

# class15

*The Spanish Inquisition*

5/21/2019

Upload packages

```
library(bio3d)
library(DESeq2)
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
```

```
##
```

```
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   anyDuplicated, 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,
##   which.max, which.min
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:base':
```

```
##
```

```
##   expand.grid
```

```
## Loading required package: IRanges
```

```

##
## Attaching package: 'IRanges'

## The following object is masked from 'package:bio3d':
##
##      trim

## The following object is masked from 'package:grDevices':
##
##      windows

## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Loading required package: SummarizedExperiment

## 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)".

## Loading required package: DelayedArray

## Loading required package: matrixStats

##
## Attaching package: 'matrixStats'

## The following objects are masked from 'package:Biobase':
##
##      anyMissing, rowMedians

## Loading required package: BiocParallel

##
## Attaching package: 'DelayedArray'

## The following objects are masked from 'package:matrixStats':
##
##      colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

## The following objects are masked from 'package:base':
##
##      aperm, apply, rowsum

## Registered S3 methods overwritten by 'ggplot2':
##      method      from
##      [.quosures   rlang
##      c.quosures   rlang
##      print.quosures rlang

```

## Trimming data

```
metaFile <- "GSE37704_metadata.csv"

countFile <- "GSE37704_featurecounts.csv"

colData = read.csv(metaFile, row.names=1)
head(colData)
```

```
##           condition
## SRR493366 control_sirna
## SRR493367 control_sirna
## SRR493368 control_sirna
## SRR493369      hoxa1_kd
## SRR493370      hoxa1_kd
## SRR493371      hoxa1_kd
```

```
countData <- read.csv(countFile, row.name = 1)
head(countData)
```

```
##           length SRR493366 SRR493367 SRR493368 SRR493369 SRR493370
## ENSG00000186092     918         0         0         0         0         0
## ENSG00000279928     718         0         0         0         0         0
## ENSG00000279457    1982        23        28        29        29        28
## ENSG00000278566     939         0         0         0         0         0
## ENSG00000273547     939         0         0         0         0         0
## ENSG00000187634    3214       124       123       205       207       212
##           SRR493371
## ENSG00000186092         0
## ENSG00000279928         0
## ENSG00000279457        46
## ENSG00000278566         0
## ENSG00000273547         0
## ENSG00000187634       258
```

```
countData <- as.matrix(countData[, -1])
head(countData)
```

```
##           SRR493366 SRR493367 SRR493368 SRR493369 SRR493370
## ENSG00000186092         0         0         0         0         0
## ENSG00000279928         0         0         0         0         0
## ENSG00000279457        23        28        29        29        28
## ENSG00000278566         0         0         0         0         0
## ENSG00000273547         0         0         0         0         0
## ENSG00000187634       124       123       205       207       212
##           SRR493371
## ENSG00000186092         0
## ENSG00000279928         0
## ENSG00000279457        46
## ENSG00000278566         0
## ENSG00000273547         0
## ENSG00000187634       258
```

## Running DESeq2

```
dds <- DESeqDataSetFromMatrix(countData = countData,
                              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

dds

## class: DESeqDataSet
## dim: 19808 6
## metadata(1): version
## assays(4): counts mu H cooks
## rownames(19808): ENSG00000186092 ENSG00000279928 ...
##      ENSG00000277475 ENSG00000268674
## rowData names(22): baseMean baseVar ... deviance maxCooks
## colnames(6): SRR493366 SRR493367 ... SRR493370 SRR493371
## colData names(2): condition sizeFactor

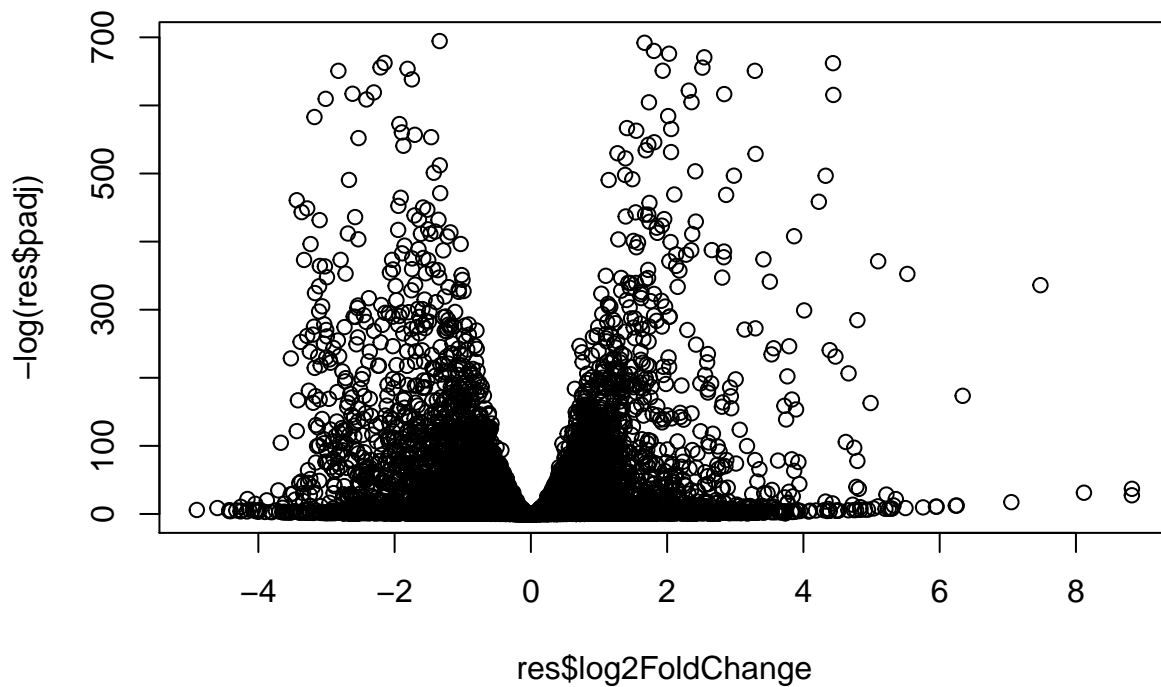
res <- results(dds, contrast = c("condition", "hoxa1_kd", "control_sirna"))

summary(res)

##
## out of 15975 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)      : 4349, 27%
## LFC < 0 (down)    : 4393, 27%
## outliers [1]      : 0, 0%
## low counts [2]     : 1221, 7.6%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

#### Data is a bit coarse as I have not removed the zero values from the original countData set.

