

Differential expression analysis between high-sugar diet vs control in fruit fly

Shuo Zhang

December 21, 2022

Load Libraries

```
library(ggplot2)
library(DESeq2)

## Warning: package 'DESeq2' was built under R version 4.3.1
## Loading required package: S4Vectors
## Loading required package: stats4
## 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, aperm, 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
##
## 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
## Warning: package 'IRanges' was built under R version 4.3.1
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
```

```

## Warning: package 'GenomeInfoDb' was built under R version 4.3.1
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Warning: package 'MatrixGenerics' was built under R version 4.3.1
## Loading required package: matrixStats
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##   colAlls, colAnyNAs, colAnys, colAvgPerRowSet, colCollapse,
##   colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##   colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##   colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##   colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##   colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##   colWeightedMeans, colWeightedMedians, colWeightedSds,
##   colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgPerColSet,
##   rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##   rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##   rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##   rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##   rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##   rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##   rowWeightedSds, rowWeightedVars
## 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)".
##
## Attaching package: 'Biobase'
## The following object is masked from 'package:MatrixGenerics':
##
##   rowMedians
## The following objects are masked from 'package:matrixStats':
##
##   anyMissing, rowMedians
library(EnhancedVolcano)

## Loading required package: ggrepel

```

read count table and sample information

```

cts <- read.table('../formatted_HSD_featureCounts.txt', header = TRUE, row.names = 1)
#keep <- rowSums(cts >= 10) >= 3

```

```
#cts <- cts[keep,]
head(cts)
```

```
##           F1_1 F1_2 F1_3 ND_1 ND_2 ND_3
## cac        5002 5055 5098 4598 5077 4850
## Cngl        4590 4478 4353 4464 4294 4531
## CG11836      82   82   81   68   79   70
## CG33096      25   22   27   17   32   28
## bam          0    0    0    0    0    0
## sisRNA:CR46364 15   21   18   16   20   14
```

read sample information

```
coldata <- read.table('../coldata.txt', header = TRUE, row.names = 1)
coldata$condition <- factor(coldata$condition, levels = c("ND", "F1"))
coldata
```

```
##      condition
## F1_1        F1
## F1_2        F1
## F1_3        F1
## ND_1        ND
## ND_2        ND
## ND_3        ND
```

quick start

```
dds <- DESeqDataSetFromMatrix(countData = cts,
                              colData = coldata,
                              design= ~ condition)
```

```
dds <- DESeq(dds)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
## fitting model and testing
```

```
resultsNames(dds) # lists the coefficients
```

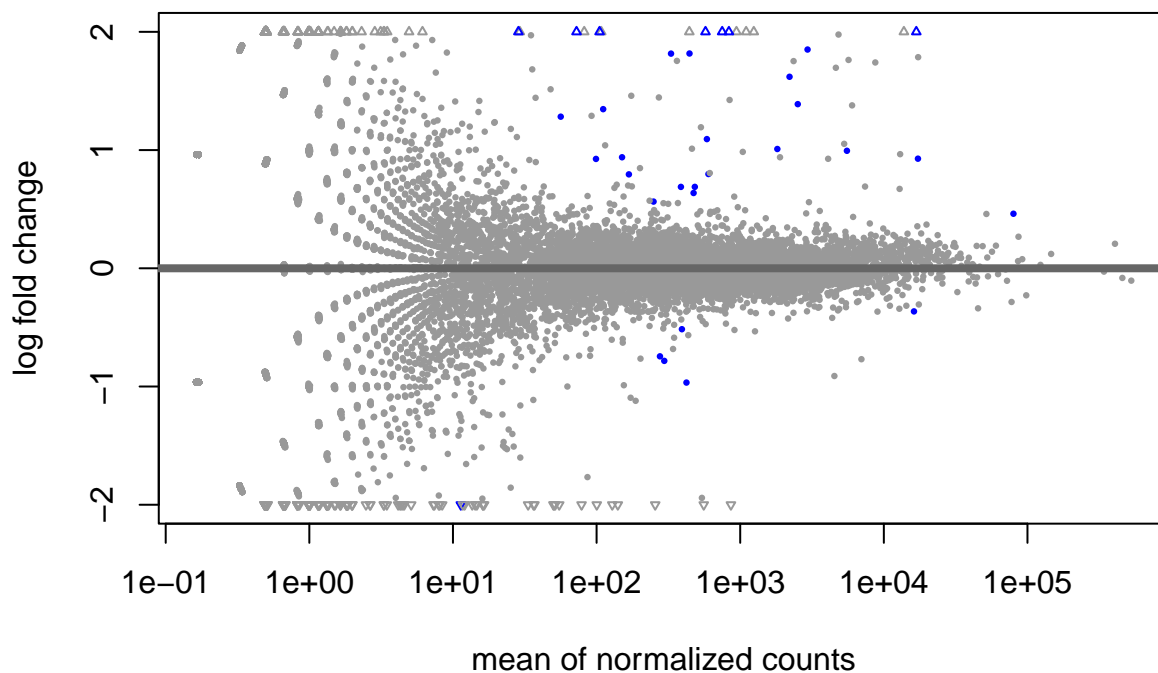
```
## [1] "Intercept"          "condition_F1_vs_ND"
```

```
# or to shrink log fold changes association with condition:
```

```
#res <- lfcShrink(dds, coef="condition_trt_vs_untrt", type="apeglm")
```

