Affymetrix GeneChip Data Analyais with LIMMA

Asad

2023-05-13

Call Libraries

```
library(affy)
```

```
## 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
```

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

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

```
library(oligo)
```

```
## Loading required package: oligoClasses
```

```
## Welcome to oligoClasses version 1.56.0
```

```
##
## Attaching package: 'oligoClasses'
```

```
## The following object is masked from 'package:affy':
##
##     list.celfiles

## Loading required package: Biostrings

## Loading required package: S4Vectors

## Loading required package: stats4

##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:utils':
##
##     findMatches

## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

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

## Loading required package: XVector

## Loading required package: GenomeInfoDb

##
## Attaching package: 'Biostrings'

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

## ========================================================================

## Welcome to oligo version 1.58.0

## ========================================================================

##
## Attaching package: 'oligo'

## The following objects are masked from 'package:affy':
##
##     intensity, MAplot, mm, mm<-, mmindex, pm, pm<-, pmindex,
##     probeNames, rma
```

```
library(GEOquery)
```

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

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

```
library(Biobase)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

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

```
library(splitstackshape)
library(arrayQualityMetrics)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:oligo':
##
##     summarize
```

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

```
## The following object is masked from 'package:GenomeInfoDb':
##
##     intersect
```

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

```
## 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
```

```
library(limma)
```

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

## The following object is masked from 'package:oligo':
##
##     backgroundCorrect

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

```
library(annotate)
```

```
## Loading required package: AnnotationDbi

##
## Attaching package: 'AnnotationDbi'

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

## Loading required package: XML
```

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

```
##
```

```
library(hgu133plus2.db)
```

```
##
```

```
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
```

```
library(umap)
library(maptools)
```

```
## Loading required package: sp
```

```
##
## Attaching package: 'sp'
```

```
## The following object is masked from 'package:oligo':
##
##     geometry
```

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

```
## The following object is masked from 'package:oligoClasses':
##
##     geometry
```

```
## Checking rgeos availability: TRUE
## Please note that 'maptools' will be retired during 2023,
## plan transition at your earliest convenience;
## some functionality will be moved to 'sp'.
```

