Comparison of E(z) mutant brain expression profiles and age-correlated probesets indicates a "younger" brain

Jason Kennerdell

9/25/2017

Input the data and metadata:

```
inpath <- "~/Desktop/brain"
outpath <- "~/Desktop/brain/_8B_hypergeo"
setwd(inpath)
files <- list.files()
htseq_files <- files[grepl("^JKL.*txt$", files)]
sampleNames <- read.csv("EzSampleNames.csv")
sampleTable <- data.frame(fileName = htseq_files,
                          stringsAsFactors=FALSE)
sampleTable$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[,c(1,3)], sampleNames, by = "seq.batch")
sampleTable$genotype <- gsub("-.*$", "", sampleTable$Sample)
sampleTable$genotype <- factor(sampleTable$genotype, levels = c("w", "Ez"))
sampleTable$Temp <- gsub("[A-Z, a-z, -]", "", sampleTable$Sample)
sampleTable$Temp <- factor(sampleTable$Temp, levels = c("25", "29"))
sampleTable$rep <- gsub("^.*-", "", sampleTable$Sample)
sampleTable$condition <- paste(sampleTable$genotype, sampleTable$Temp, sep = "-")
sampleTable
```

```
##    seq.batch          fileName    Sample batch Library genotype Temp rep
## 1    JKL10 A  JKL10-counts.txt   w-25-II     A   JKL10        w   25  II
## 2    JKL10 B JKL10b-counts.txt   w-25-II     B   JKL10        w   25  II
## 3    JKL11 A  JKL11-counts.txt  w-25-III     A   JKL11        w   25 III
## 4    JKL11 B JKL11b-counts.txt  w-25-III     B   JKL11        w   25 III
## 5    JKL12 A  JKL12-counts.txt Ez-25-III     A   JKL12       Ez   25 III
## 6    JKL12 B JKL12b-counts.txt Ez-25-III     B   JKL12       Ez   25 III
## 7    JKL13 A  JKL13-counts.txt    w-29-I     A   JKL13        w   29   I
## 8    JKL13 B JKL13b-counts.txt    w-29-I     B   JKL13        w   29   I
## 9    JKL14 A  JKL14-counts.txt   Ez-29-I     A   JKL14       Ez   29   I
## 10   JKL14 B JKL14b-counts.txt   Ez-29-I     B   JKL14       Ez   29   I
## 11   JKL15 A  JKL15-counts.txt   w-29-II     A   JKL15        w   29  II
## 12   JKL15 B JKL15b-counts.txt   w-29-II     B   JKL15        w   29  II
## 13   JKL16 A  JKL16-counts.txt  Ez-29-II     A   JKL16       Ez   29  II
## 14   JKL16 B JKL16b-counts.txt  Ez-29-II     B   JKL16       Ez   29  II
## 15   JKL17 A  JKL17-counts.txt  w-29-III     A   JKL17        w   29 III
## 16   JKL17 B JKL17b-counts.txt  w-29-III     B   JKL17        w   29 III
## 17   JKL18 A  JKL18-counts.txt Ez-29-III     A   JKL18       Ez   29 III
## 18   JKL18 B JKL18b-counts.txt Ez-29-III     B   JKL18       Ez   29 III
## 19   JKL19 A  JKL19-counts.txt   w-29-IV     A   JKL19        w   29  IV
## 20   JKL19 B JKL19b-counts.txt   w-29-IV     B   JKL19        w   29  IV
```

```
## 21   JKL20 A  JKL20-counts.txt  Ez-29-IV     A   JKL20       Ez   29  IV
## 22   JKL20 B JKL20b-counts.txt  Ez-29-IV     B   JKL20       Ez   29  IV
## 23   JKL21 A  JKL21-counts.txt    w-29-V     A   JKL21        w   29   V
## 24   JKL21 B JKL21b-counts.txt    w-29-V     B   JKL21        w   29   V
## 25   JKL22 A  JKL22-counts.txt   Ez-29-V     A   JKL22       Ez   29   V
## 26   JKL22 B JKL22b-counts.txt   Ez-29-V     B   JKL22       Ez   29   V
## 27    JKL7 A   JKL7-counts.txt    w-25-I     A    JKL7        w   25   I
## 28    JKL7 B  JKL7b-counts.txt    w-25-I     B    JKL7        w   25   I
## 29    JKL8 A   JKL8-counts.txt   Ez-25-I     A    JKL8       Ez   25   I
## 30    JKL8 B  JKL8b-counts.txt   Ez-25-I     B    JKL8       Ez   25   I
## 31    JKL9 A   JKL9-counts.txt  Ez-25-II     A    JKL9       Ez   25  II
## 32    JKL9 B  JKL9b-counts.txt  Ez-25-II     B    JKL9       Ez   25  II
##    condition
## 1       w-25
## 2       w-25
## 3       w-25
## 4       w-25
## 5      Ez-25
## 6      Ez-25
## 7       w-29
## 8       w-29
## 9      Ez-29
## 10     Ez-29
## 11      w-29
## 12      w-29
## 13     Ez-29
## 14     Ez-29
## 15      w-29
## 16      w-29
## 17     Ez-29
## 18     Ez-29
## 19      w-29
## 20      w-29
## 21     Ez-29
## 22     Ez-29
## 23      w-29
## 24      w-29
## 25     Ez-29
## 26     Ez-29
## 27      w-25
## 28      w-25
## 29     Ez-25
## 30     Ez-25
## 31     Ez-25
## 32     Ez-25
```

