

RNAseq E-MTAB-5337

Seoyeon Oh, Tobias Heyman

12/21/2021

```
library("mogene10sttranscriptcluster.db")

## Loading required package: AnnotationDbi

## 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.max, which.min

## 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: IRanges

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

## Loading required package: org.Mm.eg.db

##

##

library("ArrayExpress")
library("arrayQualityMetrics")
library("ggplot2")
library("huex10sttranscriptcluster.db")

## Loading required package: org.Hs.eg.db

##

##

library("limma")

##
## Attaching package: 'limma'

## The following object is masked from 'package:BiocGenerics':
##
##     plotMA

library("oligo")

## Loading required package: oligoClasses

## Welcome to oligoClasses version 1.54.0

## Loading required package: Biostrings

## Loading required package: XVector

## Loading required package: GenomeInfoDb

```

```

##
## Attaching package: 'Biostrings'

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

## =====

## Welcome to oligo version 1.56.0

## =====

##
## Attaching package: 'oligo'

## The following object is masked from 'package:limma':
##
##      backgroundCorrect

library("siggenes")

## Loading required package: multtest

## Loading required package: splines

library("affy")

##
## Attaching package: 'affy'

## The following objects are masked from 'package:oligo':
##
##      intensity, MAplot, mm, mm<-, mmindex, pm, pm<-, pmindex,
##      probeNames, rma

## The following object is masked from 'package:oligoClasses':
##
##      list.celfiles

#library("pd.huex.1.0.st.v2")
library("wateRmelon")

## Loading required package: matrixStats

##
## Attaching package: 'matrixStats'

```

```

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

## Loading required package: methylumi

## Loading required package: scales

## Loading required package: reshape2

## Loading required package: FDb.InfiniumMethylation.hg19

## Loading required package: GenomicFeatures

## Loading required package: GenomicRanges

## Loading required package: TxDb.Hsapiens.UCSC.hg19.knownGene

## Loading required package: minfi

## Loading required package: SummarizedExperiment

## Loading required package: MatrixGenerics

##
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':
##
##     colAlls, colAnyNAs, colAnys, colAveragesPerRowSet, colCollapse,
##     colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##     colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##     colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAveragesPerColSet,
##     rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##     rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars

## The following object is masked from 'package:Biobase':
##
##     rowMedians

## Loading required package: bumphunter

```

```

## Loading required package: foreach

## Parallel computing support for 'oligo/crlmm': Disabled
##   - Load 'ff'
##   - Load and register a 'foreach' adaptor
##       Example - Using 'multicore' for 2 cores:
##       library(doMC)
##       registerDoMC(2)
## =====
## Loading required package: iterators
## Loading required package: locfit
## locfit 1.5-9.4    2020-03-24
## Setting options('download.file.method.GEOquery'='auto')
## Setting options('GEOquery.inmemory.gpl'=FALSE)
##
## Attaching package: 'minfi'
##
## The following object is masked from 'package:oligo':
##
##   getProbeInfo
##
## The following object is masked from 'package:oligoClasses':
##
##   getM
##
## Loading required package: lumi
## No methods found in package 'RSQLite' for request: 'dbListFields' when loading 'lumi'
##
## Attaching package: 'lumi'
##
## The following objects are masked from 'package:methylumi':
##
##   estimateM, getHistory
##
## The following objects are masked from 'package:affy':
##
##   MAplot, plotDensity
##
## The following object is masked from 'package:oligo':
##
##   MAplot
##
## Loading required package: ROC
## Loading required package: IlluminaHumanMethylation450kanno.ilmn12.hg19
## Loading required package: illuminaio

```

```

library("affy")
library("arrayQualityMetrics")
library("ArrayExpress")
library("RSQLite")
library("DBI")
library("htmltools")
library("biomaRt")
library("tximport")

```

```
library("edgeR")
library("rhdf5")
```

```
setwd("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset")
```

RNAseq dataset E-MTAB-5337

RNA-seq of whole lungs from Irgm1-/- and wildtype littermates mice that were either uninfected or infected with influenza for 10 days. Here, we only used wild type 3 infected vs 3 non-infected samples. For each sample, two technical replicates were present (total 12)

```
setwd("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset")
```

```
## Get file locations
#ERR1753616 - ERR1753621 : WT, non-infected
#ERR1753622- ERR1753627: WT, infected 10 dpi
files1 <- c("/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753616.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753617.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753618.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753619.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753620.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753621.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753622.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753623.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753624.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753625.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753626.fastq",
            "/Users/seoyeon/Desktop/MSc Bioinformatics Year 1/Applied High-throughput Analysis/Project/Dataset/ERR1753627.fastq")

tx2gene1 <- read.delim("./mus_musculus_trans2gen.txt")
#tx2gene1

names(files1) <- paste0("sample", 1:12)
txi.kallisto <- tximport(files1, type = "kallisto", txOut = FALSE, tx2gene=tx2gene1)
```

```
## Note: importing 'abundance.h5' is typically faster than 'abundance.tsv'
```

```
## reading in files with read_tsv
```

```
## 1 2 3 4 5 6 7 8 9 10 11 12
## transcripts missing from tx2gene: 1673
## summarizing abundance
## summarizing counts
## summarizing length
```

```
dim(txi.kallisto$counts)
```

```
## [1] 36047 12
```

```
head(txi.kallisto$length)
```

```
##               sample1 sample2 sample3 sample4 sample5 sample6
## ENSMUSG000000000001.4 3213.0000 3213.000 3213.0000 3213.0000 3213.000 3213.0000
## ENSMUSG000000000003.15 750.5000 750.500 750.5000 750.5000 750.500 750.5000
## ENSMUSG000000000028.15 1495.8714 1480.561 1462.6724 1460.3967 1465.040 1479.9446
## ENSMUSG000000000037.16 1638.0000 2340.002 2573.9997 3276.0010 3602.508 3391.0000
## ENSMUSG000000000049.11 443.4987 433.752 443.4987 443.4987 318.000 497.3792
## ENSMUSG000000000056.7 3443.1809 3146.751 2284.8596 2217.8745 2907.992 2752.3972
##               sample7 sample8 sample9 sample10 sample11 sample12
## ENSMUSG000000000001.4 3213.0000 3213.0000 3213.000 3213.0000 3213.000 3213.000
## ENSMUSG000000000003.15 750.5000 750.5000 750.500 750.5000 750.500 750.500
## ENSMUSG000000000028.15 1540.7317 1677.0540 1743.725 1699.3468 1683.682 1591.334
## ENSMUSG000000000037.16 2808.0003 2807.9994 3651.758 2807.9998 3442.499 3510.001
## ENSMUSG000000000049.11 443.4987 497.3788 318.000 635.9999 682.000 318.000
## ENSMUSG000000000056.7 2890.5020 4067.4717 3477.942 3140.7988 3472.230 2859.385
```