MA-plot

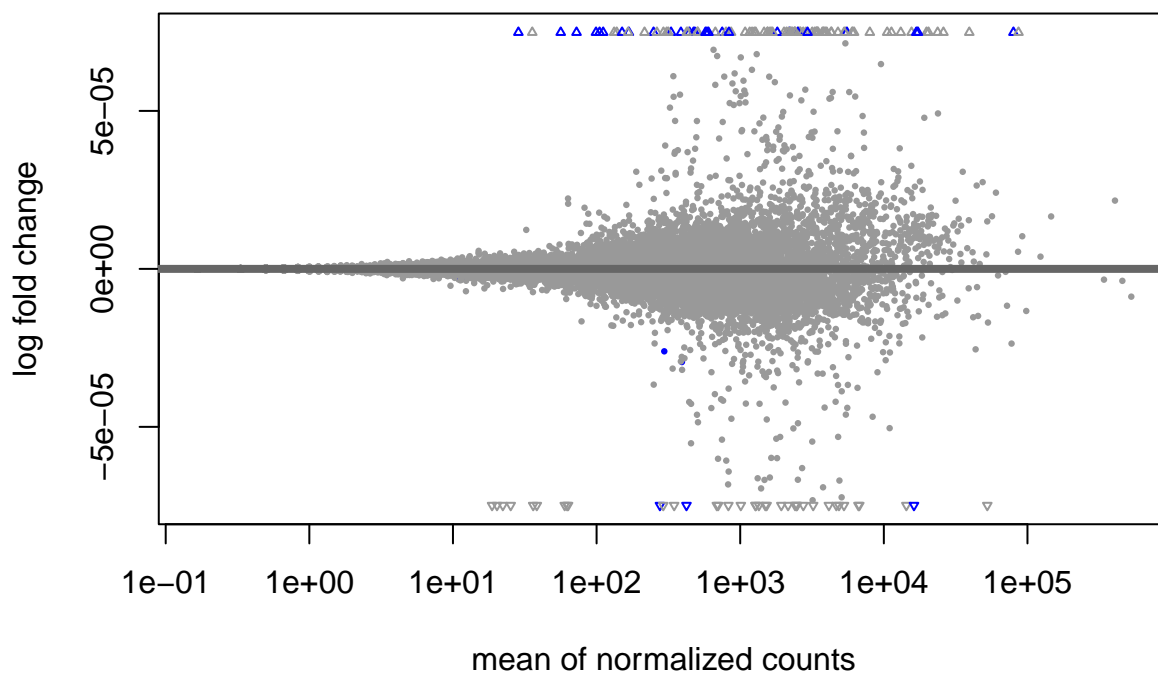
```
res <- results(dds)
plotMA(res, ylim=c(-2,2))
```



```
resLFC <- lfcShrink(dds, coef="condition_F1_vs_ND", type="apeglm")
```

```
## using 'apeglm' for LFC shrinkage. If used in published research, please cite:
##   Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for
##   sequence count data: removing the noise and preserving large differences.
##   Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895
```

