GEO2R

Differential expression analysis with limma

```
library(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")'.
```

```
library(GEOquery)
```

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

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

```
library(limma)
```

```
## Warning: package 'limma' was built under R version 3.5.1
```

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

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

```r
# load series and platform data from GEO

gset <- getGEO("GSE40551", GSEMatrix =TRUE, AnnotGPL=FALSE)
```

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

```
## GSE40551_series_matrix.txt.gz
```

```
## Parsed with column specification:
## cols(
##   ID_REF = col_character(),
##   GSM994975 = col_double(),
##   GSM995422 = col_double(),
##   GSM995435 = col_double(),
##   GSM995542 = col_double(),
##   GSM995703 = col_double(),
##   GSM995704 = col_double(),
##   GSM995705 = col_double(),
##   GSM995706 = col_double(),
##   GSM995711 = col_double(),
##   GSM995712 = col_double(),
##   GSM995774 = col_double(),
##   GSM995775 = col_double(),
##   GSM995776 = col_double(),
##   GSM995927 = col_double(),
##   GSM995928 = col_double(),
##   GSM995929 = col_double()
## )
```

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

```
## /var/folders/c8/0pj7fmr96n165tvvsg6s7q8w0000gn/T//RtmpbtrPYu/GPL16022.soft
```

```r
if (length(gset) > 1) idx <- grep("GPL16022", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]

# make proper column names to match toptable
fvarLabels(gset) <- make.names(fvarLabels(gset))

# group names for all samples
gsms <- "01010000011111110"
sml <- c()
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }

# log2 transform
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) ||
  (qx[2] > 0 && qx[2] < 1 && qx[4] > 1 && qx[4] < 2)
if (LogC) { ex[which(ex <= 0)] <- NaN
exprs(gset) <- log2(ex) }

write.csv(exprs(gset),file="data.csv")
# set up the data and proceed with analysis
sml <- paste("G", sml, sep="")    # set group names
```

```
fl <- as.factor(sml)
gset$description <- fl
design <- model.matrix(~ description + 0, gset)
colnames(design) <- levels(fl)
fit <- lmFit(gset, design)
```

```
## Warning: Partial NA coefficients for 1216 probe(s)
```

```
cont.matrix <- makeContrasts(G1-G0, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2, 0.01)
tT <- topTable(fit2, adjust="fdr", sort.by="B", number=Inf)

tT <- subset(tT, select=c("ID","adj.P.Val","P.Value","t","B","logFC","ORF","GB_ACC","SPOT_ID"))
write.csv(tT, file="diff_exp.csv", row.names=F, sep=",")
```

```
## Warning in write.csv(tT, file = "diff_exp.csv", row.names = F, sep = ","):
## attempt to set 'sep' ignored
```

Boxplot for selected GEO samples

```
library(Biobase)
library(GEOquery)

# load series and platform data from GEO

gset <- getGEO("GSE40551", GSEMatrix =TRUE, getGPL=FALSE)
```

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

```
## GSE40551_series_matrix.txt.gz
```

```
## Using locally cached version: /var/folders/c8/0pj7fmr96n165tvvsg6s7q8w0000gn/T//RtmpbtrPYu/GSE40551_s
```

