MA Plots of Differentially Expressed Genes with Age

Jason Kennerdell

8/01/2017

Input the data and metadata:

```
inpath <- "~/Desktop/brain"
outpath <- "~/Desktop/brain/_7C_aging_plotMA"
setwd(inpath)
files <- list.files()
htseq_files <- files[grepl("^JKL.*txt$", files)]
sampleNames <- read.csv("SampleNames.csv")
sampleTable <- data.frame(fileName = htseq_files,
                          stringsAsFactors=FALSE)
sampleTable$Library <- gsub("-counts.txt", "", sampleTable$fileName)
sampleTable$Library <- gsub("b", "", sampleTable$Library)
sampleTable$seq.batch <- ifelse(grepl("b", sampleTable$fileName), "B", "A")
sampleTable$seq.batch <- paste(sampleTable$Library, sampleTable$seq.batch)
sampleTable <- merge(sampleTable, sampleNames, by = "seq.batch")
sampleTable$genotype <- gsub("-.*$", "", sampleTable$Sample)
sampleTable$genotype <- gsub("JKLY.... ", "", sampleTable$genotype)
sampleTable$genotype <- gsub("Pcl\\[4\\]/", "", sampleTable$genotype)
sampleTable$genotype <- gsub("\\[5\\]", "", sampleTable$genotype)
sampleTable$genotype <- gsub("\\[1118\\]", "", sampleTable$genotype)
sampleTable$genotype <- factor(sampleTable$genotype, levels = c("w", "Ez", "Pcl"))
sampleTable$Temp <- gsub("[A-Z, a-z, -]", "", sampleTable$Sample)
sampleTable$Temp <- gsub("^11.*$", "25", sampleTable$Temp)
sampleTable$Temp <- factor(sampleTable$Temp, levels = c("25", "29"))
sampleTable$age <- gsub("JKLY.*$", "3d", sampleTable$Sample)
sampleTable$age <- gsub("^[a-zA-z].*$", "20d", sampleTable$age)
sampleTable$age <- factor(sampleTable$age)
sampleTable$age <- factor(sampleTable$age, levels = c("3d", "20d"))
sampleTable$condition <- paste(sampleTable$genotype, sampleTable$Temp, sampleTable$age, sep = "-")
# Select only the samples of interest:
sampleTable <- sampleTable[sampleTable$genotype == "w" & sampleTable$Temp == "25",]
sampleTable
```

```
##    seq.batch          fileName Library.x           Sample batch Library.y
## 3    JKL10 A  JKL10-counts.txt     JKL10          w-25-II     A     JKL10
## 4    JKL10 B JKL10b-counts.txt     JKL10          w-25-II     B     JKL10
## 5    JKL11 A  JKL11-counts.txt     JKL11         w-25-III     A     JKL11
## 6    JKL11 B JKL11b-counts.txt     JKL11         w-25-III     B     JKL11
## 23    JKL2 A   JKL2-counts.txt      JKL2 JKLY1125 w[1118]     A      JKL2
## 24    JKL2 B  JKL2b-counts.txt      JKL2 JKLY1125 w[1118]     B      JKL2
## 33    JKL4 A   JKL4-counts.txt      JKL4 JKLY1127 w[1118]     A      JKL4
## 34    JKL4 B  JKL4b-counts.txt      JKL4 JKLY1127 w[1118]     B      JKL4
## 37    JKL6 A   JKL6-counts.txt      JKL6 JKLY1129 w[1118]     A      JKL6
## 38    JKL6 B  JKL6b-counts.txt      JKL6 JKLY1129 w[1118]     B      JKL6
## 39    JKL7 A   JKL7-counts.txt      JKL7           w-25-I     A      JKL7
## 40    JKL7 B  JKL7b-counts.txt      JKL7           w-25-I     B      JKL7
##    genotype Temp age condition
```

```
## 3         w   25 20d  w-25-20d
## 4         w   25 20d  w-25-20d
## 5         w   25 20d  w-25-20d
## 6         w   25 20d  w-25-20d
## 23        w   25  3d   w-25-3d
## 24        w   25  3d   w-25-3d
## 33        w   25  3d   w-25-3d
## 34        w   25  3d   w-25-3d
## 37        w   25  3d   w-25-3d
## 38        w   25  3d   w-25-3d
## 39        w   25 20d  w-25-20d
## 40        w   25 20d  w-25-20d
```