```
setwd("D:/CancerData/Data")
```

Read celFiles

```
celFiles <- list.celfiles()
gset <- read.celfiles(celFiles)
```

```
## Loading required package: pd.hg.u133.plus.2
```

```
## Loading required package: RSQLite

## Loading required package: DBI

## Platform design info loaded.

## Reading in : GSM869667.CEL
## Reading in : GSM869668.CEL
## Reading in : GSM869669.CEL
## Reading in : GSM869670.CEL
## Reading in : GSM869671.CEL
## Reading in : GSM869672.CEL
## Reading in : GSM869673.CEL
## Reading in : GSM869674.CEL
## Reading in : GSM869675.CEL
## Reading in : GSM869676.CEL
## Reading in : GSM869678.CEL
## Reading in : GSM869679.CEL
```

Check if the expression ranges between 0-16 which means log2 transformed data

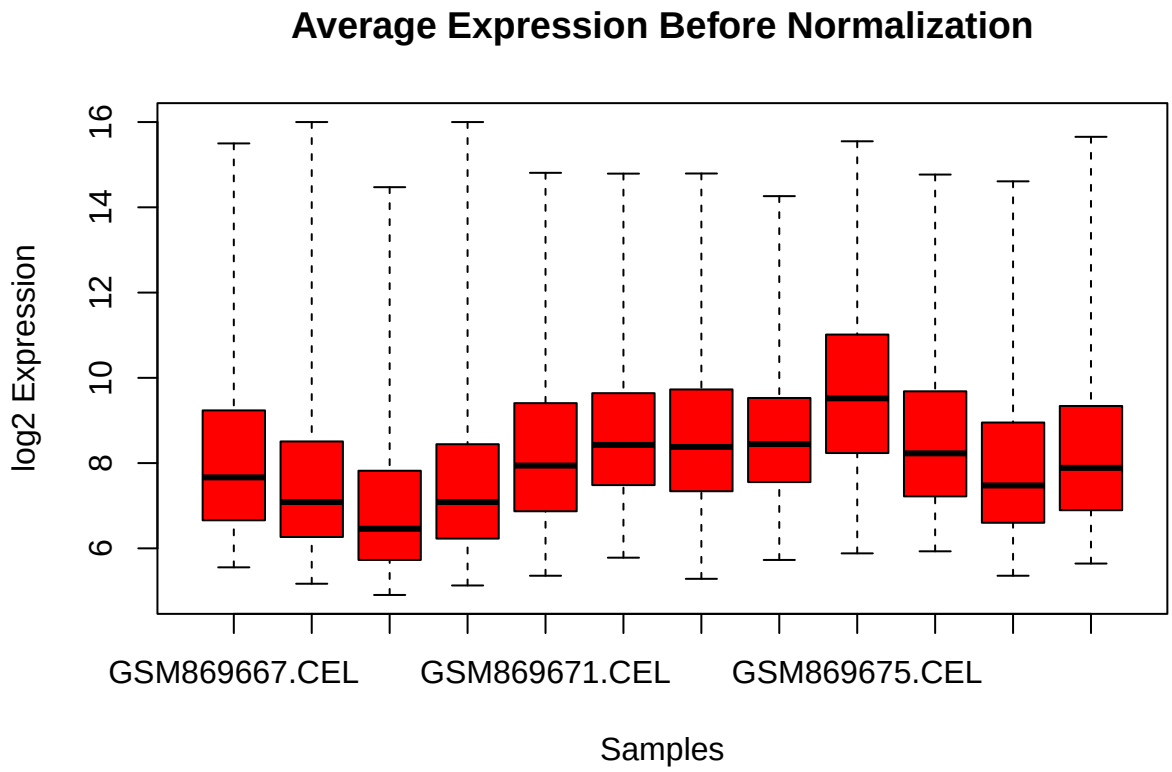
```
head(exprs(gset), 10)
```

