

# Differential Expression with Bioconductor

Code ▼

**Introduction** This tutorial and accompanying data was provided by Randy Williams. It focuses on how to perform a differential expression analysis on microarray gene expression data with Bioconductor. It will use gene expression data from a study on tobacco smoke-related transcriptome alterations in the placenta. Differential expression analysis will be performed to find out which genes in the human placenta are differentially expressed by fetal sex. The dataset is normalized with the annotation and phenotype data provided.

Hide

```
#Load Bioconductor
library(BiocManager)
```

```
Warning: package 'BiocManager' was built under R version 4.1.3
Bioconductor version 3.14 (BiocManager 1.30.17), R 4.1.2 (2021-11-01)
Bioconductor version '3.14' is out-of-date; the current release version
'3.15' is available with R version '4.2'; see
https://bioconductor.org/install
```

```
Attaching package: 'BiocManager'
```

```
The following object is masked from 'package:msa':
```

```
version
```

Hide

```
#Required BioConductor Packages
BiocManager::install(c('EnhancedVolcano', 'Biobase', 'limma', 'geneplotter', 'enrichplot'))
```

```
'getOption("repos")' replaces Bioconductor standard repositories, see  
'?repositories' for details
```

```
replacement repositories:
```

```
CRAN: https://cran.rstudio.com
```

```
Bioconductor version 3.14 (BiocManager 1.30.17), R 4.1.2 (2021-11-01)
```

```
Warning: package(s) not installed when version(s) same as current; use `force =
```

```
TRUE` to re-install: 'Biobase' 'limma'
```

```
Installing package(s) 'EnhancedVolcano', 'geneplotter', 'enrichplot'
```

```
also installing the dependencies 'beeswarm', 'vipor', 'textshaping', 'gridGraphics', 'fastmatch',  
'tweenr', 'polyclip', 'RcppEigen', 'proj4', 'ash', 'maps', 'plotly', 'Cairo', 'ggbeeswarm',  
'ragg', 'ggfun', 'ggplotify', 'patchwork', 'DO.db', 'fgsea', 'ggforce', 'tidygraph', 'graphlayouts',  
'GO.db', 'ape', 'tidytrees', 'treeio', 'ggrepel', 'ggalt', 'ggtra', 'aplot', 'DOSE', 'gggraph',  
'scatterpie', 'shadowtext', 'GOSemSim', 'ggtree', 'yulab.utils'
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/beeswarm_0.4.0.zip'
```

```
Content type 'application/zip' length 93788 bytes (91 KB)
```

```
downloaded 91 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/vipor_0.4.5.zip'
```

```
Content type 'application/zip' length 4409783 bytes (4.2 MB)
```

```
downloaded 4.2 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/textshaping_0.3.6.zip'
```

```
Content type 'application/zip' length 2168141 bytes (2.1 MB)
```

```
downloaded 2.1 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/gridGraphics_0.5-1.zip'
```

```
Content type 'application/zip' length 270628 bytes (264 KB)
```

```
downloaded 264 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/fastmatch_1.1-3.zip'
```

```
Content type 'application/zip' length 53639 bytes (52 KB)
```

```
downloaded 52 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tweenr_1.0.2.zip'
```

```
Content type 'application/zip' length 1045610 bytes (1021 KB)
```

```
downloaded 1021 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/polyclip_1.10-0.zip'
```

```
Content type 'application/zip' length 716846 bytes (700 KB)
```

```
downloaded 700 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/RcppEigen_0.3.3.9.2.zip'
```

```
Content type 'application/zip' length 2714165 bytes (2.6 MB)
```

```
downloaded 2.6 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/proj4_1.0-11.zip'
```

```
Content type 'application/zip' length 317787 bytes (310 KB)
```

```
downloaded 310 KB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ash_1.0-15.zip'
Content type 'application/zip' length 41552 bytes (40 KB)
downloaded 40 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/maps_3.4.0.zip'
Content type 'application/zip' length 3119067 bytes (3.0 MB)
downloaded 3.0 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/plotly_4.10.0.zip'
Content type 'application/zip' length 3176050 bytes (3.0 MB)
downloaded 3.0 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/Cairo_1.5-15.zip'
Content type 'application/zip' length 1041917 bytes (1017 KB)
downloaded 1017 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggbeeswarm_0.6.0.zip'
Content type 'application/zip' length 1528493 bytes (1.5 MB)
downloaded 1.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ragg_1.2.2.zip'
Content type 'application/zip' length 2388126 bytes (2.3 MB)
downloaded 2.3 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggfun_0.0.6.zip'
Content type 'application/zip' length 191419 bytes (186 KB)
downloaded 186 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggplotify_0.1.0.zip'
Content type 'application/zip' length 140530 bytes (137 KB)
downloaded 137 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/patchwork_1.1.1.zip'
Content type 'application/zip' length 3241449 bytes (3.1 MB)
downloaded 3.1 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/fgsea_1.20.0.zip'
Content type 'application/zip' length 2149842 bytes (2.1 MB)
downloaded 2.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggforce_0.3.3.zip'
Content type 'application/zip' length 2690364 bytes (2.6 MB)
downloaded 2.6 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tidygraph_1.2.1.zip'
Content type 'application/zip' length 660287 bytes (644 KB)
downloaded 644 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/graphlayouts_0.8.0.zip'
Content type 'application/zip' length 3404496 bytes (3.2 MB)
downloaded 3.2 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ape_5.6-2.zip'
Content type 'application/zip' length 3993611 bytes (3.8 MB)
downloaded 3.8 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tidytrees_0.3.9.zip'
Content type 'application/zip' length 249997 bytes (244 KB)
downloaded 244 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/treeio_1.18.1.zip'
Content type 'application/zip' length 952044 bytes (929 KB)
downloaded 929 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggrepel_0.9.1.zip'
Content type 'application/zip' length 1119267 bytes (1.1 MB)
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggalt_0.4.0.zip'
Content type 'application/zip' length 2364476 bytes (2.3 MB)
downloaded 2.3 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggrastr_1.0.1.zip'
Content type 'application/zip' length 2592102 bytes (2.5 MB)
downloaded 2.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/aplot_0.1.3.zip'
Content type 'application/zip' length 53279 bytes (52 KB)
downloaded 52 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/DOSE3_3.20.1.zip'
Content type 'application/zip' length 6754753 bytes (6.4 MB)
downloaded 6.4 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggraph_2.0.5.zip'
Content type 'application/zip' length 4762175 bytes (4.5 MB)
downloaded 4.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/scatterpie_0.1.7.zip'
Content type 'application/zip' length 411439 bytes (401 KB)
downloaded 401 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/shadowtext_0.1.2.zip'
Content type 'application/zip' length 230432 bytes (225 KB)
downloaded 225 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/GOSemSim_2.20.0.zip'
Content type 'application/zip' length 1634499 bytes (1.6 MB)
downloaded 1.6 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/ggtree_3.2.1.zip'
```

```
p'  
Content type 'application/zip' length 1009609 bytes (985 KB)  
downloaded 985 KB  
  
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/yulab.utils_0.0.4.zip'  
Content type 'application/zip' length 28369 bytes (27 KB)  
downloaded 27 KB  
  
trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/EnhancedVolcano_  
1.12.0.zip'  
Content type 'application/zip' length 2235453 bytes (2.1 MB)  
downloaded 2.1 MB  
  
trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/geneplotter_1.7  
2.0.zip'  
Content type 'application/zip' length 1608744 bytes (1.5 MB)  
downloaded 1.5 MB  
  
trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/enrichplot_1.14.  
2.zip'  
Content type 'application/zip' length 309312 bytes (302 KB)  
downloaded 302 KB
```

```
package 'beeswarm' successfully unpacked and MD5 sums checked
package 'vipor' successfully unpacked and MD5 sums checked
package 'textshaping' successfully unpacked and MD5 sums checked
package 'gridGraphics' successfully unpacked and MD5 sums checked
package 'fastmatch' successfully unpacked and MD5 sums checked
package 'tweenr' successfully unpacked and MD5 sums checked
package 'polyclip' successfully unpacked and MD5 sums checked
package 'RcppEigen' successfully unpacked and MD5 sums checked
package 'proj4' successfully unpacked and MD5 sums checked
package 'ash' successfully unpacked and MD5 sums checked
package 'maps' successfully unpacked and MD5 sums checked
package 'plotly' successfully unpacked and MD5 sums checked
package 'Cairo' successfully unpacked and MD5 sums checked
package 'ggbeeswarm' successfully unpacked and MD5 sums checked
package 'ragg' successfully unpacked and MD5 sums checked
package 'ggfun' successfully unpacked and MD5 sums checked
package 'ggplotify' successfully unpacked and MD5 sums checked
package 'patchwork' successfully unpacked and MD5 sums checked
package 'fgsea' successfully unpacked and MD5 sums checked
package 'ggforce' successfully unpacked and MD5 sums checked
package 'tidygraph' successfully unpacked and MD5 sums checked
package 'graphlayouts' successfully unpacked and MD5 sums checked
package 'ape' successfully unpacked and MD5 sums checked
package 'tidytree' successfully unpacked and MD5 sums checked
package 'treeio' successfully unpacked and MD5 sums checked
package 'ggrepel' successfully unpacked and MD5 sums checked
package 'ggalt' successfully unpacked and MD5 sums checked
package 'ggrastr' successfully unpacked and MD5 sums checked
package 'aplot' successfully unpacked and MD5 sums checked
package 'DOSE' successfully unpacked and MD5 sums checked
package 'ggraph' successfully unpacked and MD5 sums checked
package 'scatterpie' successfully unpacked and MD5 sums checked
package 'shadowtext' successfully unpacked and MD5 sums checked
package 'GOSemSim' successfully unpacked and MD5 sums checked
package 'ggtree' successfully unpacked and MD5 sums checked
package 'yulab.utils' successfully unpacked and MD5 sums checked
package 'EnhancedVolcano' successfully unpacked and MD5 sums checked
package 'geneplotter' successfully unpacked and MD5 sums checked
package 'enrichplot' successfully unpacked and MD5 sums checked
```

