

# Mouse DGEs of SACC-PHJs mono-infected with HBV or coinfecte with HBV/HDV\_factors donor time\_samples up to July 2018

## Purpose:

To determine the DGE profiles (for mouse genes), relative to uninfected controls, of self-assembling co-cultures of primary human hepatocytes (SACC-PHJs) (co-cultured with 3T3J mouse non-parenchymal cells) mono-infected with HBV or co-infected with HBV/HDV at 8 and 28 days post-infection. Here, donor and time are factors in the design. This run includes the samples sequenced in July 2018.

```
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(stringr)
library(ggplot2)
library(reshape2)
library(openxlsx)
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:dplyr':
## 
##     combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':
## 
##     IQR, mad, xtabs

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

```

## 
##   anyDuplicated, append, as.data.frame, cbind, colnames,
##   do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##   grepl, intersect, is.unsorted, lapply, lengths, 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 objects are masked from 'package:dplyr':
## 
##     first, rename

## The following objects are masked from 'package:base':
## 
##     colMeans, colSums, expand.grid, rowMeans, rowSums

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':
## 
##     collapse, desc, slice

## 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")'.

library(gplots)

##
## Attaching package: 'gplots'

## The following object is masked from 'package:IRanges':
## 
##     space

## The following object is masked from 'package:S4Vectors':
## 
##     space

## The following object is masked from 'package:stats':
## 
##     lowess

library(dplyr)
library(RColorBrewer)

```

```

library(stringr)
library(genefilter)
library(data.table)

##
## Attaching package: 'data.table'
## The following object is masked from 'package:SummarizedExperiment':
##
##     shift

## The following object is masked from 'package:GenomicRanges':
##
##     shift

## The following object is masked from 'package:IRanges':
##
##     shift

## The following objects are masked from 'package:S4Vectors':
##
##     first, second

## The following objects are masked from 'package:reshape2':
##
##     dcast, melt

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

library(genefilter)
library(ggrepel)
library(viridis)

## Loading required package: viridisLite
source("http://bioconductor.org/biocLite.R")

## Bioconductor version 3.4 (BiocInstaller 1.24.0), ?biocLite for help
## A new version of Bioconductor is available after installing the most
##   recent version of R; see http://bioconductor.org/install
biocLite("org.Mm.eg.db", suppressUpdates = TRUE)

## BioC_mirror: https://bioconductor.org
## Using Bioconductor 3.4 (BiocInstaller 1.24.0), R 3.3.3 (2017-03-06).
## Installing package(s) 'org.Mm.eg.db'
## installing the source package 'org.Mm.eg.db'
require(org.Mm.eg.db)

## Loading required package: org.Mm.eg.db
## Loading required package: AnnotationDbi
##
## Attaching package: 'AnnotationDbi'

```

```

## The following object is masked from 'package:dplyr':
##
##      select
##
Pulling in the mouse counts files
##Pulling in the counts of only mouse genes as determined previously in my
##DGE_sorting analysis.
mousecounts <- "All mouse genes"
mouse_sampleCounts <- basename(Sys.glob(file.path(mousecounts, "*.txt")))

##Function to read in the feature counts
exptcounts <- function(files) {
  d <- read.table(files)
  d
}

##Read in all of the count files
mousecounts_readin <- lapply(file.path(mousecounts, mouse_sampleCounts), exptcounts)
names(mousecounts_readin) <- sub('mousegenes', '', mouse_sampleCounts)
names(mousecounts_readin)

## [1] "BD330 HBV_HDV Day 28 sample 1.txt"
## [2] "BD330 HBV_HDV Day 28 sample 2.txt"
## [3] "BD330 HBV_HDV Day 28 sample 3.txt"
## [4] "BD330 HBV_HDV Day 8 sample 1.txt"
## [5] "BD330 HBV_HDV Day 8 sample 2.txt"
## [6] "BD330 HBV_HDV Day 8 sample 3.txt"
## [7] "BD330_Ctrl_D28.txt"
## [8] "BD330_Ctrl_D8.txt"
## [9] "BD330_HBV_D28.txt"
## [10] "BD330_HBV_D8.txt"
## [11] "BD330_HBV_HDV_D28_b.txt"
## [12] "BD330_HBV_HDV_D28.txt"
## [13] "BD330_HBV_HDV_D8_a.txt"
## [14] "BD330_HBV_HDV_D8.txt"
## [15] "BD405A HBV_HDV D28 sample 1.txt"
## [16] "BD405A HBV_HDV D28 sample 2.txt"
## [17] "BD405A HBV_HDV D28 sample 3.txt"
## [18] "BD405A HBV_HDV D8 sample 1.txt"
## [19] "BD405A HBV_HDV D8 sample 2.txt"
## [20] "BD405A HBV_HDV D8 sample 3.txt"
## [21] "BD405A_Ctrl_D28.txt"
## [22] "BD405A_Ctrl_D8.txt"
## [23] "BD405A_HBV_D28.txt"
## [24] "BD405A_HBV_D8.txt"
## [25] "BD405A_HBV_HDV_D28.txt"
## [26] "BD405A_HBV_HDV_D8.txt"
## [27] "Ctrl_D28_sample_1.txt"
## [28] "Ctrl_D28_sample_2.txt"
## [29] "Ctrl_D28_sample_3.txt"
## [30] "Ctrl_D8_sample_1.txt"
## [31] "Ctrl_D8_sample_2.txt"
## [32] "Ctrl_D8_sample_3.txt"

```

```

## [33] "HBV_D28_sample_1.txt"
## [34] "HBV_D28_sample_2.txt"
## [35] "HBV_D28_sample_3.txt"
## [36] "HBV_D8_sample_1.txt"
## [37] "HBV_D8_sample_2.txt"
## [38] "HBV_D8_sample_3.txt"
## [39] "HU1016 Ctrl D28.txt"
## [40] "HU1016 Ctrl D8.txt"
## [41] "HU1016_BD_co_D28.txt"
## [42] "HU1016_BD_co_D8.txt"
## [43] "HU1016_B_D28.txt"
## [44] "HU1016_B_D8.txt"

Now subset feature counts by “treatment” for DGE analysis.

ctrl <- names(mousecounts_readin)[grep("*Ctrl", names(mousecounts_readin))]
ctrl_counts <- mousecounts_readin[match(ctrl, names(mousecounts_readin))]
names(ctrl_counts)

## [1] "BD330_Ctrl_D28.txt"      "BD330_Ctrl_D8.txt"
## [3] "BD405A_Ctrl_D28.txt"    "BD405A_Ctrl_D8.txt"
## [5] "Ctrl_D28_sample_1.txt"   "Ctrl_D28_sample_2.txt"
## [7] "Ctrl_D28_sample_3.txt"   "Ctrl_D8_sample_1.txt"
## [9] "Ctrl_D8_sample_2.txt"    "Ctrl_D8_sample_3.txt"
## [11] "HU1016 Ctrl D28.txt"    "HU1016 Ctrl D8.txt"