Set up the statistical model

```
design <- formula(~ Temp + genotype)
```

DESEq2 Statistics

```
dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, directory = inpath, design = design)
# Combine the technical replicates (different runs) by adding the count
# totals for each gene across the two runs:
dds <- collapseReplicates(dds, groupby=dds$Library, run = dds$batch)
dds <- DESeq(dds)
# What does the data look like?
head(assay(dds)) # This is the sum of the two runs HTSeq-count output!
```

```
##             JKL10 JKL11 JKL12 JKL13 JKL14 JKL15 JKL16 JKL17 JKL18 JKL19
## FBgn0000003     0     0     0     0     0     0     0     0     0     0
## FBgn0000008  1444  1874  1687  1305  1453  1725  1529  1856  1500  1558
## FBgn0000014     0     0     0     0     0     0     1     0     1     0
## FBgn0000015     1     6     1     3     0     1     1     1     2     0
## FBgn0000017  9186 10798  9189  6877  8808  9121  9834 11313  9272  8564
## FBgn0000018   262   306   286   336   288   317   272   346   274   285
##             JKL20 JKL21 JKL22  JKL7  JKL8  JKL9
## FBgn0000003     1     0     0     0     0     1
## FBgn0000008  1353  1589  1367  1980  1861  1884
## FBgn0000014     1     1     1     0     3     0
## FBgn0000015     0     1     0     0     0     2
## FBgn0000017  8697  8196  8612 11453 11360 11139
## FBgn0000018   285   238   284   286   333   333
```

```
# What are the columns?
colData(dds)
```

