Microarray gene expression dataset analysis.

```
###Filtering
 #colnames(pData(GSE122121))
 #head(pData(GSE122121))
 pData(GSE122121)$rep <- gsub(".*rep", "rep \\1", pData(GSE122121)$title)</pre>
 pData(GSE122121) <- pData(GSE122121)[, c("title", "infection:ch1", "genotype/variation:ch1", "rep")]</pre>
 colnames(pData(GSE122121)) <- c("Title", "Infection", "Genotype", "Replicate")</pre>
 #head(pData(GSE122121))
 #colnames(fData(GSE122121))
 fData(GSE122121) <- fData(GSE122121)[, c( "Gene symbol","ID", "Gene ID")]</pre>
```

Normalization

```
7500 -
5000 -
   2500 -
                                                                    12
                                                 expression
```

exprs(GSE122121) <- normalizeBetweenArrays(exprs(GSE122121),</pre>

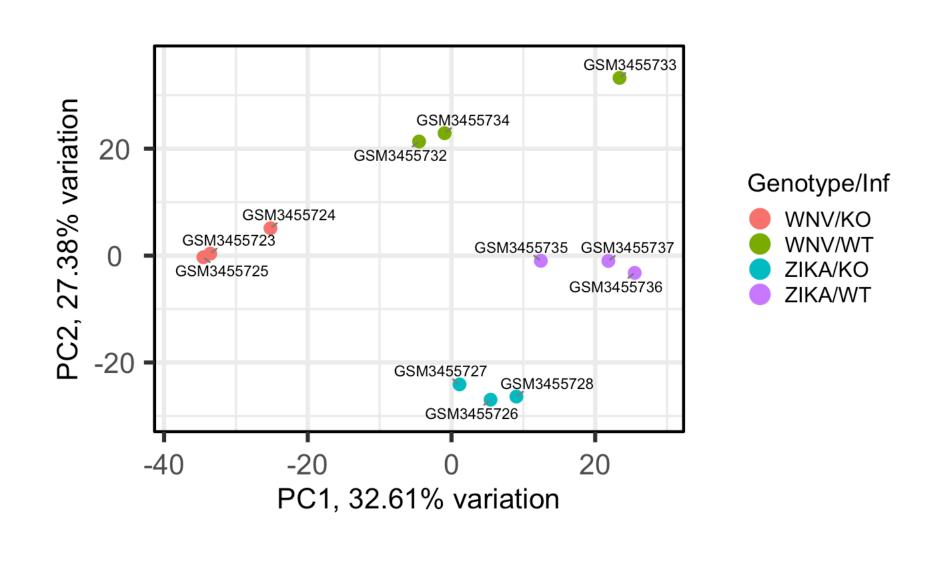
```
method="quantile")#head(exprs(GSE122121))
##Filtering
 GSE122121 <- GSE122121[!grepl("///", fData(GSE122121)$`Gene symbol`), ]
 GSE122121 <- GSE122121[fData(GSE122121)$`Gene symbol` != "", ]
 fData(GSE122121)$mean expression <- apply(exprs(GSE122121), 1, mean)</pre>
 GSE122121 <- GSE122121[order(fData(GSE122121)$mean_expression, decreasing = TRUE), ]
 GSE122121 <- GSE122121[!duplicated(fData(GSE122121)$`Gene ID`), ]
 GSE122121 <-GSE122121[,pData(GSE122121)$`Infection`!="PBS"]
```

```
#pData(GSE122121)$Infection
GSE122121 \leftarrow GSE122121[seq len(12000), ]
dim(GSE122121)
## Features Samples
```

