

Project 3 Dataset Analysis

First Step is to read in data

```
# DESeq2 readcounts
#-----
norm <- DESeqDataSetFromMatrix(countData <- merged, colData <- sampleinfo[, c(2,
  5)], design = ~Condition, tidy = TRUE)
```

Apply DESeq

```
head(reg)
```

```
## class: DESeqDataSet
## dim: 6 24
## metadata(1): version
## assays(4): counts mu H cooks
## rownames(6): Gna13 Pbsn ... Scml2 Apoh
## rowData names(30): baseMean baseVar ... deviance maxCooks
## colnames(24): let7b_KO_M2_2 let7b_KO_M1_1 ... let7b_WT_M2_1
##   let7b_WT_M2_4
## colData names(3): Sample_ID Condition sizeFactor
```

Add factor columns for WT/KO and M1/M2

```
## DataFrame with 24 rows and 5 columns
##           Sample_ID Condition sizeFactor    cellType      factor
##           <character> <factor> <numeric> <character> <character>
## 1 let7b_KO_M2_2 let7b_KO_M2_2    KO_M2  1.320879      M2      KO
## 2 let7b_KO_M1_1 let7b_KO_M1_1    KO_M1  0.876731      M1      KO
## 3 let7b_WT_M1_5 let7b_WT_M1_5    WT_M1  0.604094      M1      WT
## 4 let7b_KO_M2_5 let7b_KO_M2_5    KO_M2  0.844488      M2      KO
## 5 let7b_KO_M1_5 let7b_KO_M1_5    KO_M1  1.374687      M1      KO
## ...
##           ...
## 11 let7b_KO_M1_3 let7b_KO_M1_3    KO_M1  1.111280      M1      KO
## 12 let7b_KO_M1_2 let7b_KO_M1_2    KO_M1  1.274230      M1      KO
## 13 let7b_WT_M1_3 let7b_WT_M1_3    WT_M1  1.229895      M1      WT
## 14 let7b_WT_M2_1 let7b_WT_M2_1    WT_M2  1.045353      M2      WT
## 15 let7b_WT_M2_4 let7b_WT_M2_4    WT_M2  0.670609      M2      WT
```

Results

```
summary(res)
```

```

## 
## out of 33260 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)      : 3983, 12%
## LFC < 0 (down)    : 3780, 11%
## outliers [1]       : 115, 0.35%
## low counts [2]     : 15417, 46%
## (mean count < 1)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

```

Order by Padj scores

```

res <- res[order(res$padj),]
head(res)

## log2 fold change (MLE): Condition WT M2 vs KO M1
## Wald test p-value: Condition WT M2 vs KO M1
## DataFrame with 6 rows and 6 columns
##           baseMean log2FoldChange      lfcSE      stat     pvalue     padj
##           <numeric>      <numeric> <numeric> <numeric> <numeric> <numeric>
## Mx1      2161.22      -4.80643 0.1176778  -40.8440      0      0
## Hmox1   17193.08      -3.31714 0.0723212  -45.8668      0      0
## Sod2     7863.59      -3.84441 0.0784830  -48.9840      0      0
## Cybb    9760.80      -3.36486 0.0887481  -37.9147      0      0
## Fcgr1   3437.05      -3.55337 0.0913204  -38.9111      0      0
## Cd274   2358.74      -4.30040 0.1104621  -38.9310      0      0

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

```

```
## mean-dispersion relationship  
## final dispersion estimates  
## fitting model and testing
```

Volcano Plot

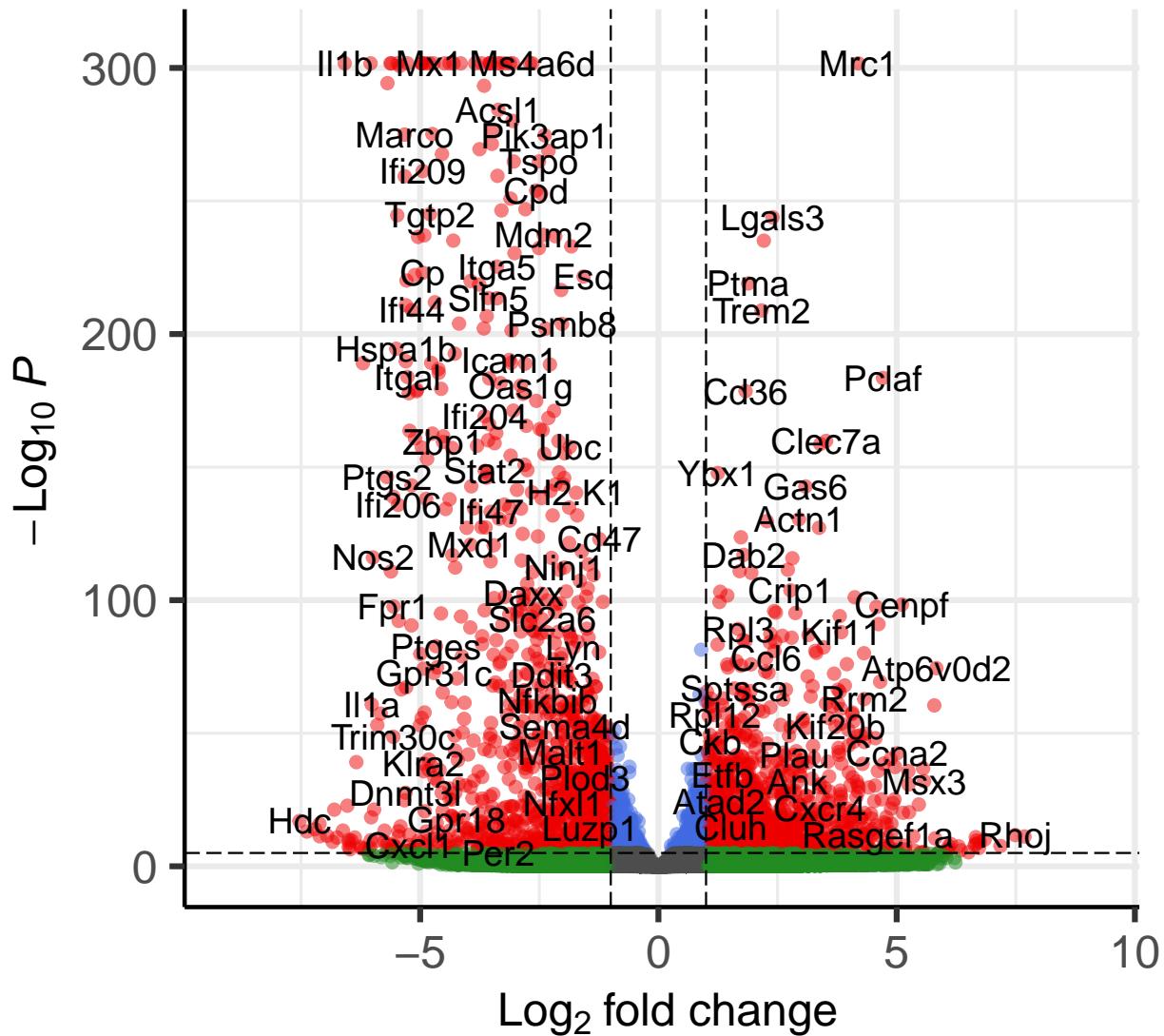
```
#All  
EnhancedVolcano(res, lab=rownames(res), x = "log2FoldChange", y = "pvalue" )
```

```
## Warning: One or more p-values is 0. Converting to 10^-1 * current lowest non-  
## zero p-value...
```