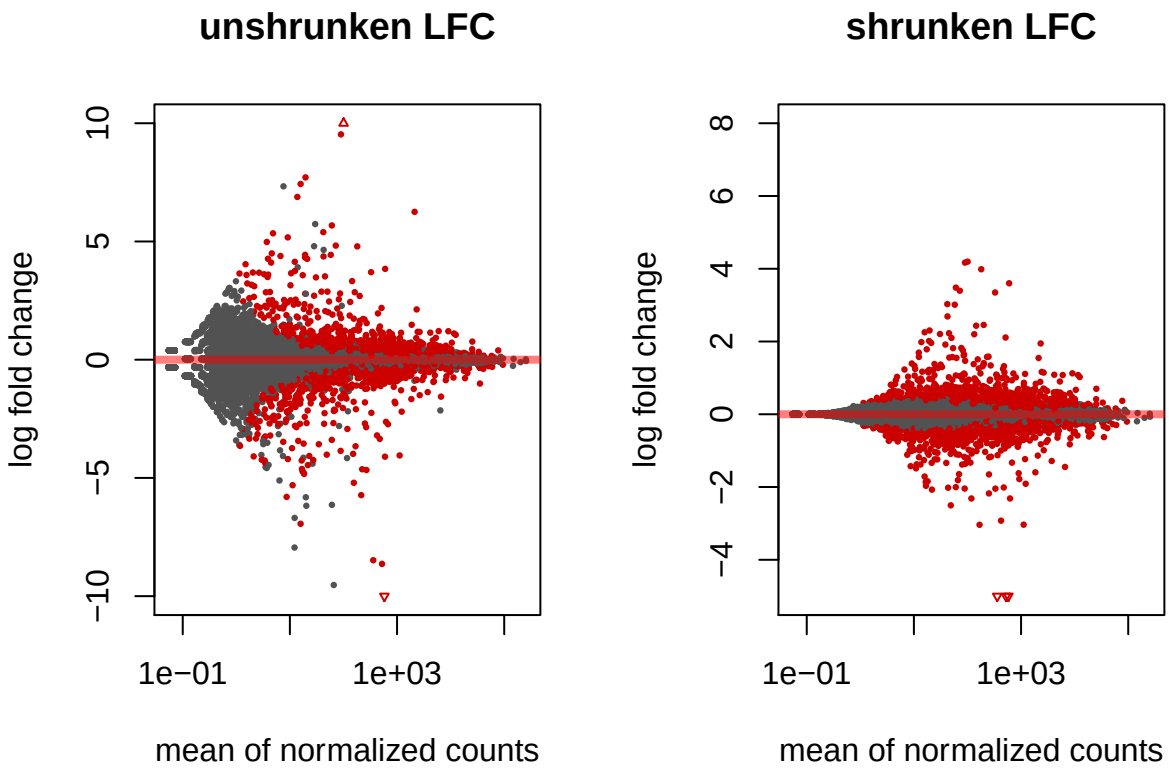
```
## DataFrame with 16 rows and 9 columns
##          Sample    batch  Library genotype     Temp         rep
##        <factor> <factor> <factor> <factor> <factor> <character>
## JKL10   w-25-II        A    JKL10        w       25          II
## JKL11  w-25-III        A    JKL11        w       25         III
## JKL12 Ez-25-III        A    JKL12       Ez       25         III
## JKL13    w-29-I        A    JKL13        w       29           I
## JKL14   Ez-29-I        A    JKL14       Ez       29           I
## ...         ...      ...      ...      ...      ...         ...
## JKL21    w-29-V        A    JKL21        w       29           V
## JKL22   Ez-29-V        A    JKL22       Ez       29           V
## JKL7     w-25-I        A     JKL7        w       25           I
## JKL8    Ez-25-I        A     JKL8       Ez       25           I
## JKL9   Ez-25-II        A     JKL9       Ez       25          II
##         condition runsCollapsed sizeFactor
##       <character>   <character>  <numeric>
## JKL10       w-25           A,B  0.9090574
## JKL11       w-25           A,B  1.1403424
## JKL12      Ez-25           A,B  1.0027251
## JKL13       w-29           A,B  1.0220507
## JKL14      Ez-29           A,B  0.9570594
## ...          ...           ...        ...
## JKL21       w-29           A,B  0.8750556
## JKL22      Ez-29           A,B  0.9618390
## JKL7        w-25           A,B  1.1341278
## JKL8       Ez-25           A,B  1.1093046
```

```
## JKL9         Ez-25            A,B   1.1303920
```

