max.print=2000)

for(i in item){
  print(grep(i, fgseaEs.all$leadingEdge))
}
```

```
##   [1]    38    89   286   601   604   605   615   723   822   869   969
##  [12]  1349  1407  1489  1491  1521  1531  1596  1657  1806  1972  2828
##  [23]  2836  3870  4509  4629  4706  4899  4923  5835  6039  6040  6349
##  [34]  6403  6966  7262  7272  7375  8150  8165  8177  8224  8359  8548
##  [45]  8554  8557  8583  8653  8654  8680  8841  8952  8963  8979  8996
##  [56]  8999  9448  9451  9455  9532  9569  9586  9613  9769  9806  9833
##  [67]  9982  9998 10291 10386 10463 10699 10720 10851 11003 11137 11148
##  [78] 11376 11428 11452 11480 11606 11639 11688 11709 11953 11976 11986
##  [89] 11995 12037 12046 12094 12108 12111 12126 12142 12146 12509 12522
## [100] 12617 12621 12653 12682 12883 12890 13152 13157 13162 13171 13531
## [111] 13759 13768 13784 13856 14180 14189 14367 14712 15075 15366 15731
## [122] 15808 15875 16114 16559 16934 16935 17059 17132 17200 17376 17477
## [133] 17506 17549 17561 17623 17720 17770 17791
## [1]  7302 16732 17001 17516 17541
## integer(0)
##   [1]   302   307   319   580   649   695   708   776   854   932   936
##  [12]  1020  1619  1631  1963  4659  4866  4932  6894  8522  8612  9035
##  [23]  9047  9065  9564  9649  9651  9659  9688  9789  9794  9796  9860
##  [34]  9921  9989 10151 10283 10625 10677 10969 11097 11196 11350 11426
##  [45] 11548 11748 11854 11947 11953 12064 12224 12416 12625 12642 13095
##  [56] 13166 13252 13275 13502 13983 14001 14002 14005 14027 14102 14214
##  [67] 14216 14312 14321 14328 14399 14400 14419 14440 14448 14471 14491
##  [78] 14496 14523 14538 14626 14636 14655 14743 14771 14784 14788 14794
##  [89] 14801 14817 14824 15185 15226 15283 15352 15354 15416 15467 15571
## [100] 15619 15690 15760 15831 16386 16387 16493 16565 16652 16671 16844
## [111] 16847 16870 16896 16962 17006 17057 17164 17204 17206 17252 17311
## [122] 17320 17485 17675
##  [1]  2801  3660  4526  4563  4568  5483  5486  8106  8616 11794
## integer(0)
## [1]    68  1188  2342  3158  3689  4002  4571 13568
## [1] 16641
## integer(0)
##  [1]   118   189   336   614  1265  1359  1562  1587  1588  1621  2340
## [12]  3419  4569  4570  4620  5244  6478  8161  9719  9901 10703 10864
## [23] 11574 11718 12824 13259 13855 14508 14800 17063 17070 17232 17545
## [34] 17640
## integer(0)
##  [1]     8    11    45   118   137   158   199   320   346   354   806
## [12]   825  1213  1815  2036  6163  6979  6981  8365  8913  8987  8991
## [23]  8996  9213  9336  9581  9878  9923  9932 10156 10183 10246 10543
## [34] 10594 10658 11240 11394 11749 11850 11924 11939 11964 11984 12046
## [45] 13083 13439 14005 14064 14180 14260 14301 14379 15052 15089 17172
## [56] 17198 17260 17360 17782 17784
## integer(0)
##   [1]   584   691   808  1251  1443  1445  1449  1536  2295  