Volcano plot

EnhancedVolcano

● NS ● Log₂ FC ● p-value ● p-value and log₂ FC



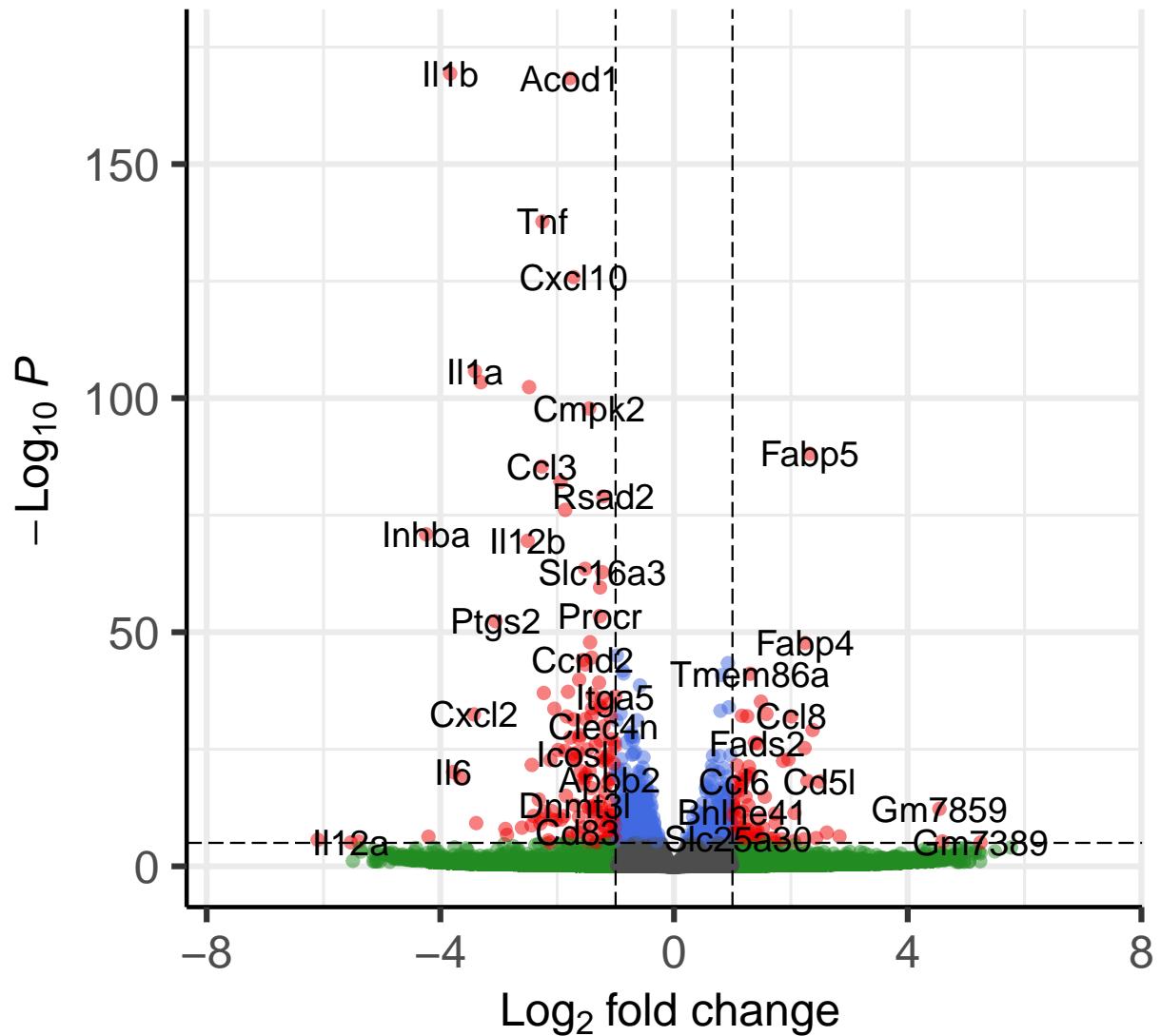
```
#M1  
EnhancedVolcano(mires, lab=rownames(m1res), x = "log2FoldChange", y = "pvalue" )
```

```
## Warning: One or more p-values is 0. Converting to 10^-1 * current lowest non-  
## zero p-value...
```

Volcano plot

EnhancedVolcano

● NS ● Log₂ FC ● p-value ● p-value and log₂ FC



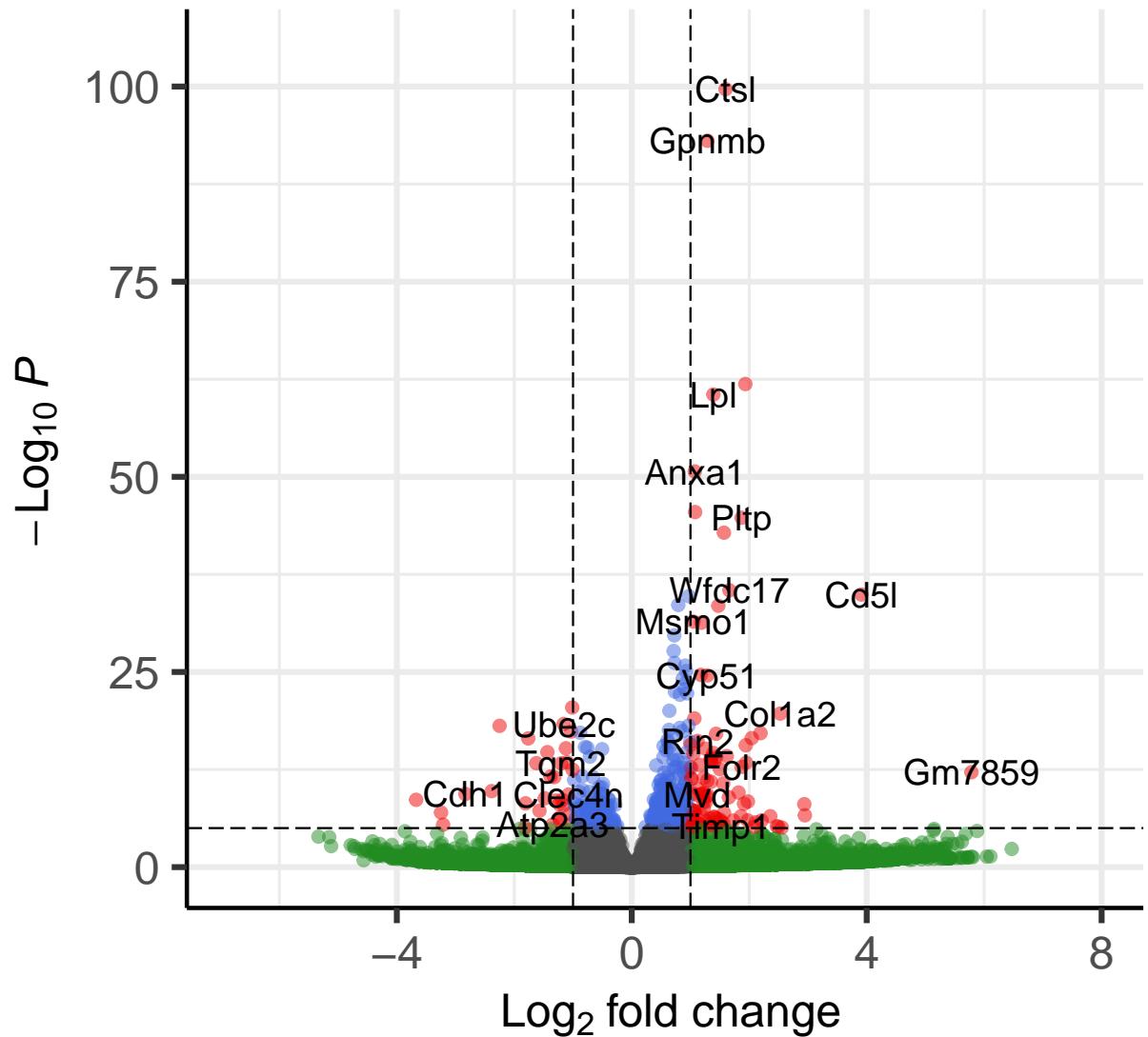
total = 33260 variables

```
#M2  
EnhancedVolcano(m2res, lab=rownames(m2res), x = "log2FoldChange", y = "pvalue" )
```

Volcano plot

EnhancedVolcano

● NS ● Log₂ FC ● p-value ● p-value and log₂ FC



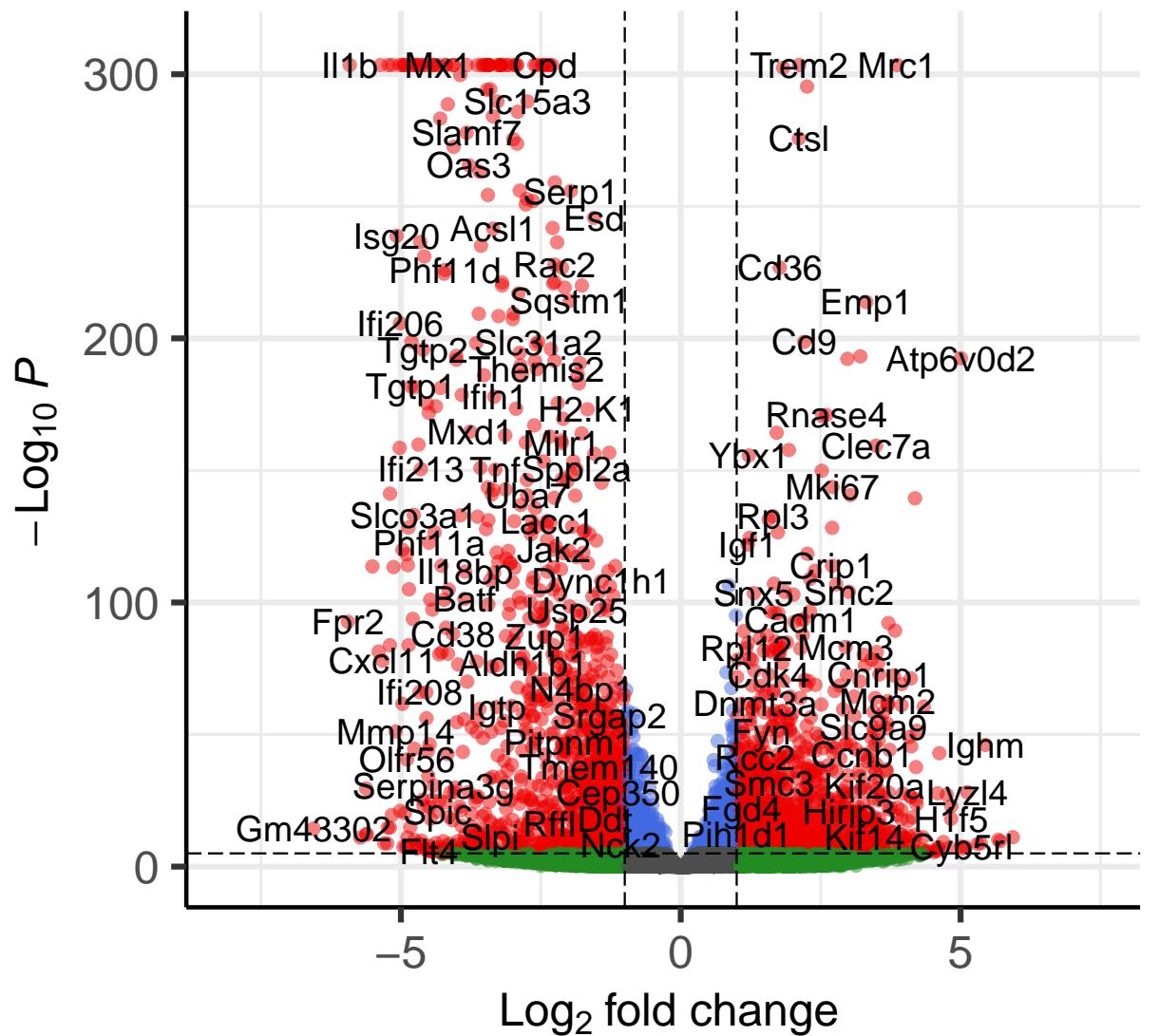
```
#UMI  
EnhancedVolcano(resUMI, lab=rownames(resUMI), x = "log2FoldChange", y = "pvalue" )
```

```
## Warning: One or more p-values is 0. Converting to 10^-1 * current lowest non-  
## zero p-value...
```