118489 features 12 samples in trascript level.36047 features in gene level.

```
sdrf <- read.delim("./RNAseq/E-MTAB-5337.sdrf.txt")
head(sdrf)
```

```
##               Source.Name Comment.ENA_SAMPLE. Comment.BioSD_SAMPLE.
## 1 Infected_Irgm1-/-_Day10_1          ERS1471225          SAMEA24832918
## 2 Infected_Irgm1-/-_Day10_1          ERS1471225          SAMEA24832918
## 3 Infected_Irgm1-/-_Day10_2          ERS1471226          SAMEA24833668
## 4 Infected_Irgm1-/-_Day10_2          ERS1471226          SAMEA24833668
## 5 Infected_Irgm1-/-_Day10_3          ERS1471227          SAMEA24834418
## 6 Infected_Irgm1-/-_Day10_3          ERS1471227          SAMEA24834418
## Characteristics.organism. Characteristics.strain. Characteristics.age.
## 1 Mus musculus          C57BL/6          8 to 12
## 2 Mus musculus          C57BL/6          8 to 12
## 3 Mus musculus          C57BL/6          8 to 12
## 4 Mus musculus          C57BL/6          8 to 12
## 5 Mus musculus          C57BL/6          8 to 12
## 6 Mus musculus          C57BL/6          8 to 12
## Unit.time.unit. Characteristics.genotype.
## 1 week          Irgm1-/- knockout
## 2 week          Irgm1-/- knockout
## 3 week          Irgm1-/- knockout
## 4 week          Irgm1-/- knockout
## 5 week          Irgm1-/- knockout
## 6 week          Irgm1-/- knockout
## Characteristics.phenotype. Characteristics.organism.part.
## 1 protected from influenza-induced mortality          lung
## 2 protected from influenza-induced mortality          lung
## 3 protected from influenza-induced mortality          lung
## 4 protected from influenza-induced mortality          lung
## 5 protected from influenza-induced mortality          lung
## 6 protected from influenza-induced mortality          lung
## Material.Type Protocol.REF Performer Protocol.REF.1 Performer.1
## 1 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
```

```

## 2 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
## 3 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
## 4 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
## 5 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
## 6 organism part P-MTAB-53254 Ashley Steed P-MTAB-53248 Ashley Steed
## Protocol.REF.2 Performer.2 Protocol.REF.3
## 1 P-MTAB-53249 Ashley Steed P-MTAB-53250
## 2 P-MTAB-53249 Ashley Steed P-MTAB-53250
## 3 P-MTAB-53249 Ashley Steed P-MTAB-53250
## 4 P-MTAB-53249 Ashley Steed P-MTAB-53250
## 5 P-MTAB-53249 Ashley Steed P-MTAB-53250
## 6 P-MTAB-53249 Ashley Steed P-MTAB-53250
## Performer.3
## 1 Genome Technology Access Center at Washington University
## 2 Genome Technology Access Center at Washington University
## 3 Genome Technology Access Center at Washington University
## 4 Genome Technology Access Center at Washington University
## 5 Genome Technology Access Center at Washington University
## 6 Genome Technology Access Center at Washington University
## Extract.Name Comment.LIBRARY_LAYOUT. Comment.LIBRARY_SELECTION.
## 1 Infected_Irgm1/-_Day10_1 SINGLE Inverse rRNA
## 2 Infected_Irgm1/-_Day10_1 SINGLE Inverse rRNA
## 3 Infected_Irgm1/-_Day10_2 SINGLE Inverse rRNA
## 4 Infected_Irgm1/-_Day10_2 SINGLE Inverse rRNA
## 5 Infected_Irgm1/-_Day10_3 SINGLE Inverse rRNA
## 6 Infected_Irgm1/-_Day10_3 SINGLE Inverse rRNA
## Comment.LIBRARY_SOURCE. Comment.LIBRARY_STRAND. Comment.LIBRARY_STRATEGY.
## 1 TRANSCRIPTOMIC not applicable RNA-Seq
## 2 TRANSCRIPTOMIC not applicable RNA-Seq
## 3 TRANSCRIPTOMIC not applicable RNA-Seq
## 4 TRANSCRIPTOMIC not applicable RNA-Seq
## 5 TRANSCRIPTOMIC not applicable RNA-Seq
## 6 TRANSCRIPTOMIC not applicable RNA-Seq
## Protocol.REF.4 Performer.4
## 1 P-MTAB-53251 Genome Technology Access Center at Washington University
## 2 P-MTAB-53251 Genome Technology Access Center at Washington University
## 3 P-MTAB-53251 Genome Technology Access Center at Washington University
## 4 P-MTAB-53251 Genome Technology Access Center at Washington University
## 5 P-MTAB-53251 Genome Technology Access Center at Washington University
## 6 P-MTAB-53251 Genome Technology Access Center at Washington University
## Assay.Name Comment.technical.replicate.group.
## 1 Infected_Irgm1/-_Day10_1 group 24
## 2 Infected_Irgm1/-_Day10_1_1 group 24
## 3 Infected_Irgm1/-_Day10_2 group 23
## 4 Infected_Irgm1/-_Day10_2_1 group 23
## 5 Infected_Irgm1/-_Day10_3 group 22
## 6 Infected_Irgm1/-_Day10_3_1 group 22
## Technology.Type Comment.ENA_EXPERIMENT.
## 1 sequencing assay ERX1820995
## 2 sequencing assay ERX1820995
## 3 sequencing assay ERX1820996
## 4 sequencing assay ERX1820996
## 5 sequencing assay ERX1820997
## 6 sequencing assay ERX1820997