HBV <- names(mousecounts_readin)[grep("*HBV_D|_B_", names(mousecounts_readin))]
HBV_counts <- mousecounts_readin[match(HBV, names(mousecounts_readin))]
names(HBV_counts)

## [1] "BD330_HBV_D28.txt"      "BD330_HBV_D8.txt"      "BD405A_HBV_D28.txt"
## [4] "BD405A_HBV_D8.txt"      "HBV_D28_sample_1.txt"  "HBV_D28_sample_2.txt"
## [7] "HBV_D28_sample_3.txt"    "HBV_D8_sample_1.txt"    "HBV_D8_sample_2.txt"
## [10] "HBV_D8_sample_3.txt"    "HU1016_B_D28.txt"     "HU1016_B_D8.txt"

coinf <- names(mousecounts_readin)[grep("*HBV_HDV|_co_", names(mousecounts_readin))]
coinf_counts <- mousecounts_readin[match(coinf, names(mousecounts_readin))]
names(coinf_counts)

## [1] "BD330 HBV_HDV Day 28 sample 1.txt"
## [2] "BD330 HBV_HDV Day 28 sample 2.txt"
## [3] "BD330 HBV_HDV Day 28 sample 3.txt"
## [4] "BD330 HBV_HDV Day 8 sample 1.txt"
## [5] "BD330 HBV_HDV Day 8 sample 2.txt"
## [6] "BD330 HBV_HDV Day 8 sample 3.txt"
## [7] "BD330_HBV_HDV_D28_b.txt"
## [8] "BD330_HBV_HDV_D28.txt"
## [9] "BD330_HBV_HDV_D8_a.txt"
## [10] "BD330_HBV_HDV_D8.txt"
## [11] "BD405A HBV_HDV D28 sample 1.txt"
## [12] "BD405A HBV_HDV D28 sample 2.txt"
## [13] "BD405A HBV_HDV D28 sample 3.txt"
## [14] "BD405A HBV_HDV D8 sample 1.txt"
## [15] "BD405A HBV_HDV D8 sample 2.txt"
## [16] "BD405A HBV_HDV D8 sample 3.txt"
## [17] "BD405A_HBV_HDV_D28.txt"
## [18] "BD405A_HBV_HDV_D8.txt"

```

```

## [19] "HU1016_BD_co_D28.txt"
## [20] "HU1016_BD_co_D8.txt"

Make files of these separated feature counts

for(i in names(ctrl_counts)) {
  filename <- paste(i, sep = "")
  write.table(ctrl_counts[i], file = file.path("Mouse_d8d28_ctrl", filename),
              col.names = FALSE, row.names=FALSE,sep="\t",quote=FALSE)
}

for(i in names(HBV_counts)) {
  filename <- paste(i, sep = "")
  write.table(HBV_counts[i], file = file.path("Mouse_d8d28_HBV", filename),
              col.names = FALSE, row.names=FALSE,sep="\t",quote=FALSE)
}

for(i in names(coinf_counts)) {
  filename <- paste(i, sep = "")
  write.table(coinf_counts[i], file = file.path("Mouse_d8d28_coinf", filename),
              col.names = FALSE, row.names=FALSE,sep="\t",quote=FALSE)
}

```

Function to perform DGE analysis with both donor and time set as factors influencing the counts. Since we already sorted out counts into folders containing the ENSEMBL IDs for mouse genes under different infection conditions, we will pull the files from these folders to perform the DGE analysis.

```

DGE_analysis <- function(sampledirectory) {
  a <- basename(Sys.glob(file.path(sampledirectory, "*.txt")))
  sample_names <- sub('.txt', '', a)
  ##Here the donors are renamed based off the Hurel names (i.e. HU___) - RNASeq reads
  ##were all named using a different ID system.
  sampleTable <- data.frame(sampleName = sample_names, sampleFile = a, treatment =
    ifelse(grepl("Ctrl", a), "mock", ifelse(grepl("*co|*HDV", a),"coinf", "HBV")),
    donor = ifelse(grepl("BD330*", a), "HU1019",
      ifelse(grepl("BD405*", a), "HU1020",
        ifelse(grepl("HU1016*", a), "HU1016", "HU1007"))),
    time = ifelse(grepl("*D8|Day 8", a), "d8", "d28"),
    replicate = ifelse(grepl("*sample_1h|*D8_ah|*D8_aa|*D8_am|*sample_1m", a), "a",
      ifelse(grepl("*sample_2h|D28_bh|D28_ba|D28_bm|*sample_2m", a), "b",
        ifelse(grepl("*sample_3h| * sample 1h| * sample 1m", a), "c",
          ifelse(grepl("* sample 2h| * sample 2m", a), "d",
            ifelse(grepl("* sample 3h| * sample 3m", a), "e", ""))))))
  dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, directory =
    sampledirectory, design = ~donor + time)
  dds
  dds@colData
  contrast <- c("time", levels(sampleTable$time))
  output_basename <- sprintf("%s-%s_vs_%s_%s_analysis", "mousegenes", contrast[2],
    contrast[3], levels(sampleTable$treatment))
  dds <- estimateSizeFactors(dds)
  dds@colData
  dds <- estimateDispersions(dds)

  plotDispEsts(dds, main=sprintf("%s Dispersion Estimates", output_basename))
  dds <- nbinomWaldTest(dds)
}

```

```

res <- results(dds, contrast=contrast)
res <- res[order(res$padj, -abs(res$log2FoldChange)),]
mcols(res, use.names=TRUE)
##Log-intensity ratios = M values, log-intensity averages = A values
##Red points indicate padj < 0.1.
plotMA(res, alpha=0.1, main=sprintf(output_basename))
attr(res, "filterThreshold")

metadata(res)$alpha
metadata(res)$filterThreshold
plot(metadata(res)$filterNumRej,
     type="b", ylab="number of rejections",
     xlab="quantiles of filter")
lines(metadata(res)$lo.fit, col="red")
abline(v=metadata(res)$filterTheta)

key = "ENSEMBL"
cols = c("ENTREZID", "SYMBOL", "GENENAME", "ALIAS", "REFSEQ", "ACCCNUM")
for (col in cols) {
  # Get annotation data for column
  annotation_data <- AnnotationDbi::select(org.Mm.eg.db, rownames(res), col,
                                              keytype=key)
  # Collapse one-to-many relationships
  tmp <- aggregate(annotation_data[col], by=annotation_data[key],
                    # to a list
                    FUN=function(x)list(x))
  # Match on key and append to results
  idx <- match(rownames(res), tmp[[key]])
  res[[col]] <- tmp[idx,col]
}

output_data <- as.data.frame(res)
LIST_COLS <- sapply(output_data, is.list)
for (COL in colnames(output_data)[LIST_COLS]) {
  output_data[COL] <-
    sapply(output_data[COL],
           function(x)sapply(x, function(y) paste(unlist(y),
                                                     collapse=", ") ) )
}

# Save data frame above as tab-separated file
write.table(output_data,file=file.path("Mouse_DGEs_donortime", paste(Sys.Date(),
"mouse_donor_time", output_basename, "_results.txt", sep='')), quote=FALSE,sep="\t",
row.names=TRUE, col.names=NA)
return(list(dds@colData, head(res)))
}

##For each infection group, determine the DGE profile when comparing
##the different times to one another (i.e. d8 versus d28).