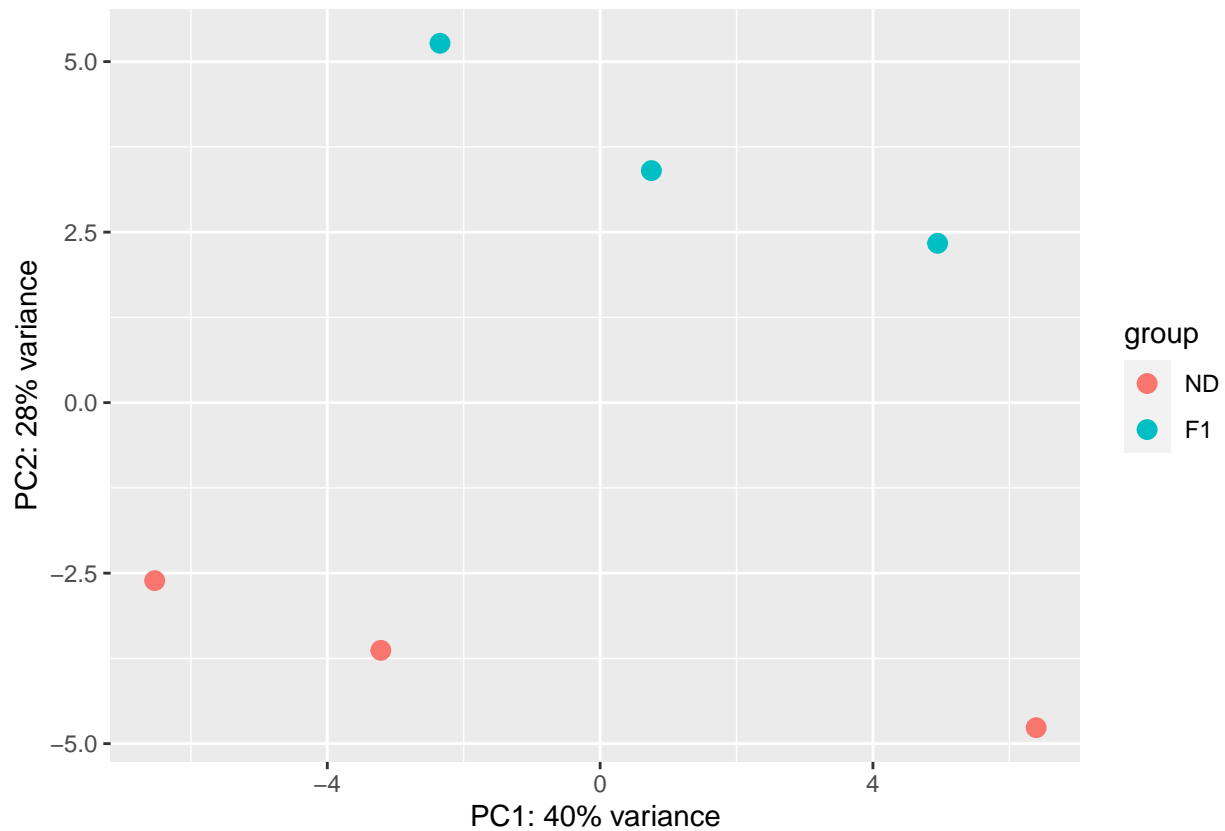
```
resLFC.Ordered <- resLFC[order(resLFC$pvalue),]
plotMA(resLFC)
```