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

Prepare 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 Annotation

```
# Add usefull gene names:
library(biomaRt)

# Add FBgn Names
#listMarts(host="aug2017.archive.ensembl.org")
mart = useMart("ENSEMBL_MART_ENSEMBL", host="aug2017.archive.ensembl.org")
#listDatasets(mart)
mart = useMart("ENSEMBL_MART_ENSEMBL", host="aug2017.archive.ensembl.org",
               dataset = "dmelanogaster_gene_ensembl")
genemap <- getBM(attributes = c("ensembl_gene_id", "entrezgene", "external_gene_name", "flybasecgid_gene
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]
```

Import Age-Correlated Genes and look for how these genes change in E(z) mutants:

```
library(xlsx)
```

```
## Loading required package: rJava
```

```
## Loading required package: xlsxjars
```

```
ageCorGenes <- read.xlsx(paste(inpath, "nature10810-s2.xls", sep="/"),
                         sheetName="Age_Gene_tbl", startRow=3, endRow=177)

# Add FBgn Names
library(biomaRt)
genemap2 <- getBM(attributes = c("affy_drosophila_2", "ensembl_gene_id", "entrezgene", "flybasecgid_gene
                 filters = "affy_drosophila_2",
                 values = as.character(ageCorGenes$Probeset.ID),
                 mart = mart)

idx2 <- match(ageCorGenes$Probeset.ID, genemap2$affy_drosophila_2)
ageCorGenes$ensembl <- genemap2$ensembl_gene_id[idx2]
ageCorGenes$geneSymbol <- genemap2$external_gene_name[idx2]
ageCorGenes$cg <- genemap2$flybasecgid_gene[idx2]

# Fix a problem getting the FBgn names
ageCorGenes[is.na(ageCorGenes$ensembl),]
```