2300  2924
##  [12]  3070  3083  3312  3335  3341  3352  3353  3358  3359  3608  4094
##  [23]  4689  4690  4769  4770  4908  4977  5314  6523  6525  6690  6692
##  [34]  7200  7603  7608  7812  8045  8172  8261  8373  8477  8479  8767
##  [45]  9285  9319  9414  9435  9445  9453 10056 10058 10060 10062 10064
##  [56] 10126 10359 10550 10710 10902 10904 10908 10910 10912 10977 11115
##  [67] 11117 11346 11396 11663 11742 11744 11839 12137 12251 12295 12301
##  [78] 12379 12394 12427 12478 12513 12867 12897 12953 12966 13347 13845
##  [89] 14092 14346 14351 14358 14364 14425 14478 15525 15732 16527 16698
## [100] 16781 16785 16807 17698 17706
##   [1]   265   277   318  1008  1019  1088  1151  1224  1272  1466  1504
##  [12]  1531  1818  2049  2113  2153  2228  2250  2281  2301  2491  2668
##  [23]  2688  3301  3407  3629  3631  3632  4190  4274  4275  5543  5568
##  [34]  5571  5713  5750  5751  6280  6281  6329  6395  6399  7230  7711
##  [45]  7720  8200  8242  8262  8269  8273  8275  8289  8311  8349  8355
##  [56]  8357  8360  8369  8389  8391  8428  8442  8544  8551  8553  8561
##  [67]  8613  8665  8711  8732  8840  8915  8924  8952  8997  8999  9010
##  [78]  9011  9144  9230  9289  9368  9369  9410  9412  9416  9418  9481
##  [89]  9498  9561  9617  9624  9643  9660  9662  9713  9735  9740  9742
## [100]  9909  9917  9922 10095 10113 10117 10141 10202 10317 10318 10390
## [111] 10406 10409 10454 10524 10531 10573 10586 10639 10656 10732 10741
## [122] 10743 10757 10769 10771 10776 10781 10791 10817 10823 10947 10974
## [133] 10998 11011 11090 11096 11126 11139 11145 11315 11399 11405 11412
## [144] 11417 11424 11428 11452 11495 11498 11502 11504 11514 11520 11537
## [155] 11539 11546 11618 11632 11677 11688 11705 11717 11737 11752 11790
## [166] 11792 11860 11886 11895 11897 11921 11925 11940 11942 11943 11963
## [177] 11973 11981 12001 12005 12037 12044 12097 12124 12195 12206 12222
## [188] 12228 12368 12380 12382 12386 12426 12428 12430 12436 12493 12536
## [199] 12538 12546 12552 12564 12615 12629 12643 12650 12693 12730 12735
## [210] 12800 12899 12908 12942 13088 13090 13207 13431 13478 13499 13772
## [221] 13781 13958 14203 14204 14206 14477 14590 14611 14709 14786 14865
## [232] 14871 14880 14885 14895 14904 14923 14935 14939 14941 14945 14949
## [243] 14952 14966 14972 14988 15035 15038 15041 15054 15060 15087 15089
## [254] 15091 15265 15331 15721 15731 15805 16755 17166 17506 17689 17743
## integer(0)
## integer(0)
##  [1]   219   585   628  1101  1567  1678  1680  1684  1688  1690  8223
## [12]  8238  8239  8246  8247  8380  8844  8901  8992  9102  9134  9147
## [23]  9150  9231  9275  9318  9320  9334  9391  9497  9893  9999 10041
## [34] 10072 10074 10121 10138 10194 10199 10204 10245 10344 10440 10441
## [45] 10636 10638 10846 10902 10906 10910 10938 10948 10954 