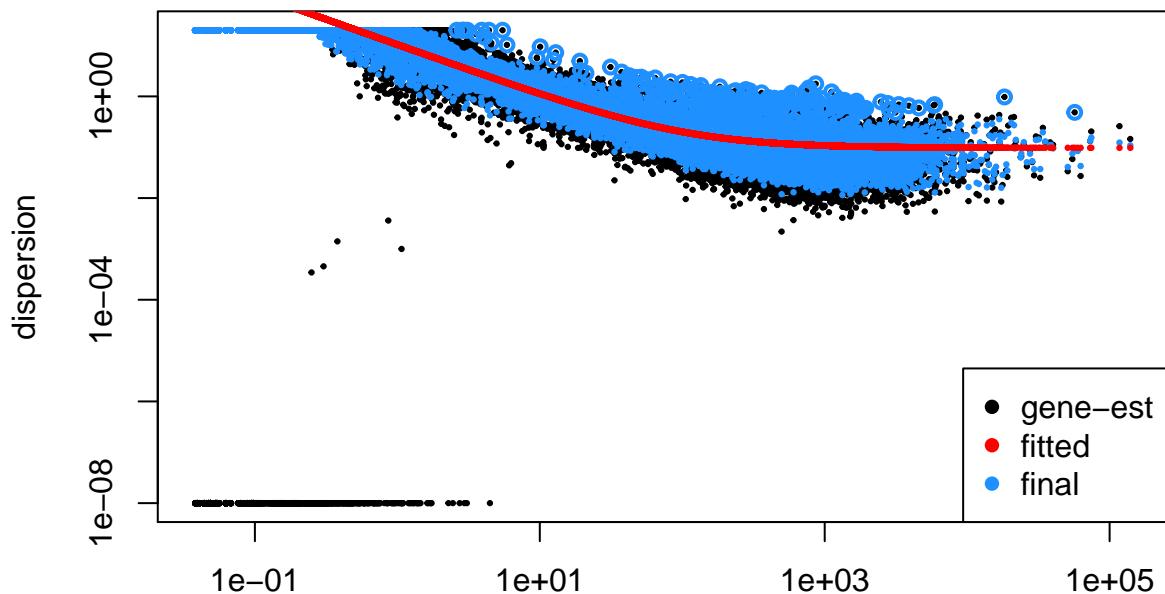
##coinfected
DGE_analysis("Mouse_d8d28_coinf")

## gene-wise dispersion estimates

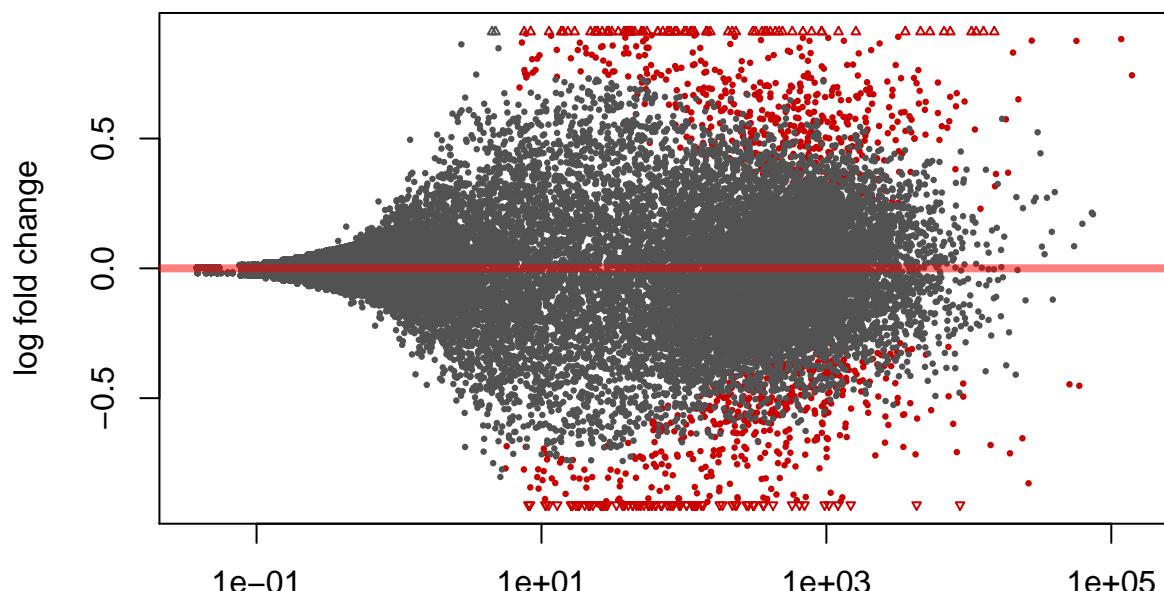
```

```
## mean-dispersion relationship  
## final dispersion estimates
```

### mousegenes-d28\_vs\_d8\_coinf\_analysis Dispersion Estimates



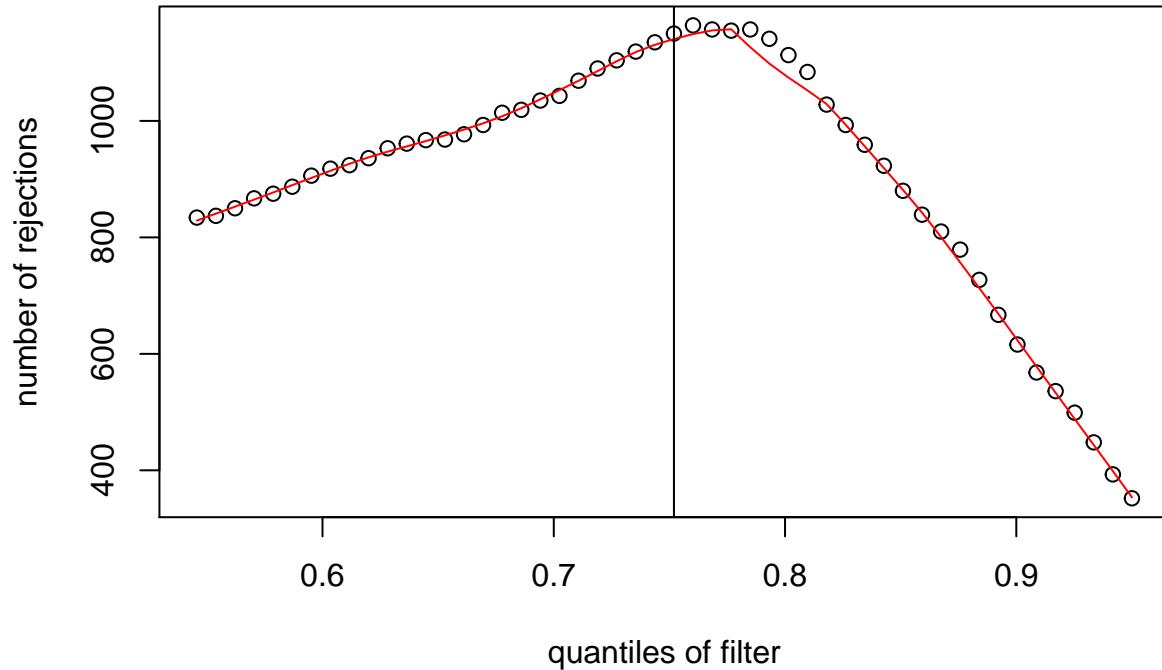
### mean of normalized counts mousegenes-d28\_vs\_d8\_coinf\_analysis



### mean of normalized counts