PCA

12000

```
for (i in c(1:12)) {
 if (pData(GSE122121)[i, "Genotype"] == 'wild type' && pData(GSE122121)[i, "Infection"] == 'Zika virus (ZIKV) strain M
R766 at an MOI of 0.1') {
    pData(GSE122121)[i, "Genotype/Inf"]="ZIKA/WT"}
   if (pData(GSE122121)[i, "Genotype"]!='wild type' && pData(GSE122121)[i, "Infection"]=='Zika virus (ZIKV) strain
MR766 at an MOI of 0.1') {
    pData(GSE122121)[i, "Genotype/Inf"]="ZIKA/KO"}
      if (pData(GSE122121)[i, "Genotype"] == 'wild type' && pData(GSE122121)[i, "Infection"]! = 'Zika virus (ZIKV) stra
in MR766 at an MOI of 0.1') {
    pData(GSE122121)[i, "Genotype/Inf"]="WNV/WT"}
      if (pData(GSE122121)[i, "Genotype"]!='wild type' && pData(GSE122121)[i, "Infection"]!='Zika virus (ZIKV) stra
in MR766 at an MOI of 0.1') {
    pData(GSE122121)[i, "Genotype/Inf"]="WNV/KO"}
#pData(GSE122121)
pcaData <- pca(exprs(GSE122121), metadata=pData(GSE122121))</pre>
biplot(pcaData, colby ="Genotype/Inf", legendPosition = "right")
```



```
##Create matrix and DE
```

fit2 <- contrasts.fit(fit, makeContrasts(Ripk3 - wild_type, levels=full_model))</pre>

4 groups observed on PCA plot instead of 2, caused by different virus types.

We see

```
#pData(GSE122121)
```

```
full_model <- model.matrix(~0 + Genotype + Infection, data = pData(GSE122121))</pre>
colnames(full model) <- c("Ripk3", "wild type", "Zika")</pre>
full_model
              Ripk3 wild_type Zika
## GSM3455723
## GSM3455724
## GSM3455725
## GSM3455726
## GSM3455727
```

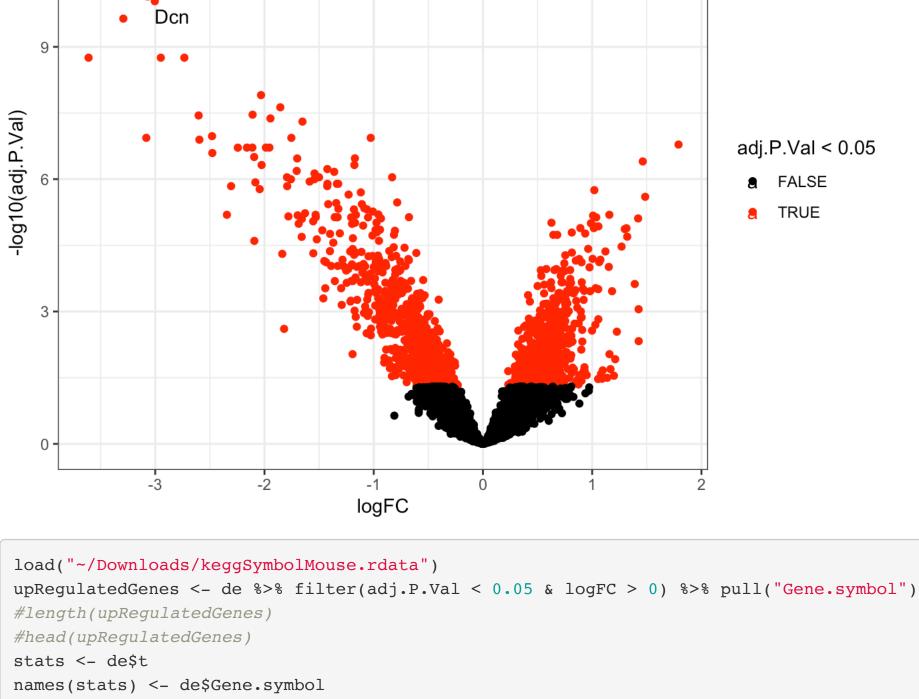
```
## GSM3455728
## GSM3455732
## GSM3455733
## GSM3455734
## GSM3455735
## GSM3455736
## GSM3455737
## attr(,"assign")
## [1] 1 1 2
## attr(,"contrasts")
## attr(,"contrasts")$Genotype
## [1] "contr.treatment"
## attr(,"contrasts")$Infection
## [1] "contr.treatment"
fit <- lmFit(GSE122121, full_model)</pre>
```

```
fit2 <- eBayes(fit2, trend = T)</pre>
de <- topTable(fit2, adjust.method="BH"</pre>
, number=Inf, sort.by = "P")
head(de$Gene.symbol,30)
                                "Col3a1"
                                                        "Vtn"
                                                                    "Dcn"
   [1] "Col1a2"
                    "Apod"
                                            "Mgp"
   [7] "Colla1"
                    "Nid1"
                                "Cxcl12"
                                            "Igf2"
                                                        "Bgn"
                                                                    "Hoxa3"
                                                                    "Myl1"
## [13] "Aldh1a2"
                    "Slc6a13"
                                "Lamb1"
                                            "H19"
                                                        "Ogn"
## [19] "Tenm1"
                    "Ptprd"
                                "Mylpf"
                                            "Erdr1"
                                                        "Col5a1"
                                                                    "Cga"
## [25] "Pcolce"
                    "Serpinf1" "Gpc3"
                                            "Tnnc2"
                                                        "Itm2a"
                                                                    "Fat4"
```

