

# BB512/BB612 - Homework III

Apr 28, 2022

Use the GSE55945 GEO dataset to perform the required analyses:

```
library(GEOquery)

## Loading required package: Biobase
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unsplit, which.max, which.min
## 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)
options(download.file.method.GEOquery = "wget")
gset <- getGEO("GSE55945", GSEMatrix = TRUE, getGPL = TRUE)

## Found 1 file(s)
## GSE55945_series_matrix.txt.gz
## Rows: 54675 Columns: 22
## -- Column specification -----
## Delimiter: "\t"
## chr (1): ID_REF
## dbl (21): GSM1348933, GSM1348934, GSM1348935, GSM1348936, GSM1348937, GSM134...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## File stored at:
##
```

```
## /var/folders/n0/hxwj6lwd21s37_b12pzrdk480000gn/T//RtmpJTKeJt/GPL570.soft
gset <- gset[[1]]
```

1. [20 pt] Transform and normalize the expression data

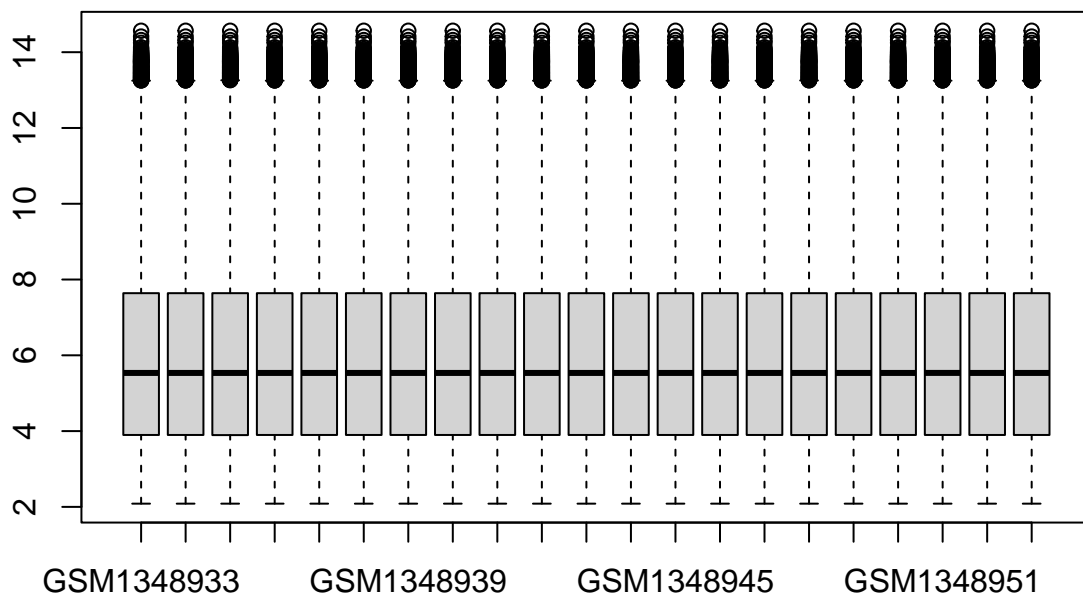
```
library(preprocessCore)

ex_org <- exprs(gset)

ex <- normalize.quantiles(ex_org)
rownames(ex) <- rownames(ex_org)
colnames(ex) <- colnames(ex_org)

ex <- log2(ex + 1)

boxplot(ex)
```



```
exprs(gset) <- ex
```

2. [40 pt] Determine the differentially expressed probes (DEGs) using `limma`, comparing “Malignant prostate tissue” vs. “Benign prostate tissue”. Filter DEGs so that  $|LFC| > 1$  and  $FDR < 0.05$ . Display results in a volcano plot.

```
library(limma)

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

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