```
## 'select()' returned 1:many mapping between keys and columns  
## 'select()' returned 1:many mapping between keys and columns
```

```
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
```



```
## [[1]]
## DataFrame with 20 rows and 5 columns
##           treatment   donor   time replicate
##           <factor> <factor> <factor> <factor>
## BD330 HBV_HDV Day 28 sample 1    coinf  HU1019    d28
## BD330 HBV_HDV Day 28 sample 2    coinf  HU1019    d28
## BD330 HBV_HDV Day 28 sample 3    coinf  HU1019    d28
## BD330 HBV_HDV Day 8 sample 1     coinf  HU1019     d8
## BD330 HBV_HDV Day 8 sample 2     coinf  HU1019     d8
## ...
## ...
## BD405A HBV_HDV D8 sample 3      coinf  HU1020     d8
## BD405A_HBV_HDV_D28              coinf  HU1020    d28
## BD405A_HBV_HDV_D8               coinf  HU1020     d8
## HU1016_BD_co_D28                coinf  HU1016    d28
## HU1016_BD_co_D8                 coinf  HU1016     d8
## sizeFactor
## <numeric>
## BD330 HBV_HDV Day 28 sample 1  0.9771958
## BD330 HBV_HDV Day 28 sample 2  1.1489916
## BD330 HBV_HDV Day 28 sample 3  1.2201694
## BD330 HBV_HDV Day 8 sample 1   0.8899125
## BD330 HBV_HDV Day 8 sample 2   0.9484827
## ...
## ...
## BD405A HBV_HDV D8 sample 3    1.2934488
## BD405A_HBV_HDV_D28            1.0293191
## BD405A_HBV_HDV_D8             0.9408521
## HU1016_BD_co_D28              1.0573426
## HU1016_BD_co_D8               1.1302832
```