```



```

##                               Scan.Name
## 1 run_1653_s_3_withindex_sequence.txt_CGAAAGT.fq.gz
## 2 run_1653_s_4_withindex_sequence.txt_CGAAAGT.fq.gz
## 3 run_1653_s_3_withindex_sequence.txt_CAGAGTC.fq.gz
## 4 run_1653_s_4_withindex_sequence.txt_CAGAGTC.fq.gz
## 5 run_1653_s_3_withindex_sequence.txt_ACTGGAT.fq.gz
## 6 run_1653_s_4_withindex_sequence.txt_ACTGGAT.fq.gz
##                               Comment.SUBMITTED_FILE_NAME. Comment.ENA_RUN.
## 1 run_1653_s_3_withindex_sequence.txt_CGAAAGT.fq.gz      ERR1753604
## 2 run_1653_s_4_withindex_sequence.txt_CGAAAGT.fq.gz      ERR1753605
## 3 run_1653_s_3_withindex_sequence.txt_CAGAGTC.fq.gz      ERR1753606
## 4 run_1653_s_4_withindex_sequence.txt_CAGAGTC.fq.gz      ERR1753607
## 5 run_1653_s_3_withindex_sequence.txt_ACTGGAT.fq.gz      ERR1753608
## 6 run_1653_s_4_withindex_sequence.txt_ACTGGAT.fq.gz      ERR1753609
##                               Comment.FASTQ_URI.
## 1 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/004/ERR1753604/ERR1753604.fastq.gz
## 2 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/005/ERR1753605/ERR1753605.fastq.gz
## 3 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/006/ERR1753606/ERR1753606.fastq.gz
## 4 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/007/ERR1753607/ERR1753607.fastq.gz
## 5 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/008/ERR1753608/ERR1753608.fastq.gz
## 6 ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR175/009/ERR1753609/ERR1753609.fastq.gz
## Protocol.REF.5                               Performer.5
## 1 P-MTAB-53252 Genome Technology Access Center at Washington University
## 2 P-MTAB-53252 Genome Technology Access Center at Washington University
## 3 P-MTAB-53252 Genome Technology Access Center at Washington University
## 4 P-MTAB-53252 Genome Technology Access Center at Washington University
## 5 P-MTAB-53252 Genome Technology Access Center at Washington University
## 6 P-MTAB-53252 Genome Technology Access Center at Washington University
## Protocol.REF.6                               Performer.6
## 1 P-MTAB-53253 Genome Technology Access Center at Washington University
## 2 P-MTAB-53253 Genome Technology Access Center at Washington University
## 3 P-MTAB-53253 Genome Technology Access Center at Washington University
## 4 P-MTAB-53253 Genome Technology Access Center at Washington University
## 5 P-MTAB-53253 Genome Technology Access Center at Washington University
## 6 P-MTAB-53253 Genome Technology Access Center at Washington University
##                               Derived.Array.Data.File
## 1 all.gene_counts_D0-and-D10-samples.txt
## 2 all.gene_counts_D0-and-D10-samples.txt
## 3 all.gene_counts_D0-and-D10-samples.txt
## 4 all.gene_counts_D0-and-D10-samples.txt
## 5 all.gene_counts_D0-and-D10-samples.txt
## 6 all.gene_counts_D0-and-D10-samples.txt
##                               Comment..Derived.ArrayExpress.FTI
## 1 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 2 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 3 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 4 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 5 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 6 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## Protocol.REF.7                               Performer.7
## 1 P-MTAB-53252 Genome Technology Access Center at Washington University
## 2 P-MTAB-53252 Genome Technology Access Center at Washington University
## 3 P-MTAB-53252 Genome Technology Access Center at Washington University
## 4 P-MTAB-53252 Genome Technology Access Center at Washington University

```

```

## 5 P-MTAB-53252 Genome Technology Access Center at Washington University
## 6 P-MTAB-53252 Genome Technology Access Center at Washington University
## Protocol.REF.8 Performer.8
## 1 P-MTAB-53253 Genome Technology Access Center at Washington University
## 2 P-MTAB-53253 Genome Technology Access Center at Washington University
## 3 P-MTAB-53253 Genome Technology Access Center at Washington University
## 4 P-MTAB-53253 Genome Technology Access Center at Washington University
## 5 P-MTAB-53253 Genome Technology Access Center at Washington University
## 6 P-MTAB-53253 Genome Technology Access Center at Washington University
## Derived.Array.Data.File.1
## 1 all.transcript_counts_D0-and-D10-samples.txt
## 2 all.transcript_counts_D0-and-D10-samples.txt
## 3 all.transcript_counts_D0-and-D10-samples.txt
## 4 all.transcript_counts_D0-and-D10-samples.txt
## 5 all.transcript_counts_D0-and-D10-samples.txt
## 6 all.transcript_counts_D0-and-D10-samples.txt
## Comment..Derived.ArrayExpress.FTP..
## 1 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 2 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 3 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 4 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 5 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 6 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## Protocol.REF.9 Performer.9
## 1 P-MTAB-53252 Genome Technology Access Center at Washington University
## 2 P-MTAB-53252 Genome Technology Access Center at Washington University
## 3 P-MTAB-53252 Genome Technology Access Center at Washington University
## 4 P-MTAB-53252 Genome Technology Access Center at Washington University
## 5 P-MTAB-53252 Genome Technology Access Center at Washington University
## 6 P-MTAB-53252 Genome Technology Access Center at Washington University
## Protocol.REF.10 Performer.10
## 1 P-MTAB-53253 Genome Technology Access Center at Washington University
## 2 P-MTAB-53253 Genome Technology Access Center at Washington University
## 3 P-MTAB-53253 Genome Technology Access Center at Washington University
## 4 P-MTAB-53253 Genome Technology Access Center at Washington University
## 5 P-MTAB-53253 Genome Technology Access Center at Washington University
## 6 P-MTAB-53253 Genome Technology Access Center at Washington University
## Derived.Array.Data.File.2
## 1 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## 2 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## 3 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## 4 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## 5 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## 6 all.gene_moderated_log2cpm_D0-and-D10-samples.txt
## Comment..Derived.ArrayExpress.FTP..
## 1 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 2 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 3 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 4 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 5 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## 6 ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/MTAB/E-MTAB-5337/E-MTAB-5337.processed
## Factor.Value.genotype. Factor.Value.infect. Factor.Value.sampling.time.point.
## 1 Irgm1-/- knockout influenza A (H1N1) 10
## 2 Irgm1-/- knockout influenza A (H1N1) 10