```
##
```

```
## plotMA
```

```
head(pData(gset))
```

```
##               title geo_accession                status
## GSM1348933 Prostate Cancer_MS 36D6   GSM1348933 Public on Mar 18 2014
## GSM1348934 Prostate Cancer_MS 36C1   GSM1348934 Public on Mar 18 2014
## GSM1348935 Prostate Cancer_MS 36D7   GSM1348935 Public on Mar 18 2014
## GSM1348936 Prostate Cancer_MS 36D8   GSM1348936 Public on Mar 18 2014
## GSM1348937 Prostate Cancer_MS 36A6   GSM1348937 Public on Mar 18 2014
## GSM1348938 Prostate Cancer_MS 36C4   GSM1348938 Public on Mar 18 2014
##      submission_date last_update_date type channel_count
## GSM1348933      Mar 17 2014      Mar 18 2014  RNA             1
## GSM1348934      Mar 17 2014      Mar 18 2014  RNA             1
## GSM1348935      Mar 17 2014      Mar 18 2014  RNA             1
## GSM1348936      Mar 17 2014      Mar 18 2014  RNA             1
## GSM1348937      Mar 17 2014      Mar 18 2014  RNA             1
## GSM1348938      Mar 17 2014      Mar 18 2014  RNA             1
##               source_name_ch1 organism_ch1
## GSM1348933 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
## GSM1348934 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
## GSM1348935 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
## GSM1348936 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
## GSM1348937 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
## GSM1348938 malignant tissue_TMPRSS2:ERG fusion negative Homo sapiens
##      characteristics_ch1
## GSM1348933 tissue type: Malignant prostate tissue
## GSM1348934 tissue type: Malignant prostate tissue
## GSM1348935 tissue type: Malignant prostate tissue
## GSM1348936 tissue type: Malignant prostate tissue
## GSM1348937 tissue type: Malignant prostate tissue
## GSM1348938 tissue type: Malignant prostate tissue
##      characteristics_ch1.1
## GSM1348933 tmprss2: ERG fusion status: negative
## GSM1348934 tmprss2: ERG fusion status: negative
## GSM1348935 tmprss2: ERG fusion status: negative
## GSM1348936 tmprss2: ERG fusion status: negative
## GSM1348937 tmprss2: ERG fusion status: negative
## GSM1348938 tmprss2: ERG fusion status: negative
##      treatment_protocol_ch1
## GSM1348933 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
## GSM1348934 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
## GSM1348935 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
## GSM1348936 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
## GSM1348937 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
## GSM1348938 Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min.
##      growth_protocol_ch1
## GSM1348933 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
## GSM1348934 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
## GSM1348935 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
## GSM1348936 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
## GSM1348937 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
## GSM1348938 A biopsy punch was used to select the PCa tissues from the OCT prostate sample blocks.
```

```

##          molecule_ch1
## GSM1348933      total RNA
## GSM1348934      total RNA
## GSM1348935      total RNA
## GSM1348936      total RNA
## GSM1348937      total RNA
## GSM1348938      total RNA
##
## GSM1348933 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
## GSM1348934 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
## GSM1348935 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
## GSM1348936 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
## GSM1348937 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
## GSM1348938 Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spect:
##          label_ch1
## GSM1348933      biotin
## GSM1348934      biotin
## GSM1348935      biotin
## GSM1348936      biotin
## GSM1348937      biotin
## GSM1348938      biotin
##
## GSM1348933 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
## GSM1348934 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
## GSM1348935 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
## GSM1348936 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
## GSM1348937 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
## GSM1348938 Biotin-UTP was incorporated during the overnight in vitro transcription step according to
##          taxid_ch1
## GSM1348933      9606
## GSM1348934      9606
## GSM1348935      9606
## GSM1348936      9606
## GSM1348937      9606
## GSM1348938      9606
##
## GSM1348933 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
## GSM1348934 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
## GSM1348935 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
## GSM1348936 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
## GSM1348937 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
## GSM1348938 cRNA (15 g) was fragmented and hybridized to arrays according to the manufacturer's (Ambi
##          scan_protocol description
## GSM1348933 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36D6
## GSM1348934 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36C1
## GSM1348935 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36D7
## GSM1348936 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36D8
## GSM1348937 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36A6
## GSM1348938 GeneArray Scanner G2500A (Hewlett-Packard)      MS 36C4
##          data_processing platform_id      contact_name
## GSM1348933 BioConductor R packages      GPL570 M. Simo,,Arredouani
## GSM1348934 BioConductor R packages      GPL570 M. Simo,,Arredouani
## GSM1348935 BioConductor R packages      GPL570 M. Simo,,Arredouani
## GSM1348936 BioConductor R packages      GPL570 M. Simo,,Arredouani

```