The downloaded binary packages are in

C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages

```
installing the source packages 'D0.db', 'G0.db'
```

```
trying URL 'https://bioconductor.org/packages/3.14/data/annotation/src/contrib/D0.db_2.9.tar.gz'  
Content type 'application/x-gzip' length 1769978 bytes (1.7 MB)  
downloaded 1.7 MB
```

```
trying URL 'https://bioconductor.org/packages/3.14/data/annotation/src/contrib/G0.db_3.14.0.tar.  
gz'  
Content type 'application/x-gzip' length 30254515 bytes (28.9 MB)  
downloaded 28.9 MB
```

```

* installing *source* package 'DO.db' ...
** using staged installation
** R
** inst
** byte-compile and prepare package for lazy loading
Warning message:
package 'S4Vectors' was built under R version 4.1.3
** help
*** installing help indices
    converting help for package 'DO.db'
      finding HTML links ... done
      DOANCESTOR                      html
      DOBASE                          html
      DOCHILDREN                      html
      DOMAPCOUNTS                   html
      DOOBSOLETE                      html
      DOOFFSPRING                    html
      DOPARENTS                      html
      DOSYNONYM                      html
      DOTERM                         html
      DOTerms-class                   html
      DOTermsAnnDbBimap               html
      DO_dbconn                       html
** building package indices
** testing if installed package can be loaded from temporary location
*** arch - i386
Warning: package 'S4Vectors' was built under R version 4.1.3
*** arch - x64
Warning: package 'S4Vectors' was built under R version 4.1.3
** testing if installed package can be loaded from final location
*** arch - i386
Warning: package 'S4Vectors' was built under R version 4.1.3
*** arch - x64
Warning: package 'S4Vectors' was built under R version 4.1.3
** testing if installed package keeps a record of temporary installation path
* DONE (DO.db)
* installing *source* package 'GO.db' ...
** using staged installation
** R
** inst
** byte-compile and prepare package for lazy loading
Warning message:
package 'S4Vectors' was built under R version 4.1.3
** help
*** installing help indices
    converting help for package 'GO.db'
      finding HTML links ... done
      GOBASE                          html
      GOBPANCESTOR                    html
      GOBPCHILDREN                     html
      GOBPOFFSPRING                   html
      GOBPPARENTS                     html

```



GOCCANCESTOR	html
GOCCCHILDREN	html
GOCCOFFSPRING	html
GOCCPARENTS	html
GOMAPCOUNTS	html
GOMFANCESTOR	html
GOMFCHILDREN	html
GOMFOFFSPRING	html
GOMFPARENTS	html
GOOBSOLETE	html
GOSYNONYM	html
GOTERM	html
GO_dbconn	html

```

** building package indices
** testing if installed package can be loaded from temporary location
*** arch - i386
Warning: package 'S4Vectors' was built under R version 4.1.3
*** arch - x64
Warning: package 'S4Vectors' was built under R version 4.1.3
** testing if installed package can be loaded from final location
*** arch - i386
Warning: package 'S4Vectors' was built under R version 4.1.3
*** arch - x64
Warning: package 'S4Vectors' was built under R version 4.1.3
** testing if installed package keeps a record of temporary installation path
* DONE (GO.db)

```

The downloaded source packages are in  
 'C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages'  
 Installation paths not writeable, unable to update packages  
 path: C:/Program Files/R/R-4.1.2/library  
 packages:  
 class, cluster, foreign, MASS, Matrix, mgcv, nlme, nnet, rpart,  
 spatial, survival  
 Old packages: 'cli', 'dplyr', 'ggplot2', 'httr', 'MSnbase', 'RSQLite',  
 'tibble'

Update all/some/none? [a/s/n]:

Hide

a

There are binary versions available but the source versions  
 are later:

	<b>binary</b> <chr>	<b>source</b> <chr>	<b>needs_compilation</b> <lg >
ggplot2	3.3.5	3.3.6	FALSE
httr	1.4.2	1.4.3	FALSE
MSnbase	2.20.1	2.20.4	TRUE
tibble	3.1.6	3.1.7	TRUE
4 rows			

Binaries will be installed

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/cli\_3.3.0.zip'  
Content type 'application/zip' length 1260519 bytes (1.2 MB)  
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/dplyr\_1.0.9.zip'  
Content type 'application/zip' length 1392479 bytes (1.3 MB)  
downloaded 1.3 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/MSnbase\_2.20.1.zip'  
Content type 'application/zip' length 7739582 bytes (7.4 MB)  
downloaded 7.4 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/RSQLite\_2.2.13.zip'  
Content type 'application/zip' length 2566673 bytes (2.4 MB)  
downloaded 2.4 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tibble\_3.1.6.zip'  
Content type 'application/zip' length 872156 bytes (851 KB)  
downloaded 851 KB

package 'cli' successfully unpacked and MD5 sums checked

Warning: cannot remove prior installation of package 'cli'  
Warning in file.copy(savedcopy, lib, recursive = TRUE) :  
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86\_64-w64-mingw32\00LOCK\cli\libs\x64\cli.dll to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86\_64-w64-mingw32\cli\libs\x64\cli.dll: Permission denied  
Warning: restored 'cli'

package 'dplyr' successfully unpacked and MD5 sums checked

```
Warning: cannot remove prior installation of package 'dplyr'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\dplyr\libs\x64\dplyr.dll to
C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\dplyr\libs\x64\dplyr.dll: Permission denied
Warning: restored 'dplyr'
```

```
package 'MSnbase' successfully unpacked and MD5 sums checked
package 'RSQLite' successfully unpacked and MD5 sums checked
```

```
Warning: cannot remove prior installation of package 'RSQLite'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\RSQLite\libs\x64\RSQLite.dll
to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\RSQLite\libs\x64\RSQLite.dll: Permission denied
Warning: restored 'RSQLite'
```

```
package 'tibble' successfully unpacked and MD5 sums checked
```

```
Warning: cannot remove prior installation of package 'tibble'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\tibble\libs\x64\tibble.dll to
C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\tibble\libs\x64\tibble.dll: Permission denied
Warning: restored 'tibble'
```

```
The downloaded binary packages are in
  C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded_packages
```

```
installing the source packages 'ggplot2', 'httr'

trying URL 'https://cran.rstudio.com/src/contrib/ggplot2_3.3.6.tar.gz'
Content type 'application/x-gzip' length 3061989 bytes (2.9 MB)
downloaded 2.9 MB

trying URL 'https://cran.rstudio.com/src/contrib/httr_1.4.3.tar.gz'
Content type 'application/x-gzip' length 161122 bytes (157 KB)
downloaded 157 KB
```

```

* installing *source* package 'ggplot2' ...
** package 'ggplot2' successfully unpacked and MD5 sums checked
** using staged installation
** R
** data
*** moving datasets to lazyload DB
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
    converting help for package 'ggplot2'
      finding HTML links ... done
      absoluteGrob                                html
      add_theme                                    html
      aes                                           html
      aes_                                          html
      aes_all                                       html
      aes_auto                                      html
      aes_colour_fill_alpha                        html
      aes_eval                                      html
      aes_group_order                             html
      aes_linetype_size_shape                     html
      aes_position                                html
      annotate                                      html
      annotation_custom                           html
      annotation_logticks                         html
      annotation_map                               html
      annotation_raster                           html
      as.list.ggproto                             html
      as_labeller                                  html
      autolayer                                    html
      autoplot                                     html
      benchplot                                    html
      bidirection                                  html
      binned_scale                                html
      borders                                      html
      calc_element                                html
      combine_vars                                 html
      continuous_scale                            html
      coord_cartesian                             html
      coord_fixed                                  html
      coord_flip                                   html
      coord_map                                    html
      coord_munch                                  html
      coord_polar                                  html
      coord_trans                                  html
      cut_interval                                html
      diamonds                                      html
      discrete_scale                              html
      draw_key                                     html
      economics                                    html
      element                                      html