```

```
## 3      Irgm1-/- knockout    influenza A (H1N1)      10
## 4      Irgm1-/- knockout    influenza A (H1N1)      10
## 5      Irgm1-/- knockout    influenza A (H1N1)      10
## 6      Irgm1-/- knockout    influenza A (H1N1)      10
##      Unit.time.unit..1
## 1              day
## 2              day
## 3              day
## 4              day
## 5              day
## 6              day
```

```
print(sdrf[,c("Source.Name", "Technology.Type", "Characteristics.phenotype.", "Characteristics.organism.")
```

```
##      Source.Name  Technology.Type
## 1  Infected_Irgm1-/-_Day10_1 sequencing assay
## 2  Infected_Irgm1-/-_Day10_1 sequencing assay
## 3  Infected_Irgm1-/-_Day10_2 sequencing assay
## 4  Infected_Irgm1-/-_Day10_2 sequencing assay
## 5  Infected_Irgm1-/-_Day10_3 sequencing assay
## 6  Infected_Irgm1-/-_Day10_3 sequencing assay
## 7  Uninfected_Irgm1-/-_Day0_1 sequencing assay
## 8  Uninfected_Irgm1-/-_Day0_1 sequencing assay
## 9  Uninfected_Irgm1-/-_Day0_2 sequencing assay
## 10 Uninfected_Irgm1-/-_Day0_2 sequencing assay
## 11 Uninfected_Irgm1-/-_Day0_3 sequencing assay
## 12 Uninfected_Irgm1-/-_Day0_3 sequencing assay
## 13      Uninfected_WT_Day0_1 sequencing assay
## 14      Uninfected_WT_Day0_1 sequencing assay
## 15      Uninfected_WT_Day0_2 sequencing assay
## 16      Uninfected_WT_Day0_2 sequencing assay
## 17      Uninfected_WT_Day0_3 sequencing assay
## 18      Uninfected_WT_Day0_3 sequencing assay
## 19      Infected_WT_Day10_1 sequencing assay
## 20      Infected_WT_Day10_1 sequencing assay
## 21      Infected_WT_Day10_2 sequencing assay
## 22      Infected_WT_Day10_2 sequencing assay
## 23      Infected_WT_Day10_3 sequencing assay
## 24      Infected_WT_Day10_3 sequencing assay
## 25  Infected_Irgm1-/-_Day3_1 sequencing assay
## 26  Infected_Irgm1-/-_Day3_1 sequencing assay
## 27  Infected_Irgm1-/-_Day3_2 sequencing assay
## 28  Infected_Irgm1-/-_Day3_2 sequencing assay
## 29  Infected_Irgm1-/-_Day3_3 sequencing assay
## 30  Infected_Irgm1-/-_Day3_3 sequencing assay
## 31  Infected_Irgm1-/-_Day6_1 sequencing assay
## 32  Infected_Irgm1-/-_Day6_1 sequencing assay
## 33  Infected_Irgm1-/-_Day6_2 sequencing assay
## 34  Infected_Irgm1-/-_Day6_2 sequencing assay
## 35  Infected_Irgm1-/-_Day6_3 sequencing assay
## 36  Infected_Irgm1-/-_Day6_3 sequencing assay
## 37      Infected_WT_Day3_1 sequencing assay
## 38      Infected_WT_Day3_1 sequencing assay
## 39      Infected_WT_Day3_2 sequencing assay
```

```

## 40      Infected_WT_Day3_2 sequencing assay
## 41      Infected_WT_Day3_3 sequencing assay
## 42      Infected_WT_Day3_3 sequencing assay
## 43      Infected_WT_Day6_1 sequencing assay
## 44      Infected_WT_Day6_1 sequencing assay
## 45      Infected_WT_Day6_2 sequencing assay
## 46      Infected_WT_Day6_2 sequencing assay
## 47      Infected_WT_Day6_3 sequencing assay
## 48      Infected_WT_Day6_3 sequencing assay
##          Characteristics.phenotype. Characteristics.organism.part.
## 1  protected from influenza-induced mortality          lung
## 2  protected from influenza-induced mortality          lung
## 3  protected from influenza-induced mortality          lung
## 4  protected from influenza-induced mortality          lung
## 5  protected from influenza-induced mortality          lung
## 6  protected from influenza-induced mortality          lung
## 7  protected from influenza-induced mortality          lung
## 8  protected from influenza-induced mortality          lung
## 9  protected from influenza-induced mortality          lung
## 10 protected from influenza-induced mortality          lung
## 11 protected from influenza-induced mortality          lung
## 12 protected from influenza-induced mortality          lung
## 13 susceptible to influenza-induced mortality          lung
## 14 susceptible to influenza-induced mortality          lung
## 15 susceptible to influenza-induced mortality          lung
## 16 susceptible to influenza-induced mortality          lung
## 17 susceptible to influenza-induced mortality          lung
## 18 susceptible to influenza-induced mortality          lung
## 19 susceptible to influenza-induced mortality          lung
## 20 susceptible to influenza-induced mortality          lung
## 21 susceptible to influenza-induced mortality          lung
## 22 susceptible to influenza-induced mortality          lung
## 23 susceptible to influenza-induced mortality          lung
## 24 susceptible to influenza-induced mortality          lung
## 25 protected from influenza-induced mortality          lung
## 26 protected from influenza-induced mortality          lung
## 27 protected from influenza-induced mortality          lung
## 28 protected from influenza-induced mortality          lung
## 29 protected from influenza-induced mortality          lung
## 30 protected from influenza-induced mortality          lung
## 31 protected from influenza-induced mortality          lung
## 32 protected from influenza-induced mortality          lung
## 33 protected from influenza-induced mortality          lung
## 34 protected from influenza-induced mortality          lung
## 35 protected from influenza-induced mortality          lung
## 36 protected from influenza-induced mortality          lung
## 37 susceptible to influenza-induced mortality          lung
## 38 susceptible to influenza-induced mortality          lung
## 39 susceptible to influenza-induced mortality          lung
## 40 susceptible to influenza-induced mortality          lung
## 41 susceptible to influenza-induced mortality          lung
## 42 susceptible to influenza-induced mortality          lung
## 43 susceptible to influenza-induced mortality          lung
## 44 susceptible to influenza-induced mortality          lung

```

```
## 45 susceptible to influenza-induced mortality      lung
## 46 susceptible to influenza-induced mortality      lung
## 47 susceptible to influenza-induced mortality      lung
## 48 susceptible to influenza-induced mortality      lung
```

```
#get annotation data
mart = useMart(host="useast.ensembl.org",
               biomart="ENSEMBL_MART_ENSEMBL",
               dataset="mmusculus_gene_ensembl")

mmusculus <- getBM(attributes=c('ensembl_transcript_id',
                               'ensembl_gene_id',
                               'external_gene_name'),
                  mart = mart)

head(mmusculus)
```

```
##   ensembl_transcript_id   ensembl_gene_id external_gene_name
## 1 ENSMUST00000082387 ENSMUSG00000064336      mt-Tf
## 2 ENSMUST00000082388 ENSMUSG00000064337      mt-Rnr1
## 3 ENSMUST00000082389 ENSMUSG00000064338      mt-Tv
## 4 ENSMUST00000082390 ENSMUSG00000064339      mt-Rnr2
## 5 ENSMUST00000082391 ENSMUSG00000064340      mt-Tl1
## 6 ENSMUST00000082392 ENSMUSG00000064341      mt-Nd1
```

```
# What are the available attributes
atr <- listAttributes(mart)

data <- getBM(attributes = c('ensembl_gene_id', 'ensembl_transcript_id',
                             'external_gene_name'),
              mart = mart)
```

Which samples we are using, and not using: We will only be using samples involving susceptible (Wild Type) after 10 days of infection with influenza.