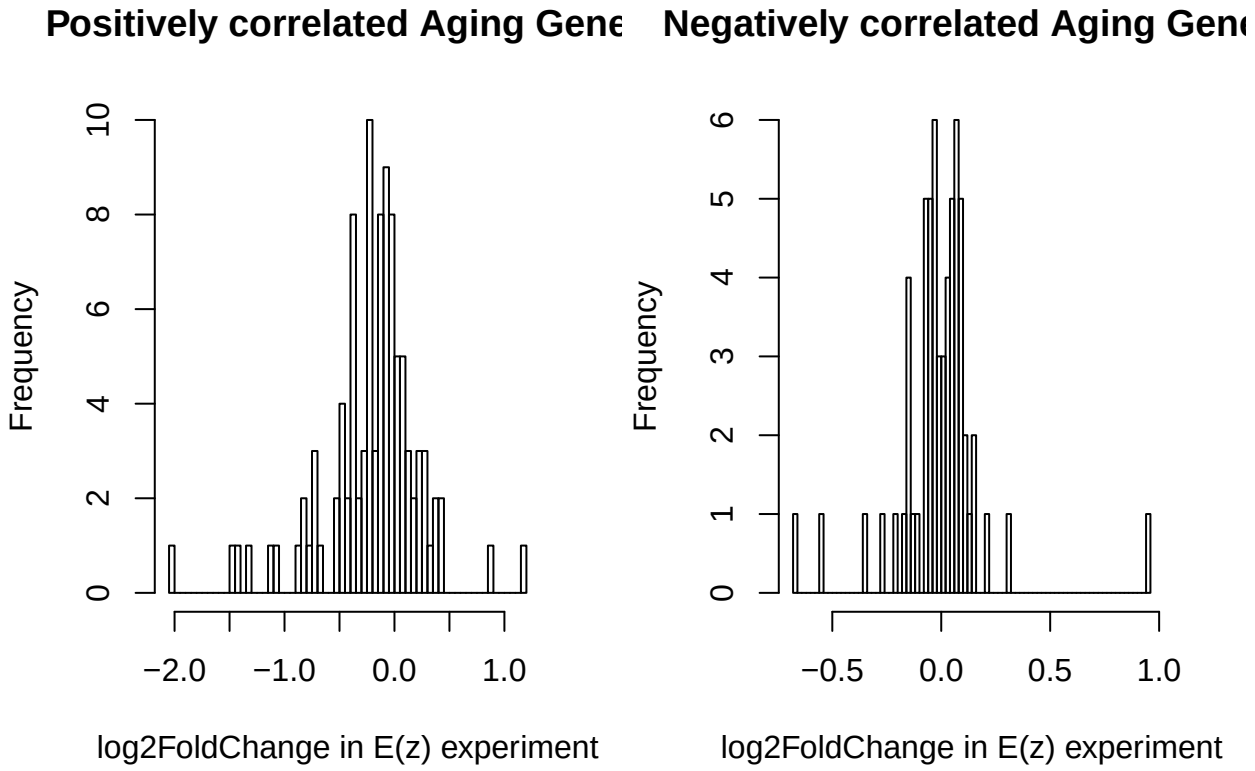
```
##      Probeset.ID Fly.gene.ID Fly.gene.   Beta1  p.value FDR..q.value.
## 50   1632683_s_at        <NA>      <NA>  58.816 4.48e-05        0.0159
## 64    1631089_at        <NA>      <NA>  31.825 2.18e-04        0.0324
## 120 1624543_s_at        <NA>      <NA>  36.057 8.12e-05        0.0203
## 140 1638075_a_at     CG10494      <NA> -73.867 4.75e-04        0.0472
##     control.log2.20d.3d.. mir.34.....log2.20d.3d. DIR2         dd3
## 50              0.2535215               0.5540302    1  0.30050871
## 64              0.7630460               0.9473263    1  0.18428027
## 120             0.5560126               0.5036011    1 -0.05241152
## 140            -0.7401696              -0.5919468   -1  0.14822273
##         dd.dir4 ensembl   cg
## 50   0.30050871    <NA> <NA>
## 64   0.18428027    <NA> <NA>
## 120 -0.05241152    <NA> <NA>
## 140 -0.14822273    <NA> <NA>
```

```
# of CG10494:
ageCorGenes[140, 12] <- "FBgn0034634"
# CG5953 Or copia/GIP
ageCorGenes[50, 12] <- "FBgn0032587" # OR FBgn0013437
# Two others that are not associated with FBgn
ageCorGenes[is.na(ageCorGenes$ensembl),]
```

```
##      Probeset.ID Fly.gene.ID Fly.gene.  Beta1  p.value FDR..q.value.
## 64    1631089_at        <NA>      <NA> 31.825 2.18e-04        0.0324
## 120 1624543_s_at        <NA>      <NA> 36.057 8.12e-05        0.0203
##     control.log2.20d.3d.. mir.34.....log2.20d.3d. DIR2         dd3
## 64              0.7630460               0.9473263    1  0.18428027
## 120             0.5560126               0.5036011    1 -0.05241152
##         dd.dir4 ensembl   cg
## 64   0.18428027    <NA> <NA>
## 120 -0.05241152    <NA> <NA>
```

```
ageCorGenesPos <- ageCorGenes[ageCorGenes$DIR2==1,]
ageCorGenesNeg <- ageCorGenes[ageCorGenes$DIR2==-1,]

par(mfrow=c(1,2))
hist(results[rownames(results)%in%ageCorGenesPos$ensembl,]$log2FoldChange, breaks=100,
     main="Positively correlated Aging Genes", xlab="log2FoldChange in E(z) experiment")
hist(results[rownames(results)%in%ageCorGenesNeg$ensembl,]$log2FoldChange, breaks=100,
     main="Negatively correlated Aging Genes", xlab="log2FoldChange in E(z) experiment")
```




```
########################Prepare Publication Plot#############################
jpeg(file=paste(outpath, "Fig8b_Positively_age_Cor_in_Ez_MLE.jpg",
                sep="/"),
     quality=100,
     res=300,
     width=1440,
     height=960)
resultsMLE$ageCorGenPos <- rownames(resultsMLE) %in% ageCorGenesPos$ensembl
quartzFonts(arial=c("Arial", "Arial Italic", "Arial Bold", "Arial Bold Italic"))
par(mai=c(0,0,0,0), mar=c(2,4.5,3,1)+.5, family="arial")
plotMA(resultsMLE, alpha = 0.05, 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$ageCorGenPos==TRUE,],
     {points(baseMean, lfcMLE, col = "red", cex=1, lwd=1)})
dev.off()
```