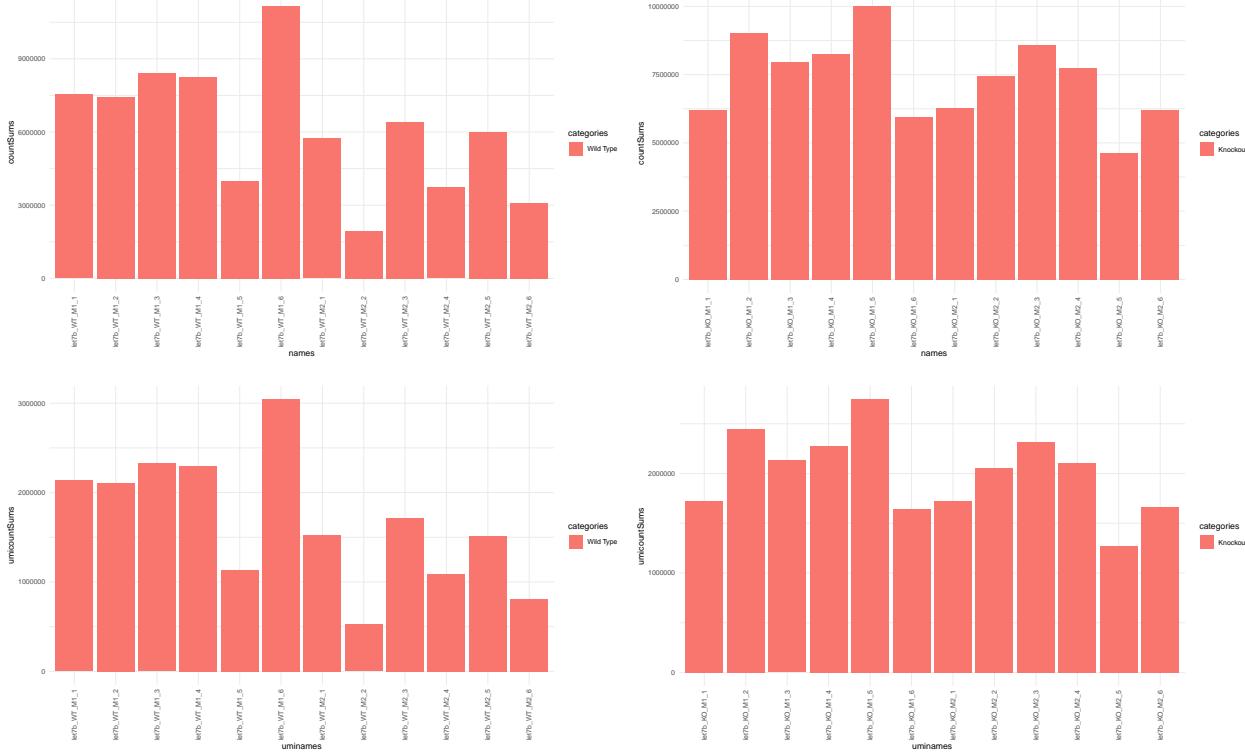
Volcano plot

Enhanced Volcano

● NS ● Log₂ FC ● p-value ● p-value and log₂ FC



Plot count distributions

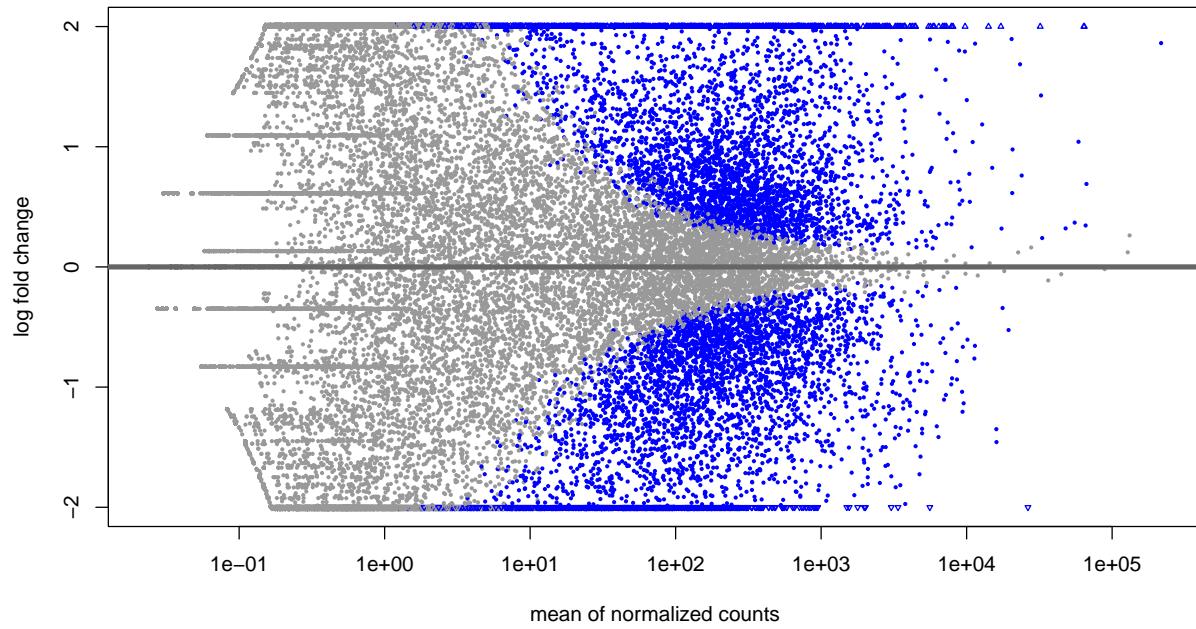


```
## using 'apeglm' for LFC shrinkage. If used in published research, please cite:
##      Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for
##      sequence count data: removing the noise and preserving large differences.
##      Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895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

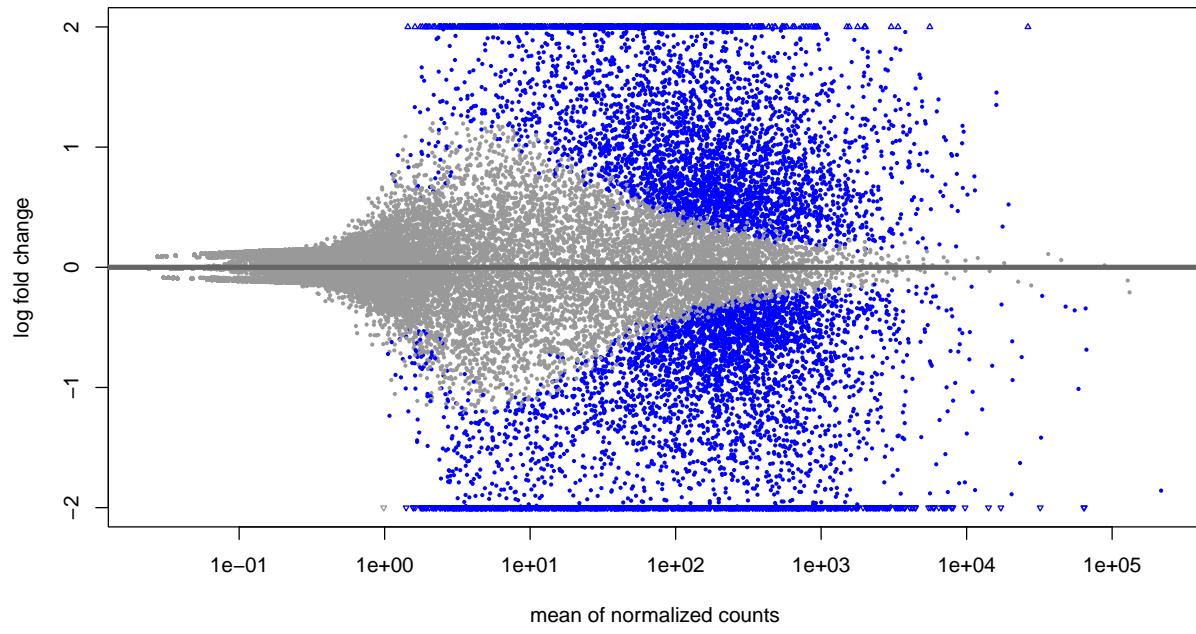
```
## using 'normal' for LFC shrinkage, the Normal prior from Love et al (2014).
##
## Note that type='apeglm' and type='ashr' have shown to have less bias than type='normal'.
## See ?lfcShrink for more details on shrinkage type, and the DESeq2 vignette.
## Reference: https://doi.org/10.1093/bioinformatics/bty895
```

MA Plots

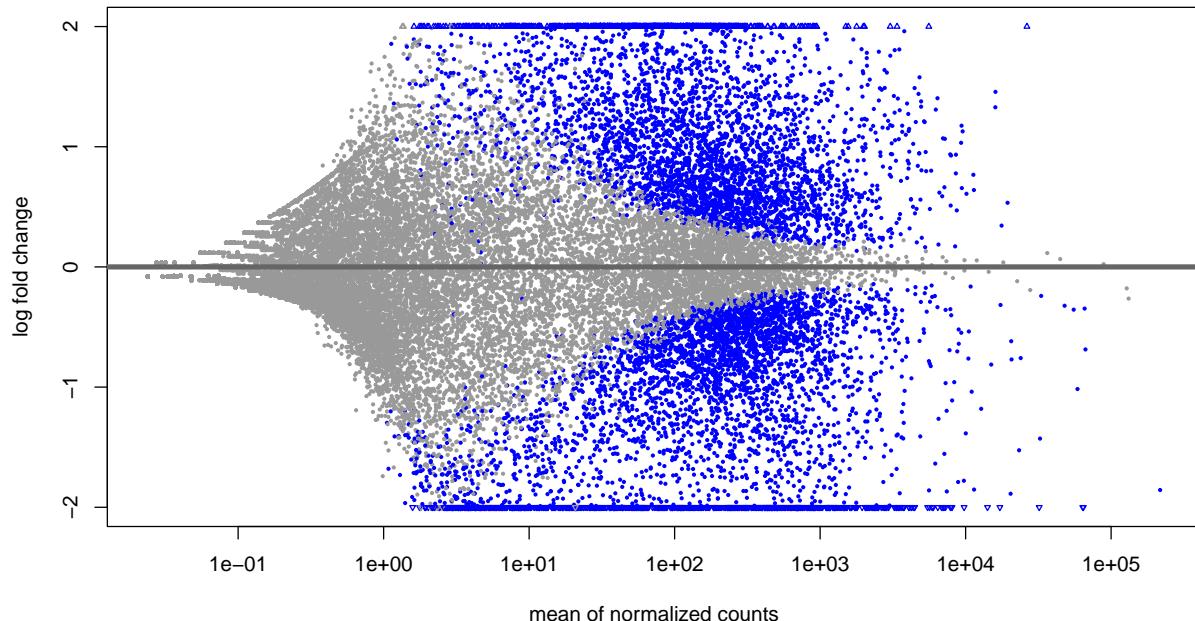
```
#Base MA plot
#-----
Ma <- plotMA(res, ylim = c( -2, 2))
```