```

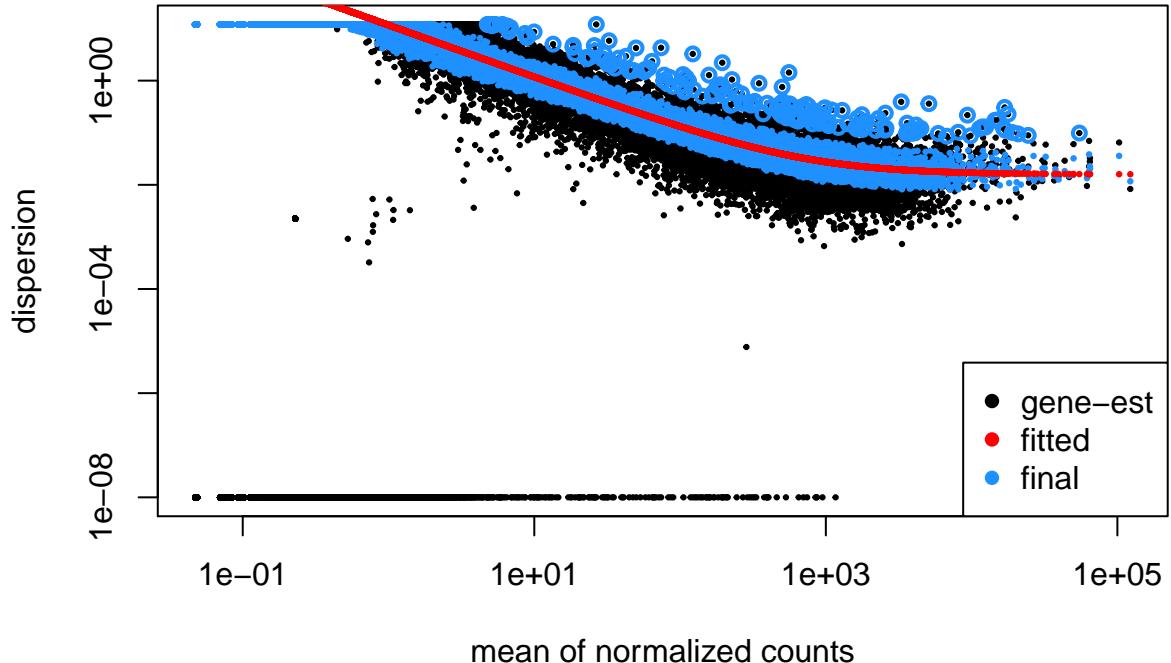
## 
## [[2]]
## log2 fold change (MAP): time d28 vs d8
## Wald test p-value: time d28 vs d8
## DataFrame with 6 rows and 12 columns
##           baseMean log2FoldChange      lfcSE      stat
##           <numeric>     <numeric> <numeric> <numeric>
## ENSMUSG00000026576    229.4038    -2.275039 0.2651809 -8.579197
## ENSMUSG00000020644    647.1975    -1.469670 0.1773931 -8.284820
## ENSMUSG00000047246   121.9555    -2.060951 0.2674337 -7.706397
## ENSMUSG00000024164  12676.0319     1.090027 0.1430412  7.620366
## ENSMUSG00000029231  10434.3852     1.099436 0.1530003  7.185845
## ENSMUSG00000024087   574.2816    -1.532941 0.2257476 -6.790510
##           pvalue      padj ENTREZID      SYMBOL
##           <numeric>     <numeric> <list>     <list>
## ENSMUSG00000026576 9.553755e-18 1.266637e-13    11931    Atp1b1
## ENSMUSG00000020644 1.182881e-16 7.841318e-13    15902     Id2
## ENSMUSG00000047246 1.294194e-14 5.719473e-11   319179  Hist1h2be
## ENSMUSG00000024164 2.529576e-14 8.384281e-11   12266      C3
## ENSMUSG00000029231 6.679295e-13 1.771082e-09   18595    Pdgfra
## ENSMUSG00000024087 1.117378e-11 2.469033e-08   13078    Cyp1b1
##                                     GENENAME
##                                     <list>
## ENSMUSG00000026576      ATPase, Na+/K+ transporting, beta 1 polypeptide
## ENSMUSG00000020644      inhibitor of DNA binding 2
## ENSMUSG00000047246      histone cluster 1, H2be
## ENSMUSG00000024164      complement component 3
## ENSMUSG00000029231 platelet derived growth factor receptor, alpha polypeptide
## ENSMUSG00000024087      cytochrome P450, family 1, subfamily b, polypeptide 1
##                                     ALIAS
##                                     <list>
## ENSMUSG00000026576      Atp4b,Atpb,Atpb-1, ...
## ENSMUSG00000020644      AI255428,C78922,Idb2, ...
## ENSMUSG00000047246      Hist1h2be
## ENSMUSG00000024164      AI255234,ASP,HSE-MSF, ...
## ENSMUSG00000029231      AI115593,CD140a,Pdgfr-2, ...
## ENSMUSG00000024087      CP1B,P4501b1,Cyp1b1
##                                     REFSEQ
##                                     <list>
## ENSMUSG00000026576      NM_009721,NP_033851
## ENSMUSG00000020644      NM_010496,NP_034626
## ENSMUSG00000047246      NM_001177653,NM_001290530,NM_178194, ...
## ENSMUSG00000024164      NM_009778,NP_033908,XM_011246258, ...
## ENSMUSG00000029231      NM_001083316,NM_011058,NP_001076785, ...
## ENSMUSG00000024087      NM_009994,NP_034124,XM_006523577, ...
##                                     ACCNUM
##                                     <list>
## ENSMUSG00000026576      AAH27319,AAH94070,ABS83510, ...
## ENSMUSG00000020644      AAA79771,AAH06921,AAH53699, ...
## ENSMUSG00000047246      AAH69889,AA006246,AK030546, ...
## ENSMUSG00000024164      AAA37339,AAA37378,AAC42013, ...
## ENSMUSG00000029231      AAA39733,AAA39904,AAH37467, ...
## ENSMUSG00000024087      AAC52131,AAC52141,AAH50063, ...

```

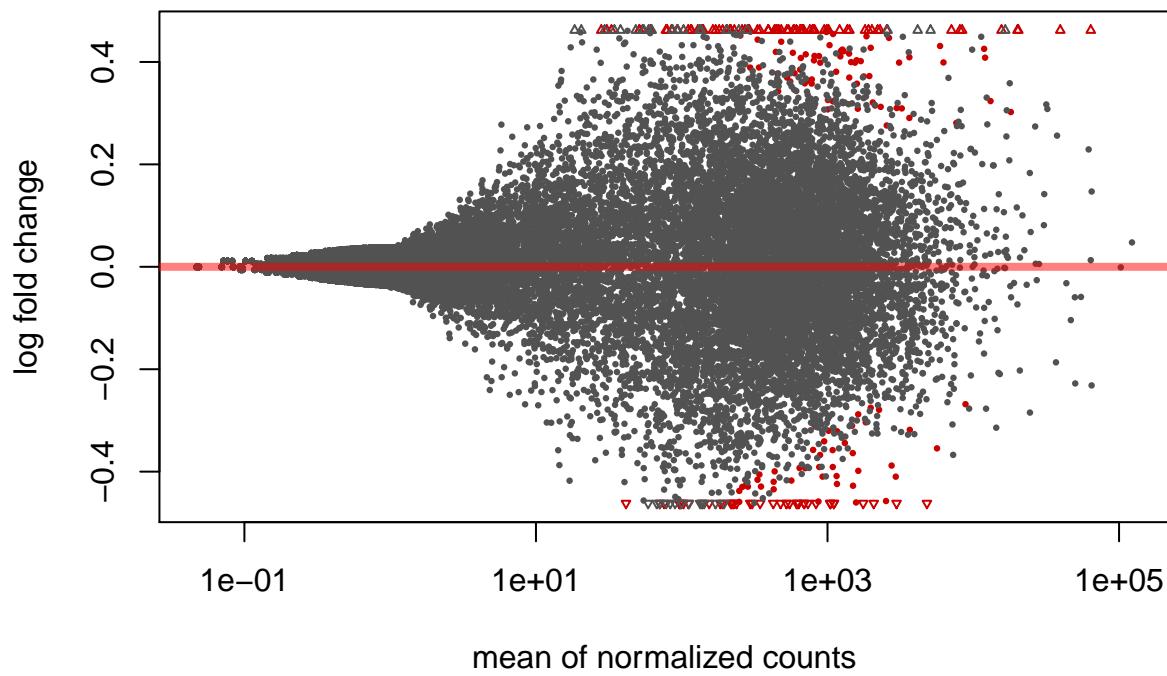
```
##monoinfected with HBV  
DGE_analysis("Mouse_d8d28_HBV")
```

```
## gene-wise dispersion estimates  
## mean-dispersion relationship  
## final dispersion estimates
```

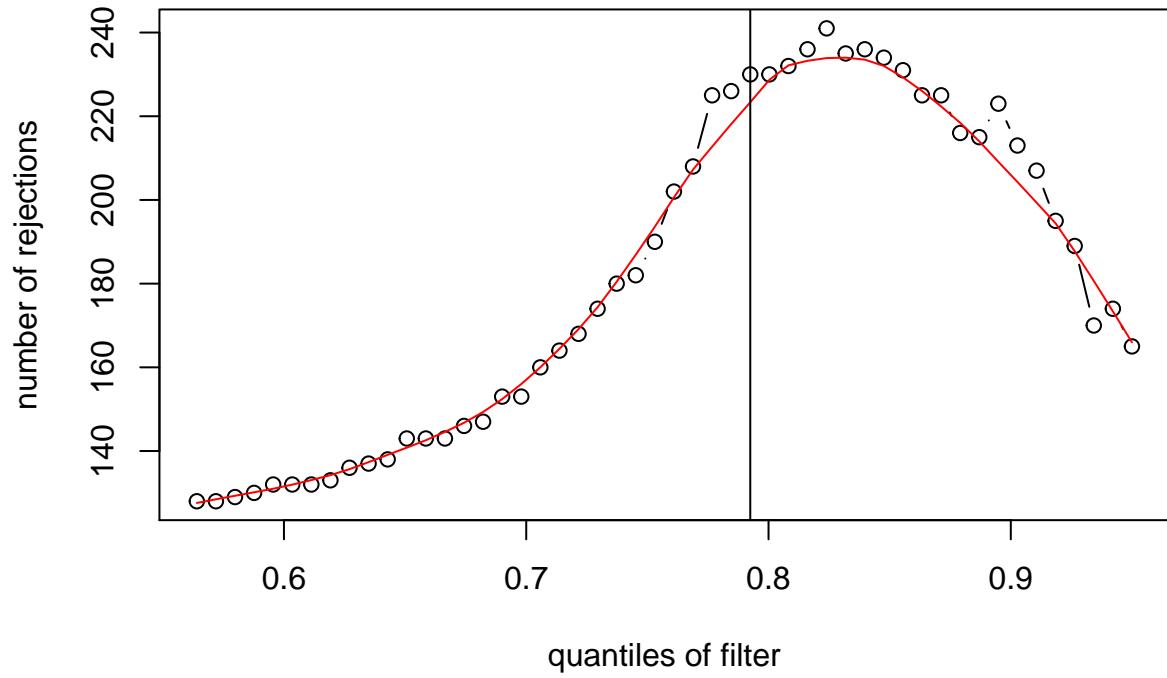
### mousegenes-d28\_vs\_d8\_HBV\_analysis Dispersion Estimates



## mousegenes-d28\_vs\_d8\_HBV\_analysis