11062 11064
## [56] 11478 11479 11573 11929 12045 12054 12096 12136 12180 12201 12245
## [67] 12298 12351 12396 12688 12691 12718 12874 12883 13205 13262 13325
## [78] 14241 15057 15312 15414 15762 15814 16893 17121 17302 17524 17626
## [89] 17698
##   [1]     5    72   196   302   307   319   626   628   762  1194  1287
##  [12]  1329  1334  1342  1393  1517  1536  1566  1643  1713  1737  1829
##  [23]  1833  1836  1876  1997  2015  2577  5136  5575  5715  6008  6138
##  [34]  7143  7477  7884  7894  7897  7899  8385  8579  8784  8836  8895
##  [45]  8943  9278  9298  9615  9622  9626  9682  9738  9761  9793  9823
##  [56] 10000 10014 10176 10210 10213 10235 10237 10375 10432 10484 10533
##  [67] 10808 10848 10995 11000 11032 11084 11117 11172 11237 11244 11254
##  [78] 11881 11883 11893 12063 12103 12287 12373 12377 12397 12401 12413
##  [89] 12439 12448 12464 12522 12624 12656 12829 12918 12964 12967 13016
## [100] 13066 13410 13433 13505 13787 13791 13858 13875 14058 14062 14119
## [111] 14233 14235 14315 15052 15204 15209 15210 15211 15212 15275 15305
## [122] 15306 15311 15314 15315 15316 15414 15420 15428 15443 15452 15972
## [133] 16462 16522 16528 16591 16742 16744 16790 16836 16867 16876 16877
## [144] 16890 16903 17072 17121 17135 17225 17312 17316 17354 17382 17392
## [155] 17479 17530 17585 17610 17643 17657 17796
##  [1]  1438  1666  1669  2516  2569  2692  3016  3368  5341  7109  7110
## [12]  7111  7112  7113  7215  7377  8513  9444 11624 12810 12864 13257
## [23] 13446 13707 15236 16764 16799 17612
## integer(0)
##  [1]    67   361  1264  2046  2051  2072  2118  2140  2151  2172  2173
## [12]  2176  3028  3032  3033  3468  3717  3795  3796  3799  4405  4409
## [23]  4412  4551  4552  4573  5231  5480  5485  5538  6284  7832  7833
## [34]  7908  7909  8112  8285  8295  8306  8685  8688  8769  8866  9508
## [45]  9530  9798  9987 10136 10475 10844 10855 10857 10948 11244 11303
## [56] 11701 11792 11933 11935 12259 12606 12787 12795 12809 12843 12851
## [67] 13007 13127 13367 13498 13624 13690 13908 13974 14064 14154 14482
## [78] 14528 14764 15325 16155 16561 16757 17292 17447 17629
## [1] 3237
## [1]  2329  7325 14622
## integer(0)
## [1] 15111
##  [1]   188   248   292  1236  1245  1542  1562  1600  1675  1863  2524
## [12]  2628  2630  2800  2840  3485  3776  3878  4184  4548  5193  5194
## [23]  5209  5580  5726  5887  5987  5988  6002  6184  6494  6933  7103
## [34]  7238  7277  7348  7378  8170  8171  8382  8748  8770  8783  8858
## [45]  8860  8870  8872  9249  9481  9594  9628  9792  9808  9927 10719
## [56] 10782 10976 11079 11249 11755 11924 12586 12838 13190 13251 13448
## [67] 13667 13700 14004 14533 14554 14584 14585 14627 14634 15224 15227
## [78] 15818 16333 16574 16833 17114 17372 17435 17617 17660
## integer(0)
## integer(0)
## [1] 748
## integer(0)
```