```

element_grob	html
element_render	html
expand_limits	html
expansion	html
facet_grid	html
facet_null	html
facet_wrap	html
faithfuld	html
find_panel	html
fortify-multcomp	html
fortify	html
fortify.lm	html
fortify.map	html
fortify.sp	html
geom_abline	html
geom_bar	html
geom_bin_2d	html
geom_blank	html
geom_boxplot	html
geom_contour	html
geom_count	html
geom_density	html
geom_density_2d	html
geom_dotplot	html
geom_errorbarh	html
geom_function	html
geom_hex	html
geom_histogram	html
geom_jitter	html
geom_linerange	html
geom_map	html
geom_path	html
geom_point	html
geom_polygon	html
geom_qq	html
geom_quantile	html
geom_ribbon	html
geom_rug	html
geom_segment	html
geom_smooth	html
geom_spoke	html
geom_text	html
geom_tile	html
geom_violin	html
get_alt_text	html
gg-add	html
gg_dep	html
ggplot	html
ggplot2-ggproto	html

REDIRECT:topic Previous alias or file overwritten by alias: C:/Users/samen/Desktop/Bioinformatics Projects/Bioconductor tools for Mass Spectrometry/Bioconductor/renv/library/R-4.1/x86\_64-w64-mingw32/00LOCK-ggplot2/00new/ggplot2/help/Stat.html

finding level-2 HTML links ... done

ggplot2-package	html
ggplotGrob	html
ggplot_add	html
ggplot_build	html
ggplot_gtable	html
ggproto	html
ggsave	html
ggsf	html
ggtheme	html
graphical-units	html
guide-exts	html
guide_axis	html
guide_bins	html
guide_colourbar	html
guide_coloursteps	html
guide_legend	html
guide_none	html
guides	html
hmisc	html
is.Coord	html
is.facet	html
is.ggplot	html
is.rel	html
is.theme	html
label_bquote	html
labeller	html
labellers	html
labs	html
last_plot	html
layer	html
layer_sf	html
limits	html
lims	html
luv_colours	html
map_data	html
max_height	html
mean_se	html
merge_element	html
midwest	html
mpg	html
msleep	html
position_dodge	html
position_identity	html
position_jitter	html
position_jitterdodge	html
position_nudge	html
position_stack	html
presidential	html
print.ggplot	html
print.ggproto	html

qplot	html
reexports	html
register_theme_elements	html
remove_missing	html
render_axes	html
render_strips	html
resolution	html
scale_alpha	html
scale_binned	html
scale_brewer	html
scale_colour_continuous	html
scale_colour_discrete	html
scale_continuous	html
scale_date	html
scale_discrete	html
scale_gradient	html
scale_grey	html
scale_hue	html
scale_identity	html
scale_linetype	html
scale_manual	html
scale_shape	html
scale_size	html
scale_steps	html
scale_type	html
scale_viridis	html
seals	html
sec_axis	html
set_last_plot	html
sf_transform_xy	html
should_stop	html
standardise_aes_names	html
stat_ecdf	html
stat_ellipse	html
stat_identity	html
stat_sf_coordinates	html
stat_summary	html
stat_summary_2d	html
stat_unique	html
summarise_plot	html
summary.ggplot	html
theme	html
theme_get	html
tidyeval	html
transform_position	html
translate_qplot_ggplot	html
translate_qplot_lattice	html
txhousing	html
update_defaults	html
update_labels	html
vars	html
waiver	html

```

    wrap_dims                                html
    zeroGrob                                 html
*** copying figures
** building package indices
** installing vignettes
** testing if installed package can be loaded from temporary location
*** arch - i386
*** arch - x64
** testing if installed package can be loaded from final location
*** arch - i386
*** arch - x64
** testing if installed package keeps a record of temporary installation path
* DONE (ggplot2)
* installing *source* package 'httr' ...
** package 'httr' successfully unpacked and MD5 sums checked
** using staged installation
** R
** demo
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
    converting help for package 'httr'
      finding HTML links ... done
    BROWSE                                html
    DELETE                                html
    GET                                   html
    HEAD                                  html
    PATCH                                html
    POST                                  html
    PUT                                   html
    RETRY                                html
    Token-class                           html
    VERB                                  html
    add_headers                           html
    authenticate                           html
    cache_info                            html
    config                                html
    content                                html
    content_type                           html
    cookies                                html
    get_callback                           html
    guess_media                            html
    handle                                 html
    handle_pool                            html
    has_content                            html
    headers                                html
    hmac_sha1                             html
    http_condition                         html
    http_error                             html
    http_status                            html
    http_type                             html

```



httr-package	html
httr_dr	html
httr_options	html
init_oauth1.0	html
init_oauth2.0	html
insensitive	html
jwt_signature	html
modify_url	html
oauth1.0_token	html
oauth2.0_token	html
oauth_app	html
oauth_callback	html
oauth_endpoint	html
oauth_endpoints	html
oauth_exchanger	html
oauth_listener	html
oauth_service_token	html
oauth_signature	html
parse_http_date	html
parse_media	html
parse_url	html
progress	html
response	html
revoke_all	html
set_config	html
set_cookies	html
sha1_hash	html
sign_oauth	html
status_code	html
stop_for_status	html
timeout	html
upload_file	html
use_proxy	html
user_agent	html
verbose	html
with_config	html
write_disk	html
write_function	html
write_stream	html

```
** building package indices
** installing vignettes
** testing if installed package can be loaded from temporary location
*** arch - i386
*** arch - x64
** testing if installed package can be loaded from final location
*** arch - i386
*** arch - x64
** testing if installed package keeps a record of temporary installation path
* DONE (httr)
```

The downloaded source packages are in  
 'C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages'

Hide

```
BiocManager::install('clusterProfiler')
```

'getOption("repos")' replaces Bioconductor standard repositories, see  
 '?repositories' for details

replacement repositories:

CRAN: <https://cran.rstudio.com>

Bioconductor version 3.14 (BiocManager 1.30.17), R 4.1.2 (2021-11-01)

Installing package(s) 'clusterProfiler'

also installing the dependency 'downloader'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/downloader\_0.4.zip'

Content type 'application/zip' length 24996 bytes (24 KB)

downloaded 24 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/clusterProfiler\_4.2.2.zip'

Content type 'application/zip' length 777312 bytes (759 KB)

downloaded 759 KB

package 'downloader' successfully unpacked and MD5 sums checked

package 'clusterProfiler' successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages

Installation paths not writeable, unable to update packages

path: C:/Program Files/R/R-4.1.2/library

packages:

class, cluster, foreign, MASS, Matrix, mgcv, nlme, nnet, rpart,  
 spatial, survival

Old packages: 'cli', 'dplyr', 'MSnbase', 'RSQLite', 'tibble'

Update all/some/none? [a/s/n]:

Hide

a

There are binary versions available but the source versions are later:

	binary <chr>	source <chr>	needs_compilation <lgl>
MSnbase	2.20.1	2.20.4	TRUE
tibble	3.1.6	3.1.7	TRUE
2 rows			

Binaries will be installed

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/cli_3.3.0.zip'
```

```
Content type 'application/zip' length 1260519 bytes (1.2 MB)
```

```
downloaded 1.2 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/dplyr_1.0.9.zip'
```