```
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
```



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

```

## DataFrame with 12 rows and 5 columns
##          treatment   donor    time replicate sizeFactor
##          <factor> <factor> <factor> <factor> <numeric>
## BD330_HBV_D28      HBV  HU1019    d28        0.5235705
## BD330_HBV_D8       HBV  HU1019     d8        0.8934511
## BD405A_HBV_D28     HBV  HU1020    d28        0.8188988
## BD405A_HBV_D8     HBV  HU1020     d8        0.8670608
## HBV_D28_sample_1   HBV  HU1007    d28        1.0131104
## ...
## ...
## HBV_D8_sample_1    HBV  HU1007     d8        1.7664648
## HBV_D8_sample_2    HBV  HU1007     d8        1.1915286
## HBV_D8_sample_3    HBV  HU1007     d8        1.7129827
## HU1016_B_D28      HBV  HU1016    d28        1.0908481
## HU1016_B_D8       HBV  HU1016     d8        0.9953479
##
## [[2]]
## log2 fold change (MAP): time d28 vs d8
## Wald test p-value: time d28 vs d8
## DataFrame with 6 rows and 12 columns
##          baseMean log2FoldChange      lfcSE      stat
##          <numeric>     <numeric> <numeric> <numeric>
## ENSMUSG00000031503 15621.3787      0.5638417 0.07811890 7.217737
## ENSMUSG00000027460  128.9846      1.1237249 0.15822425 7.102103
## ENSMUSG00000044786  391.2283      0.8280038 0.12393916 6.680728
## ENSMUSG00000032350  660.1810      0.8043660 0.11995728 6.705438
## ENSMUSG00000024940  1910.3683      0.5645423 0.08444322 6.685466
## ENSMUSG00000035673 1414.4223      0.6461405 0.09962567 6.485682
##          pvalue      padj ENTREZID SYMBOL
##          <numeric>     <numeric> <list> <list>
## ENSMUSG00000031503 5.285995e-13 5.865340e-09 12827 Col4a2
## ENSMUSG00000027460 1.228724e-12 6.816962e-09 11602 Angpt4
## ENSMUSG00000044786 2.377579e-11 5.276324e-08 22695 Zfp36
## ENSMUSG00000032350 2.008035e-11 5.276324e-08 14629 Gclc
## ENSMUSG00000024940 2.301906e-11 5.276324e-08 16998 Ltbp3
## ENSMUSG00000035673 8.833126e-11 1.633539e-07 216161 Sbno2
##
##          GENENAME
##          <list>
## ENSMUSG00000031503           collagen, type IV, alpha 2
## ENSMUSG00000027460           angiopoietin 4
## ENSMUSG00000044786           zinc finger protein 36
## ENSMUSG00000032350           glutamate-cysteine ligase, catalytic subunit
## ENSMUSG00000024940           latent transforming growth factor beta binding protein 3
## ENSMUSG00000035673           strawberry notch homolog 2 (Drosophila)
##          ALIAS
##          <list>
## ENSMUSG00000031503           Col4a-2,Col4a2
## ENSMUSG00000027460           ANG-3,ANG-4,Agpt4,...
## ENSMUSG00000044786           Gos24,Nup475,TIS11D,...
## ENSMUSG00000032350           D9Wsu168e,GLCL-H,Ggcs-hs,...
## ENSMUSG00000024940           Ltbp2,mFLJ00070,Ltbp3
## ENSMUSG00000035673           BC019206,F730040C21,Sno,...
##          REFSEQ
##          <list>
## ENSMUSG00000031503           NM_009932,NP_034062

```

```

## ENSMUSG00000027460          NM_009641, NP_033771
## ENSMUSG00000044786          NM_011756, NP_035886
## ENSMUSG00000032350 NM_010295, NP_034425, XM_006510812, ...
## ENSMUSG00000024940 NM_008520, NP_032546, XM_006531668, ...
## ENSMUSG00000035673 NM_183426, NP_906271, XM_006513492, ...
##                                     ACCNUM
##                                     <list>
## ENSMUSG00000031503 AAA37341, AAA37438, AAA50293, ...
## ENSMUSG00000027460 AAD21586, AAI29966, AAI29967, ...
## ENSMUSG00000044786 AAA39837, AAA40498, AAA72947, ...
## ENSMUSG00000032350 AAB42020, AAB52542, AAF00505, ...
## ENSMUSG00000024940 AAB53015, AAI68401, AK007774, ...
## ENSMUSG00000035673 AAH56369, AAH64113, AJ512611, ...

##ctrl
DGE_analysis("Mouse_d8d28_ctrl")