```
#sink()
```

```
# Have to do a lot of cleaning of the data before importing it as csv
#(to make all values in each cell separately inside one column):
# getting only unique values from all numbers, because one gene may
#overlap with other, we only want the unique  #row numbers:

new_numbers <- read.csv("C://Users//Natalia//Desktop//ITMO//SystemBiology//RNAseq_analysis//RNAseq_anal

unique_vals <- data.frame(as.integer(unique(unlist(new_numbers))))
colnames(unique_vals) <- c('row_number')

new_unique_vals <- na.omit(unique_vals)

pathways.final <- subset(fgseaEs.all, rownames(fgseaEs.all) %in% new_unique_vals$row_number)

View(pathways.final)
```

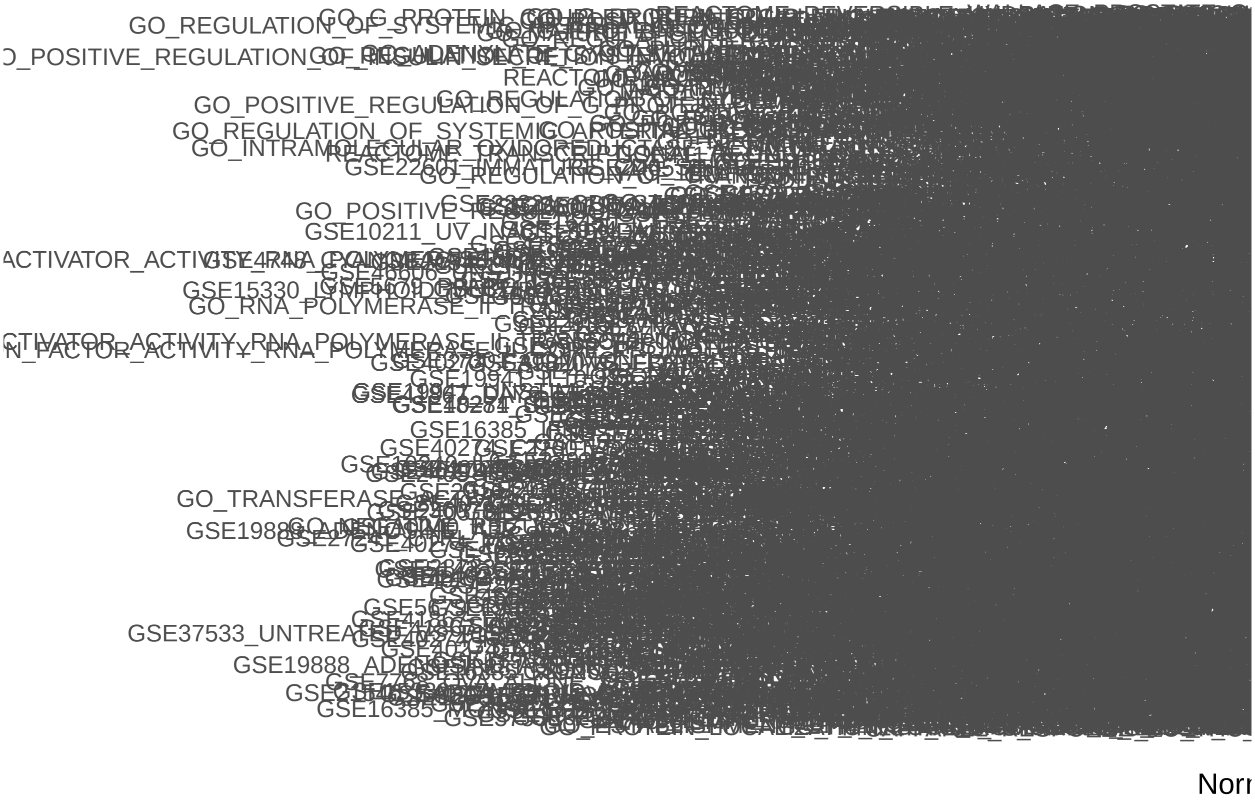
```
# Show the first few pathways, and within those, show only the first few genes:

pathways.final %>%
  head() %>%
  lapply(head)
```

```
## $pathway
## [1] "AAANWWTGC_UNKNOWN"
## [2] "AACATTC_MIR4093P"
## [3] "AACTGGA_MIR145"
## [4] "ACATTCC_MIR1_MIR206"
## [5] "ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_DN"
## [6] "ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_UP"
##
## $pval
## [1] 0.05194805 0.85714286 0.63436563 0.98001998 0.42857143 0.48451548
##
## $padj
## [1] 0.5433700 0.9885751 0.8581658 1.0000000 0.7223087 0.7578166
##
## $ES
## [1] 0.9991592 0.9778229 0.9870500 0.9770472 0.9907564 0.9871795
##
## $NES
## [1] 1.0119701 0.9903889 0.9976878 0.9868581 1.0042264 1.0043827
```