```
##    GSM869667.CEL GSM869668.CEL GSM869669.CEL GSM869670.CEL GSM869671.CEL
## 1            229           108           109           124           175
## 2          10821          6408          6348          7954         19795
## 3            281           120           191           197           229
## 4          11322          6574          6440          8379         20164
## 5             84            97            64            94           149
## 6            189            96            87            96           138
## 7          10640          6409          6237          8195         20049
## 8            261           127           142           157           215
## 9          10548          6369          6215          8078         19846
## 10           237           104           127           130           210
##    GSM869672.CEL GSM869673.CEL GSM869674.CEL GSM869675.CEL GSM869676.CEL
## 1            175           206           190           430           236
## 2          21528         22435         18942         27267          9115
## 3            613           717           460           916           358
## 4          22583         22919         19122         28664          9391
## 5            140           117           179           149           174
## 6            164           192           181           258           237
## 7          21700         22731         19600         29211          9493
## 8            386           589           484           634           345
## 9          20858         22756         19066         29337          9252
## 10           339           558           335           384           292
##    GSM869678.CEL GSM869679.CEL
## 1            138           143
## 2           8686          9857
## 3            214           215
## 4           9018          9944
## 5            181           135
```

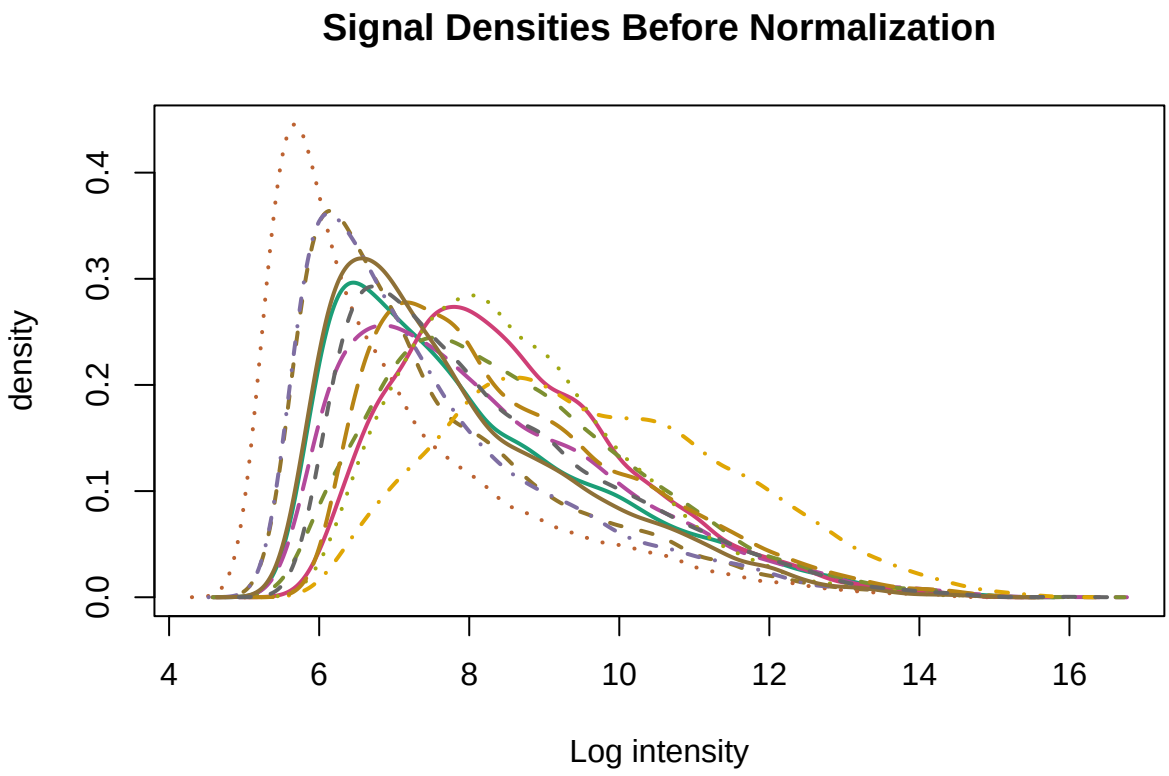
```
## 6              94           122
## 7            8503         10106
## 8             186           176
## 9            8491         10031
## 10            159           154
```

Check sample distribution and probe intensity plot before normalization

```
boxplot(gset, xlab='Samples', ylab='log2 Expression', col='Red',
        main= 'Average Expression Before Normalization')
```




```
hist(gset, lwd=2,xlab='Log intensity',
     main="Signal Densities Before Normalization")
```



#Generate pseudo-image of chip intensity for individual sample (CEL file) #image(affyRaw[,1])

#Perform normalization #RMA Normalization