Set up the statistical model:

```
design <- formula(~ age)
```

```
dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, directory = inpath, design = design)
dds <- collapseReplicates(dds, groupby=dds$Library.x, run = dds$batch)
dds <- DESeq(dds)
```

```
## estimating size factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
```

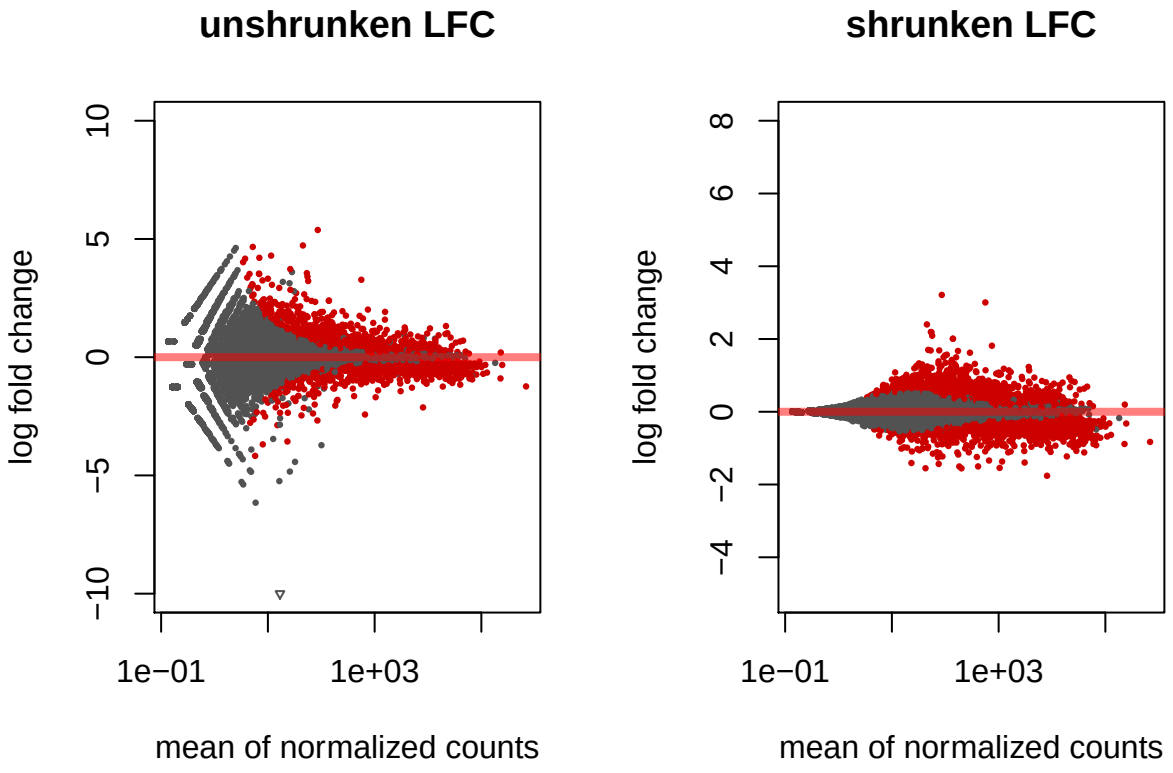
```
results <- results(dds, alpha=0.05)
results$ensembl <- rownames(results)
summary(results)
```

```
##
## out of 14387 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)     : 1662, 12%
## LFC < 0 (down)   : 2115, 15%
## outliers [1]     : 19, 0.13%
## low counts [2]   : 3860, 27%
## (mean count < 3)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

Prepare MA Plots

```
# Some MA plots:
# for Maximum likelihood estimates:
resultsMLE <- results(dds, addMLE=TRUE, alpha = 0.05)
par(mfrow=c(1,2))
```

```
plotMA(resultsMLE, MLE=TRUE, alpha = 0.05, main="unshrunken LFC", ylim=c(-10,10))
plotMA(results, alpha = 0.05, main="shrunken LFC", ylim=c(-5,8))
```