```
##
## $nMoreExtreme
## [1]  51 857 634 980 428 484
##
## $size
## [1] 107 106 158 229  95  66
##
## $leadingEdge
## $leadingEdge[[1]]
## [1] "EBF1" "CD14"
##
## $leadingEdge[[2]]
## [1] "KCNMA1"   "NR4A2"    "MYLK"     "AKAP12"   "HS3ST3B1"
##
## $leadingEdge[[3]]
## [1] "PHLDB2" "MAF"    "AKAP12"
##
## $leadingEdge[[4]]
## [1] "GAS2L1"   "NR4A2"    "MYLK"     "AKAP12"   "PSD3"     "HS3ST3B1"
##
## $leadingEdge[[5]]
## [1] "ALAS2"    "ID1"      "SELENBP1"
##
## $leadingEdge[[6]]
## [1] "LPL"    "GNPDA1"
```

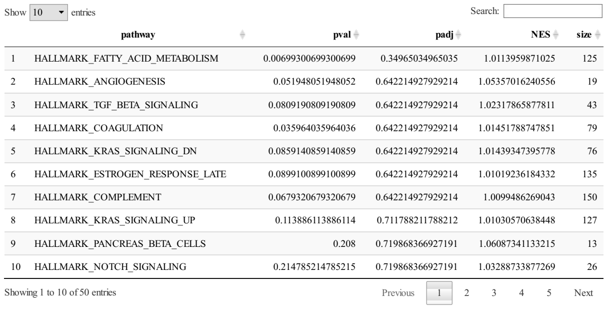
```r
final <- data.frame(pathways.final)
# running the fgsea algorithm on final pathways
# Let's look at the plot
```

```r
# ggplot for final pathways:

library(ggplot2)

 #pdf('final_pathways.pdf', width=15, height = 120 )
ggplot(final, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=pval<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Selected genes from the study") +
  theme_minimal()
```



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

```
# install.packages('DT')
library(DT)