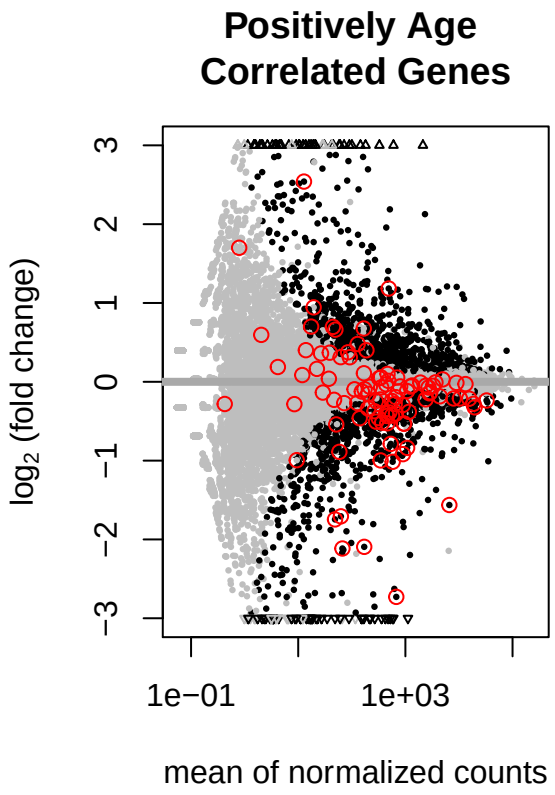
```
## pdf
##   2
```

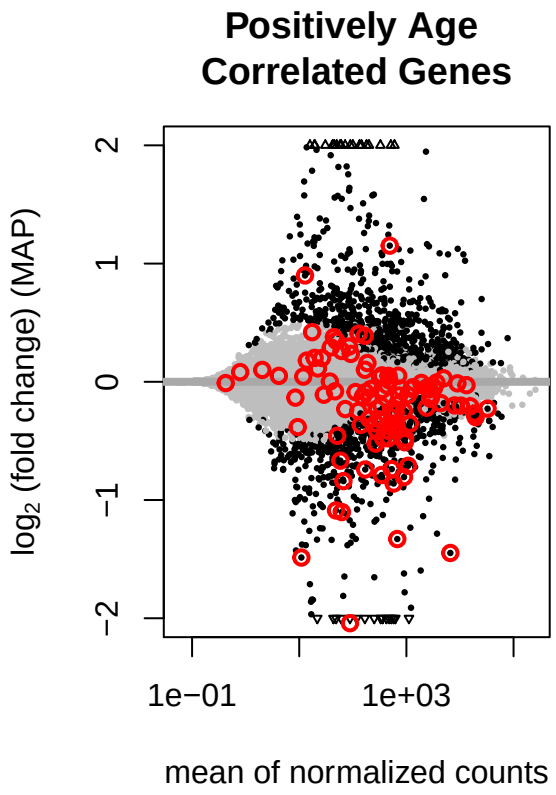
```
################################################################################
plotMA(resultsMLE, alpha = 0.05, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       colLine = "darkgrey", MLE=TRUE,
       ylab=expression("log"[2]*" (fold change)"),
       main="Positively Age \nCorrelated Genes")
with(resultsMLE[resultsMLE$ageCorGenPos==TRUE,], {points(baseMean, lfcMLE,
                                                         col = "red", cex=1, lwd=1)})

jpeg(file=paste(outpath, "Positively_age_Cor_in_Ez_MAP.jpg", sep="/"),
     width=360, height=240)
results$ageCorGenPos <- rownames(results) %in% ageCorGenesPos$ensembl
plotMA(results, alpha = 0.01, ylim=c(-2,2), colNonSig = "gray", colSig = "black", colLine = "darkgrey",
       ylab=expression("log"[2]*" (fold change) (MAP)"),
       main="Positively Age \nCorrelated Genes")
with(results[results$ageCorGenPos==TRUE,], {points(baseMean,
                                                   log2FoldChange,
                                                   col = "red", cex=1, lwd=2)})
dev.off()
```