```
Content type 'application/zip' length 1392479 bytes (1.3 MB)
```

```
downloaded 1.3 MB
```

```
trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/MSnbase_2.20.1.zip'
```

```
Content type 'application/zip' length 7739582 bytes (7.4 MB)
```

```
downloaded 7.4 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/RSQLite_2.2.13.zip'
```

```
Content type 'application/zip' length 2566673 bytes (2.4 MB)
```

```
downloaded 2.4 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tibble_3.1.6.zip'
```

```
Content type 'application/zip' length 872156 bytes (851 KB)
```

```
downloaded 851 KB
```

package 'cli' successfully unpacked and MD5 sums checked

Warning: cannot remove prior installation of package 'cli'

Warning in file.copy(savedcopy, lib, recursive = TRUE) :

```
problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\cli\libs\x64\cli.dll to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\cli\libs\x64\cli.dll: Permission denied
```

Warning: restored 'cli'

package 'dplyr' successfully unpacked and MD5 sums checked

```
Warning: cannot remove prior installation of package 'dplyr'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\dplyr\libs\x64\dplyr.dll to
  C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\dplyr\libs\x64\dplyr.dll: Permission denied
Warning: restored 'dplyr'
```

```
package 'MSnbase' successfully unpacked and MD5 sums checked
package 'RSQLite' successfully unpacked and MD5 sums checked
```

```
Warning: cannot remove prior installation of package 'RSQLite'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\RSQLite\libs\x64\RSQLite.dll
  to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\RSQLite\libs\x64\RSQLite.dll: Permission denied
Warning: restored 'RSQLite'
```

```
package 'tibble' successfully unpacked and MD5 sums checked
```

```
Warning: cannot remove prior installation of package 'tibble'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\tibble\libs\x64\tibble.dll to
  C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\tibble\libs\x64\tibble.dll: Permission denied
Warning: restored 'tibble'
```

```
The downloaded binary packages are in
  C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded_packages
```

[Hide](#)

```
BiocManager::install('pheatmap')
```

'getOption("repos")' replaces Bioconductor standard repositories, see  
'?repositories' for details

replacement repositories:

CRAN: <https://cran.rstudio.com>

Bioconductor version 3.14 (BiocManager 1.30.17), R 4.1.2 (2021-11-01)

Installing package(s) 'pheatmap'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/pheatmap\_1.0.12.zip'

Content type 'application/zip' length 78594 bytes (76 KB)

downloaded 76 KB

package 'pheatmap' successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages

Installation paths not writeable, unable to update packages

path: C:/Program Files/R/R-4.1.2/library

packages:

class, cluster, foreign, MASS, Matrix, mgcv, nlme, nnet, rpart,  
spatial, survival

Old packages: 'cli', 'dplyr', 'MSnbase', 'RSQLite', 'tibble'

Update all/some/none? [a/s/n]:

Hide

a

Warning: package 'dplyr' is in use and will not be installed

There are binary versions available but the source versions  
are later:

	binary <chr>	source <chr>	needs_compilation <lgl>
MSnbase	2.20.1	2.20.4	TRUE
tibble	3.1.6	3.1.7	TRUE

2 rows

Binaries will be installed

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/cli_3.3.0.zip'
Content type 'application/zip' length 1260519 bytes (1.2 MB)
downloaded 1.2 MB
```

```
trying URL 'https://bioconductor.org/packages/3.14/bioc/bin/windows/contrib/4.1/MSnbase_2.20.1.zip'
Content type 'application/zip' length 7739582 bytes (7.4 MB)
downloaded 7.4 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/RSQLite_2.2.13.zip'
Content type 'application/zip' length 2566673 bytes (2.4 MB)
downloaded 2.4 MB
```

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/tibble_3.1.6.zip'
Content type 'application/zip' length 872156 bytes (851 KB)
downloaded 851 KB
```

package 'cli' successfully unpacked and MD5 sums checked

```
Warning: cannot remove prior installation of package 'cli'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\cli\libs\x64\cli.dll to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\cli\libs\x64\cli.dll: Permission denied
Warning: restored 'cli'
```

package 'MSnbase' successfully unpacked and MD5 sums checked  
package 'RSQLite' successfully unpacked and MD5 sums checked

```
Warning: cannot remove prior installation of package 'RSQLite'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\RSQLite\libs\x64\RSQLite.dll to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\RSQLite\libs\x64\RSQLite.dll: Permission denied
Warning: restored 'RSQLite'
```

package 'tibble' successfully unpacked and MD5 sums checked

```
Warning: cannot remove prior installation of package 'tibble'
Warning in file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\00LOCK\tibble\libs\x64\tibble.dll to C:\Users\samen\Desktop\Bioinformatics Projects\Bioconductor tools for Mass Spectrometry\Bioconductor\renv\library\R-4.1\x86_64-w64-mingw32\tibble\libs\x64\tibble.dll: Permission denied
Warning: restored 'tibble'
```

The downloaded binary packages are in  
C:\Users\samen\AppData\Local\Temp\Rtmp6vPix9\downloaded\_packages

[Hide](#)

```
#loading all Cran and Bioconductor Packages
library(Biobase)
library(limma)
library(RColorBrewer)
library(dplyr)
library(ggplot2)
library(geneplotter)
library(pheatmap)
```

Warning: package ‘pheatmap’ was built under R version 4.1.3

[Hide](#)

```
library(enrichplot)
```

Registered S3 method overwritten by 'data.table':

```
method      from
print.data.table
```

Registered S3 method overwritten by 'ggtree':

```
method      from
identify.gg ggfun
```

Attaching package: ‘enrichplot’

The following object is masked from ‘package:lattice’:

```
dotplot
```

[Hide](#)

```
library(tidyr)
```

Warning: package ‘tidyr’ was built under R version 4.1.3

Attaching package: ‘tidyr’

The following object is masked from ‘package:S4Vectors’:

```
expand
```

[Hide](#)

```
library(EnhancedVolcano)
```

Loading required package: ggrepel

Warning: package 'ggrepel' was built under R version 4.1.3

Registered S3 methods overwritten by 'ggalt':

method	from
grid.draw.absoluteGrob	ggplot2
grobHeight.absoluteGrob	ggplot2
grobWidth.absoluteGrob	ggplot2
grobX.absoluteGrob	ggplot2
grobY.absoluteGrob	ggplot2

[Hide](#)

```
library(clusterProfiler)
```

clusterProfiler v4.2.2 For help: <https://yulab-smu.top/biomedical-knowledge-mining-book/>

If you use clusterProfiler in published research, please cite:

T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021, 2(3):100141

Attaching package: 'clusterProfiler'

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

select

The following object is masked from 'package:lattice':

dotplot

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

slice

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

slice

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

rename

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

filter

First, I load the normalized expression assay, the phenotype data and the feature annotation data for this dataset



```
GSE27272Norm_exprs <- read.delim('C:/Users/samen/Desktop/Bioinformatics Projects/Differential Expression Analysis with Bioconductor/GSE27272Norm_exprs.txt')

GSE27272Norm_phenoData <- read.delim('C:/Users/samen/Desktop/Bioinformatics Projects/Differential Expression Analysis with Bioconductor/GSE27272Norm_phenoData.txt')

GSE27272Norm_featureData <- read.delim('C:/Users/samen/Desktop/Bioinformatics Projects/Differential Expression Analysis with Bioconductor/GSE27272Norm_featureData.txt')
```

The file “GSE27272Norm\_exprs” contains the quantile normalized expression values for each sample in the study. The rows represent genes, which are labeled with an ensemble ids. The columns are denoted with the sample ids. Each entry is an expression value for a gene in a given tissue sample.

Hide

```
head(GSE27272Norm_exprs)
```

	GSM674...	GSM674...	GSM674...	GSM674...	GSM674...	GSM674...	GSM674...
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
ENSG00000196205	6.642863	6.601257	6.199341	6.270765	6.304746	6.599808	6.722213
ENSG00000233927	3.406199	4.947477	4.023500	3.513543	3.792942	4.218166	3.934279
ENSG00000117408	5.995206	5.899045	6.016128	6.524552	6.139264	5.660865	5.309342
ENSG00000167513	8.146764	8.246858	8.739746	8.197775	8.324278	8.194702	8.023971
ENSG00000145012	3.085944	2.478971	3.092929	3.346263	3.348481	3.747339	4.133915
ENSG00000197451	8.179314	8.253950	8.267293	8.129182	7.935259	7.770380	7.777080

6 rows | 1-9 of 51 columns

The file “GSE27272Norm\_phenoData” contains the phenotype data, which contains metadata information for each sample in the study. Each column is a descriptive variable, which records characteristics like sex, bmi, and smoking status. Each row index represents a particular sample.

Hide