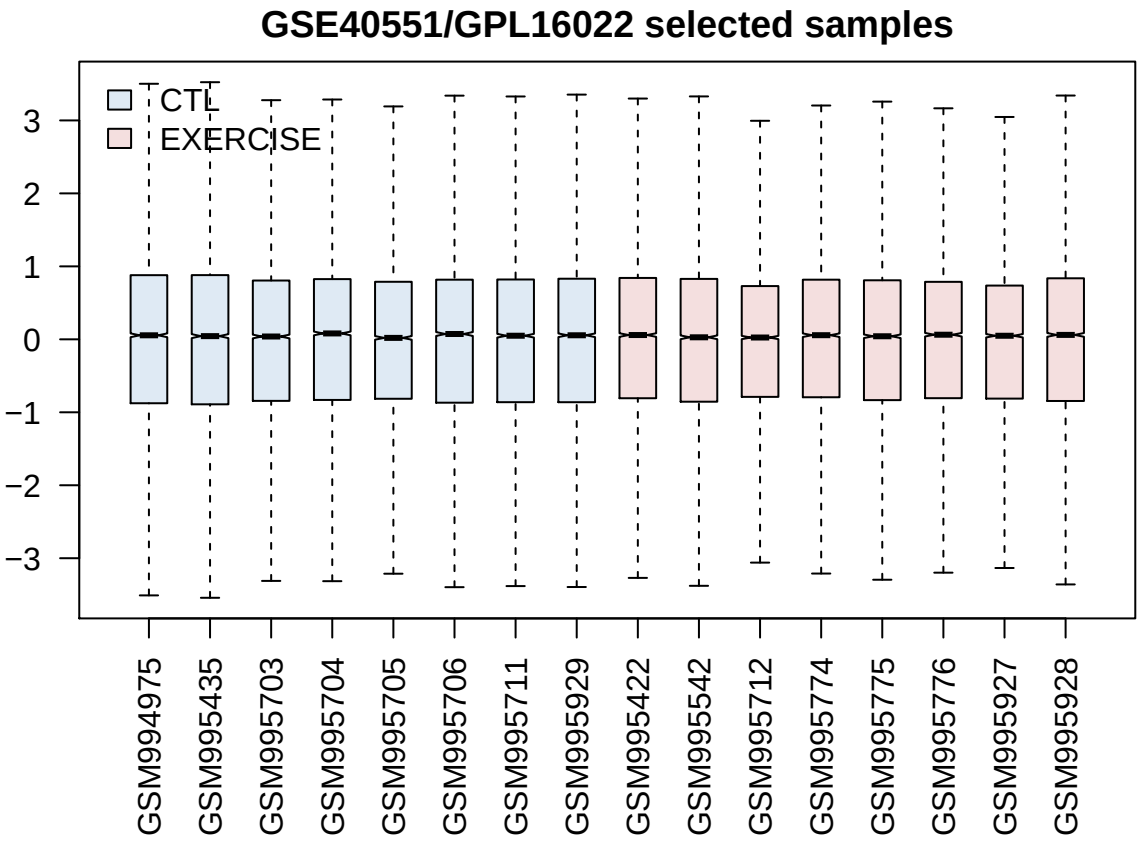
```
## Parsed with column specification:
## cols(
##   ID_REF = col_character(),
##   GSM994975 = col_double(),
##   GSM995422 = col_double(),
##   GSM995435 = col_double(),
##   GSM995542 = col_double(),
##   GSM995703 = col_double(),
##   GSM995704 = col_double(),
##   GSM995705 = col_double(),
##   GSM995706 = col_double(),
##   GSM995711 = col_double(),
##   GSM995712 = col_double(),
##   GSM995774 = col_double(),
##   GSM995775 = col_double(),
##   GSM995776 = col_double(),
##   GSM995927 = col_double(),
##   GSM995928 = col_double(),
##   GSM995929 = col_double()
## )
```

```
if (length(gset) > 1) idx <- grep("GPL16022", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]
```

```
# group names for all samples in a series
gsms <- "0101000001111110"
sml <- c()
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }
sml <- paste("G", sml, sep="")

# order samples by group
ex <- exprs(gset)[ , order(sml)]
sml <- sml[order(sml)]
fl <- as.factor(sml)
labels <- c("CTL","EXERCISE")

# set parameters and draw the plot
palette(c("#dfeaf4","#f4dfdf", "#AABBCC"))
#dev.new(width=4+dim(gset)[[2]]/5, height=6)
par(mar=c(2+round(max(nchar(sampleNames(gset)))/2),4,2,1))
title <- paste ("GSE40551", '/', annotation(gset), " selected samples", sep ='')
boxplot(ex, boxwex=0.6, notch=T, main=title, outline=FALSE, las=2, col=fl)
legend("topleft", labels, fill=palette(), bty="n")
```



HEATMAP

```
file <- read.csv("heatmap.csv",row.names = 1)
col_names <- colnames(file)
groups <- c(rep("Control",8),rep("Exercise",8))
```

```
coloursSamples <- factor(groups, levels=c("Control","Exercise"))
coloursSamples <- colorRampPalette(c("royalblue", "orange"))(length(unique(coloursSamples)))[factor(colo

library(ggplot2)
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
heatmap.2(as.matrix(file), scale="row", trace="none", Colv=F, ColSideColors=coloursSamples,dendrogram="
```

```
## Warning in heatmap.2(as.matrix(file), scale = "row", trace = "none", Colv
## = F, : Using scale="row" or scale="column" when breaks arespecified can
## produce unpredictable results.Please consider using only one or the other.
```