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

```
plotMA(results, alpha = 0.01, ylim=c(-2,2), colNonSig = "gray", colSig = "black", colLine = "darkgrey",
       ylab=expression("log"[2]*" (fold change) (MAP)"),
       main="Positively Age \nCorrelated Genes")
with(results[results$ageCorGenPos==TRUE,], {points(baseMean,
                                                   log2FoldChange,
                                                   col = "red", cex=1, lwd=2)})
```






```
#########################Prepare Publication Plot#############################
jpeg(file=paste(outpath, "FigS5e_Negatively_age_Cor_in_Ez_MLE.jpg", sep="/"),
     width=360, height=240)
resultsMLE$ageCorGenNeg <- rownames(resultsMLE) %in% ageCorGenesNeg$ensembl
par(mai=c(0,0,0,0), mar=c(2,4,3,3)+.5)
plotMA(resultsMLE, alpha = 0.05, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       colLine = "darkgrey", MLE=TRUE, las=1,
       ylab=expression("log"[2]*" (fold change)"),xlab="", xaxp=c(1,2,1), xaxt="n",
       #main="Negatively Correlated Age \nGenes are Unchanged"
       main=""
       )
axis(1, at=c(1, 10, 100, 1000, 10000), labels=c(1, 10, 100, 1000, 10000))
with(resultsMLE[resultsMLE$ageCorGenNeg==TRUE,], {points(baseMean, lfcMLE,
                                                         col = "red", cex=1, lwd=1)})
dev.off()
```