```
#View(head(GSE27272Norm_phenoData))
head(GSE27272Norm_phenoData)
```

geo_accession	title	status	submission_date	last_
<chr>	<chr>	<chr>	<chr>	<chr>
GSM674282 GSM674282	non-smoker 302PL	Public on Apr 01 2011	Feb 13 2011	Apr 0
GSM674283 GSM674283	smoker 303PL	Public on Apr 01 2011	Feb 13 2011	Apr 0
GSM674284 GSM674284	non-smoker 305PL	Public on Apr 01 2011	Feb 13 2011	Apr 0

<b>geo_accession</b> <chr>	<b>title</b> <chr>	<b>status</b> <chr>	<b>submission_date</b> <chr>	<b>last_</b> <chr>
GSM674285	GSM674285	non-smoker 308PL	Public on Apr 01 2011	Feb 13 2011
GSM674286	GSM674286	smoker 313PL	Public on Apr 01 2011	Feb 13 2011
GSM674287	GSM674287	non-smoker 314PL	Public on Apr 01 2011	Feb 13 2011

6 rows | 1-6 of 60 columns

The file “GSE27272Norm\_featureData” contains the annotation data, which is used to characterized the genes in the expression data. Each row index is a particular gene that is also found in the expression data. The columns Probe IDs, Symbols, entrez IDs and ensembl IDs are all different labels to identify the features. The column “CHR” indicates the chromosome the gene is located on. “Start.Pos” and “End.Pos”, shows the position where the coding sequence of a gene starts and stops.

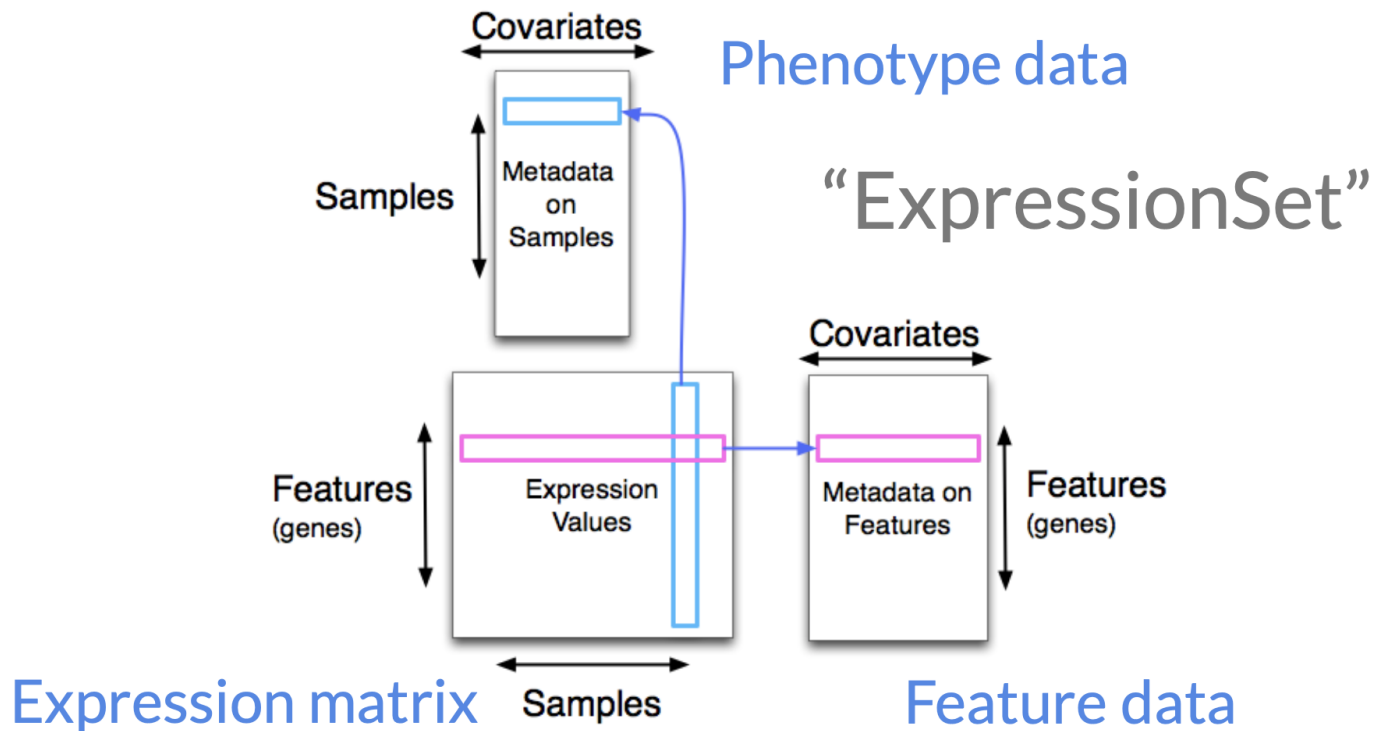
[Hide](#)

```
#View(head(GSE27272Norm_featureData))
head(GSE27272Norm_featureData)
```

<b>Probe_IDs</b> <chr>	<b>Symbols</b> <chr>	<b>Entrez_IDs</b> <int>	<b>...</b> <chr>	<b>Ensembl_IDs</b> <chr>	<b>Start.Pos</b> <int>	<b>End.Pos</b> <int>
1 ILMN_1343291	EEF1A1P5	NA	9	ENSG00000196205	133019486	133020874
2 ILMN_1651228	RPS28	6234	19	ENSG00000233927	8321158	8323340
3 ILMN_1651229	IPO13	9670	1	ENSG00000117408	43946950	43968022
4 ILMN_1651237	CDT1	81620	16	ENSG00000167513	88803789	88809258
5 ILMN_1651254	LPP	4026	3	ENSG00000145012	188153284	188890671
6 ILMN_1651262	HNRNPAB	3182	5	ENSG00000197451	178204533	178211163

6 rows

Now that the data is loaded, I will create an Expression Set with the expressino assay phenotype and the feature annotation data. An ExpressionSet is a standardized data structure in Bioconductor (via the BioBase library) which combines several different sources of information conveniently into one object



Hide

```
#Creating an ExpressionSet object with all attributes

GSE27272_Eset <- ExpressionSet(as.matrix(GSE27272Norm_exprs))

pData(GSE27272_Eset) <- GSE27272Norm_phenoData

featureData(GSE27272_Eset) <- as(GSE27272Norm_featureData, "AnnotatedDataFrame")
```

Exploratory Graph Before applying hypothesis testing on the data we should examine exploratory graphs like PCA and heatmaps to assess our data

Hide

```
#function configuring graph size
fig <- function(width, height){
  options(repr.plot.width = width, repr.plot.height = height)
}
```

I will create PCA plot examine the variation in the data by the phenotype variable of interest

Hide

```

fig(12,8)
GSE27272Norm_exprs <- Biobase::exprs(GSE27272_Eset)
PCA <- prcomp(t(GSE27272Norm_exprs), scale = FALSE )