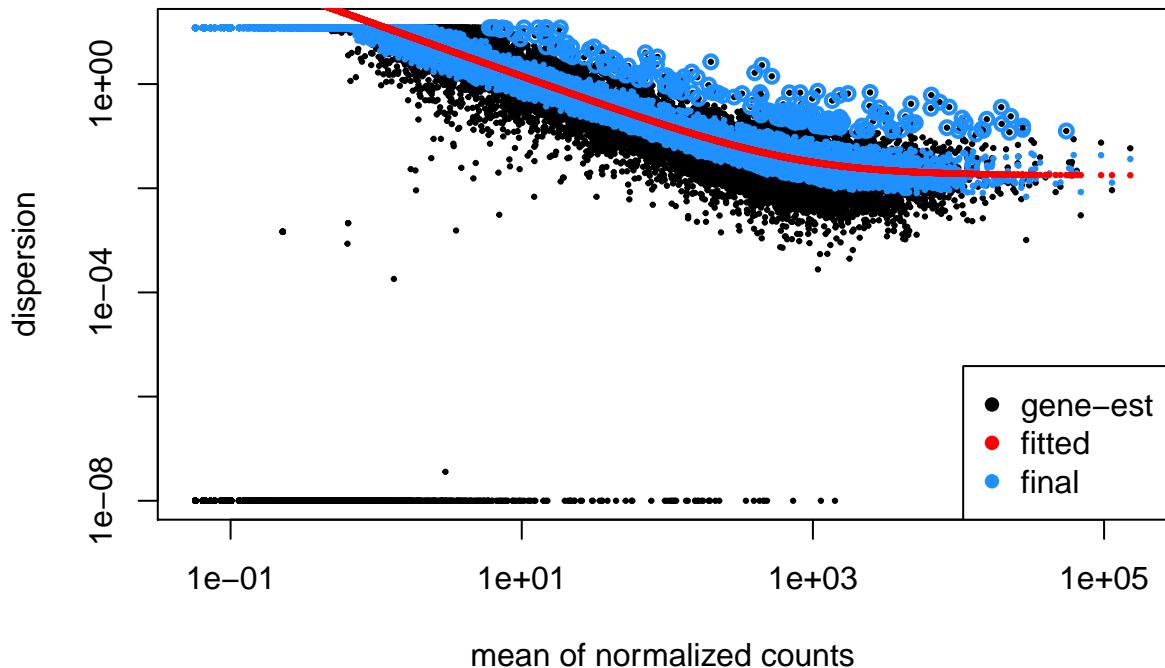
```

```

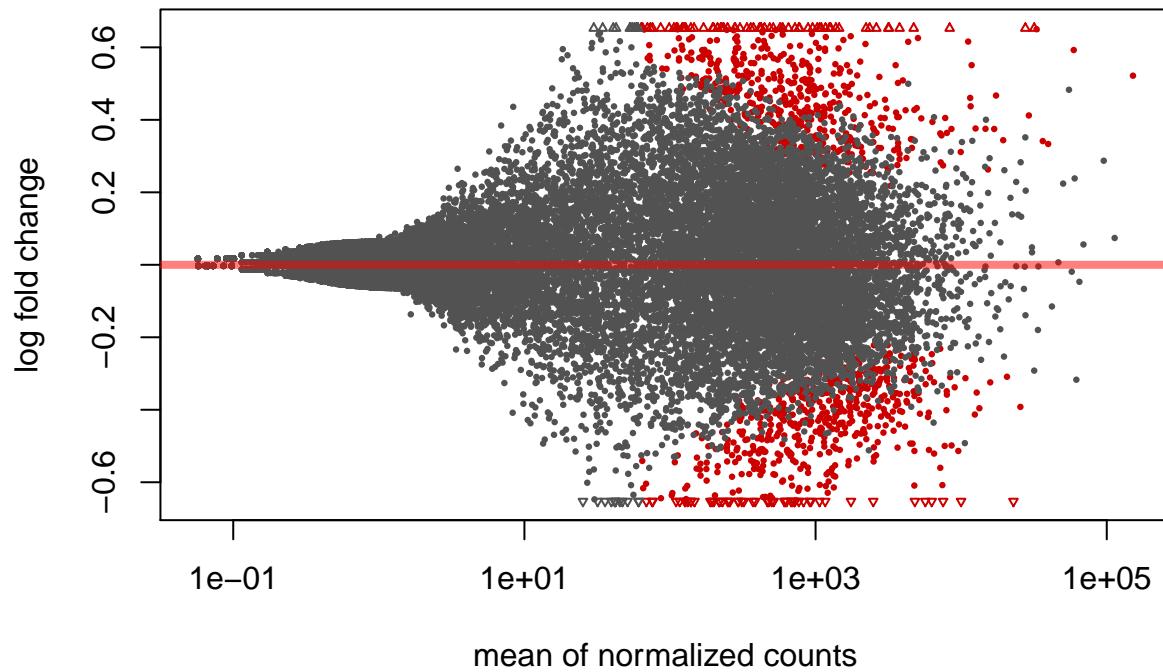
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates

```

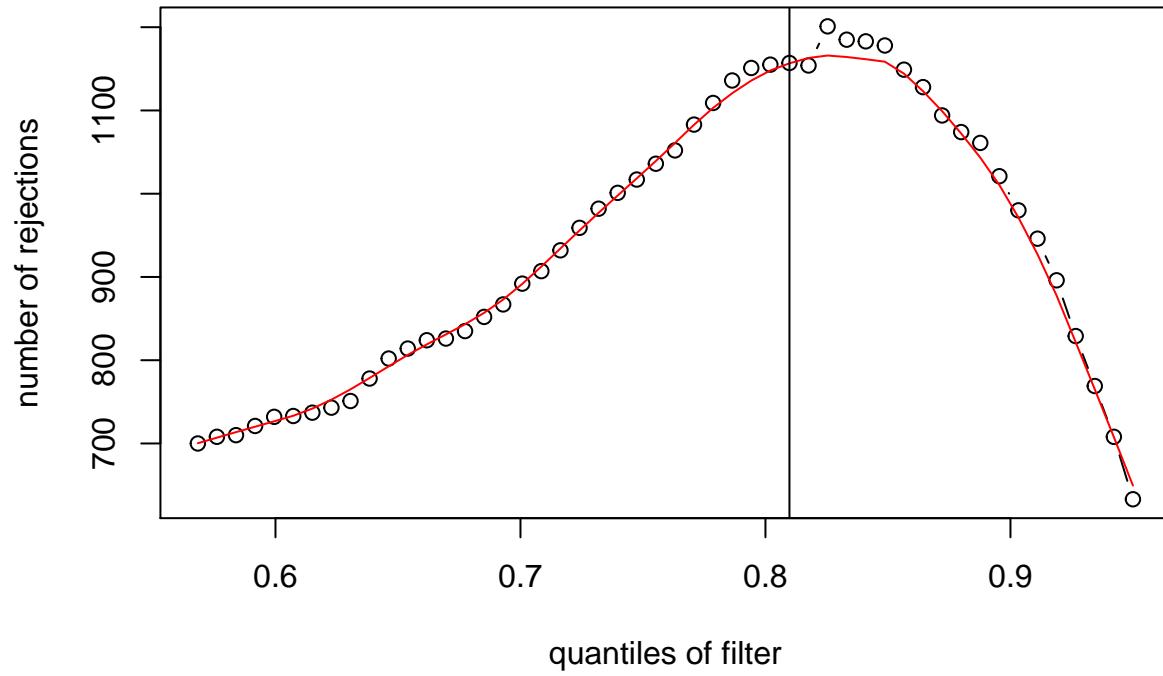
## mousegenes-d28\_vs\_d8\_mock\_analysis Dispersion Estimates



## mousegenes-d28\_vs\_d8\_mock\_analysis



```
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
## 'select()' returned 1:many mapping between keys and columns
```



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