```
gset <- rma(gset)
```

```
## Background correcting
## Normalizing
## Calculating Expression
```

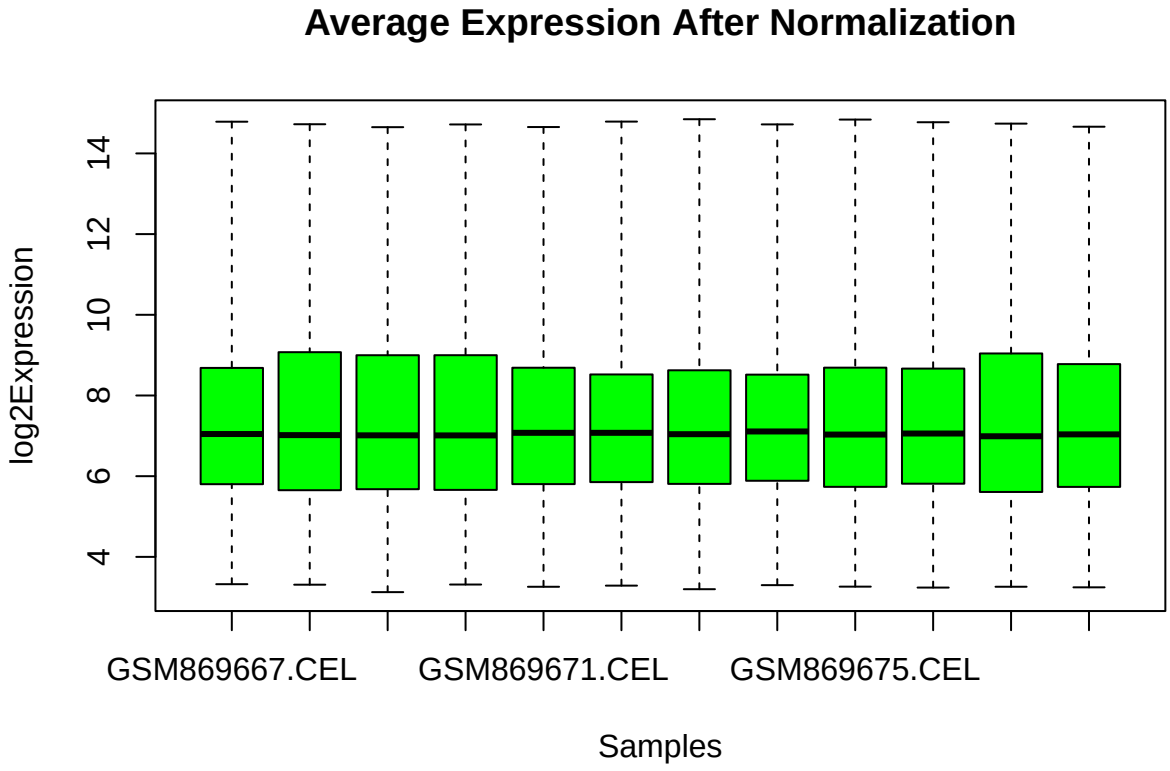
#Quantile normalization #gset<- normalize(gset, method='quantile')

Apply log2 transformation if Required