```

## GSM1348937 BioConductor R packages      GPL570 M. Simo,,Arredouani
## GSM1348938 BioConductor R packages      GPL570 M. Simo,,Arredouani
##          contact_email contact_department
## GSM1348933 simarred@gmail.com            Surgery
## GSM1348934 simarred@gmail.com            Surgery
## GSM1348935 simarred@gmail.com            Surgery
## GSM1348936 simarred@gmail.com            Surgery
## GSM1348937 simarred@gmail.com            Surgery
## GSM1348938 simarred@gmail.com            Surgery
##          contact_institute contact_address contact_city
## GSM1348933 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
## GSM1348934 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
## GSM1348935 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
## GSM1348936 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
## GSM1348937 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
## GSM1348938 Beth Israel Deaconess Medical Center 330 Brookline Ave.      Boston
##          contact_state contact_zip/postal_code contact_country
## GSM1348933          MA          02215          USA
## GSM1348934          MA          02215          USA
## GSM1348935          MA          02215          USA
## GSM1348936          MA          02215          USA
## GSM1348937          MA          02215          USA
## GSM1348938          MA          02215          USA
##
## GSM1348933 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348933/suppl/GSM1348933_011508_HGU1
## GSM1348934 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348934/suppl/GSM1348934_110607_HGU1
## GSM1348935 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348935/suppl/GSM1348935_011508_HGU1
## GSM1348936 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348936/suppl/GSM1348936_011508_HGU1
## GSM1348937 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348937/suppl/GSM1348937_110807_HGU1
## GSM1348938 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1348nnn/GSM1348938/suppl/GSM1348938_110607_HGU1
##          data_row_count relation          tissue type:ch1
## GSM1348933          54675      Malignant prostate tissue
## GSM1348934          54675      Malignant prostate tissue
## GSM1348935          54675      Malignant prostate tissue
## GSM1348936          54675      Malignant prostate tissue
## GSM1348937          54675      Malignant prostate tissue
## GSM1348938          54675      Malignant prostate tissue
##          tmprss2:ch1
## GSM1348933 ERG fusion status: negative
## GSM1348934 ERG fusion status: negative
## GSM1348935 ERG fusion status: negative
## GSM1348936 ERG fusion status: negative
## GSM1348937 ERG fusion status: negative
## GSM1348938 ERG fusion status: negative
table(gset$`tissue type:ch1`)

##
##      Benign prostate tissue Malignant prostate tissue
##              8              13

gset$group <- ifelse(pData(gset)$`tissue type:ch1` == "Malignant prostate tissue", "malignant", "benign")

design <- model.matrix(~0 + group, gset)
colnames(design) <- sub("group", "", colnames(design))

```

```

fit <- lmFit(gset, design)

cont.matrix <- makeContrasts(contrasts = "malignant-benign", levels = design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2, 0.01)

tT <- topTable(fit2, adjust="fdr", sort.by="p", number = Inf)

tT_final <- tT[abs(tT$logFC) > 1 & tT$adj.P.Val < 0.05, ]

library(EnhancedVolcano)

