Prepare a heatmap of the positively age correlated genes for publication.

Jason Kennerdell

8/12/2018

Input the data and metadata:

```
inpath <- "~/Desktop/brain"
outpath <- "~/Desktop/brain/_8A_ageCorGenes_htmp"

colSet <- brewer.pal(12, "Paired")
files <- list.files(inpath)
htseq_files <- files[grepl("^JKL.*txt$", files)]
sampleNames <- read.csv(file.path(inpath, "SampleNames.csv"))
sampleTable <- data.frame(fileName = htseq_files,
                          stringsAsFactors=FALSE)
sampleTable$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[,1:3], by = "seq.batch")
sampleTable$genotype <- gsub("-.*$", "", sampleTable$Sample)
sampleTable$genotype <- gsub("JKLY.... ", "", sampleTable$genotype)
sampleTable$genotype <- gsub("\\[1118\\]", "", sampleTable$genotype)
sampleTable$genotype <- factor(sampleTable$genotype, levels = c("w", "Ez"))
sampleTable <- sampleTable[!is.na(sampleTable$genotype),]
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$age, levels = c("3d", "20d"))
sampleTable$condition <- paste(sampleTable$genotype, sampleTable$Temp,
                               sampleTable$age, sep = "-")
sampleTable$color <- c(rep(colSet[7], 4), rep(colSet[9], 2),
                       rep(colSet[8], 2), rep(colSet[10], 2),
                       rep(colSet[8], 2), rep(colSet[10], 2),
                       rep(colSet[8], 2), rep(colSet[10], 2),
                       rep(colSet[8], 2), rep(colSet[1], 2),
                       rep(colSet[10], 2), rep(colSet[8], 2),
                       rep(colSet[10], 2), rep(colSet[1], 4),
                       rep(colSet[7], 2), rep(colSet[9], 4))
sampleTable
```

```
##    seq.batch          fileName Library          Sample batch genotype
## 3    JKL10 A  JKL10-counts.txt   JKL10          w-25-II     A        w
## 4    JKL10 B JKL10b-counts.txt   JKL10          w-25-II     B        w
## 5    JKL11 A  JKL11-counts.txt   JKL11         w-25-III     A        w
## 6    JKL11 B JKL11b-counts.txt   JKL11         w-25-III     B        w
```

```
## 7    JKL12 A  JKL12-counts.txt   JKL12        Ez-25-III     A       Ez
## 8    JKL12 B JKL12b-counts.txt   JKL12        Ez-25-III     B       Ez
## 9    JKL13 A  JKL13-counts.txt   JKL13           w-29-I     A        w
## 10   JKL13 B JKL13b-counts.txt   JKL13           w-29-I     B        w
## 11   JKL14 A  JKL14-counts.txt   JKL14          Ez-29-I     A       Ez
## 12   JKL14 B JKL14b-counts.txt   JKL14          Ez-29-I     B       Ez
## 13   JKL15 A  JKL15-counts.txt   JKL15          w-29-II     A        w
## 14   JKL15 B JKL15b-counts.txt   JKL15          w-29-II     B        w
## 15   JKL16 A  JKL16-counts.txt   JKL16         Ez-29-II     A       Ez
## 16   JKL16 B JKL16b-counts.txt   JKL16         Ez-29-II     B       Ez
## 17   JKL17 A  JKL17-counts.txt   JKL17         w-29-III     A        w
## 18   JKL17 B JKL17b-counts.txt   JKL17         w-29-III     B        w
## 19   