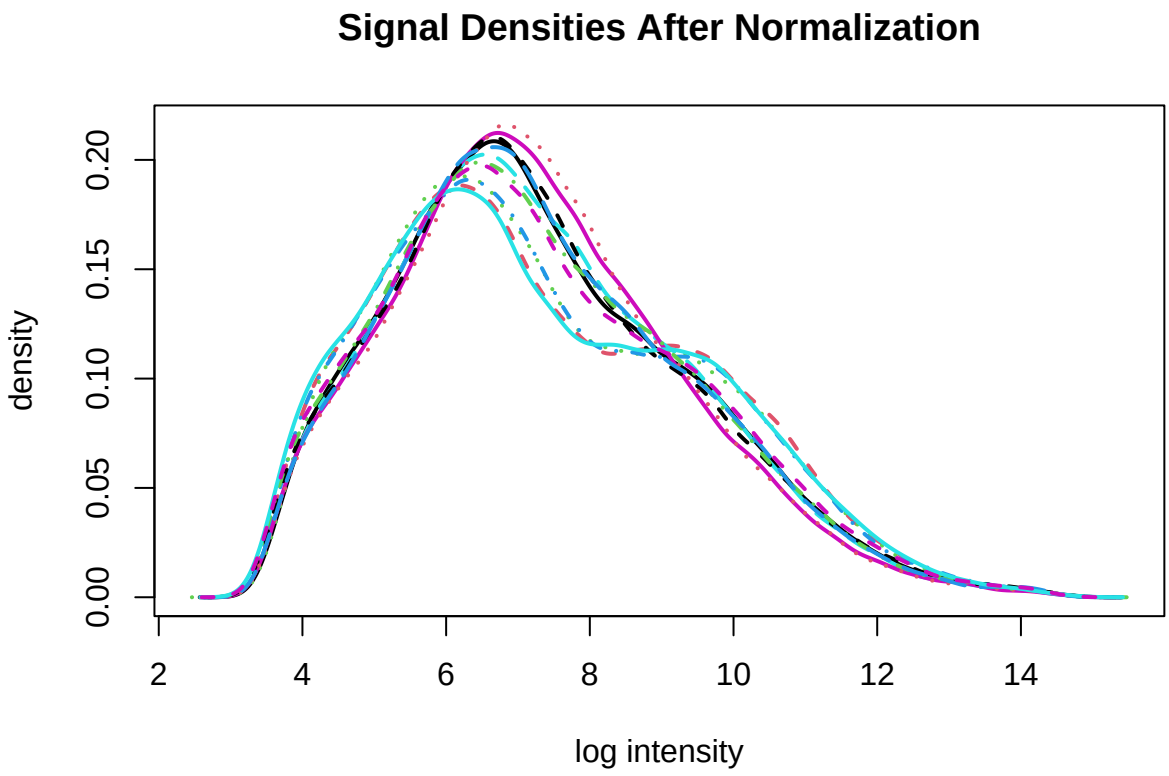
```
ex <- exprs(gset)
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
  (qx[6]-qx[1] > 50 && qx[2] > 0)
if (LogC) { ex[which(ex <= 0)] <- NaN
exprs(gset) <- log2(ex) }
```

Check probe intensity plot after normalization

