Plot heatmaps from expXXXX-RNAseqCounts.xlsx

SP©Nucleomics Core

9 Aug 2016

```
# load data from expXXXX-RNAseqCounts.xlsx
# generate heatmap

# dependencies. Please install them if not yet available on your machine
library("openxlsx"); # read from excel files
library("pheatmap"); # plot pretty heatmaps (different module than used at the NC)
library("RColorBrewer"); # build nice color palettes

# edit following lines to point to the folder and file
basedir <- ("~/git_repos/Nucleomics-VIB/plotting-tools/fpkm2heatmap/data")
data.file <- "expXXXX-RNAseqCounts.xlsx"

# load data from file
# the first worksheet contains raw counts while the second worksheet contains FPKM data
setwd(basedir)
excel.data <- read.xlsx(data.file, sheet=2)
last <- length(excel.data)-9

# keep only useful columns
fpkm.data <- excel.data[,c(2,1,3:last)]

# remove some columns
row.names(fpkm.data) <- paste(fpkm.data[,1], fpkm.data[,2], sep=":")
fpkm.data <- fpkm.data[,-1]

# kick useless part of names for samples
colnames(fpkm.data) <- sub("@.*", "", colnames(fpkm.data))

# custom signatures comment/uncomment line pairs to create each in turn
sig.name <- "test"
signature <- c("ENSG00000000003", "ENSG00000000005", "ENSG00000000419",
               "ENSG00000000457", "ENSG00000000460", "ENSG00000000938",
               "ENSG00000000971", "ENSG00000001036", "ENSG00000001084", "ENSG00000001167")

# select only signature rows and discard Gene.ID column to keep only FPKM in data.frame
selection <- fpkm.data[fpkm.data$Gene.ID %in% signature, 2:length(fpkm.data)]

# convert 0 to NA to avoid log-normalization error
selection[selection == 0] <- NA

# remove full NA rows
selection <- selection[rowSums(is.na(selection[,2:length(selection)]))
                       < length(selection[,2:length(selection)]),]

# prepare output
filename <- paste(sig.name,"-signature-heatmap.pdf",sep="")
outfile <- paste(basedir, filename, sep="/")
```

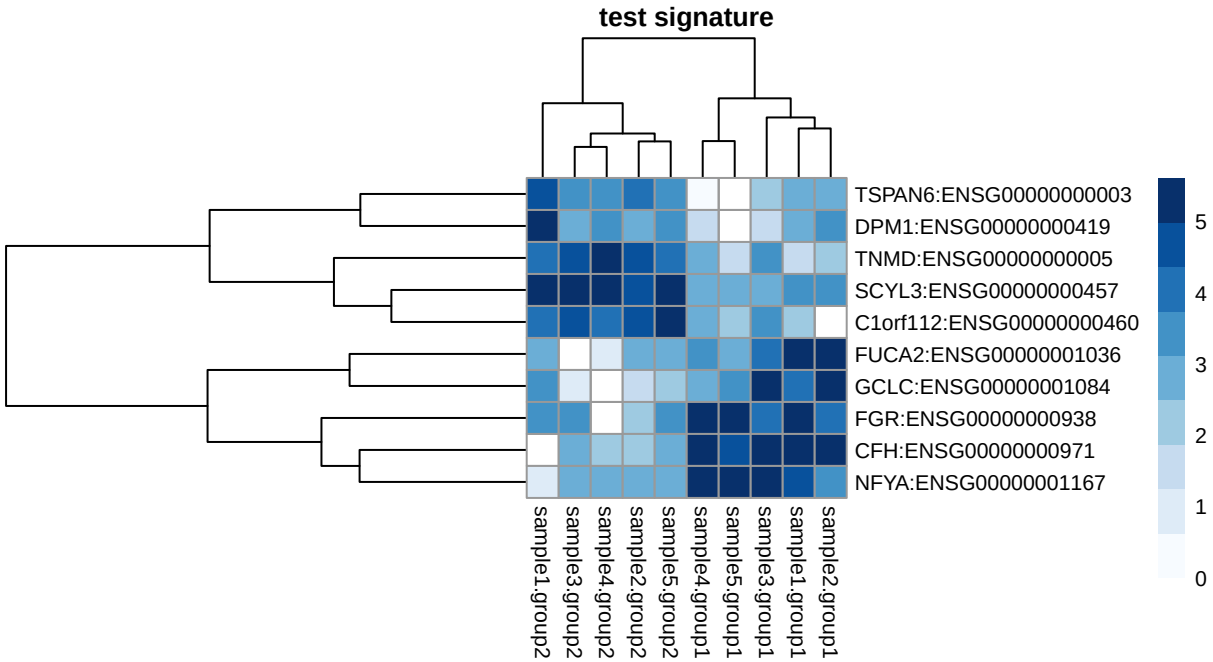
```
main.title <- paste(sig.name,"signature", sep=" ")

# color pallet
col.pal <- brewer.pal(9,"Blues")

# define metrics for clustering
drows <- "euclidean"
dcols <- "euclidean"
clustmet <- "average"

# create heatmap
# type "?pheatmap()" for more help
hm.data <- log(selection, 2)
hm.parameters <- list(hm.data,
                      color = col.pal,
                      fontsize = 8,
                      cellwidth = 12, cellheight = 12, scale = "none",
                      treeheight_row = 200,
                      kmeans_k = NA,
                      show_rownames = T,
                      show_colnames = T,
                      main = main.title,
                      clustering_method = clustmet,
                      cluster_rows = TRUE,
                      cluster_cols = TRUE,
                      clustering_distance_rows = drows,
                      clustering_distance_cols = dcols)

# To draw the heatmap on screen (comment-out if you run the script from terminal)
do.call("pheatmap", hm.parameters)
```



```
# To draw to file (you may want to adapt the sizes)
do.call("pheatmap", c(hm.parameters, filename=outfile, width=10, height=7))
```