JKL18 A  JKL18-counts.txt   JKL18        Ez-29-III     A       Ez
## 20   JKL18 B JKL18b-counts.txt   JKL18        Ez-29-III     B       Ez
## 21   JKL19 A  JKL19-counts.txt   JKL19          w-29-IV     A        w
## 22   JKL19 B JKL19b-counts.txt   JKL19          w-29-IV     B        w
## 23    JKL2 A   JKL2-counts.txt    JKL2 JKLY1125 w[1118]     A        w
## 24    JKL2 B  JKL2b-counts.txt    JKL2 JKLY1125 w[1118]     B        w
## 25   JKL20 A  JKL20-counts.txt   JKL20         Ez-29-IV     A       Ez
## 26   JKL20 B JKL20b-counts.txt   JKL20         Ez-29-IV     B       Ez
## 27   JKL21 A  JKL21-counts.txt   JKL21           w-29-V     A        w
## 28   JKL21 B JKL21b-counts.txt   JKL21           w-29-V     B        w
## 29   JKL22 A  JKL22-counts.txt   JKL22          Ez-29-V     A       Ez
## 30   JKL22 B JKL22b-counts.txt   JKL22          Ez-29-V     B       Ez
## 33    JKL4 A   JKL4-counts.txt    JKL4 JKLY1127 w[1118]     A        w
## 34    JKL4 B  JKL4b-counts.txt    JKL4 JKLY1127 w[1118]     B        w
## 37    JKL6 A   JKL6-counts.txt    JKL6 JKLY1129 w[1118]     A        w
## 38    JKL6 B  JKL6b-counts.txt    JKL6 JKLY1129 w[1118]     B        w
## 39    JKL7 A   JKL7-counts.txt    JKL7           w-25-I     A        w
## 40    JKL7 B  JKL7b-counts.txt    JKL7           w-25-I     B        w
## 41    JKL8 A   JKL8-counts.txt    JKL8          Ez-25-I     A       Ez
## 42    JKL8 B  JKL8b-counts.txt    JKL8          Ez-25-I     B       Ez
## 43    JKL9 A   JKL9-counts.txt    JKL9         Ez-25-II     A       Ez
## 44    JKL9 B  JKL9b-counts.txt    JKL9         Ez-25-II     B       Ez
##    Temp age condition   color
## 3    25 20d  w-25-20d #FDBF6F
## 4    25 20d  w-25-20d #FDBF6F
## 5    25 20d  w-25-20d #FDBF6F
## 6    25 20d  w-25-20d #FDBF6F
## 7    25 20d Ez-25-20d #CAB2D6
## 8    25 20d Ez-25-20d #CAB2D6
## 9    29 20d  w-29-20d #FF7F00
## 10   29 20d  w-29-20d #FF7F00
## 11   29 20d Ez-29-20d #6A3D9A
## 12   29 20d Ez-29-20d #6A3D9A
## 13   29 20d  w-29-20d #FF7F00
## 14   29 20d  w-29-20d #FF7F00
## 15   29 20d Ez-29-20d #6A3D9A
## 16   29 20d Ez-29-20d #6A3D9A
## 17   29 20d  w-29-20d #FF7F00
## 18   29 20d  w-29-20d #FF7F00
## 19   29 20d Ez-29-20d #6A3D9A
## 20   29 20d Ez-29-20d #6A3D9A
## 21   29 20d  w-29-20d #FF7F00
```

```
## 22   29 20d  w-29-20d #FF7F00
## 23   25  3d   w-25-3d #A6CEE3
## 24   25  3d   w-25-3d #A6CEE3
## 25   29 20d Ez-29-20d #6A3D9A
## 26   29 20d Ez-29-20d #6A3D9A
## 27   29 20d  w-29-20d #FF7F00
## 28   29 20d  w-29-20d #FF7F00
## 29   29 20d Ez-29-20d #6A3D9A
## 30   29 20d Ez-29-20d #6A3D9A
## 33   25  3d   w-25-3d #A6CEE3
## 34   25  3d   w-25-3d #A6CEE3
## 37   25  3d   w-25-3d #A6CEE3
## 38   25  3d   w-25-3d #A6CEE3
## 39   25 20d  w-25-20d #FDBF6F
## 40   25 20d  w-25-20d #FDBF6F
## 41   25 20d Ez-25-20d #CAB2D6
## 42   25 20d Ez-25-20d #CAB2D6
## 43   25 20d Ez-25-20d #CAB2D6
## 44   25 20d Ez-25-20d #CAB2D6
```