```
boxplot(gset, xlab='Samples', ylab='log2Expression',col='Green',
        main= 'Average Expression After Normalization')
```




```
hist(gset, lwd=2,xlab='log intensity',
     main="Signal Densities After Normalization")
```



Naming the Probe IDs to Gene IDs

```
ID<- featureNames(gset)
Symbol <- getSYMBOL(ID,'hgu133plus2.db')
fData(gset) <- data.frame(Symbol=Symbol)
```

Perform DEG Analysis

```
Groups<- factor(c("Normal", "Normal", "Normal","Normal", "Normal",
                  "Normal","Tumor","Tumor","Tumor","Tumor","Tumor","Tumor"))
design<- model.matrix(~Groups-1)
colnames(design)<- c("Normal","Tumor")
head(design)
```

```
##   Normal Tumor
## 1      1     0
## 2      1     0
## 3      1     0
## 4      1     0
## 5      1     0
## 6      1     0
```

```
fit<- lmFit(gset, design)
contrast.mat<- makeContrasts("Tumor-Normal", levels = design)
fit2<- contrasts.fit(fit, contrast.mat)
fit3<-eBayes(fit2)
options(digits = 2)
topTable(fit3)
```

```
##                 Symbol logFC AveExpr   t P.Value adj.P.Val      B
## 200806_s_at      HSPD1  1.70    12.0 6.0 4.6e-05      0.56  0.780
## 204666_s_at      SIKE1  1.28     6.9 5.9 6.2e-05      0.56  0.623
## 215512_at      MARCHF6  1.70     6.0 5.8 6.4e-05      0.56  0.600
## 1555033_a_at     RGS12  1.04     5.9 5.6 9.9e-05      0.56  0.359
## 214513_s_at      CREB1  1.34     8.0 5.4 1.3e-04      0.56  0.220
## 210950_s_at      FDFT1  1.07    11.1 5.4 1.3e-04      0.56  0.214
## 235819_at       BTF3L4  1.38     6.7 5.2 1.7e-04      0.56  0.030
## 220220_at    LRRC37A4P  1.03     5.7 5.2 1.8e-04      0.56  0.024
## 230134_s_at      RC3H2  0.94     7.7 5.1 2.1e-04      0.56 -0.071
## 242019_at        CERS6  1.32     6.3 5.1 2.3e-04      0.56 -0.127
```

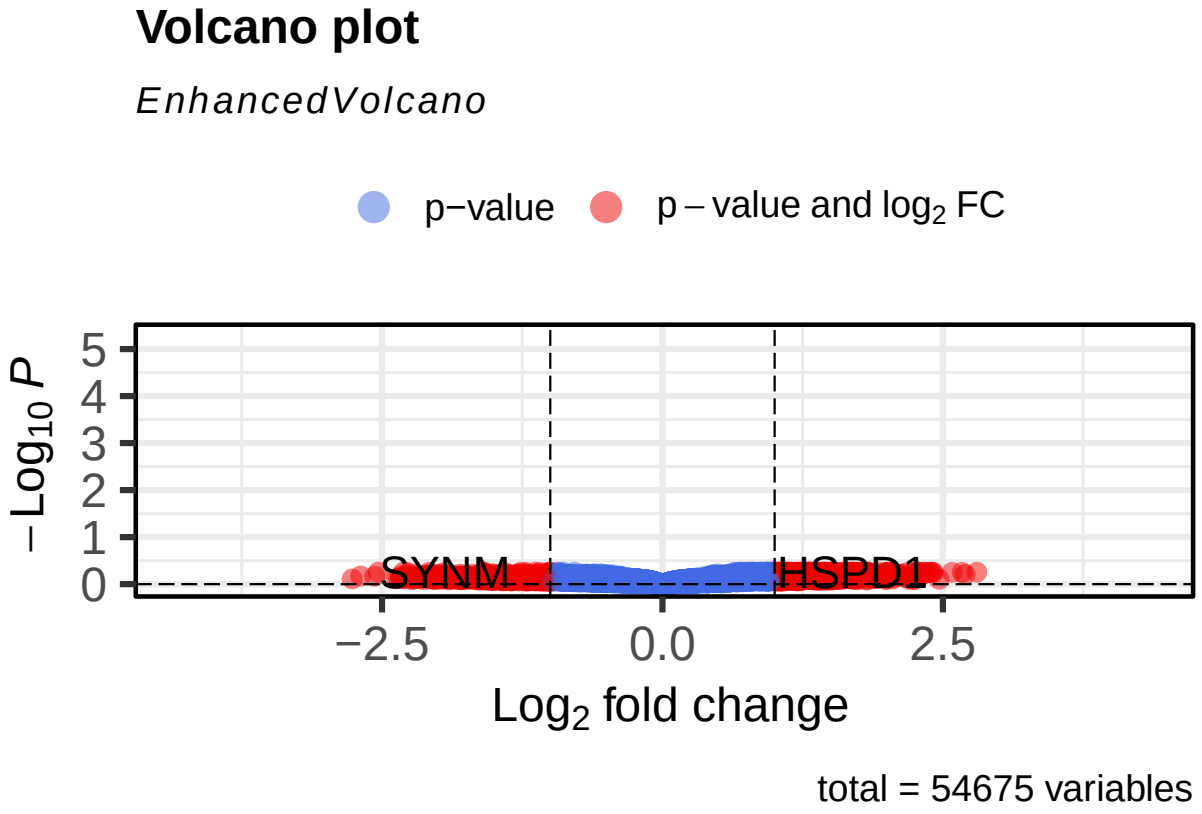
```
DEGs<- topTable(fit3, n=Inf, adjust="BH")
write.csv(DEGs, file="DEGs.csv")
```

Stratify