```
# Add usefull gene names:
library(biomaRt)
# Get the archived version:
mart = useMart("ENSEMBL_MART_ENSEMBL",
               host="aug2017.archive.ensembl.org")
#x <- listDatasets(mart)
mart = useMart("ENSEMBL_MART_ENSEMBL", host="aug2017.archive.ensembl.org",
               dataset = "dmelanogaster_gene_ensembl")
#listAttributes(mart)
genemap <- getBM(attributes = c("ensembl_gene_id", "entrezgene", "flybasecgid_gene",
                                "external_gene_name"),
                 filters = "ensembl_gene_id",
                 values = results$ensembl,
                 mart = mart)

idx <- match(results$ensembl, genemap$ensembl_gene_id)
results$entrez <- genemap$entrezgene[idx]
results$geneSymbol <- genemap$external_gene_name[idx]
results$cg <- genemap$flybasecgid_gene[idx]
```

Prepare a plot highlighting changes in small hsp genes:

```
small_hsp <- read.delim(file.path(inpath, "FBgg0000507.txt"))
colnames(small_hsp) <- c("FlyBase", "NAME", "SYMBOL")
```

```
jpeg(file=paste(outpath, "aging_sHSP_MAP.jpg", sep="/"), width=240, height=240)
# Small HSP Family
results$shsp <- results$ensembl %in% small_hsp$FlyBase
plotMA(results, alpha = 0.05, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       ylab=expression("log"[2]*" (fold change)"), main="Small Heat Shock Proteins are \ndownregulated
with(results[results$shsp==TRUE,], {points(baseMean, log2FoldChange, col = "red", cex=1, lwd=2)})
dev.off()
```

```
## pdf
##   2
```

Try without shrinkage:

```
# Small HSP Family
######################Prepare Publication Plot###################################
jpeg(file=paste(outpath, "Fig7c_aging_sHSP_MLE.jpg", sep="/"),
     quality=100,
     res=300,
     width=960,
     height=960)
resultsMLE$shsp <- rownames(resultsMLE) %in% small_hsp$FlyBase
par(mai=c(0,0,0,0), mar=c(2,4.5,3,1) + 0.5)
plotMA(resultsMLE, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       MLE=TRUE, colLine = "darkgrey", las=1, cex.lab=1, cex.axis=1, cex.main=1.5,
       ylab=expression(""), xlab = "", xaxp=c(1,2,1), xaxt="n",
       main="")
axis(1, at=c(1,10,100,1000,10000), labels=c(1,10,100,1000,10000), cex.axis=1)
with(resultsMLE[resultsMLE$shsp==TRUE,], {points(baseMean, lfcMLE, col = "red", cex=1, lwd=3)})
dev.off()
```

```
## pdf
##   2
```

```
# Now as Cairo pdf:
cairo_pdf(file=paste(outpath, "Fig7c_aging_sHSP_MLE.pdf", sep="/"))
resultsMLE$shsp <- rownames(resultsMLE) %in% small_hsp$FlyBase
par(mai=c(0,0,0,0), mar=c(0,4.5,3,1) + 0.5)
plotMA(resultsMLE, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       MLE=TRUE, colLine = "darkgrey", las=1, cex.lab=4, cex.axis=4, cex.main=1.5,
       ylab=expression(""), xlab = "", xaxp=c(1,2,1), xaxt="n",
       main="")
axis(1, at=c(1,10,100,1000,10000), labels=c(1,10,100,1000,10000), cex.axis=3, outer=T)
with(resultsMLE[resultsMLE$shsp==TRUE,], {points(baseMean, lfcMLE, col = "red", cex=1, lwd=3)})
dev.off()
```

```
## pdf
##   2
```

```
#################################################################################

plotMA(resultsMLE, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       MLE=TRUE, colLine = "darkgrey",
```

```
       ylab=expression("log"[2]*" (fold change)"), xlab = "Mean expression",
       main="HSP20 Family is \ndownregulated with age")
with(resultsMLE[resultsMLE$shsp==TRUE,], {points(baseMean, lfcMLE, col = "red", cex=1, lwd=2)})
```