```
plot(res$log2FoldChange, -log(res$padj))
```



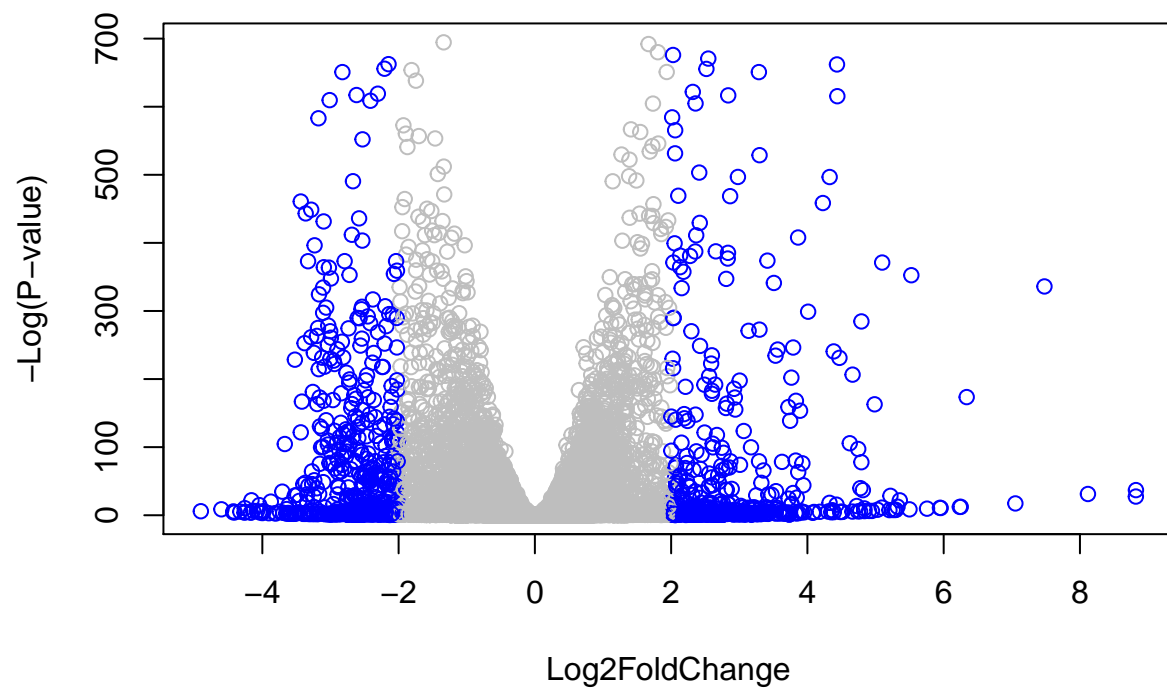
```
mycols <- rep("gray", nrow(res))

#color red the genes with absolute fold change above 2.
mycols[abs(res$log2FoldChange) > 2] <- "red"

#color blue those with p-value less than 0.01.

inds <- pnorm(res$log2FoldChange) & (abs(res$log2FoldChange) > 2)
mycols[inds] <- "blue"

plot(res$log2FoldChange, -log(res$padj), col = mycols,
      xlab = "Log2FoldChange",
      ylab = "-Log(P-value)")
```



```
library("AnnotationDbi")
library("org.Hs.eg.db")
```

```
##
```

```
columns(org.Hs.eg.db)
```

```
## [1] "ACCNUM"      "ALIAS"       "ENSEMBL"     "ENSEMBLPROT"
## [5] "ENSEMBLTRANS" "ENTREZID"    "ENZYME"      "EVIDENCE"
## [9] "EVIDENCEALL"  "GENENAME"    "GO"          "GOALL"
## [13] "IPI"         "MAP"         "OMIM"        "ONTOLOGY"
## [17] "ONTOLOGYALL" "PATH"        "PFAM"        "PMID"
## [21] "PROSITE"     "REFSEQ"      "SYMBOL"      "UCSCKG"
## [25] "UNIGENE"     "UNIPROT"
```