```
Significant_DEGs<- filter(DEGs,logFC>1|logFC< -1 & adj.P.Val<0.7)
Significant_UP<- filter(Significant_DEGs,logFC> 1)
Significant_Down<- filter(Significant_DEGs,logFC< -1)
```

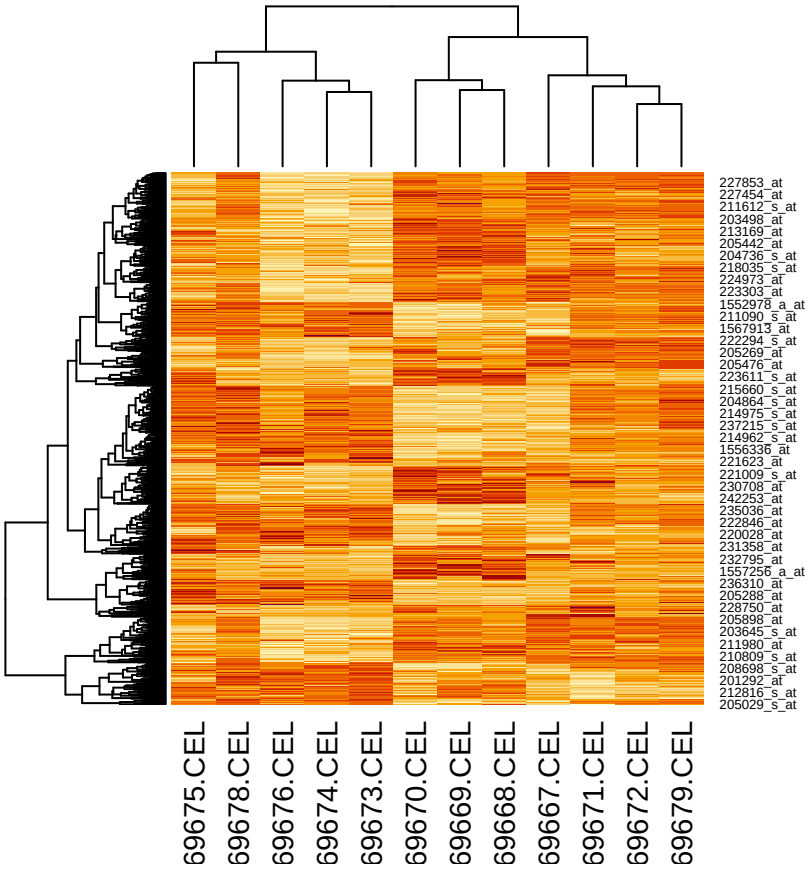
Create basic volcano plot

```
EnhancedVolcano(DEGs,
                lab = DEGs$Symbol,
                x = 'logFC',
                y = 'adj.P.Val',
                pCutoff = 1,
                FCcutoff = 1,
                pointSize = 3.0,
                labSize = 6.0,
                border='full')
```



Visualizing DEG of Interest

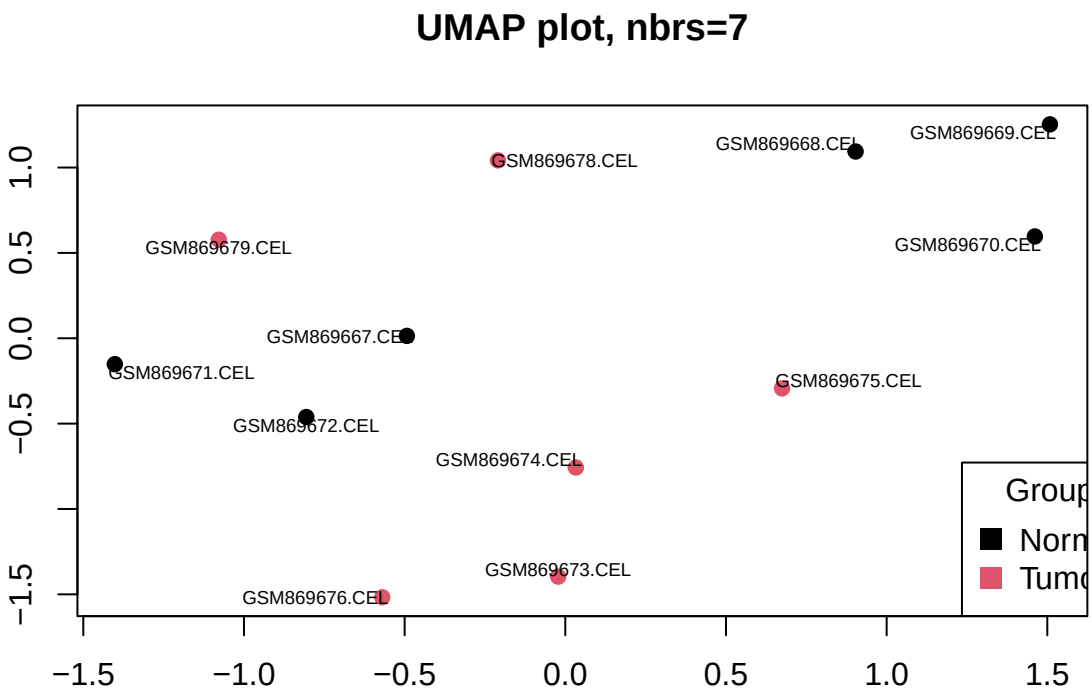
```
DEG_of_Interest <- data.frame(topTable(fit3, number=Inf,  lfc=1, p.value = 1))
Genes_of_interest <- gset[rownames(DEG_of_Interest),]
heatmap(exprs(Genes_of_interest))
```



UMAP plot (dimensionality reduction)