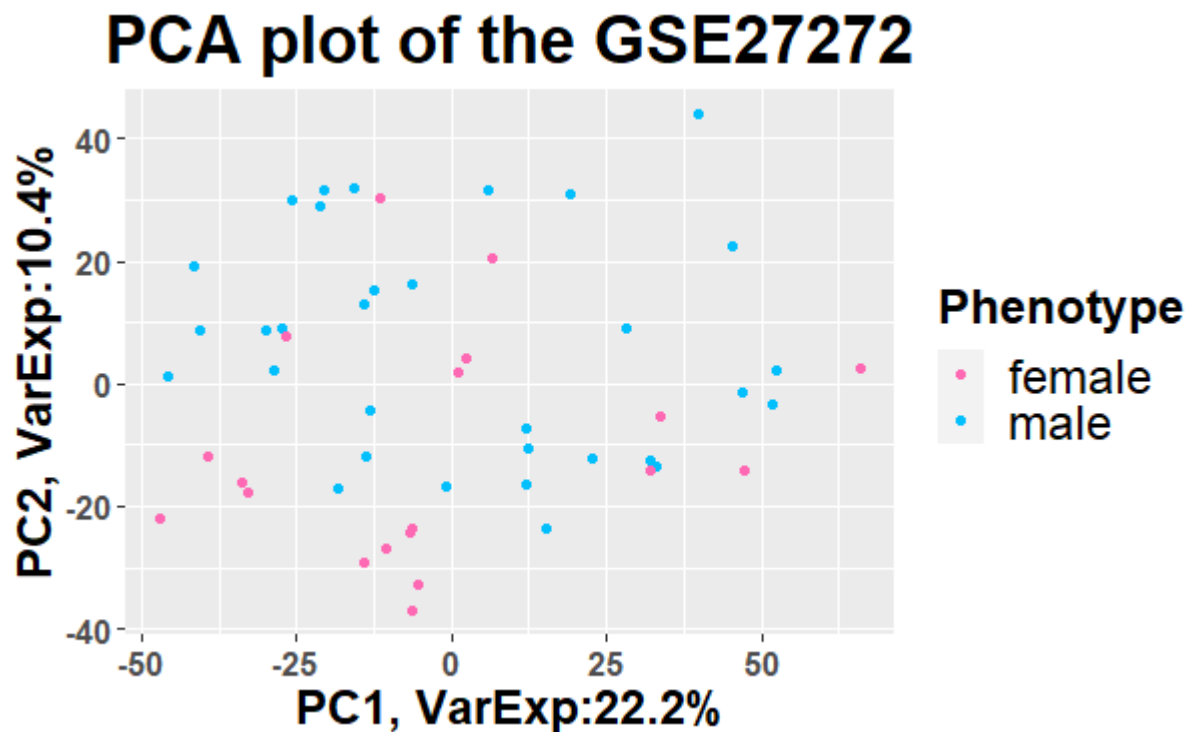
percentVar <- round(100*PCA$sdev^2/sum(PCA$sdev^2), 1)

sd_ratio <- sqrt(percentVar[2] / percentVar[1])

dataGG <- data.frame(PC1 = PCA$x[,1], PC2 = PCA$x[,2], Phenotype = Biobase::pData(GSE27272_Eset)
$sex)

ggplot(dataGG, aes(PC1, PC2)) + geom_point(aes(colour = Phenotype))+ ggtitle("PCA plot of the GS
E27272") + xlab(paste0("PC1, VarExp:", percentVar[1], "%")) + ylab(paste0("PC2, VarExp:", percen
tVar[2], "%"))+
  theme(plot.title = element_text(hjust = 0.5, size = 25, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y =element_text(size = 12, face = 'bold'),
        axis.title.x = element_text(size = 18, face = 'bold'),
        axis.title.y = element_text(size = 18, face = 'bold'),
        legend.title = element_text(size = 18, face = 'bold'),
        legend.text = element_text(size = 18)) +scale_color_manual(values = c("hotpink", "deepsk
yblue"))

```



Now I will be plotting a heatmap to examine the sample to sample distances and to see how well the samples cluster to sex.

[Hide](#)

```

annotation_for_heatmap <- data.frame(Phenotype = Biobase:: pData(GSE27272_Eset)$sex)

row.names(annotation_for_heatmap) <- row.names(pData(GSE27272_Eset))

dists<- as.matrix(dist(t(GSE27272Norm_exprs), method = "manhattan"))

rownames(dists) <- row.names(pData(GSE27272_Eset))
hmccl <- rev(colorRampPalette(RColorBrewer::brewer.pal(9, "YlOrRd"))(255))

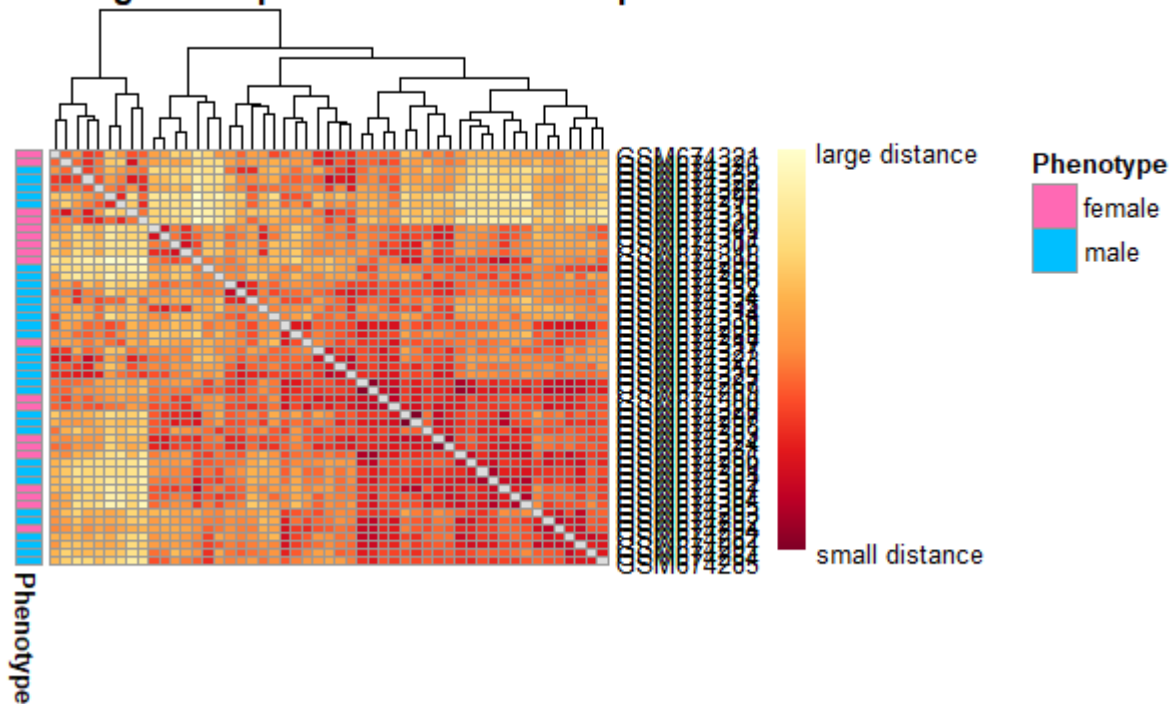
colnames(dists) <- NULL
diag(dists) <- NA

ann_colors <- list(Phenotype = c(female = "hotpink", male = "deepskyblue"))

pheatmap(dists, col = (hmccl), annotation_row = annotation_for_heatmap, annotation_colors = ann_
colors, legend = TRUE, treeheight_row = 0, legend_breaks = c(min(dists, na.rm = TRUE), max(dist
s, na.rm = TRUE)), legend_labels = (c("small distance", "large distance")), main = "Clustering h
eatmap for the GSE2727 samples")

```

### Clustering heatmap for the GSE2727 samples



**Filtering Data** Sometimes when performing a differential expression analysis we have to subset the genes we are testing based off the annotation data. For example, if we are doing a differential expression analysis by sex it would make sense to filter out the genes on the Y chromosome. Biologically, a male has a X and Y sex chromosome while a female has two X chromosomes. Features on the Y chromosome should have no expression for females because they have no Y chromosome. Therefore, we cannot compare the difference in expression between males and females for Y-linked genes.

[Hide](#)

```
# Filters the ExpressionSet (which includes the feature data and the expression data)
# to the genes that are not present in the Y chromosome

GSE27272_noY <- GSE27272_Eset[GSE27272_Eset@featureData@data$CHR != "Y",]
```

**HYPOTHESIS TESTING** Our next aim is to perform hypothesis testing on all of our genes after filtering the data. We do this by fitting a linear model for every gene and defining contrasts to test our hypotheses. In our case, our contrasts are “female” and “male” because we are interested in finding genes in the placenta are differentially expressed by the sex of the fetus. The first step is to create a design matrix for variable of interest.

Hide