```
#Apeglm shrinkage estimator
#-----
Ma <- plotMA(resLFC, ylim = c( -2, 2))
```



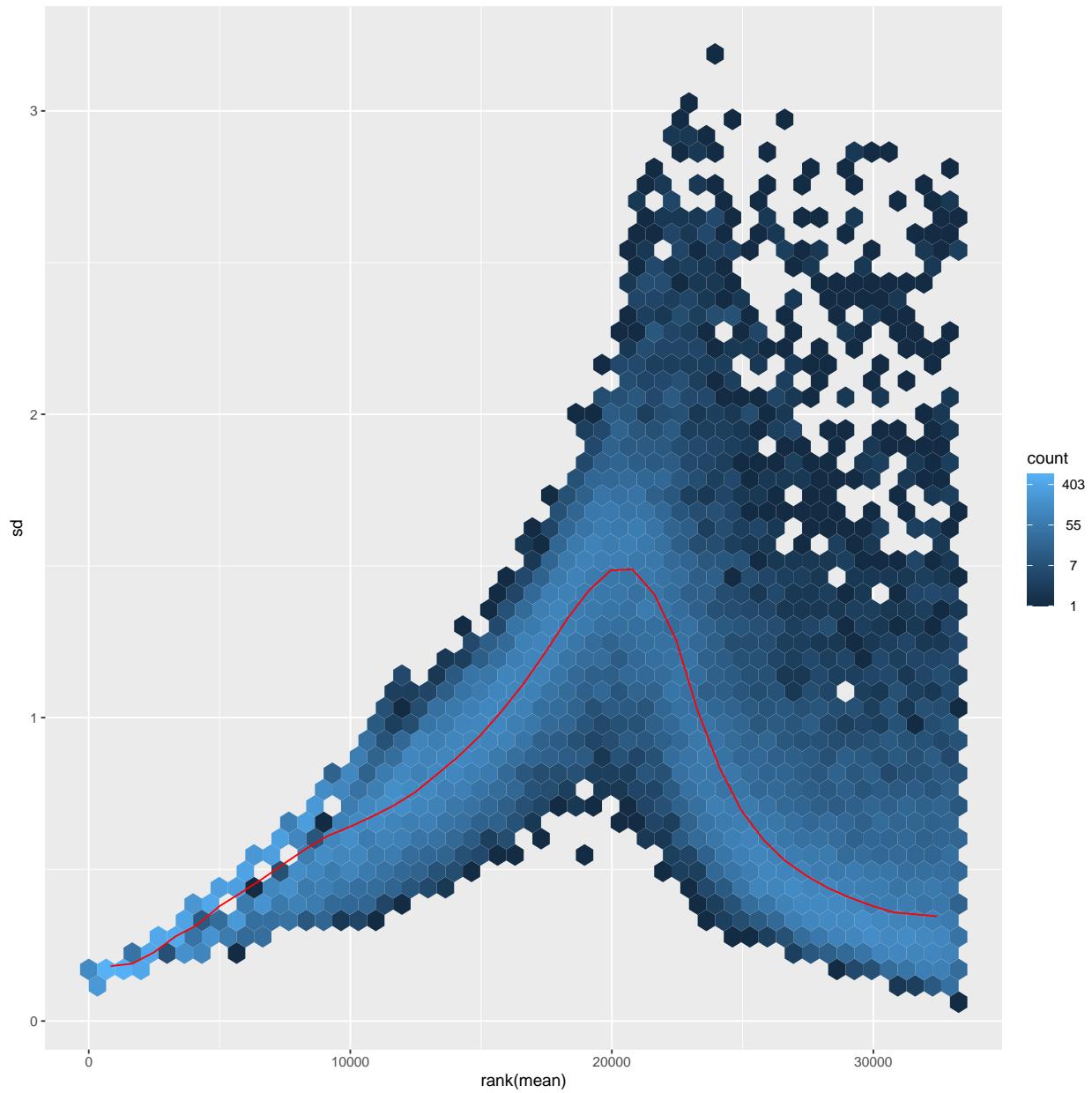
```
#DESeq2 normal shrinkage estimator
#-----
Ma <- plotMA(resNorm, ylim = c( -2, 2))
```



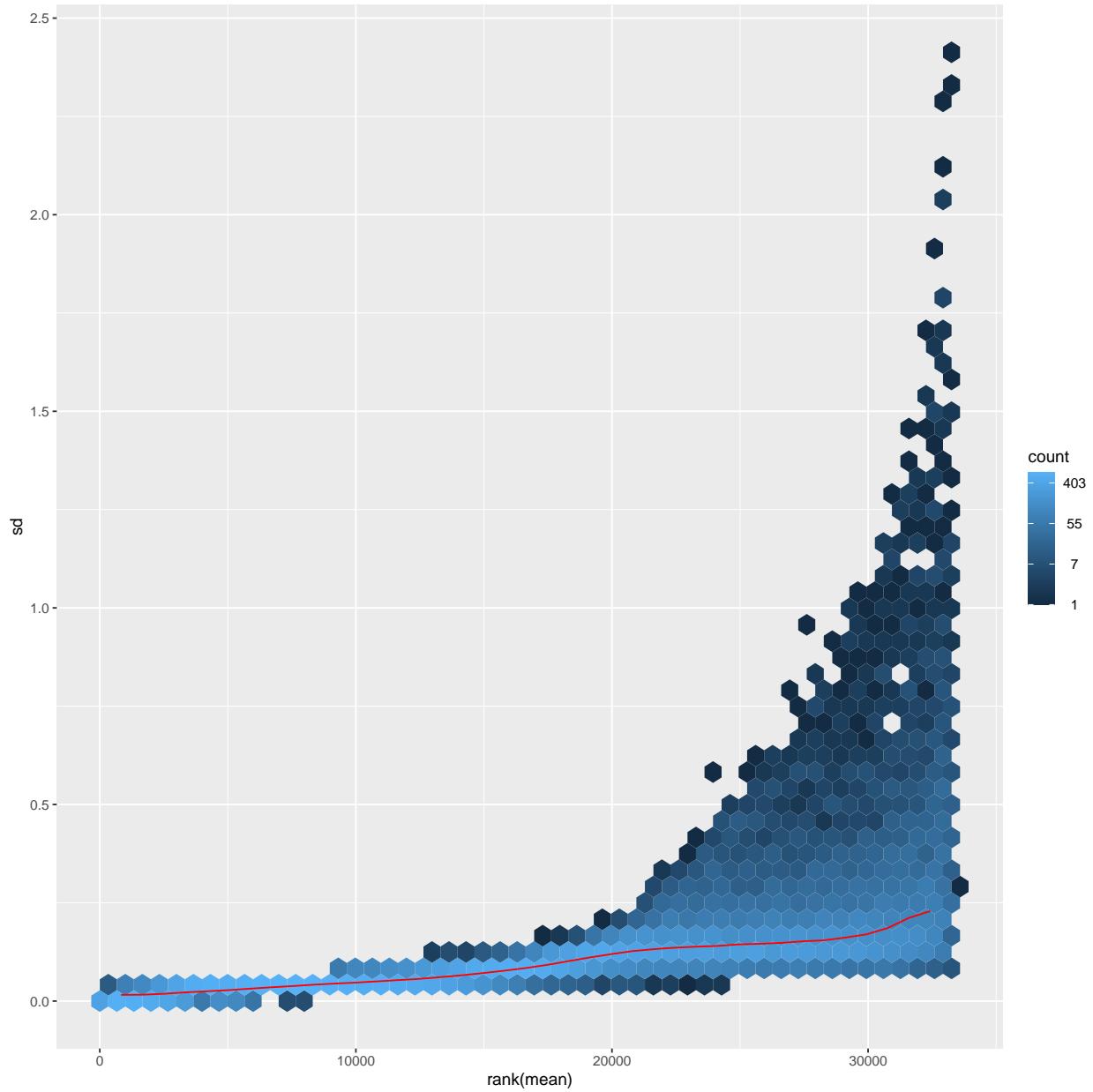
Variance Plots