Check for duplicate rows

```
# no duplicate rows
sum(duplicated(rownames(txi.kallisto$counts)))
```

```
## [1] 0
```

```
setwd("./RNAseq_output")
```

```
## Make tpm values compatible with edgeR
cts <- txi.kallisto$counts
normMat <- txi.kallisto$length
head(normMat)
```

```
##           sample1 sample2 sample3 sample4 sample5 sample6
## ENSMUSG00000000001.4 3213.0000 3213.000 3213.0000 3213.0000 3213.000 3213.0000
```

```
## ENSMUSG000000000003.15 750.5000 750.500 750.5000 750.5000 750.500 750.5000
## ENSMUSG0000000000028.15 1495.8714 1480.561 1462.6724 1460.3967 1465.040 1479.9446
## ENSMUSG0000000000037.16 1638.0000 2340.002 2573.9997 3276.0010 3602.508 3391.0000
## ENSMUSG0000000000049.11 443.4987 433.752 443.4987 443.4987 318.000 497.3792
## ENSMUSG0000000000056.7 3443.1809 3146.751 2284.8596 2217.8745 2907.992 2752.3972
## sample7 sample8 sample9 sample10 sample11 sample12
## ENSMUSG000000000001.4 3213.0000 3213.0000 3213.000 3213.0000 3213.000 3213.000
## ENSMUSG000000000003.15 750.5000 750.5000 750.500 750.5000 750.500 750.500
## ENSMUSG0000000000028.15 1540.7317 1677.0540 1743.725 1699.3468 1683.682 1591.334
## ENSMUSG0000000000037.16 2808.0003 2807.9994 3651.758 2807.9998 3442.499 3510.001
## ENSMUSG0000000000049.11 443.4987 497.3788 318.000 635.9999 682.000 318.000
## ENSMUSG0000000000056.7 2890.5020 4067.4717 3477.942 3140.7988 3472.230 2859.385
```

```
# Obtaining per-observation scaling factors for length, adjusted to avoid
# changing the magnitude of the counts.
normMat <- normMat/exp(rowMeans(log(normMat)))
normCts <- cts/normMat
head(normCts)
```

```
## sample1 sample2 sample3 sample4 sample5
## ENSMUSG000000000001.4 686.0000 618.000000 730.0000 679.000000 618.000000
## ENSMUSG000000000003.15 0.0000 0.000000 0.0000 0.000000 0.000000
## ENSMUSG0000000000028.15 153.4060 121.642167 231.9512 204.354916 178.065788
## ENSMUSG0000000000037.16 12.4827 6.241345 12.4827 6.241352 4.054054
## ENSMUSG0000000000049.11 0.0000 2.044942 0.0000 0.000000 2.789300
## ENSMUSG0000000000056.7 158.4009 176.174112 222.8767 225.495887 150.201141
## sample6 sample7 sample8 sample9 sample10
## ENSMUSG000000000001.4 586.000000 603.000000 610.000000 1038.00000 991.00000
## ENSMUSG000000000003.15 0.000000 0.000000 0.000000 0.00000 0.00000
## ENSMUSG0000000000028.15 223.130622 106.616330 55.786146 170.16046 181.10422
## ENSMUSG0000000000037.16 4.306915 6.241344 12.482701 11.19825 12.48269
## ENSMUSG0000000000049.11 1.783343 0.000000 1.783344 1.39465 1.39465
## ENSMUSG0000000000056.7 133.225844 202.129337 125.939701 188.81850 203.66169
## sample11 sample12
## ENSMUSG000000000001.4 584.000000 621.00000
## ENSMUSG000000000003.15 0.000000 0.00000
## ENSMUSG0000000000028.15 147.9495762 125.32669
## ENSMUSG0000000000037.16 6.7879752 12.48269
## ENSMUSG0000000000049.11 0.6502914 1.39465
## ENSMUSG0000000000056.7 144.0585049 162.63249
```

```
# Computing effective library sizes from scaled counts, to account for
# composition biases between samples.
eff.lib <- calcNormFactors(normCts) * colSums(normCts)
head(eff.lib)
```

```
## sample1 sample2 sample3 sample4 sample5 sample6
## 6809954 6625648 7294221 7122479 6301408 6145094
```

```
# Combining effective library sizes with the length factors, and calculating
# offsets for a log-link GLM.
```

```
#merge every two columns which correspond to each 2 technical replicates
```

```
merged_normMat <- NULL
for (x in seq(1,12,2)){
  merged_normMat <- cbind(merged_normMat,normMat[,x]+normMat[,x+1])
}

normMat <- merged_normMat
normMat <- sweep(normMat, 2, eff.lib, "*")
```

```
## Warning in sweep(normMat, 2, eff.lib, "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
```

```
normMat <- log(normMat)

new_cts <- NULL
for (c in seq(1,12,2)){
  new_cts <- cbind(new_cts, cts[,c]+cts[,c+1])
}
cts <- new_cts

y <- DGEList(cts)
y <- scaleOffset(y, normMat)
head(y)
```

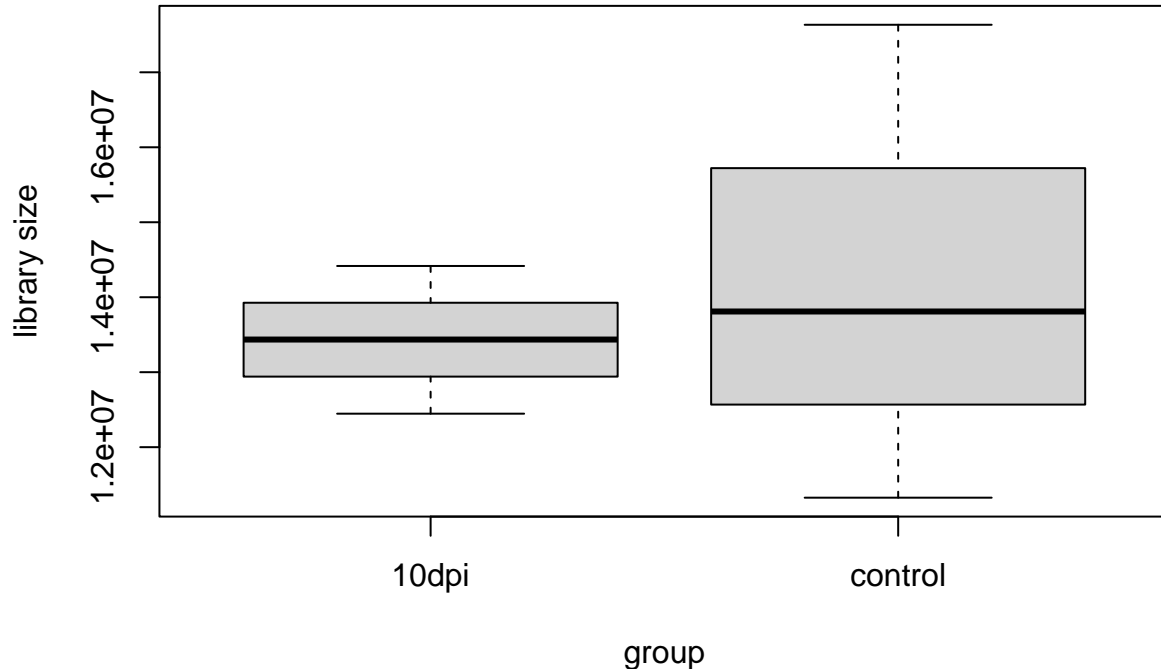
```
## An object of class "DGEList"
## $counts
##
##      Sample1      Sample2      Sample3      Sample4      Sample5
## ENSMUSG00000000001.4 1304.0000 1409.0000 1204.0000 1213.0000 2029.000000
## ENSMUSG00000000003.15  0.00000  0.00000   0.0000   0.0000   0.000000
## ENSMUSG000000000028.15 262.26604 408.34856 378.5002 165.0942 387.066300
## ENSMUSG000000000037.16 12.00001 18.00001 10.0000 18.0000 26.000006
## ENSMUSG000000000049.11  2.00000  0.00000  4.0000  2.0000  2.999999
## ENSMUSG000000000056.7 364.97800 334.97230 266.6449 363.8936 430.216440
##
##      Sample6
## ENSMUSG00000000001.4 1205.0000
## ENSMUSG00000000003.15  0.0000
## ENSMUSG000000000028.15 287.2151
## ENSMUSG000000000037.16 23.0000
## ENSMUSG000000000049.11  2.0000
## ENSMUSG000000000056.7 320.3270
##
## $samples
##      group lib.size norm.factors
## Sample1    1 13413997          1
## Sample2    1 14385696          1
## Sample3    1 12362925          1
## Sample4    1 13767983          1
## Sample5    1 18090430          1
## Sample6    1 11097351          1
##
## $offset
##
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## ENSMUSG00000000001.4 16.44816 16.42072 16.51685 16.49303 16.37055 16.34543
```

```
## ENSMUSG000000000003.15 16.42970 16.40437 16.66884 16.65421 16.22957 16.20804
## ENSMUSG0000000000028.15 16.39974 16.35421 16.45782 16.52258 16.46777 16.39261
## ENSMUSG0000000000037.16 16.04006 16.40039 16.84341 16.60941 16.32474 16.37672
## ENSMUSG0000000000049.11 16.41103 16.39464 16.40659 16.52592 16.41729 16.43926
## ENSMUSG0000000000056.7 16.51539 16.10921 16.60249 16.79426 16.31963 16.25375
```

```
## Library sizes
infection <- factor(c("10dpi","10dpi","10dpi","control", "control", "control"))
lib <- NULL
n <- 1
for (x in seq(1,12,2)){
  lib[n] <- eff.lib[x]+ eff.lib[x+1]
  n <- n + 1
}
lib
```

```
## [1] 13435602 14416700 12446502 13809810 17633891 11326350
```

```
#jpeg("library_sizes_musmusculus.jpg")
boxplot(lib~as.factor(infection),xlab="group",ylab="library size")
```



```
#dev.off()
wilcox.test(lib~as.factor(infection))
```