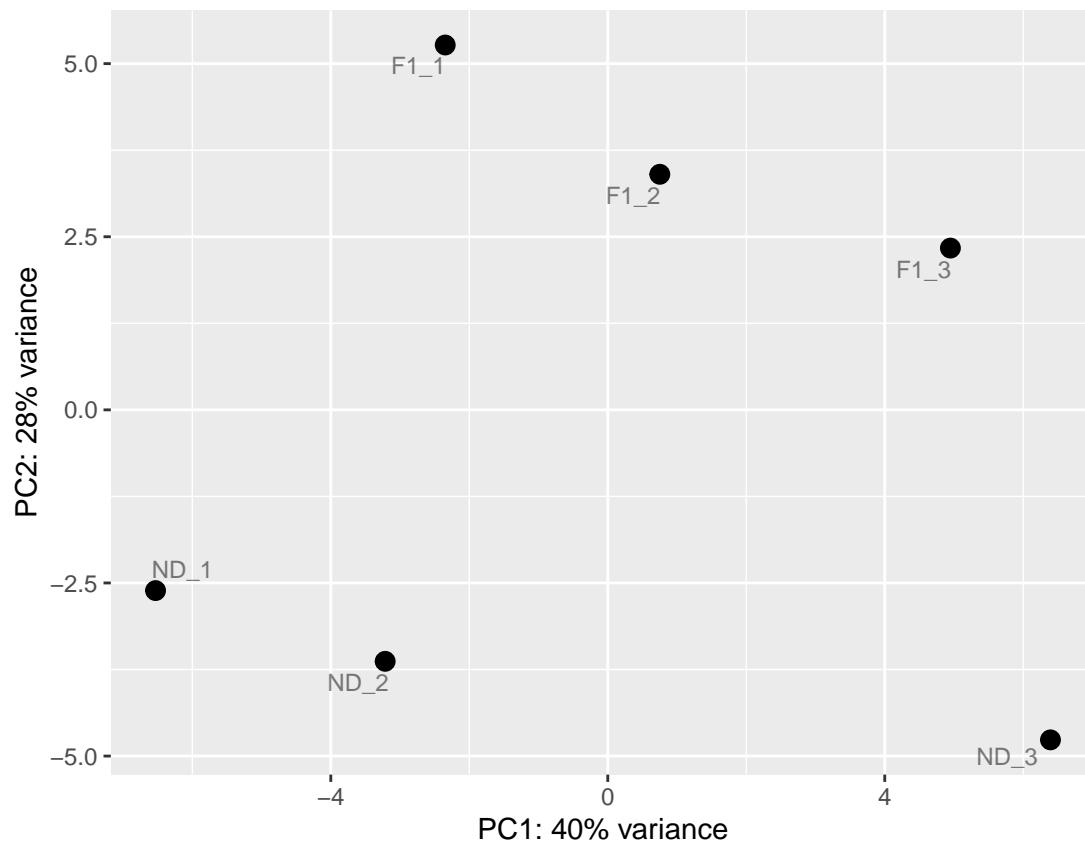
PCA

#

```
rld <- rlog(dds, blind = FALSE)
plotPCA(rld, intgroup=c("condition"))
```



```
pcaData <- plotPCA(rld, intgroup = c("condition"), returnData=TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))
ggplot(pcaData, aes(PC1, PC2, label = name)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  coord_fixed() + geom_text_repel(alpha=0.5, size=3)
```

