

#### **Preparation**



• Create the DESeq2 object

```
library(DESeq2)
mr$Group <- factor(mr$Group)
d <- DESeqDataSetFromMatrix(countData=cf,colData=mr,design=~Group)
d</pre>
```

```
## class: DESeqDataSet
## dim: 10573 6
## metadata(1): version
## assays(1): counts
## rownames(10573): ENSMUSG00000098104 ENSMUSG000000033845 ...
## ENSMUSG00000063897 ENSMUSG00000095742
## rowData names(0):
## colnames(6): DSSd00_1 DSSd00_2 ... DSSd07_2 DSSd07_3
## colData names(7): SampleName SampleID ... Group Replicate
```

- Categorical variables must be factors
- Building GLM models: ~var , ~covar+var

#### Size factors



• Normalisation factors are computed

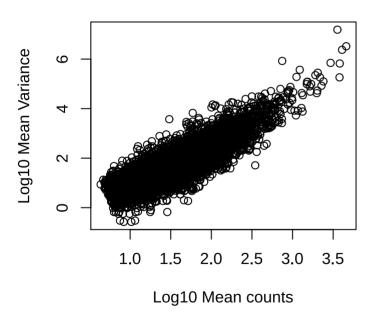
```
d <- DESeq2::estimateSizeFactors(d,type="ratio")
sizeFactors(d)

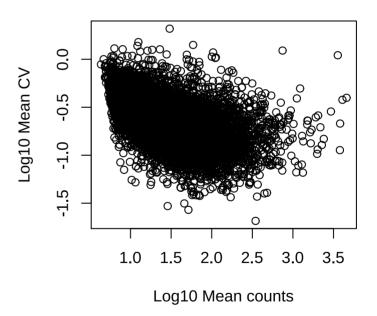
## DSSd00_1 DSSd00_2 DSSd00_3 DSSd07_1 DSSd07_2 DSSd07_3
## 1.0136617 0.9570561 0.9965245 1.0354178 1.0780855 1.0017753</pre>
```

# **Dispersion**



- Dispersion is a measure of variability in gene expression for a given mean
- Dispersion is unreliable for low mean counts





### **Dispersion**



- Genes with similar mean values must have similar dispersion
- Estimate likely (ML) dispersion for each gene based on counts
- Fit a curve through the gene-wise estimates
- Shrink dispersion towards the curve

d <- DESeq2::estimateDispersions(d)</pre>



Log2 fold changes changes are computed after GLM fitting FC = counts group B / counts group A

```
dg <- nbinomWaldTest(d)
resultsNames(dg)

## [1] "Intercept" "Group_day07_vs_day00"</pre>
```

- Use results() to customise/return results
  - Set coefficients using contrast or name
  - Filtering results by fold change using <a href="life">lfcThreshold</a>
  - cooksCutoff removes outliers
  - independentFiltering removes low count genes
  - pAdjustMethod sets method for multiple testing correction
  - o alpha set the significance threshold



```
res <- results(dg,name="Group_day07_vs_day00",alpha=0.05)
summary(res)

##

## out of 10573 with nonzero total read count

## adjusted p-value < 0.05

## LFC > 0 (up) : 193, 1.8%

## LFC < 0 (down) : 238, 2.3%

## outliers [1] : 1, 0.0095%

## low counts [2] : 4920, 47%

## (mean count < 21)

## [1] see 'cooksCutoff' argument of ?results

## [2] see 'independentFiltering' argument of ?results
```

Alternative way to specify contrast

```
results(dg,contrast=c("Group","day07","day00"),alpha=0.05)
```



#### head(res)

```
## log2 fold change (MLE): Group day07 vs day00
## Wald test p-value: Group day07 vs day00
## DataFrame with 6 rows and 6 columns
                       baseMean log2FoldChange
##
                                                  lfcSE
                                                             stat
                                                                     pvalue
##
                      <numeric>
                                     <numeric> <numeric> <numeric> <numeric>
  ENSMUSG00000098104
                       18.8505
                                     0.205656 0.401543
                                                         0.512164 0.6085362
  FNSMUSG00000033845
                      23.3333
                                     0.653565 0.379627
                                                         1.721596 0.0851426
                      37, 1016
  ENSMUSG00000025903
                                     0.672348 0.298923
                                                         2.249232 0.0244977
  ENSMUSG00000033793
                       33.3673
                                     0.144833 0.305139
                                                         0.474646 0.6350394
## ENSMUSG00000025907
                       22.3875
                                     0.821006 0.376414
                                                         2.181125 0.0291742
                       21.1485
## ENSMUSG00000051285
                                     0.452451 0.378725 1.194669 0.2322163
                          padi
##
                      <numeric>
##
  ENSMUSG00000098104
                            NA
  ENSMUSG00000033845
                       0.377432
  ENSMUSG00000025903
                      0.177491
## ENSMUSG00000033793
                       0.886264
## ENSMUSG00000025907
                       0.201741
## FNSMUSG00000051285
                            NA
```



- Use <a href="left">IfcShrink()</a> to correct fold changes for genes with high dispersion or low counts
- Does not change number of DE genes