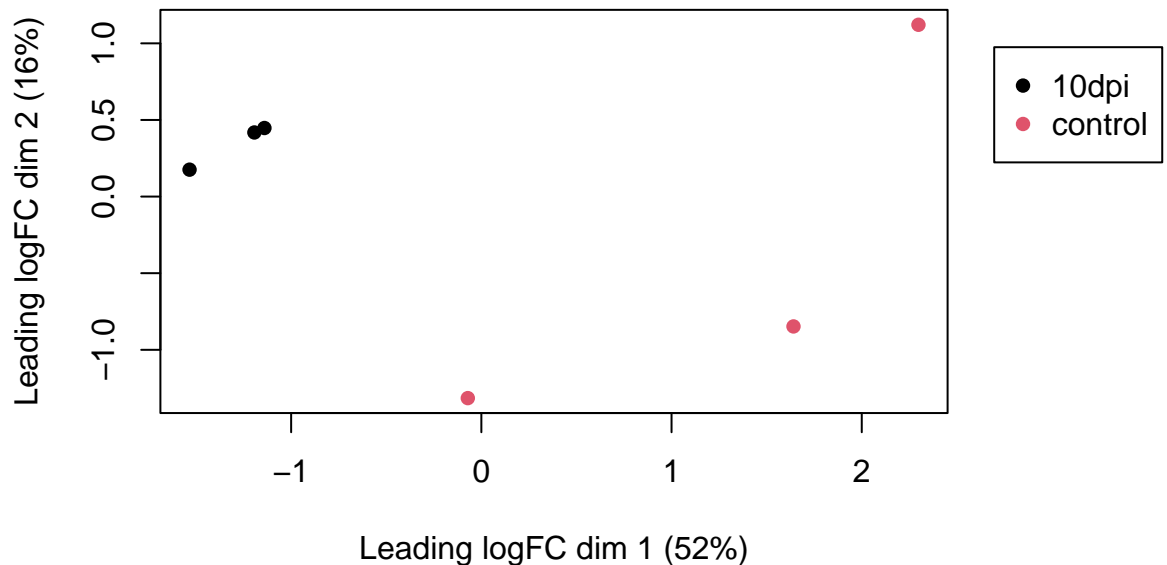
```
##
## Wilcoxon rank sum exact test
##
## data: lib by as.factor(infection)
## W = 4, p-value = 1
## alternative hypothesis: true location shift is not equal to 0

## Optionally filter on counts mean
cutoff <- 3/(mean(y$samples$lib.size)/1000000)
keep <- rowSums(cpm(y)>cutoff) >= 3
y <- y[keep, ,keep.lib.sizes=FALSE]
summary(keep) #FALSE 15672 TRUE 20375

## Mode FALSE TRUE
## logical 15672 20375

## MDS plot
#jpeg("plotMDS_mmusculus.jpg")
par(mar=c(6,6,6,6))
plotMDS(y,col=as.numeric(as.factor(infection)), pch=16, main="3 infected vs 3 controls")
par(xpd=T)
legend(par("usr")[2]*1.1,par("usr")[4]*0.8,sort(unique(infection)),
      pch=c(16),col=as.double(as.factor(sort(unique(infection)))))
```

3 infected vs 3 controls



```
#dev.off()
```

```
## Differential expression analysis
```