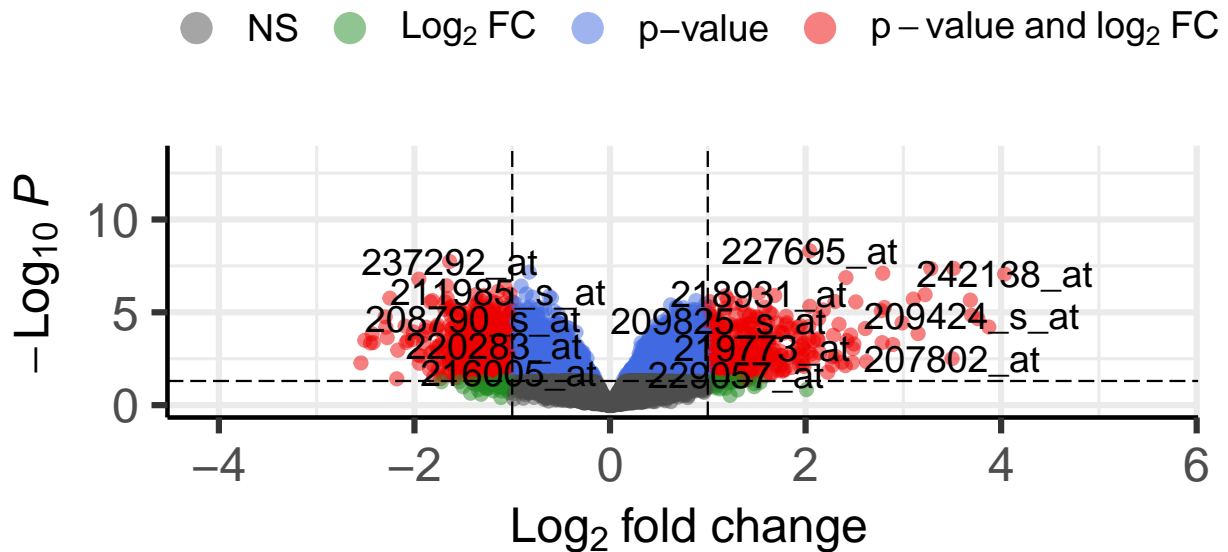
## Loading required package: ggplot2
## Loading required package: ggrepel
## Registered S3 methods overwritten by 'ggalt':
##   method                from
##   grid.draw.absoluteGrob ggplot2
##   grobHeight.absoluteGrob ggplot2
##   grobWidth.absoluteGrob  ggplot2
##   grobX.absoluteGrob      ggplot2
##   grobY.absoluteGrob      ggplot2
EnhancedVolcano(tT,
  lab = rownames(tT),
  x = "logFC",
  y = "P.Value",
  pCutoff = 0.05,
  FCcutoff = 1)

## Warning: Ignoring unknown parameters: xlim, ylim

```

## Volcano plot

*Enhanced Volcano*



3. [40 pt] Using the filtered DEGs, perform enrichment analysis on KEGG pathways.

```
library(enrichR)
```

```
## Welcome to enrichR
## Checking connection ...

## Enrichr ... Connection is Live!
## FlyEnrichr ... Connection is available!
## WormEnrichr ... Connection is available!
## YeastEnrichr ... Connection is available!
## FishEnrichr ... Connection is available!
```

```
res <- enrichr(tT_final$Gene.Symbol, databases = "KEGG_2021_Human")
```

```
## Uploading data to Enrichr... Done.
## Querying KEGG_2021_Human... Done.
## Parsing results... Done.
```

```
head(res$KEGG_2021_Human)
```

##	Term	Overlap	P.value	Adjusted.P.value
## 1	Vascular smooth muscle contraction	14/133	1.5093e-06	0.00036072
## 2	Dilated cardiomyopathy	10/96	5.1989e-05	0.00621265
## 3	cGMP-PKG signaling pathway	13/167	9.0254e-05	0.00719024
## 4	Focal adhesion	14/201	1.6064e-04	0.00959843
## 5	Salivary secretion	9/93	2.1621e-04	0.01033464
## 6	Insulin secretion	8/86	6.2879e-04	0.02504685
##	Old.P.value	Old.Adjusted.P.value	Odds.Ratio	Combined.Score

## 1	0	0	5.3433	71.621
## 2	0	0	5.2412	51.701
## 3	0	0	3.8182	35.558
## 4	0	0	3.3884	29.602
## 5	0	0	4.8187	40.666
## 6	0	0	4.6036	33.936
##				Genes
## 1	PPP1R14A;EDN3;PRKCB;ITPR1;CACNA1D;ADRA1A;MYLK;ADCY5;CALD1;KCNMB1;MYH11;CALM1;MYL9;PRKG1			
## 2	SGCD;PLN;TGFB3;TPM2;TPM1;ITGA7;CACNA1D;SLC8A1;SGCG;ADCY5			
## 3	ITPR1;ATP2B4;CACNA1D;ATP1A2;ADRA1A;SLC8A1;MYLK;ADCY5;PLN;KCNMB1;CALM1;MYL9;PRKG1			
## 4	PRKCB;CAV2;CAV1;THBS4;MYLK;CCND2;RAP1A;COL4A6;FLNA;ITGA7;PIP5K1B;PAK3;MYL9;VCL			
## 5	CHRM3;PRKCB;ITPR1;ATP2B4;ATP1A2;CALM1;ADRA1A;PRKG1;ADCY5			
## 6	CHRM3;ADCYAP1R1;PRKCB;KCNMB1;ATP1A2;CACNA1D;ADCY5;TRPM4			

```
plotEnrich(res$KEGG_2021_Human)
```