```
#Mean/SD Normal transform
#-----
meanSdPlot(assay(ntd))
```

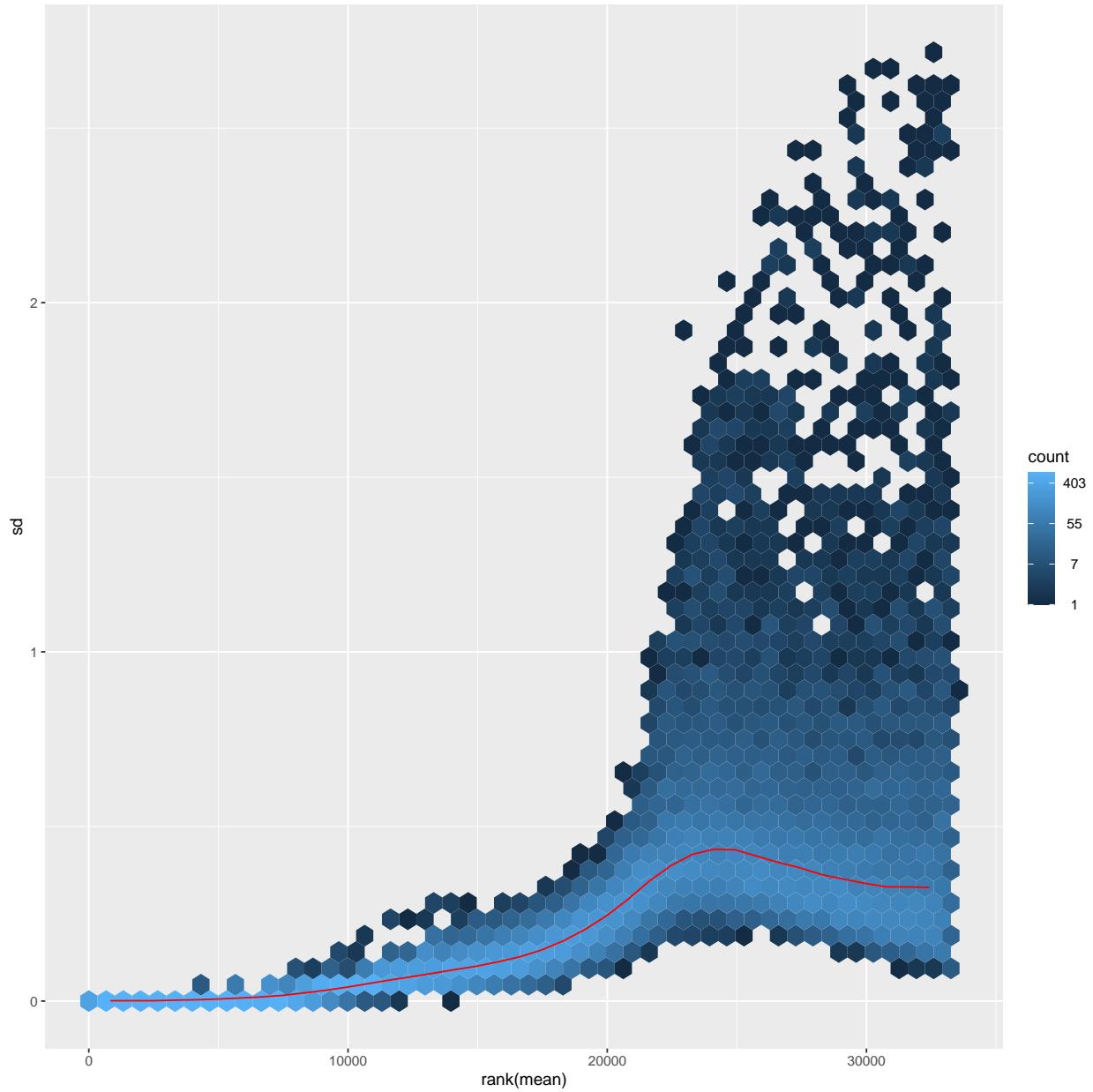
```
## Found more than one class "simpleUnit" in cache; using the first, from namespace 'ggbio'
## Also defined by 'hexbin'
```



```
#Mean/SD Variance stabilizing transformation  
#-----  
meanSdPlot(assay(vsd))
```



```
#Mean/SD Regularized log transform  
#-----  
meanSdPlot(assay(rld))
```



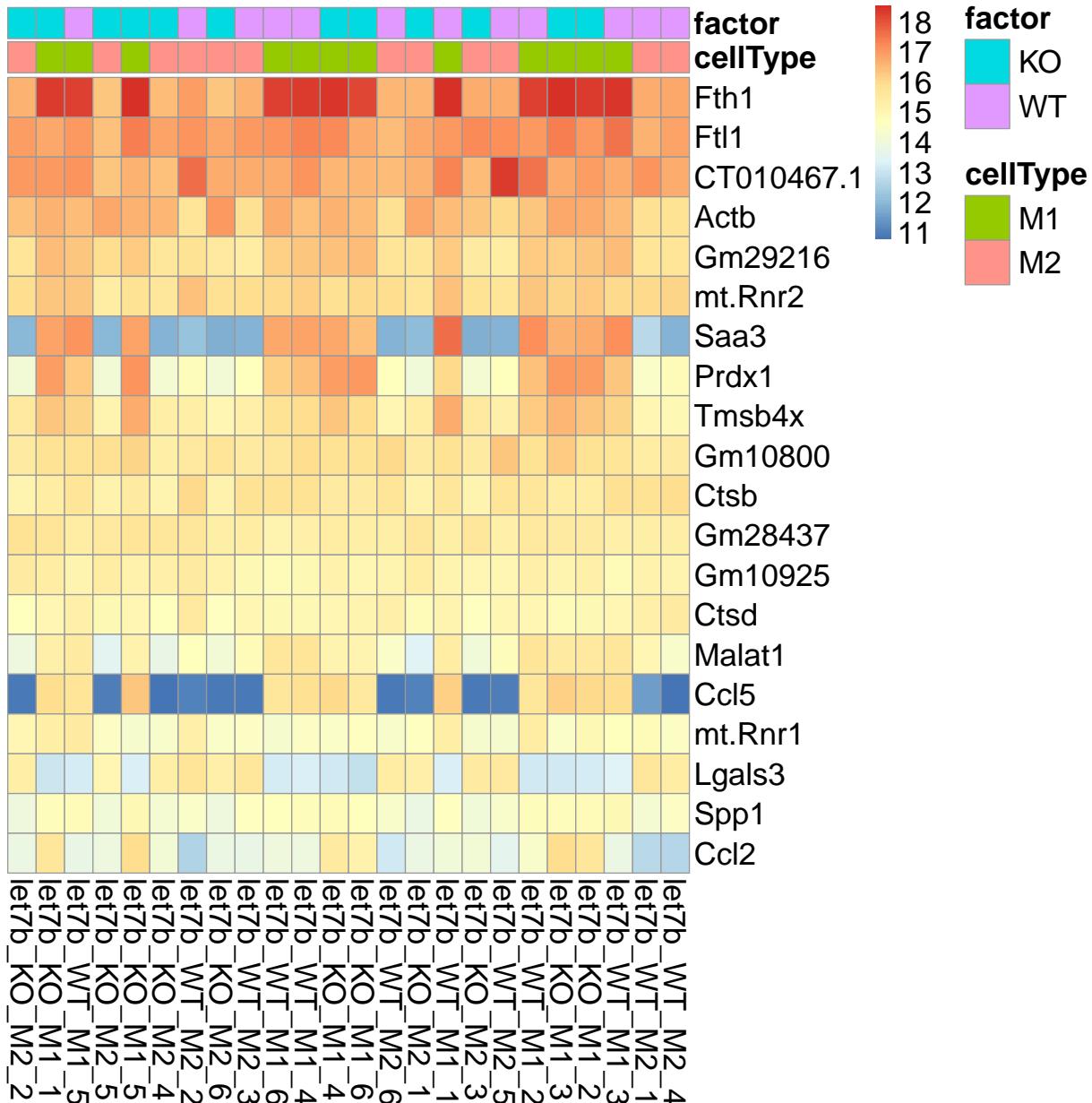
Heatmaps