Set up the statistical model to test for Differentially Expressed genes in E(z) mutants:

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

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

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

```
ageCorGenes <- read.xlsx(paste(inpath, "nature10810-s2.xls", sep="/"),
                         sheetName="Age_Gene_tbl", startRow=3, endRow=177)
# 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")
#listAttributes(mart)
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]
# Four probesets have NA ensembl names
# 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

ageCorGenesPos <- ageCorGenes[ageCorGenes$DIR2==1,]
ageCorGenesNeg <- ageCorGenes[ageCorGenes$DIR2==-1,]
ageCorGenesPos[is.na(ageCorGenesPos$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>
```

```
ageCorGenesNeg[is.na(ageCorGenesNeg$ensembl),]
```

```
##  [1] Probeset.ID             Fly.gene.ID
##  [3] Fly.gene.               Beta1
##  [5] p.value                 FDR..q.value.
##  [7] control.log2.20d.3d..   mir.34.....log2.20d.3d.
##  [9] DIR2                    dd3
## [11] dd.dir4                 ensembl
## [13] cg
## <0 rows> (or 0-length row.names)
```

Prepare rlog transformed data:

```
rld <- rlog(dds, blind=FALSE)
```

Get E(z) data statistics data without the influence of 3d data in the model:

```
design2 <- formula(~ Temp + genotype)
dds2 <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable[sampleTable$age != "3d",],
                                   directory = inpath, design = design2)
# Combine the technical replicates (different runs) by adding the count
# totals for each gene across the two runs:
dds2 <- collapseReplicates(dds2, groupby=dds2$Library, run = dds2$batch)
dds2 <- DESeq(dds2)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
## fitting model and testing
```

```
# What does the data look like?
head(assay(dds2))
```

```
##             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(dds2)
```

```
## DataFrame with 16 rows and 10 columns
##           Library    Sample    batch genotype     Temp      age
##       <character>  <factor> <factor> <factor> <factor> <factor>
## JKL10       JKL10   w-25-II        A        w       25      20d
## JKL11       JKL11  w-25-III        A        w       25      20d
## JKL12       JKL12 Ez-25-III        A       Ez       25      20d
## JKL13       JKL13    w-29-I        A        w       29      20d
## JKL14       JKL14   Ez-29-I        A       Ez       29      20d
## ...           ...       ...      ...      ...      ...      ...
## JKL21       JKL21    w-29-V        A        w       29      20d
## JKL22       JKL22   Ez-29-V        A       Ez       29      20d
## JKL7         JKL7    w-25-I        A        w       25      20d
## JKL8         JKL8   Ez-25-I        A       Ez       25      20d
## JKL9         JKL9  Ez-25-II        A       Ez       25      20d
##         condition       color runsCollapsed sizeFactor
##       <character> <character>   <character>  <numeric>
## JKL10    w-25-20d     #FDBF6F           A,B  0.9090574
## JKL11    w-25-20d     #FDBF6F           A,B  1.1403424
## JKL12   Ez-25-20d     #CAB2D6           A,B  1.0027251
## JKL13    w-29-20d     #FF7F00           A,B  1.0220507
## JKL14   Ez-29-20d     #6A3D9A           A,B  0.9570594
## ...           ...         ...           ...        ...
## JKL21    w-29-20d     #FF7F00           A,B  0.8750556
## JKL22   Ez-29-20d     #6A3D9A           A,B  0.9618390
## JKL7     w-25-20d     #FDBF6F           A,B  1.1341278
## JKL8    Ez-25-20d     #CAB2D6           A,B  1.1093046
## JKL9    Ez-25-20d     #CAB2D6           A,B  1.1303920
```

```
datEz <- results(dds2, alpha=0.05)
datEz$ensembl <- rownames(datEz)
```

Heatmap for age-correlated genes only:

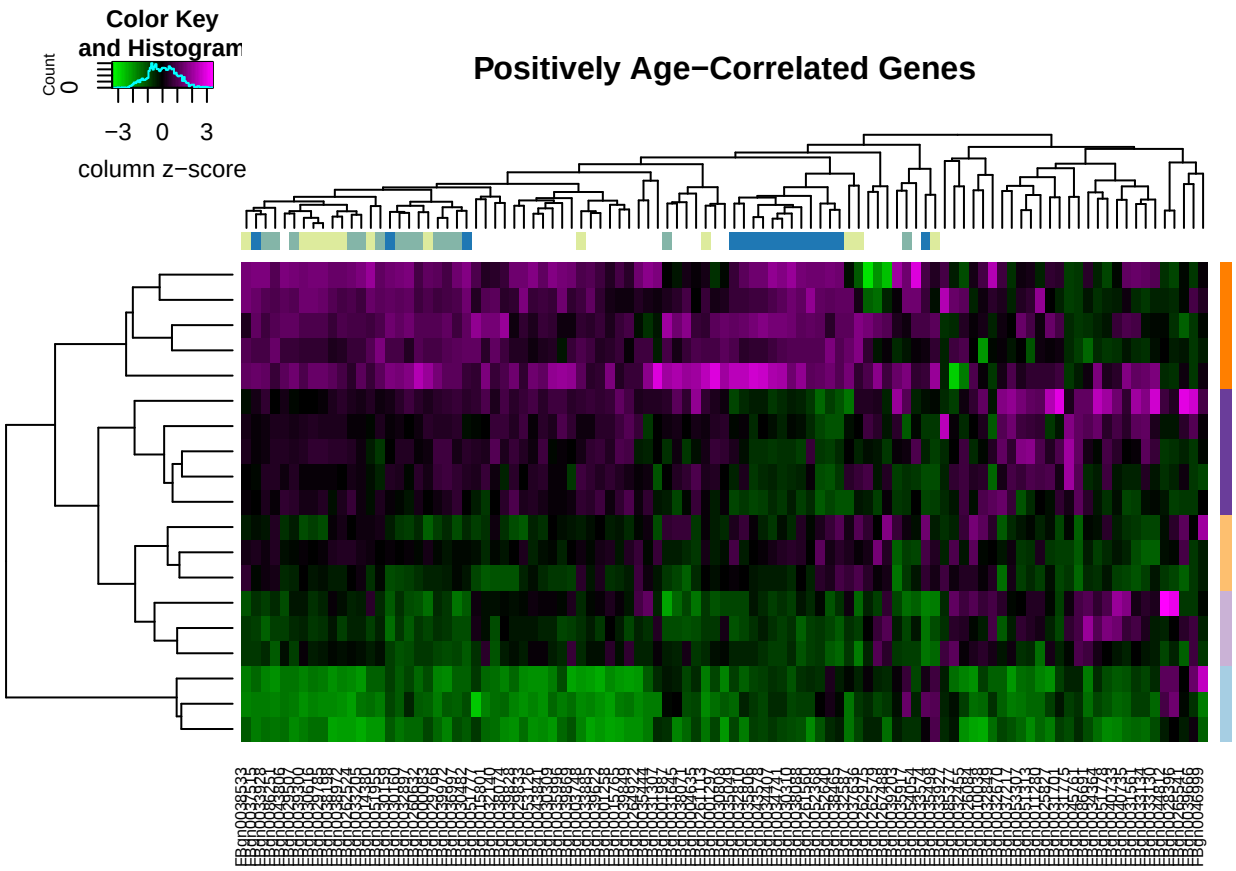
```
# Make a function to assign genes to categories, shaded by log2FoldChange:
calledEz <- subset(datEz, padj < 0.05 & abs(log2FoldChange) > 0.5)
relaxedEz <- subset(datEz, padj < 0.05 & abs(log2FoldChange) > 0.3)
relaxedrelaxedEz <- subset(datEz, padj < 0.05 & abs(log2FoldChange) > 0.15)
Ezassign <- function(x){
  y <- rep(NA, length(x))
  for(i in 1:length(x))
    if(x[i] %in% calledEz[calledEz$log2FoldChange<0,7])
      #y[i] <- colSet[2]
      y[i] <- "#1F78B4"
    else if(x[i] %in% relaxedEz[relaxedEz$log2FoldChange<0,7])
      #y[i] <- colSet[12]
      y[i] <- "#84B5A8"
    else if(x[i] %in% relaxedrelaxedEz[relaxedrelaxedEz$log2FoldChange<0,7])
      #y[i] <- colSet[11]
      y[i] <- "#DDEB9D"
    else
```

```
      y[i] <- NA
  y
}

ageCorPosIndex <- which(rownames(rld) %in% ageCorGenesPos$ensembl)
# Remove one gene that is not expressed in brains (Hsp22)
ageCorPosIndex <- ageCorPosIndex[-2]
dat  <- scale(t(assay(rld)[ageCorPosIndex,]))
attr(dat, "color") <- colData(dds)$color
attr(dat, "condition") <- colData(dds)$condition
gene <- attr(dat, which = "dimnames")[[2]]
myColors <- Ezassign(gene)
table(myColors)
```

```
## myColors
## #1F78B4 #84B5A8 #DDEB9D
##      16      14      13
```

```
# Print out a figure
par(cex.main=1)
htmp <- heatmap.2(dat,
          lmat=rbind(c(6,5,0), c(0,2,0), c(4,3,1)),
          lwid=c(1, 4, .1),
          lhei=c(1.5, 0.2, 4),
          cexRow=1.2, cexCol=0.75, scale="none", offsetRow=-1, srtRow=45,
          col=colorpanel(75, "green", "black", "magenta"),
          trace="none",
          RowSideColors = attr(dat, which = "color"),
          ColSideColors = myColors,
          margins=c(5,0.5),
          key.par=list(cex.main=1),
          key=T, main="Positively Age-Correlated Genes",
          key.xlab = "column z-score",
          labRow=NA
          )
```




```
jpeg(file=paste(outpath, "Fig_8a_AgeCorPosGreenMag.jpg", sep="/"),
     quality=100,
     res=300,
     height=1920,
     width=3840)
par(cex.main=1, bg="transparent")
heatmap.2(dat,
          lmat=rbind(c(6,5,0), c(0,2,0), c(4,3,1)),
          lwid=c(1, 6, .1),
          lhei=c(1.5, 0.3, 4),
          cexRow=1.2, cexCol=0.9, scale="none",
          offsetRow=-1,