# Show in a table for all pathways:

fgseaEsTidy %>%
  dplyr::select(-leadingEdge, -ES, -nMoreExtreme) %>%
  arrange(padj) %>%
  DT::datatable()
```



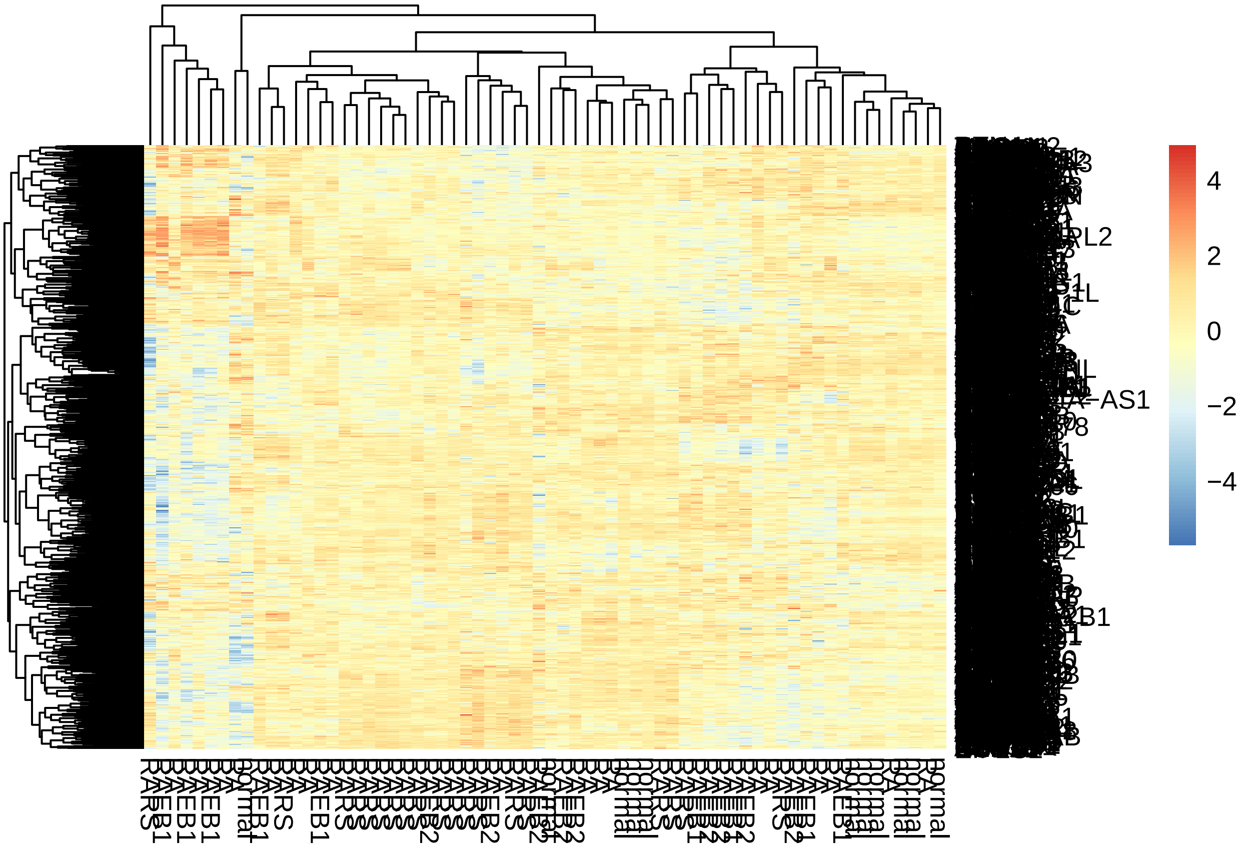
Show 10 entries Search:

| | pathway | pval | padj | NES | size |
|---|---|---|---|---|---|
| 1 | HALLMARK_FATTY_ACID_METABOLISM | 0.00699300699300699 | 0.34965034965035 | 1.0113959871025 | 125 |
| 2 | HALLMARK_ANGIOGENESIS | 0.051948051948052 | 0.642214927929214 | 1.05357016240556 | 19 |
| 3 | HALLMARK_TGF_BETA_SIGNALING | 0.0809190809190809 | 0.642214927929214 | 1.02317865877811 | 43 |
| 4 | HALLMARK_COAGULATION | 0.035964035964036 | 0.642214927929214 | 1.01451788747851 | 79 |
| 5 | HALLMARK_KRAS_SIGNALING_DN | 0.0859140859140859 | 0.642214927929214 | 1.01439347395778 | 76 |
| 6 | HALLMARK_ESTROGEN_RESPONSE_LATE | 0.0899100899100899 | 0.642214927929214 | 1.01019236184332 | 135 |
| 7 | HALLMARK_COMPLEMENT | 0.0679320679320679 | 0.642214927929214 | 1.0099486269043 | 150 |
| 8 | HALLMARK_KRAS_SIGNALING_UP | 0.113886113886114 | 0.711788211788212 | 1.01030570638448 | 127 |
| 9 | HALLMARK_PANCREAS_BETA_CELLS | 0.208 | 0.719868366927191 | 1.06087341133215 | 13 |
| 10 | HALLMARK_NOTCH_SIGNALING | 0.214785214785215 | 0.719868366927191 | 1.03288733877269 | 26 |

Showing 1 to 10 of 50 entries