```
# Heatmaps
#-----
# NORM
select <- order(rowMeans(counts(reg, normalized = TRUE)), decreasing = TRUE)[1:20]
df <- as.data.frame(colData(reg)[, c("cellType", "factor")])
pheatmap(assay(ntd)[select, ], cluster_rows = FALSE, show_rownames = TRUE, cluster_cols = FALSE,
         annotation_col = df, fontsize = 20)

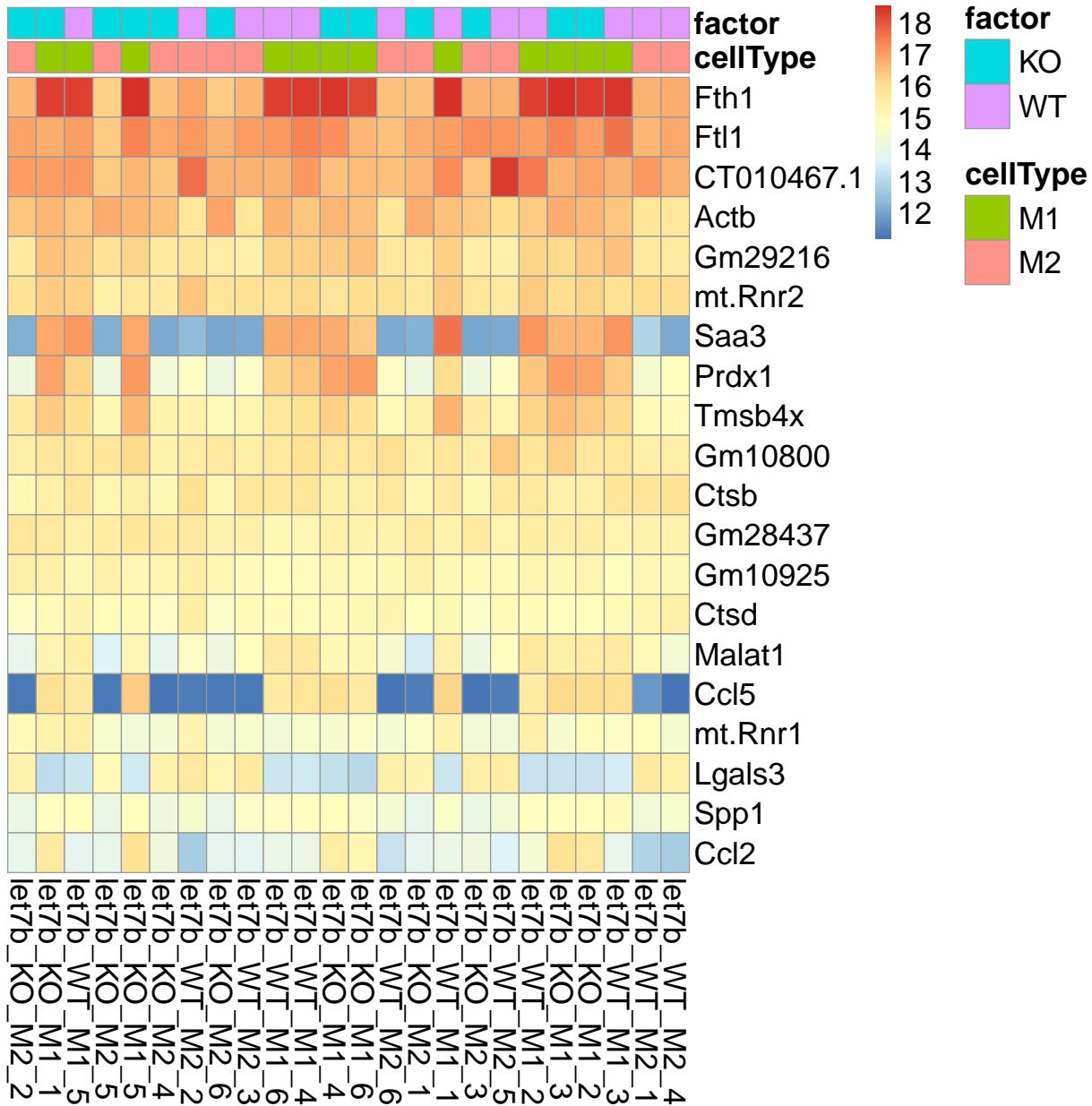
## Found more than one class "unit" in cache; using the first, from namespace 'ggbio'