          offsetCol = -0.5,
          srtRow=45,
          col=colorpanel(75, "green", "black", "magenta"),
          trace="none",
          RowSideColors = attr(dat, which = "color"),
          ColSideColors = myColors,
          margins=c(5,0.5),
          key.par=list(cex.main=1),
          key=T, main="", key.title="",
          key.xlab = "", key.ylab="",
          labRow=NA
          )
dev.off()
```

```
## pdf
```

```
##   2
```

```
# Which are these 16 genes?
hits <- intersect(calledEz[calledEz$log2FoldChange<0,7],ageCorGenesPos$ensembl)
calledEz[calledEz$ensembl %in% hits,]
```

```
## log2 fold change (MAP): genotype Ez vs w
## Wald test p-value: genotype Ez vs w
## DataFrame with 16 rows and 7 columns
##               baseMean log2FoldChange      lfcSE       stat       pvalue
##              <numeric>      <numeric>  <numeric>  <numeric>    <numeric>
## FBgn0030159  903.40013     -0.8058498  0.1068034  -7.545168 4.517053e-14
## FBgn0030310   62.52550     -1.1017723  0.1526684  -7.216768 5.323750e-13
## FBgn0030482  578.84903     -0.8604131  0.1137411  -7.564666 3.888610e-14
## FBgn0032810   58.59347     -0.6669497  0.1374054  -4.853882 1.210677e-06
## FBgn0033574   48.60368     -1.0880644  0.1511080  -7.200573 5.995998e-13
## ...                ...            ...        ...        ...          ...
## FBgn0038465 6612.20978     -1.4478882 0.08271313 -17.504938 1.313689e-68
## FBgn0043578  170.52543     -0.7415507 0.14974260  -4.952170 7.339060e-07
## FBgn0052368  676.41495     -1.3299033 0.15802778  -8.415630 3.907882e-17
## FBgn0052640   10.96169     -1.4864518 0.15807399  -9.403520 5.276775e-21
## FBgn0261560  344.98792     -0.7884363 0.13302505  -5.926977 3.085627e-09
##                    padj     ensembl
##                <numeric> <character>
## FBgn0030159 1.781031e-12 FBgn0030159
## FBgn0030310 1.927359e-11 FBgn0030310
## FBgn0030482 1.553753e-12 FBgn0030482
## FBgn0032810 1.962545e-05 FBgn0032810
## FBgn0033574 2.151177e-11 FBgn0033574
## ...                 ...         ...
## FBgn0038465 3.138928e-66 FBgn0038465
## FBgn0043578 1.233189e-05 FBgn0043578
## FBgn0052368 1.992009e-15 FBgn0052368
## FBgn0052640 3.444898e-19 FBgn0052640
## FBgn0261560 7.492681e-08 FBgn0261560
```