```
ex <- na.omit(ex) # eliminate rows with NAs
ex <- ex[!duplicated(ex), ]  # remove duplicates
ump <- umap(t(ex), n_neighbors = 7, random_state = 123)
#par(mar=c(3,3,2,6), xpd=F)
plot(ump$layout, main="UMAP plot, nbrs=7", xlab="", ylab="", col=Groups, pch=20, cex=1.5)
legend('bottomright', inset=c(-0.05,0), legend=levels(Groups), pch=15,
       col=1:nlevels(Groups), title="Group", pt.cex=1.5)
pointLabel(ump$layout, labels = rownames(ump$layout), method="SANN", cex=0.6)
```

```
## Warning: Function moved to the car package because maptools is retiring in 2023
```



#Saving normalized expression

```
head(exprs(gset),10)
```

```
##           GSM869667.CEL GSM869668.CEL GSM869669.CEL GSM869670.CEL GSM869671.CEL
## 1007_s_at          11.6          13.0          13.2          12.4          12.4
## 1053_at             9.4           9.6           8.9           9.3           8.9
## 117_at              8.4           7.9           8.4           9.8           8.2
## 121_at              9.9           8.7           8.5           8.7           9.7
## 1255_g_at           5.7           4.7           4.7           4.9           4.8
## 1294_at             8.8           8.1           8.1           8.7           9.2
## 1316_at             6.7           7.3           8.3           7.9           6.5
## 1320_at             6.3           6.1           6.6           6.0           6.9
## 1405_i_at           7.4           5.3           6.6           6.2           7.8
## 1431_at             5.2           5.0           5.5           5.3           5.0
##           GSM869672.CEL GSM869673.CEL GSM869674.CEL GSM869675.CEL GSM869676.CEL
## 1007_s_at          12.8          13.1          12.8          13.3          12.4
## 1053_at             9.8           9.4           9.9           8.6           9.4
## 117_at              8.7           7.9           8.3           7.4           8.7
## 121_at             10.1          10.0          10.4           9.9           9.6
## 1255_g_at           4.9           4.9           4.8           7.0           5.2
## 1294_at             8.4           8.2           8.0           8.2           7.3
## 1316_at             6.4           6.4           7.0           6.4           7.0
## 1320_at             6.6           6.7           6.8           7.0           6.5
## 1405_i_at           6.8           5.6           5.5           6.1           6.3
## 1431_at             5.3           5.2           5.1           5.4           5.1
```

```
##           GSM869678.CEL GSM869679.CEL
## 1007_s_at          13.0          12.4
## 1053_at            11.0           9.7
## 117_at              8.9           8.5
## 121_at              9.3          10.0
## 1255_g_at           4.5           4.7
## 1294_at             9.2           9.0
## 1316_at             6.0           6.4
## 1320_at             7.7           6.9
## 1405_i_at           7.3           8.8
## 1431_at             4.5           5.1
```

```
norm.exprs<- data.frame(exprs(gset))
norm.exprs$Symbol <-  getSYMBOL(ID,'hgu133plus2.db')
norm.exprs<-norm.exprs %>% relocate(Symbol)
write.csv(norm.exprs, 'normalized_expression.csv')
```