## Also defined by 'hexbin'
```

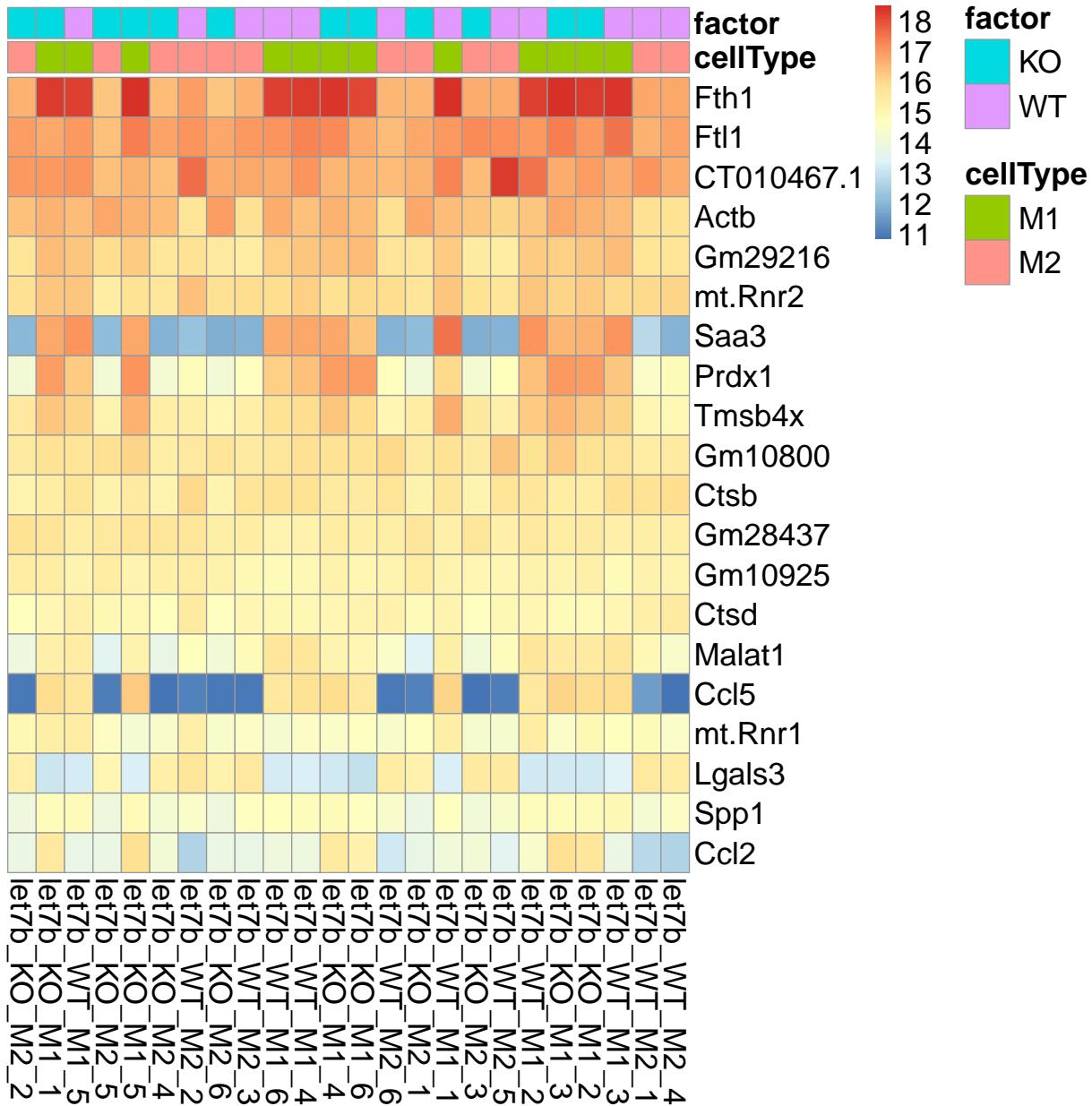
```
## Found more than one class "simpleUnit" in cache; using the first, from namespace 'ggbio'  
## Also defined by 'hexbin'
```



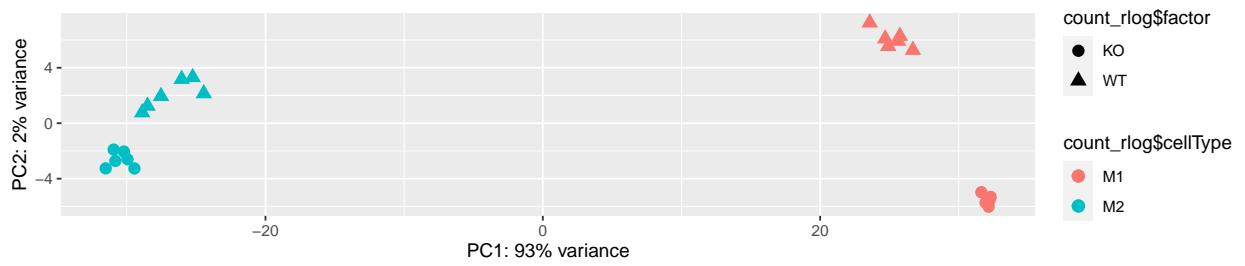
```
# VSD
select <- order(rowMeans(counts(reg, normalized = TRUE)), decreasing = TRUE)[1:20]
df <- as.data.frame(colData(reg)[, c("cellType", "factor")])
pheatmap(assay(vsd)[select, ], cluster_rows = FALSE, show_rownames = TRUE, cluster_cols = FALSE,
         annotation_col = df, fontsize = 20)
```



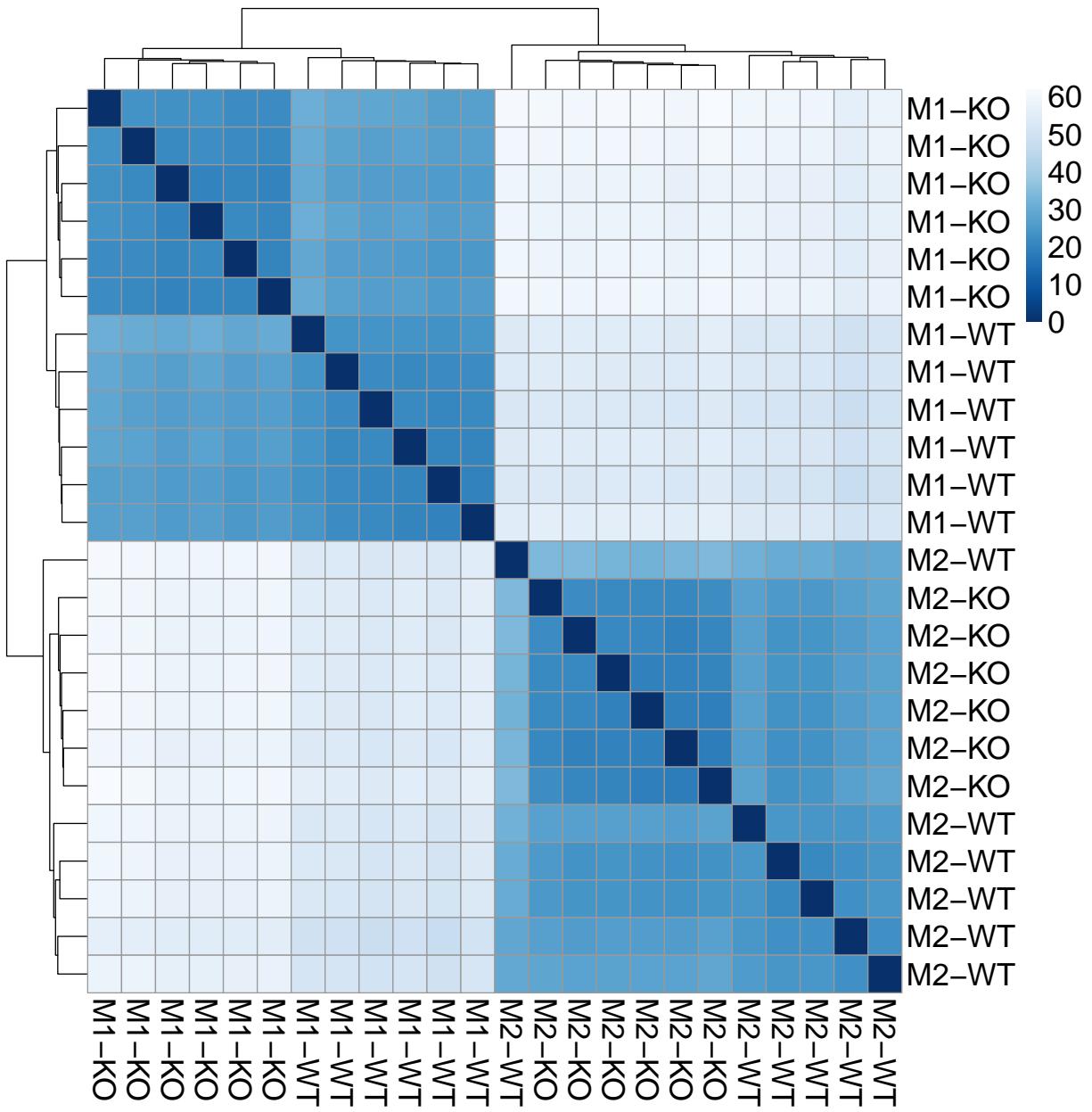
```
# RLD
select <- order(rowMeans(counts(reg, normalized = TRUE)), decreasing = TRUE)[1:20]
df <- as.data.frame(colData(reg)[, c("cellType", "factor")])
pheatmap(assay(rld)[select, ], cluster_rows = FALSE, show_rownames = TRUE, cluster_cols = FALSE,
         annotation_col = df, fontsize = 20)
```



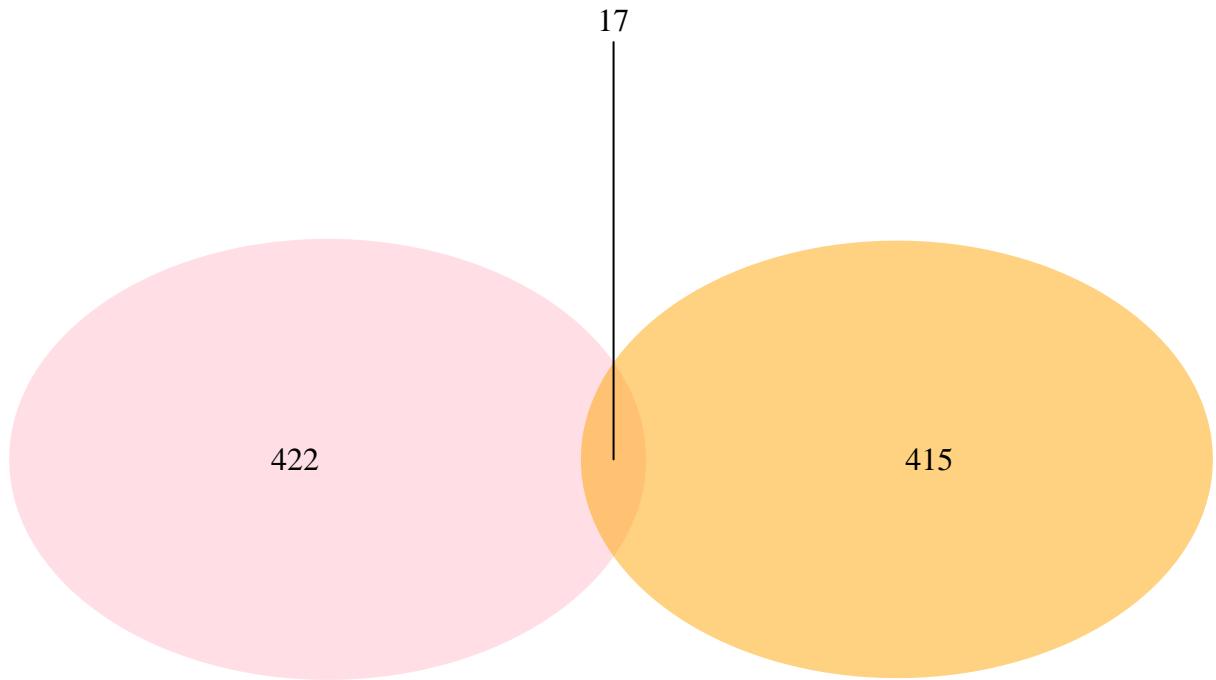
```
ggplot(pcaplot, aes(PC1, PC2, color=count_rlog$cellType, shape=count_rlog$factor)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  coord_fixed()
```



Sample distances