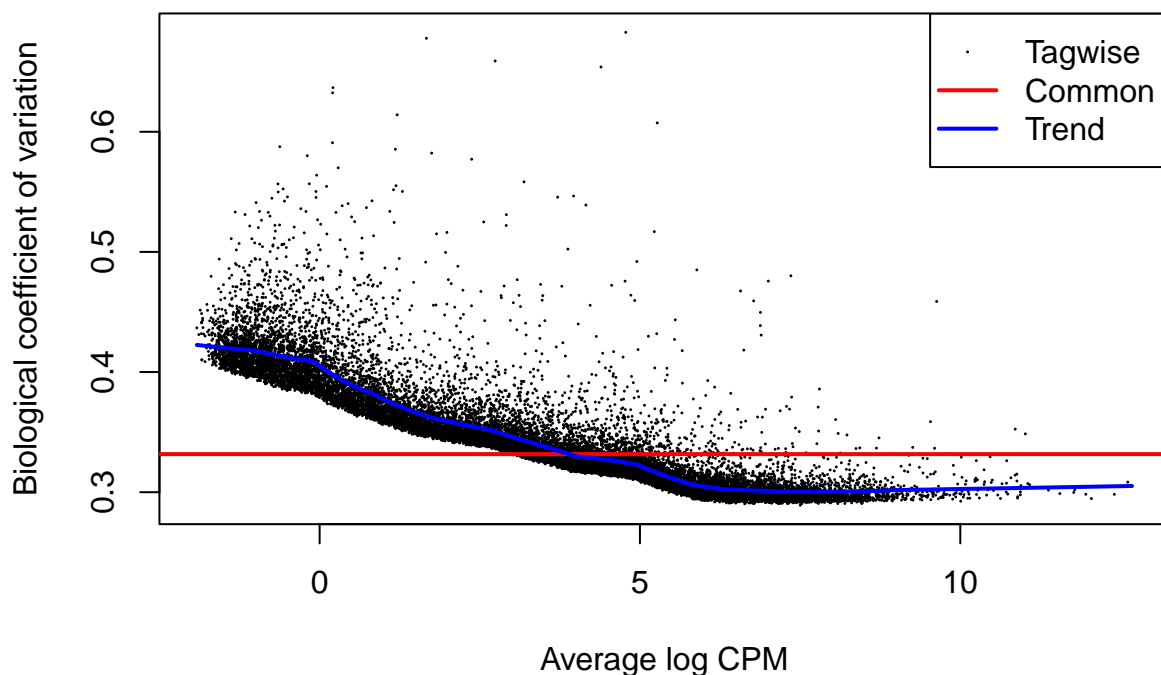
```
design <- model.matrix(~infection)
```

```
rownames(design) <- colnames(y)
```

```
y <- estimateDisp(y,design)
```

```
#jpeg("BCVplot_mmusculus.jpg")
```

```
plotBCV(y)
```



```
#dev.off()
```

```
#there is a trend!
```

```
fit_edgeR <- glmQLFit(y,design)
```

```
qlf_edgeR <- glmQLFTest(fit_edgeR,coef=2)
```

```
# coef=2 => infected vs control, see "design" object
```

```
# note: standard, the last column is considered as the contrast of interest
```

```
res_edgeR <- topTags(qlf_edgeR,n=nrow(qlf_edgeR$table))$table
```

```
## Add gene symbols
```

```
data_sorted_edgeR <- data[sort(data$ensembl_transcript_id,index.return=T)$ix,]
```

```
data_sorted_edgeR <- data_sorted_edgeR[duplicated(data_sorted_edgeR$ensembl_gene_id)==F,]
```

```
res_edgeR <- cbind(rownames(res_edgeR),res_edgeR)
```

```
colnames(res_edgeR)[1] <- "Ensembl_gene_id"

res_edgeR$diffexpressed <- "NO"
res_edgeR$diffexpressed[res_edgeR$logFC > 0 & res_edgeR$FDR < 0.05] <- "UP"
res_edgeR$diffexpressed[res_edgeR$logFC < 0 & res_edgeR$FDR < 0.05] <- "DOWN"

res_edgeR_sorted <- res_edgeR[sort(res_edgeR$Ensembl_gene_id, index.return=T)$ix,]
head(res_edgeR_sorted)
```

```
##               Ensembl_gene_id      logFC  logCPM      F
## ENSMUSG00000000001.4  ENSMUSG00000000001.4  0.2742049  6.650000  0.5884430
## ENSMUSG000000000028.15 ENSMUSG000000000028.15 -0.3937051  4.530961  1.0743186
## ENSMUSG000000000037.16 ENSMUSG000000000037.16  0.7040886  0.522503  1.8979925
## ENSMUSG000000000056.7  ENSMUSG000000000056.7  0.1279611  4.663505  0.1152170
## ENSMUSG000000000058.6  ENSMUSG000000000058.6 -0.5199073  7.047578  2.0858873
## ENSMUSG000000000078.7  ENSMUSG000000000078.7  0.1639739  7.977682  0.2053453
##               PValue      FDR diffexpressed
## ENSMUSG00000000001.4  0.4467346  0.8271373      NO
## ENSMUSG000000000028.15 0.3051047  0.7431571      NO
## ENSMUSG000000000037.16 0.1746180  0.6024115      NO
## ENSMUSG000000000056.7  0.7357444  0.9338949      NO
## ENSMUSG000000000058.6  0.1550858  0.5732253      NO
## ENSMUSG000000000078.7  0.6524583  0.9108295      NO
```

```
original <- res_edgeR$Ensembl_gene_id
#substr(original[1], 1, 18)
n <- 1
for (id in original){
  res_edgeR$Ensembl_gene_id[n] <- substr(id, 1, 18)
  n <- n + 1
}
head(res_edgeR$Ensembl_gene_id)
```

```
## [1] "ENSMUSG00000076612" "ENSMUSG000000100131" "ENSMUSG00000029417"
## [4] "ENSMUSG00000076613" "ENSMUSG000000094708" "ENSMUSG00000095937"
```

```
ids <- res_edgeR_sorted$Ensembl_gene_id
#substr(original[1], 1, 18)
n <- 1
for (id in ids){
  res_edgeR_sorted$Ensembl_gene_id[n] <- substr(id, 1, 18)
  n <- n + 1
}
#gsub
data_sorted_edgeR <- data_sorted_edgeR[data_sorted_edgeR$ensembl_gene_id%in%res_edgeR_sorted$Ensembl_gene_id,]
res_edgeR_sorted <- res_edgeR_sorted[res_edgeR_sorted$Ensembl_gene_id%in%data_sorted_edgeR$ensembl_gene_id,]

dim(res_edgeR_sorted) #19174
```

```
## [1] 20321      7
```

```
dim(data_sorted_edgeR) #19174
```

```
## [1] 20321      3
```

```
#which(res_edgeR_sorted$Ensembl_gene_id == data_sorted_edgeR$ensembl_transcript_id)
```

```
res_edgeR_sorted$Gene_symbol <- data_sorted_edgeR$external_gene_name
```

```
## Resort and save results
```

```
res_edgeR <- res_edgeR_sorted[sort(res_edgeR_sorted$PValue,index.return=T)$ix,]  
head(res_edgeR[,c(1,7,2,5,6)],10) #geneID, symbol, logFC, pvalue, FDR
```