```
design <- model.matrix(~0+GSE27272Norm_phenoData$sex)

colnames(design) <- c("female", "male")
GSE27272_samples <- as.character(GSE27272Norm_phenoData$geo_accession)

rownames(design) <- GSE27272_samples

#levels(GSE27272Norm_phenoData$characteristics_ch1.1) <- c("non_smoker", "smoker")
#design <- model.matrix(~0+GSE27272Norm_phenoData$characteristics_ch1.1)
#colnames(design) <- c("non_smoker", "smoker")
#GSE27272_samples <-
#  as.character(GSE27272Norm_phenoData$geo_accession)
#rownames(design) <- GSE27272_samples
```

Next, we create a contrast matrix that uses the function ‘makeContrasts’ from the ‘limma’ package that will take our design matrix as an input for the levels. Afterwards, we fit a linear model with function `contrast.fit()` from the ‘limma’ package in order to examine the relationship gene expression and our variable of interest. We use the function `eBayes()` on our linear model to get moderated t-test statistics. The `eBayes()` function performs the empirical Bayes method to squeeze the gene-wise residual variance towards a pooled estimate. Moderating the test-statistics with empirical Bayes method increases the statistical power of the differential expression analysis.

Hide

```
contrast_matrix <- makeContrasts(female-male, levels= design)
#contrast_matrix <- makeContrasts(non-smoke-smoker, levels = design)

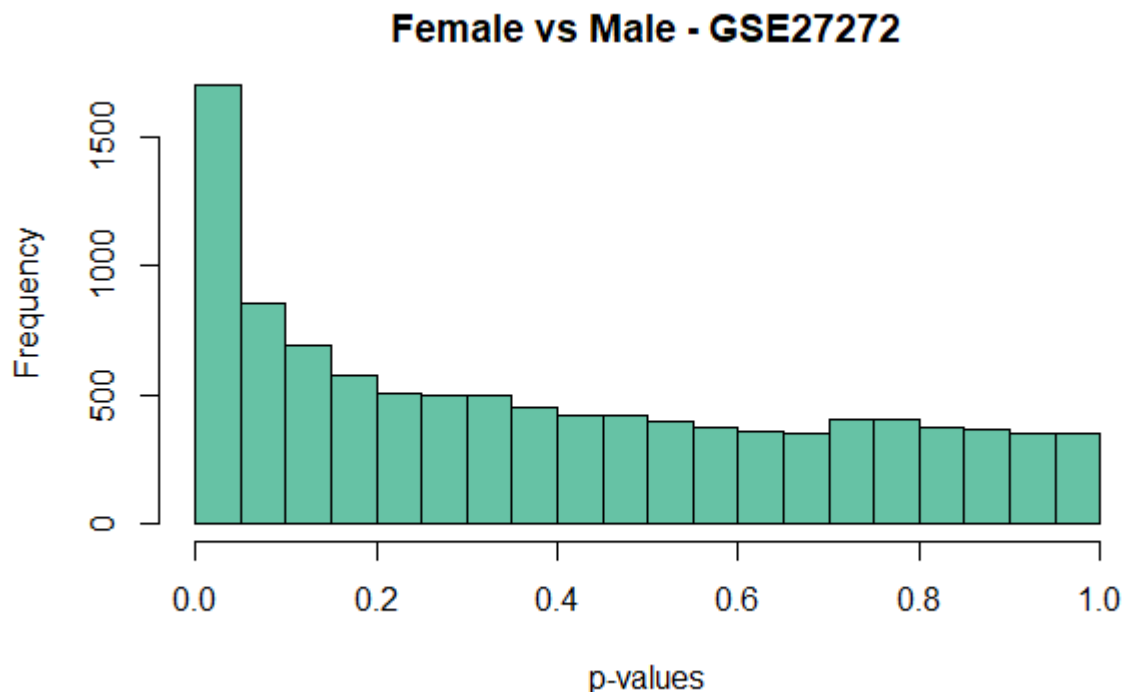
GSE27272_fit <- eBayes(contrasts.fit (lmFit(GSE27272_noY, design = design),
                                         contrast_matrix))
```

Next, we make a histogram showing the frequency of pvalues. Ideally, the histogram should be a right or positively skewed distribution. If not, this could be because we did not consider batch effects or failed consider possible covariates that we need to adjust for in the hypothesis testing. Furthermore, this would lead to quality loss in later analyses.

Hide

```
table_GSE27272 <- topTable(GSE27272_fit, number = Inf, confint = TRUE)

hist(table_GSE27272$P.Value, col = brewer.pal(3, name = "Set2") [1],
     main = "Female vs Male - GSE27272", xlab = "p-values")
```



Below I will make a table of the results of from differential expression analysis. The columns first three columns are Ensembl\_IDs, Entrez\_IDs, and Symbol, which are gene labels. Next we have the columns pvalue and adj.pvalue (adjusted pvalue), which are calculated using the limma package functions. We need to have adjusted p-values because we are doing mutiple independent tests, which could result in a large number of false positives. Another column is Log2FC, which stands for Log2 fold change. The fold change measures the effect size and informs us of the direction of the causal relationship. I will talk more about fold changes in the next section called Volcano Plot. The left and right end of the confidence intervals are denoted with CI.L and CI.R . Finally, there is the t-test statistic that is respresented with the column t

[Hide](#)

```
GSE27272_Results <- data.frame(Ensembl_IDs = table_GSE27272$Ensembl_IDs,
                               Entrez_IDs = table_GSE27272$Entrez_IDs,
                               Symbol = table_GSE27272$Symbols,
                               Log2FC = table_GSE27272$logFC,
                               pvalue = table_GSE27272$P.Value,
                               adj.pvalue = table_GSE27272$adj.P.Val,
                               CI.R = table_GSE27272$CI.R,
                               t = table_GSE27272$t,
                               stringsAsFactors = FALSE)

head(GSE27272_Results)
```

Ensembl_IDs	Entrez_IDs	Sym...	Log2FC	pvalue	adj.pvalue	CI.R
<chr>	<int>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>

Ensembl_IDs <chr>	Entrez_IDs <int>	Sym... <chr>	Log2FC <dbl>	pvalue <dbl>	adj.pvalue <dbl>	CI.R <dbl>
1 ENSG00000279231	NA		-7.074435	9.859400e-53	1.018279e-48	-6.865831 -68
2 ENSG00000112033	5467	PPARD	-2.453566	5.743812e-35	2.966104e-31	-2.292001 -30
3 ENSG00000057757	57095	PITHD1	-4.340124	3.755695e-33	1.292960e-29	-4.029060 -27
4 ENSG00000120696	84078	KBTBD7	-1.751614	1.063224e-19	2.745244e-16	-1.505696 -14
5 ENSG00000169490	83877	TM2D2	-2.464005	1.333513e-19	2.754504e-16	-2.116191 -14
6 ENSG00000219200	440400	RNASEK	-1.750475	2.085473e-19	3.589794e-16	-1.500714 -14
6 rows						

### Volcano Plots

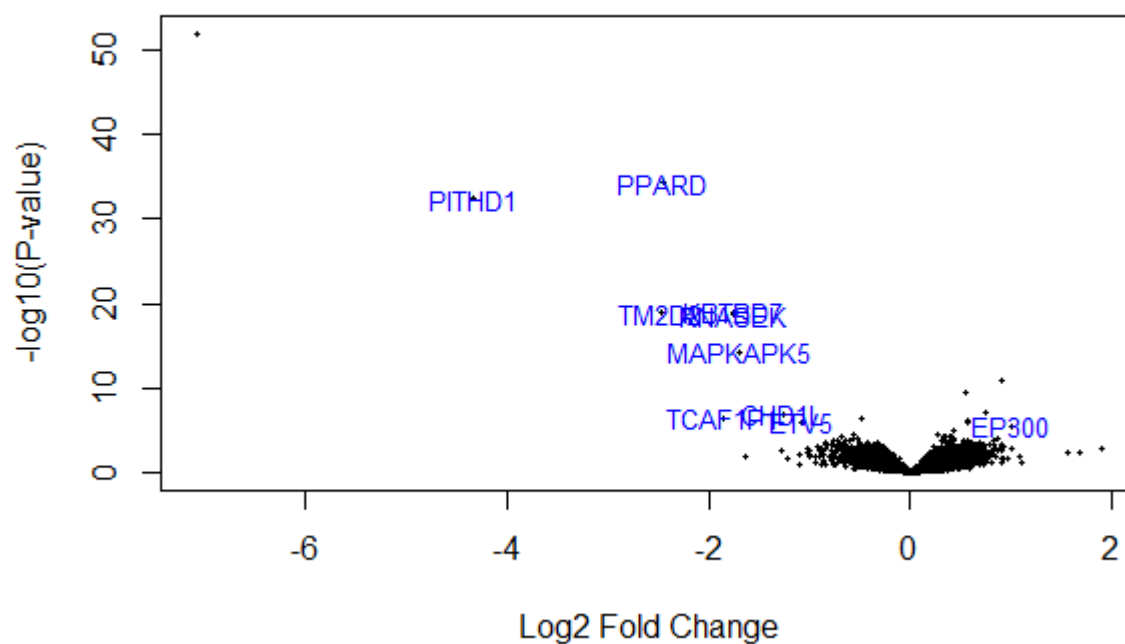
After performing hypothesis testing, it is ideal for us to try visual our results. The most common graph made are volcano plots, which are scatter plots of the fold change versus the p-value for each gene. Pvalues are usually transformed on the  $-\log_{10}$  scale. This means the more significant the p-value is (or the smaller the p-value is), the larger the value for the  $-\log_{10}$  pvalue. A  $-\log_{10}$  pvalue greater than  $-\log_{10}(0.05)$  is statistically significant.

[Hide](#)