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

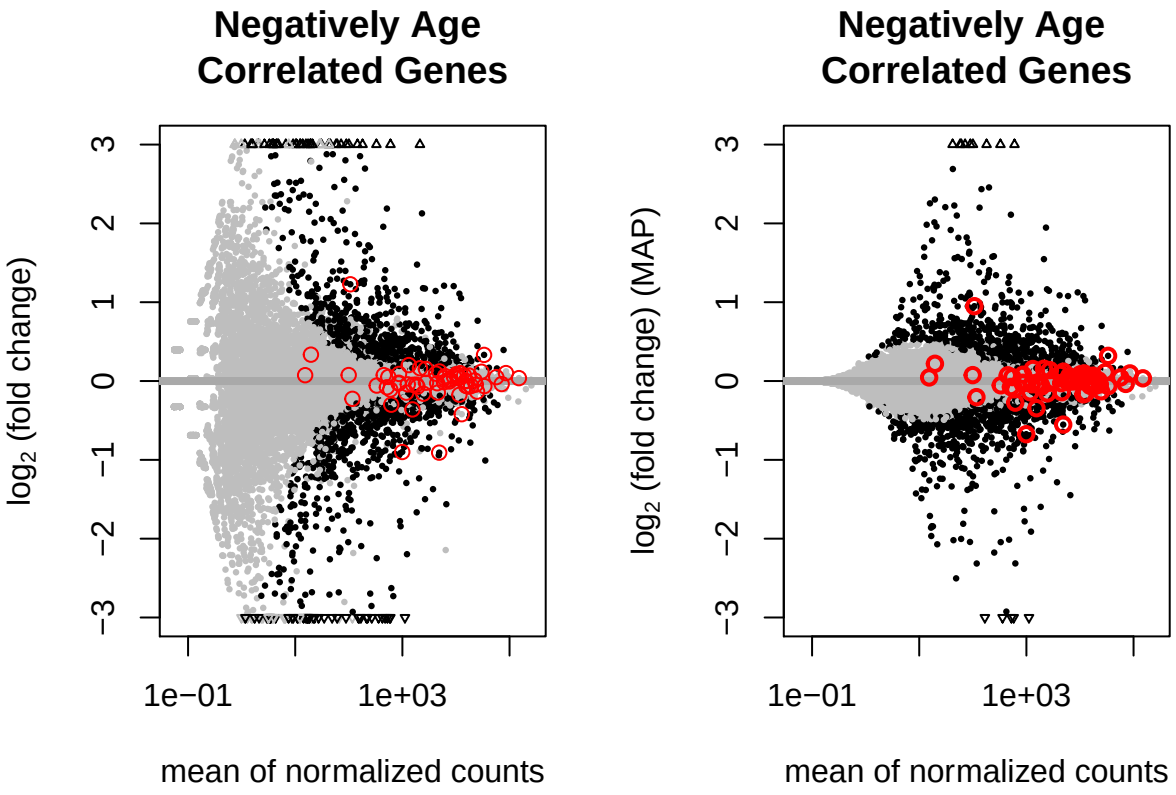
```
###############################################################################
plotMA(resultsMLE, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       colLine = "darkgrey", MLE=TRUE,
       ylab=expression("log"[2]*" (fold change)"),
       main="Negatively Age \nCorrelated Genes")
with(resultsMLE[resultsMLE$ageCorGenNeg==TRUE,], {points(baseMean, lfcMLE,
                                                         col = "red", cex=1, lwd=1)})

jpeg(file=paste(outpath, "Negatively_age_Cor_in_Ez_MAP.jpg", sep="/"),
     width=360, height=240)
results$ageCorGenNeg <- rownames(results) %in% ageCorGenesNeg$ensembl
```

```
plotMA(results, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       colLine = "darkgrey",
       ylab=expression("log"[2]*" (fold change) (MAP)"),
       main="Negatively Age \nCorrelated Genes")
with(results[results$ageCorGenNeg==TRUE,], {points(baseMean, log2FoldChange,
                                                   col = "red", cex=1, lwd=2)})
dev.off()
```

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

```
plotMA(results, alpha = 0.01, ylim=c(-3,3), colNonSig = "gray", colSig = "black",
       colLine = "darkgrey",
       ylab=expression("log"[2]*" (fold change) (MAP)"),
       main="Negatively Age \nCorrelated Genes")
with(results[results$ageCorGenNeg==TRUE,], {points(baseMean, log2FoldChange,
                                                   col = "red", cex=1, lwd=2)})
```



Is this result significant? Using the hypergeometric test

```
ageCorResults <- results[results$ageCorGenPos==TRUE,]
ageCorResults <- ageCorResults[!is.na(ageCorResults$padj),]
myGenes <- ageCorResults[ageCorResults$log2FoldChange < -0.5 & ageCorResults$padj < 0.05,]
# Get number of Age Correlated Genes that are downregulated in E(z) mutants
dim(myGenes)[1]
```

```
## [1] 16
```

```r
# Hypergeometric test for over-representation:
phyper(15, 108, 18845, 317, lower.tail=FALSE)
```

```
## [1] 3.268416e-11
```

```r
#      upInEz*and*AgeRegulated - 1 = 15
#      number of age correlated genes from microarray data = 108
#      number of genes NOT age correlated genes from microarray data = 18953-108
#      SampleSize = 317 (genes downregulated in E(z) mutants)


# Hypergeometric test for under-representation:
phyper(16, 108, 18845, 317, lower.tail=TRUE)
```

```
## [1] 1
```

```r
# Fisher Exact test for over-representation:
contingency.matrix <- rbind(c(16,92), c(301, 18544))
contingency.matrix
```

```
##      [,1]  [,2]
## [1,]   16    92
## [2,]  301 18544
```

```r
fisher.test(contingency.matrix, alternative="greater")$p.value
```

```
## [1] 3.268416e-11
```

Save image

```r
save.image(file="hyperGeo.RData")
```