##		Ensembl_gene_id	diffexpressed	logFC	PValue
##	ENSMUSG00000076612.8	ENSMUSG00000076612	UP	6.251235	3.580705e-17
##	ENSMUSG00000100131.1	ENSMUSG00000100131	DOWN	-12.409971	6.470052e-17
##	ENSMUSG00000029417.9	ENSMUSG00000029417	UP	7.421845	6.322303e-16
##	ENSMUSG00000076613.4	ENSMUSG00000076613	UP	5.750282	1.958567e-15
##	ENSMUSG00000094708.1	ENSMUSG00000094708	DOWN	-9.590055	4.942463e-15
##	ENSMUSG00000095937.1	ENSMUSG00000095937	DOWN	-9.514044	6.419399e-15
##	ENSMUSG00000099875.1	ENSMUSG00000099875	UP	9.619127	1.212535e-14
##	ENSMUSG00000031972.5	ENSMUSG00000031972	UP	8.506408	8.774228e-14
##	ENSMUSG00000034855.13	ENSMUSG00000034855	UP	6.094653	2.401049e-13
##	ENSMUSG00000042385.14	ENSMUSG00000042385	UP	6.964233	2.638372e-13
##		FDR			
##	ENSMUSG00000076612.8	6.591366e-13			
##	ENSMUSG00000100131.1	6.591366e-13			
##	ENSMUSG00000029417.9	4.293898e-12			
##	ENSMUSG00000076613.4	9.976453e-12			
##	ENSMUSG00000094708.1	2.014054e-11			
##	ENSMUSG00000095937.1	2.179921e-11			
##	ENSMUSG00000099875.1	3.529343e-11			
##	ENSMUSG00000031972.5	2.234686e-10			
##	ENSMUSG00000034855.13	5.375683e-10			
##	ENSMUSG00000042385.14	5.375683e-10			

```
write.table(res_edgeR,file="res_edgeR_mmusculus.txt",col.names=T,row.names=T,sep="\t",quote=F)  
# top 1000 loci, also contain non-significant loci (filtered out in next step)  
res_edgeR_sign <- res_edgeR[res_edgeR$FDR<0.05,]  
dim(res_edgeR_sign)
```

```
## [1] 1329      8
```

```
# last column of res (ncol(res)) contains FDRs => filtering at 5% level
```

```
## MA plot
```

```
jpeg("res_edgeR_mmusculus_MA.png")
```

```
## MA-plot
```

```
with(res_edgeR,plot(logCPM,logFC,pch=16,cex=0.2))
```

```

# MAplot: all data points
with(res_edgeR, points(logCPM[FDR<0.05], logFC[FDR<0.05], pch=16, col="red", cex=0.6))
# MA-plot: significant loci
abline(0,0)
# X-axis
dev.off()

```

```

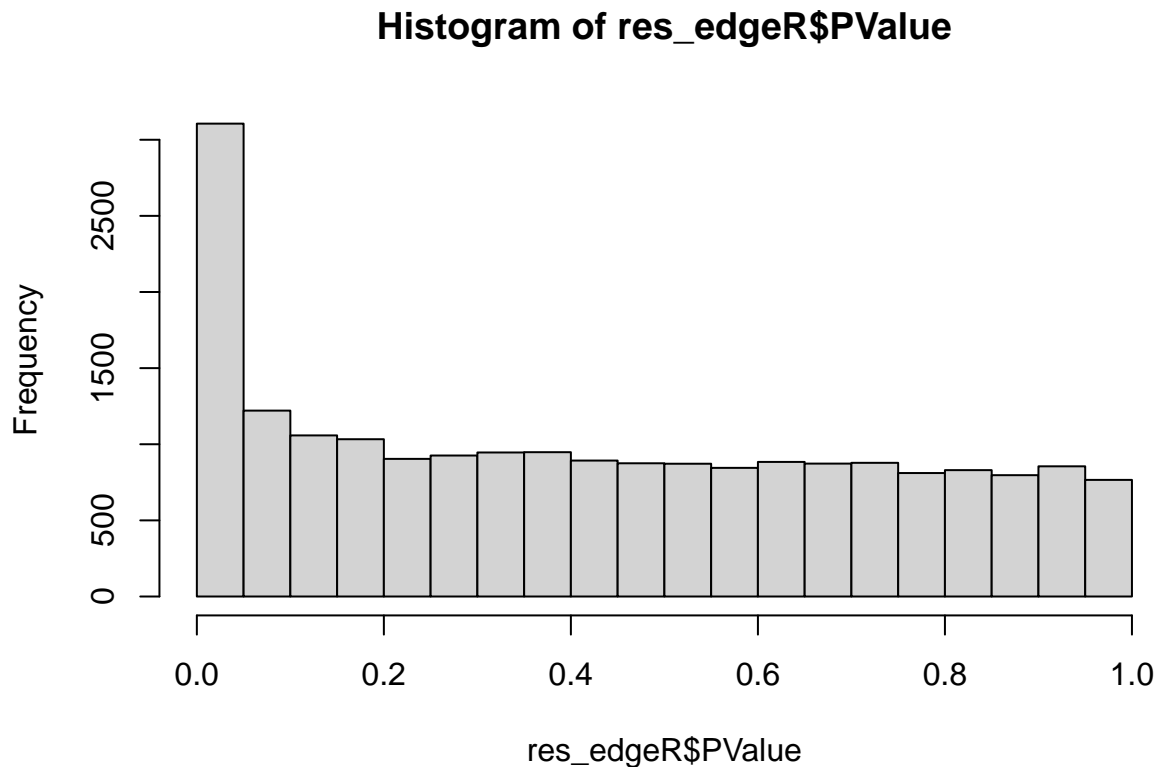
## pdf
## 2

```

```

## Pvalue distribution
#jpeg("res_edgeR_mmusculus_histogram.jpg")
hist(res_edgeR$PValue)

```



```

#dev.off()

## Plot boxplots for top 20 loci (first make the folder where you want to put them!)
setwd("./RNAseq_output/Boxplots/")
counts_k <- txi.kallisto$counts[keep,]
for (i in 1:20){
  counts_part <- as.numeric(cpm(y)[rownames(counts_k)==rownames(res_edgeR)[i],])
  dat_boxplot <- data.frame(counts=counts_part, group=infection)
  jpeg(paste(i, "_", rownames(res_edgeR)[i], ".jpg", sep=""))
}

```

```

if (res_edgeR$Gene_symbol[i]!=""){
  boxplot(counts~group,dat_boxplot,main=paste(rownames(res_edgeR)[i], " (",res_edgeR$Gene_symbol[i],")",
} else {
  boxplot(counts~group,dat_boxplot,main=paste(rownames(res_edgeR)[i], " (NA)",sep=""))
}
dev.off()
}

```

```

RNAseq_DEgenes <- res_edgeR_sign$Gene_symbol
head(RNAseq_DEgenes)

```

```

## [1] "Thy1"      "Hmgb1-ps8" "Rhbdd3"    "Uhrf1bp1"  "Gm17235"   "Hspe1-ps3"

```

```

length(RNAseq_DEgenes)

```

```

## [1] 1329

```

```

length(which(res_edgeR$diffexpressed=="UP")) #99

```

```

## [1] 1230

```

```

length(which(res_edgeR$diffexpressed=="DOWN")) #1230

```

```

## [1] 99

```

```

RNA_upregulated <- res_edgeR$Gene_symbol[res_edgeR$diffexpressed=="UP"]
RNA_downregulated <- res_edgeR$Gene_symbol[res_edgeR$diffexpressed=="DOWN"]

```