```
draw.pairwise.venn(area1 = nrow(da_best_M1), area2 = nrow(da_best_M2), cross.area = nrow(intersection),
```

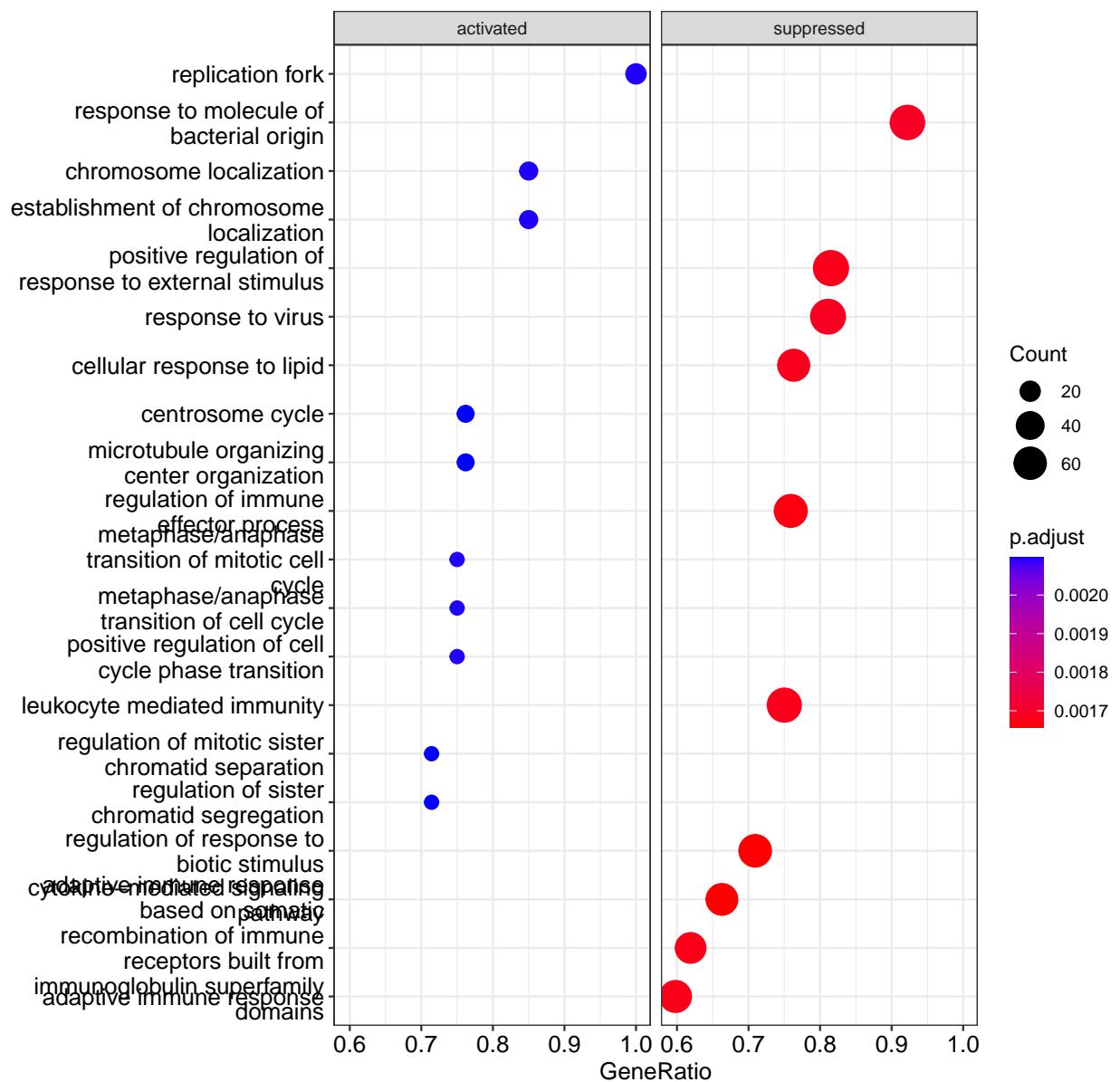


```
## (polygon[GRID.polygon.790], polygon[GRID.polygon.791], polygon[GRID.polygon.792], polygon[GRID.polygon.793])
```

Cluster Profiler

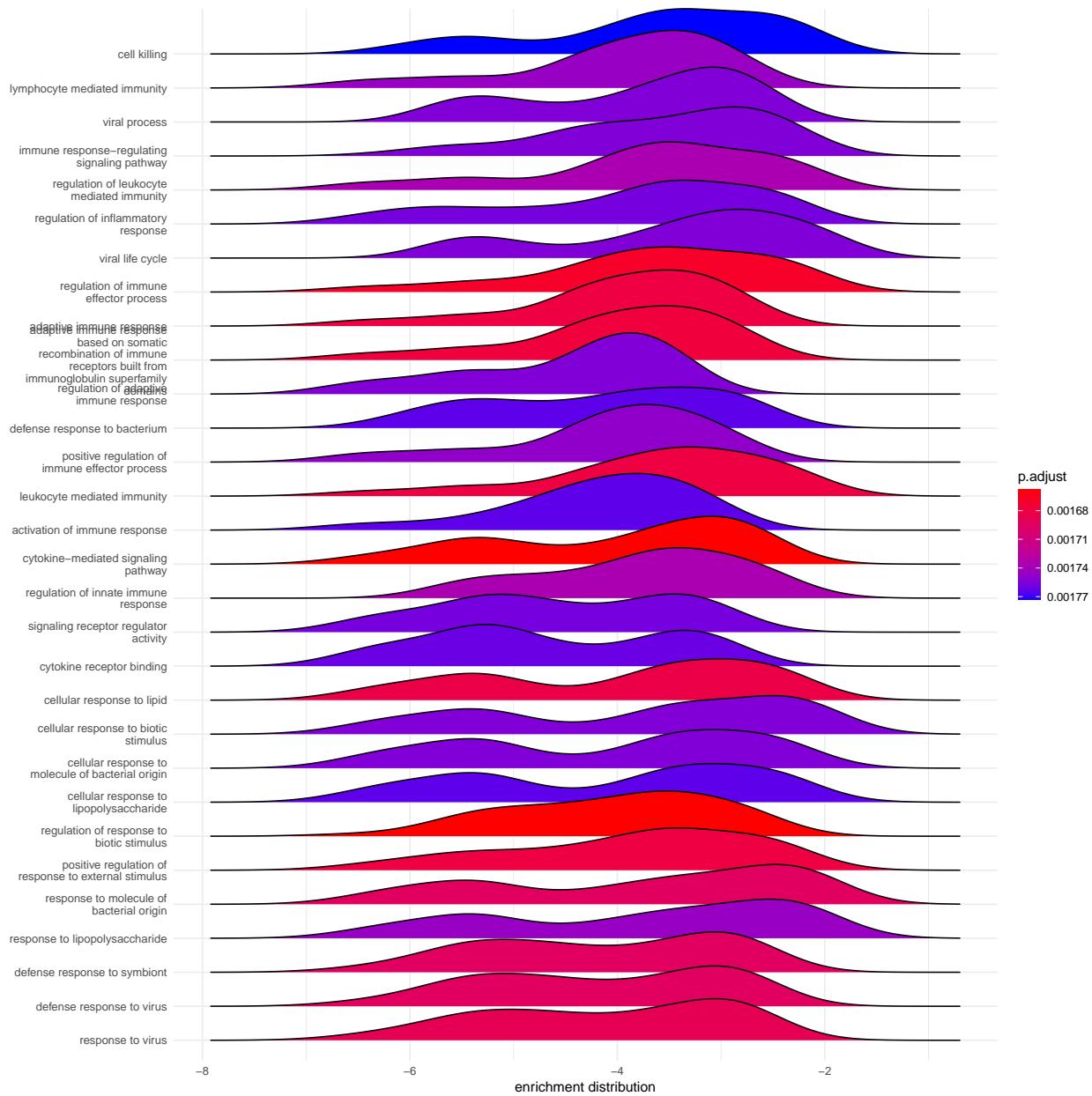
```
gse <- gseGO(geneList=gene_list,
               ont = "ALL",
               keyType = "ENSEMBL",
               nPerm = 1000,
               minGSSize = 5,
               maxGSSize = 100,
               pvalueCutoff = 0.05,
               verbose = TRUE,
               OrgDb = "org.Mm.eg.db",
               pAdjustMethod = "none")

dotplot(gse, showCategory=10, split=".sign") + facet_grid(.~.sign)
```



```
ridgeplot(gse) + labs(x = "enrichment distribution") + theme_minimal()
```

```
## Picking joint bandwidth of 0.445
```



```
gseaplot(gse, by = "all", title = gse$Description[1], geneSetID = 1)
```

regulation of response to biotic stimulus