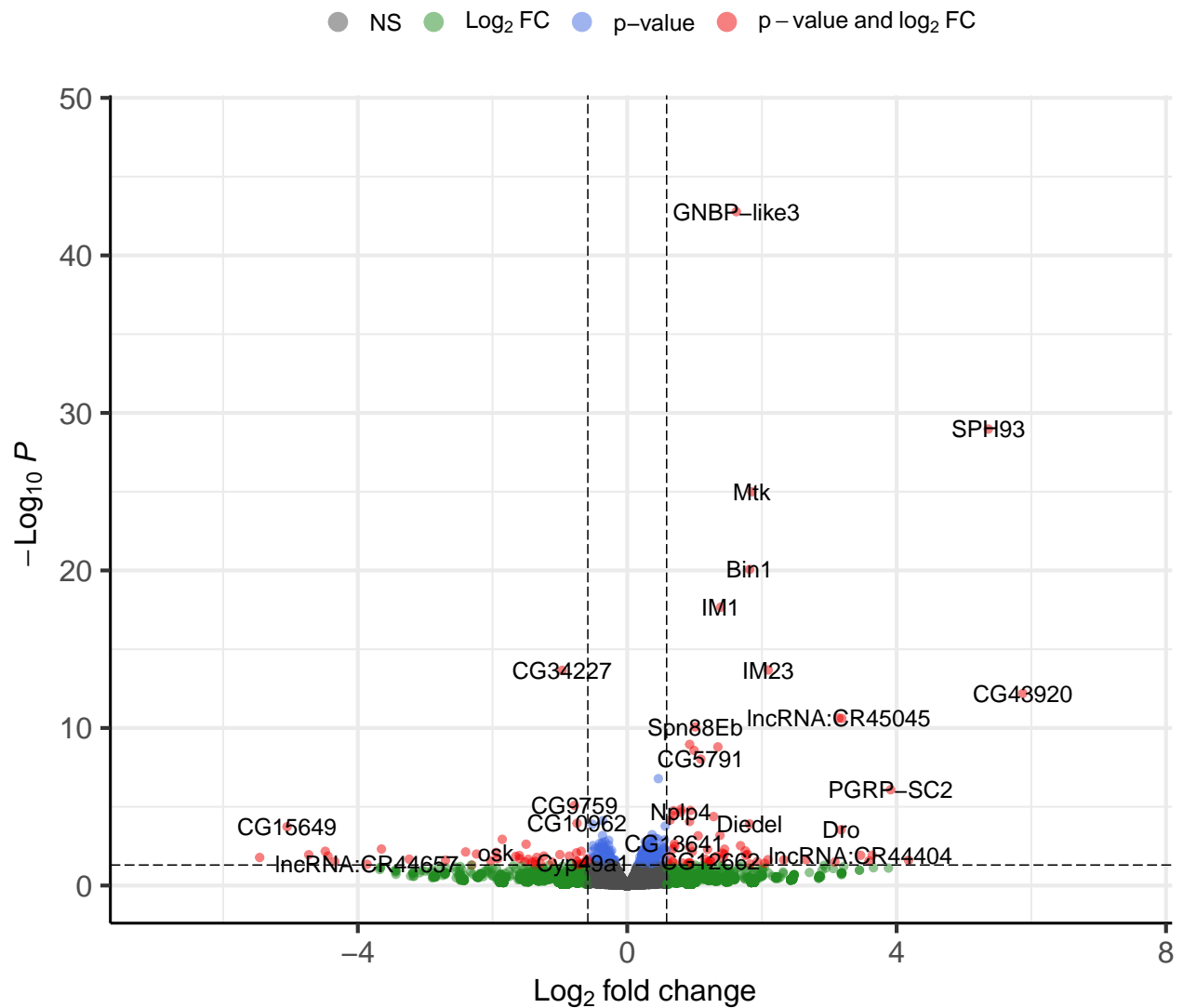


volcano plot

```
EnhancedVolcano(res,  
  lab = rownames(res),  
  x = 'log2FoldChange',  
  y = 'pvalue',  
  pCutoff = 0.05,  
  FCcutoff = 0.5849625  
)
```

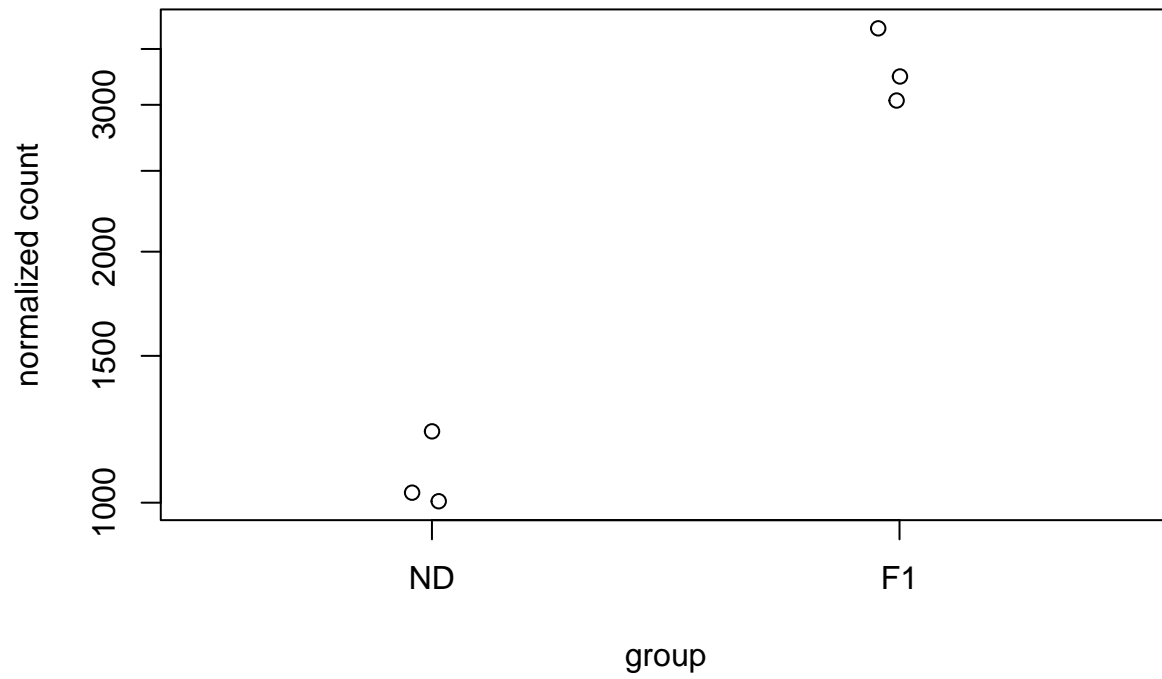
Volcano plot

EnhancedVolcano



```
plotCounts(dds, gene=which.min(res$padj), intgroup="condition")
```

GGBP-like3



```
plotCounts(dds, gene='cad', intgroup="condition")
```

cad