ggplot(de, aes(x=logFC, y=-log10(adj.P.Val), color=adj.P.Val < 0.05)) +</pre> geom_point() + theme_bw() + scale_color_manual(values=c("black", "red")) + geom_text_repel(data=de %>% dplyr::filter(adj.P.Val < 1e-10), aes(label=Gene.symbol, color=NULL))</pre>

##Volkano Plot

```
12 Col1a2 Apod
     Col3a1
  Mgp
Vtn
```



leadingEdge

```
fgseaResults <- fgseaMultilevel(keggSymbolMouse, stats, minSize = 15, maxSize = 500)</pre>
head(fgseaResults, 3)
##
                                                                          pathway
## 1:
                                         ABC transporters - Mus musculus (mouse)
## 2: AGE-RAGE signaling pathway in diabetic complications - Mus musculus (mouse)
## 3:
                                   AMPK signaling pathway - Mus musculus (mouse)
                          padj log2err
                                                           NES size
                                                 ES
             pval
## 1: 2.558140e-01 0.4443084455 0.1301056 0.3713572 1.146222 24
## 2: 1.097427e-05 0.0003621509 0.5933255 -0.5506316 -2.082490 81
## 3: 3.815029e-01 0.5855625756 0.1172497 0.2370674 1.024453 103
```

2: Col1a2, Col3a1, Col1a1, Thbd, Col4a1, Col4a2,... ## 3: Cpt1a,G6pc3,Prkag2,Foxo1,Ppp2r3a,Srebf1,...

Abca1, Abcg1, Abcc3, Abcb8, Abca2, Abca5, ...

```
###Top Pathways
 topPathwaysUp <- fgseaResults[ES > 0, ][head(order(pval), n=5), pathway]
 topPathwaysDown <- fgseaResults[ES < 0, ][head(order(pval), n=5), pathway]</pre>
 topPathways <- c(topPathwaysUp, rev(topPathwaysDown))</pre>
```

topPathwaysUp

[1] "Lysosome - Mus musculus (mouse)"

#Top PathwaysUP

##

1:

```
## [2] "Fatty acid degradation - Mus musculus (mouse)"
 ## [3] "Peroxisome - Mus musculus (mouse)"
 ## [4] "Oxidative phosphorylation - Mus musculus (mouse)"
 ## [5] "PPAR signaling pathway - Mus musculus (mouse)"
#Top PathwaysDown
 topPathwaysDown
```

```
## [1] "ECM-receptor interaction - Mus musculus (mouse)"
## [2] "Protein digestion and absorption - Mus musculus (mouse)"
## [3] "Axon guidance - Mus musculus (mouse)"
## [4] "Amoebiasis - Mus musculus (mouse)"
      "PI3K-Akt signaling pathway - Mus musculus (mouse)"
plotGseaTable(keggSymbolMouse[topPathways], stats, fgseaResults, gseaParam = 0.5)
```

```
Pathway
           Gene ranks
                NES
                     padj
                  pval
```

```
Peroxisome - Mus musculus (mouse)
hosphorylation - Mus musculus (mouse)
naling pathway - Mus musculus (mouse)
naling pathway - Mus musculus (mouse)
```

Amoebiasis - Mus musculus (mouse)

Axon guidance - Mus musculus (mouse)

cid degradation - Mus musculus (mouse) | | 2.43 | 3.9·10⁻⁸ | 3.5·10⁻⁶

```
and absorption - Mus musculus (mouse) \parallel - \parallel
ptor interaction - Mus musculus (mouse) | | | | | | | | | | -2.42 | 2.1·10<sup>-8</sup> | 2.7·10<sup>-6</sup>
```

3000 6000 9000 12000