```
# heatmap
library(pheatmap)
```

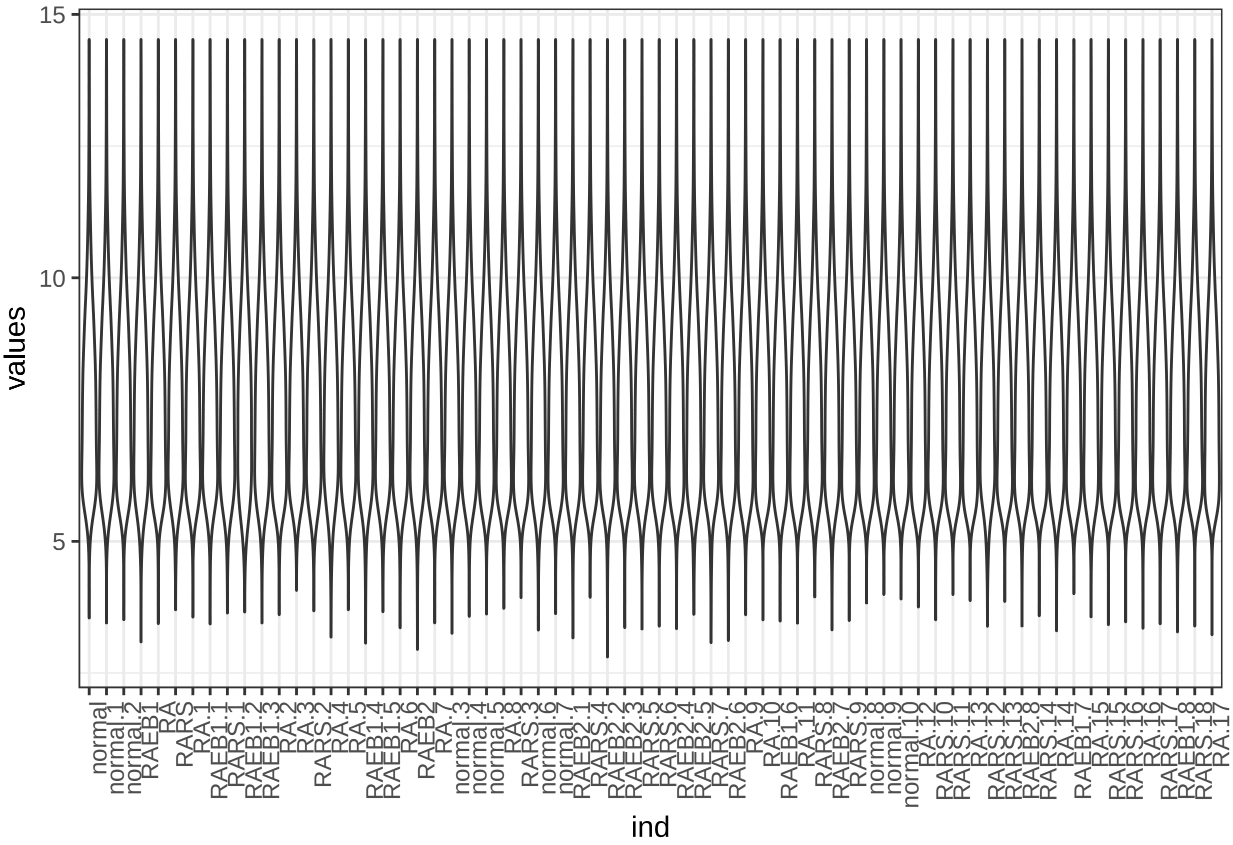
```
#scale rows
xt <-t(as.matrix(es.qnorm.top12K)) # this is a matrix of normalised 12k genes

# To get a heatmap of 1000 genes:

xts <-scale(xt)
xtst <-t(xts)
xtst <- na.omit(xtst)
colnames(xtst) <- es$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()
```

```
# To make a boxplot of the data:

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

```
#pdf('box_dataset.pdf', height = 5, width = 30)


gt <- t(xt) # taking xt from the heatmap and transposing it
colnames(gt)<- es$condition # now giving it labels from condition
ggplot(stack(data.frame(gt)), aes(x = ind, y = values)) +
  geom_violin() + theme_bw() + theme(axis.text.x = element_text(angle=90, hjust=1))
```


15
10
5
values
normal
normal.1
normal.2
RAEB1
RA
RARS
RA.1
RAEB1.1
RARS.1
RAEB1.2
RAEB1.3
RA.2
RA.3
RARS.2
RA.4
RA.5
RAEB1.4
RAEB1.5
RA.6
RAEB2
RA.7
normal.3
normal.4
normal.5
RA.8
RARS.3
normal.6
normal.7
RAEB2.1
RARS.4
RAEB2.2
RAEB2.3
RARS.5
RARS.6
RAEB2.4
RAEB2.5
RARS.7
RAEB2.6
RA.9
RA.10
RAEB1.6
RA.11
RARS.8
RAEB2.7
RARS.9
normal.8
normal.9
normal.10
RA.12
RARS.10
RARS.11
RA.13
RARS.12
RARS.13
RAEB2.8
RARS.14
RA.14
RAEB1.7
RA.15
RARS.15
RARS.16
RA.16
RARS.17
RAEB1.8
RARS.18
RA.17
ind

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