```

## DataFrame with 12 rows and 5 columns
##           treatment   donor    time replicate sizeFactor
##           <factor> <factor> <factor> <factor> <numeric>
## BD330_Ctrl_D28      mock  HU1019    d28     1.0721445
## BD330_Ctrl_D8       mock  HU1019     d8      1.1436568
## BD405A_Ctrl_D28     mock  HU1020    d28     0.9257930
## BD405A_Ctrl_D8     mock  HU1020     d8      0.8538694
## Ctrl_D28_sample_1   mock  HU1007    d28     0.5211684
## ...                 ...    ...     ...     ...     ...
## Ctrl_D8_sample_1   mock  HU1007     d8      1.2924433
## Ctrl_D8_sample_2   mock  HU1007     d8      1.4495022
## Ctrl_D8_sample_3   mock  HU1007     d8      1.2460616
## HU1016 Ctrl D28    mock  HU1016    d28     1.4589246
## HU1016 Ctrl D8     mock  HU1016     d8      0.9707685
##
## [[2]]
## log2 fold change (MAP): time d28 vs d8
## Wald test p-value: time d28 vs d8
## DataFrame with 6 rows and 12 columns
##           baseMean log2FoldChange      lfcSE      stat
##           <numeric>      <numeric> <numeric> <numeric>
## ENSMUSG00000022816  7515.5679     -0.8717800 0.09670291 -9.015034
## ENSMUSG00000050578  1026.5627      1.1861706 0.14639913  8.102307
## ENSMUSG00000089774  9986.2411     -1.1114599 0.13925882 -7.981253
## ENSMUSG00000067279  619.0564      -0.9344345 0.13023736 -7.174858
## ENSMUSG00000045414  1164.3689     -0.7987976 0.11640456 -6.862254
## ENSMUSG00000027460  199.1397      1.3516356 0.20603344  6.560273
##           pvalue      padj ENTREZID      SYMBOL
##           <numeric>      <numeric> <list>      <list>
## ENSMUSG00000022816  1.968100e-19 2.001558e-15  14314      Fstl1
## ENSMUSG00000050578  5.392669e-16 2.742172e-12  17386      Mmp13
## ENSMUSG00000089774  1.448546e-15 4.910572e-12  53881      Slc5a3
## ENSMUSG00000067279  7.238221e-13 1.840318e-09  53412      Ppp1r3c
## ENSMUSG00000045414  6.778252e-12 1.378696e-08  68861 1190002N15Rik
## ENSMUSG00000027460  5.370950e-11 9.103761e-08  11602      Angpt4
##           GENENAME
##           <list>
## ENSMUSG00000022816          follistatin-like 1
## ENSMUSG00000050578          matrix metallopeptidase 13
## ENSMUSG00000089774 solute carrier family 5 (inositol transporters), member 3
## ENSMUSG00000067279 protein phosphatase 1, regulatory (inhibitor) subunit 3C
## ENSMUSG00000045414          RIKEN cDNA 1190002N15 gene
## ENSMUSG00000027460          angiopoietin 4
##           ALIAS      REFSEQ
##           <list>      <list>
## ENSMUSG00000022816 AI316791,AW107808,Fstl,... NM_008047,NP_032073
## ENSMUSG00000050578 Clg,MMP-13,Mmp1,... NM_008607,NP_032633
## ENSMUSG00000089774 AA623876,BF642829,Smit1,... NM_017391,NP_059087
## ENSMUSG00000067279 PTG,Ppp1r5,Ppp1r3c NM_016854,NP_058550
## ENSMUSG00000045414 GoPro49,HASF,1190002N15Rik NM_001033145,NP_001028317
## ENSMUSG00000027460 ANG-3,ANG-4,Agpt4,... NM_009641,NP_033771
##           ACCNUM
##           <list>
## ENSMUSG00000022816 AAC37633,AAH28921,AK049440,...

```

```

## ENSMUSG00000050578 AAI25321,AAI25323,AAT46404,...
## ENSMUSG00000089774 AAF43668,AAI40983,AAI40984,...
## ENSMUSG00000067279 AAB49689,AAH52769,AAH54739,...
## ENSMUSG00000045414 AK004457,AK162197,AK189994,...
## ENSMUSG00000027460 AAD21586,AAI29966,AAI29967,...

Session Info

sessionInfo()

## R version 3.3.3 (2017-03-06)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel stats4 stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] org.Mm.eg.db_3.4.0 AnnotationDbi_1.36.2
## [3] BiocInstaller_1.24.0 viridis_0.5.1
## [5] viridisLite_0.3.0 ggrepel_0.8.0
## [7] data.table_1.11.4 genefilter_1.56.0
## [9] RColorBrewer_1.1-2 gplots_3.0.1
## [11] DESeq2_1.14.1 SummarizedExperiment_1.4.0
## [13] Biobase_2.34.0 GenomicRanges_1.26.4
## [15] GenomeInfoDb_1.10.3 IRanges_2.8.2
## [17] S4Vectors_0.12.2 BiocGenerics_0.20.0
## [19] openxlsx_4.1.0 reshape2_1.4.3
## [21] ggplot2_3.0.0 stringr_1.3.1
## [23] dplyr_0.7.6
##
## loaded via a namespace (and not attached):
## [1] bit64_0.9-7 splines_3.3.3 gtools_3.8.1
## [4] Formula_1.2-3 assertthat_0.2.0 latticeExtra_0.6-28
## [7] blob_1.1.1 yaml_2.2.0 pillar_1.3.0
## [10] RSQLite_2.1.1 backports_1.1.2 lattice_0.20-35
## [13] glue_1.3.0 digest_0.6.15 XVector_0.14.1
## [16] checkmate_1.8.5 colorspace_1.3-2 htmltools_0.3.6
## [19] Matrix_1.2-8 plyr_1.8.4 XML_3.98-1.12
## [22] pkgconfig_2.0.1 zlibbioc_1.20.0 purrr_0.2.5
## [25] xtable_1.8-2 scales_0.5.0 gdata_2.18.0
## [28] BiocParallel_1.8.2 tibble_1.4.2 htmlTable_1.12
## [31] annotate_1.52.1 withr_2.1.2 nnet_7.3-12
## [34] lazyeval_0.2.1 survival_2.42-6 magrittr_1.5
## [37] crayon_1.3.4 memoise_1.1.0 evaluate_0.11
## [40] foreign_0.8-71 tools_3.3.3 locfit_1.5-9.1
## [43] munsell_0.5.0 cluster_2.0.5 zip_1.0.0
## [46] bindrcpp_0.2.2 caTools_1.17.1.1 rlang_0.2.1
## [49] grid_3.3.3 RCurl_1.95-4.11 rstudioapi_0.7
## [52] htmlwidgets_1.2 bitops_1.0-6 base64enc_0.1-3
## [55] rmarkdown_1.10 gtable_0.2.0 DBI_1.0.0

```

```
## [58] R6_2.2.2           gridExtra_2.3      knitr_1.20
## [61] bit_1.1-14          bindr_0.1.1        Hmisc_4.1-1
## [64] rprojroot_1.3-2    KernSmooth_2.23-15 stringi_1.2.4
## [67] Rcpp_0.12.18        geneplotter_1.52.0 rpart_4.1-13
## [70] acepack_1.4.1       tidyselect_0.2.4
```