```
volcano_names <- ifelse(abs(GSE27272_fit$coefficients) >= 1,
                        as.character(GSE27272_fit$genes$Symbols),
                        NA)

volcanoplot(GSE27272_fit, coef = 1L, style = "p-value", highlight = 100,
            names = volcano_names,
            xlab = "Log2 Fold Change",
            ylab = NULL,
            pch = 16,
            cex = 0.35)
```





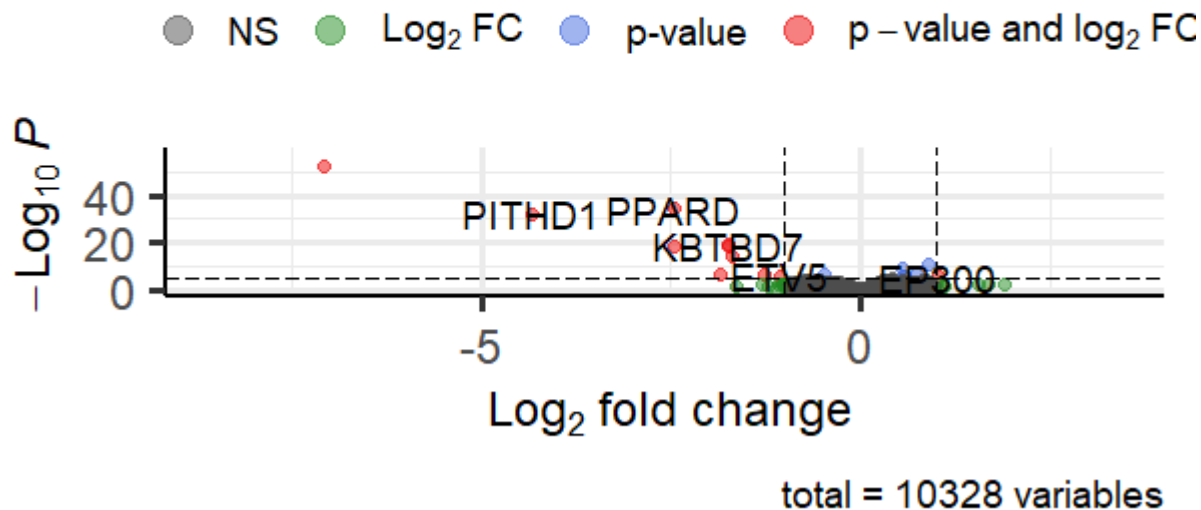
The graph above is an output of the basic volcano plot function from the limma package. This a good plot to do if you want a quick visualization of the differentially expressed genes. However, if you would want a graph that looks worthy of putting in a publication, then 'EnhancedVolcano' library should be used. The code and output of a graph from EnhancedVolcano is shown below.

[Hide](#)

```
EnhancedVolcano(GSE27272_Results,
  lab = as.character(table_GSE27272$Symbols),
  x = 'Log2FC',
  title = "GSE27272 Volcano Plot Female Vs Male",
  y = 'pvalue')
```

## GSE27272 Volcano Plot Female Vs Male

*EnhancedVolcano*



A gene is statistically significant after multiple testing correction if it has an adjusted p value less than .05. The log<sub>2</sub> fold change determines if a gene has a biologically significant change in expression (or a gene having a meaningful effect size) between both groups. For this example, Log<sub>2</sub> fold change determines whether a gene is upregulated or downregulated with respect to the reference group, which is placenta of female fetuses. If log<sub>2</sub> fold change is greater than 0, then the gene expression is higher in placenta of female fetuses compared placenta of a male fetuses. On the other hand, if the log<sub>2</sub> fold change is less than 0, then the gene expression is lower in the placentas of female fetus compared to male. If a gene's log<sub>2</sub> fold change is too close to zero in either direction, it is hard to claim that there is a biologically meaningful effect in the genes expression with respect to fetal sex, regardless of the statistical significance of the gene. Therefore, it's important to set a threshold to determine whether gene has a biological significant effect size. For our purposes, we are setting the criteria to be a log<sub>2</sub> fold change that has an absolute value greater than 1.

For the volcano plot produced using EnhancedVolcano, all of the genes shown in gray are non-significant both statistically and in effect size. The genes in blue have statistically significant p-values but didn't a log<sub>2</sub> fold change that suggests biological significance. The green labeled genes aren't statistically significant but has a biologically meaningful log<sub>2</sub> fold change. The genes colored in red are both statistically significant and has a biologically meaningful effect size.

Gene Enrichment Analysis and Over-Representation Analysis Next, I would like to assess if the differentially expressed genes were associate with certain biological processes by performing a functional enrichment analysis. One type of enrichment analysis for determining pathways is an Over-representation analysis. For an over-representation analysis, we find which differentially expressed genes are annotated to a specific gene set of a database, then we test whether the overlap of these genes is significantly over-represented in the gene set with a p-value calculated from a hypergeometric distribution. The p-values are correct for multiple testing correction.

Below is the list of statistically genes after multiple testing correction that had a meaningful effect size. These are the genes which I will test for enrichment analyses.

Hide

```
sigGenes <- GSE27272_Results[ GSE27272_Results$adj.pvalue < 0.05 & !is.na(GSE27272_Results$adj.pvalue) & abs(GSE27272_Results$Log2FC) >1, ]
```

```
sigGenes
```

	Ensembl_IDs <chr>	Entrez_IDs <int>	Symbol <chr>	Log2FC <dbl>	pvalue <dbl>	adj.pvalue <dbl>	
1	ENSG00000279231	NA		-7.074435	9.859400e-53	1.018279e-48	-6.865
2	ENSG00000112033	5467	PPARD	-2.453566	5.743812e-35	2.966104e-31	-2.292
3	ENSG00000057757	57095	PITHD1	-4.340124	3.755695e-33	1.292960e-29	-4.029
4	ENSG00000120696	84078	KBTBD7	-1.751614	1.063224e-19	2.745244e-16	-1.505
5	ENSG00000169490	83877	TM2D2	-2.464005	1.333513e-19	2.754504e-16	-2.116
6	ENSG00000219200	440400	RNASEK	-1.750475	2.085473e-19	3.589794e-16	-1.500
7	ENSG00000234608	NA	MAPKAPK5- AS1	-1.690772	5.234261e-15	7.722779e-12	-1.378
11	ENSG00000131778	9557	CHD1L	-1.269681	1.231929e-07	1.156669e-04	-0.853
13	ENSG00000223459	NA	TCAF1P1	-1.847744	4.153758e-07	3.300001e-04	-1.206
15	ENSG00000244405	2119	ETV5	-1.072775	1.086652e-06	7.481962e-04	-0.682
1-10 of 11 rows   1-8 of 8 columns					Previous	1	2
					Next		

Now we will use the function `enrichKEGG` that will perform an enrichment analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

[Hide](#)

```
sigGenes <- GSE27272_Results$Entrez_IDs[GSE27272_Results$adj.pvalue < 0.05 &
!is.na(GSE27272_Results$adj.pvalue) >1 ]
```

```
sigGenes <- na.exclude(sigGenes)
```

```
kk <- enrichKEGG(gene = sigGenes, organism = 'hsa' )
```

Reading KEGG annotation online:

Reading KEGG annotation online:

[Hide](#)

```
head(kk, n =10)
```

0 rows

Our gene list is significant after multiple testing correction on the Wnt signaling pathway for the genes PPARG and EP300. Wnt signal pathway is a group of signal transduction pathways with proteins that pass signals into a cell through cell receptor surfaces. Another significant pathway after multiple testing correction is Prostate Cancer for genes ETV5 and PPARG.