script2.R

vlad

2019-10-21

```
setwd("/home/vlad/Documents/programming_ls/group-project/")
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: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,
##     which.max, which.min
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:base':
##
##     expand.grid
```

```
## Loading required package: IRanges
```

```
## 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")'.

## Loading required package: DelayedArray

## Loading required package: matrixStats

##
## Attaching package: 'matrixStats'

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

## Loading required package: BiocParallel

##
## Attaching package: 'DelayedArray'

## The following objects are masked from 'package:matrixStats':
##
##     colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

## The following objects are masked from 'package:base':
##
##     aperm, apply, rowsum
```

```
reads <- read.table("raw_countstdl.tsv", header = TRUE, sep = '\t')
reads_nonzero <- subset(reads, ctl1 != 0 | ctl2 != 0 | ctl3 != 0 | ctl4 != 0 | treat1 != 0 | treat2 != 0

head(reads_nonzero)
```

```
##                  gene ctl1 ctl2 ctl3 ctl4 treat1 treat2 treat3 treat4
## 2  ENSMUSG00000064842    0    0    0    0      1      0      0      0
## 3  ENSMUSG00000051951    2    0    1    1     11      2      1     24
## 10 ENSMUSG00000103147    0    1    0    0      0      0      0      0
## 11 ENSMUSG00000103161    0    0    0    0      1      0      0      0
## 12 ENSMUSG00000102331    0    0    0    0      0      0      1      3
## 17 ENSMUSG00000102948    2    0    1    1      0      0      0      0
```

```
library ( magrittr ) # this will allow us to string commands together in a UNIX -pipe - like fashion us

row.names(reads_nonzero) <- reads_nonzero$gene

read_counts <- reads_nonzero [ , c(2:9)]

head(read_counts)
```

```
##                    ctl1 ctl2 ctl3 ctl4 treat1 treat2 treat3 treat4
## ENSMUSG00000064842    0    0    0    0      1      0      0      0
## ENSMUSG00000051951    2    0    1    1     11      2      1     24
## ENSMUSG00000103147    0    1    0    0      0      0      0      0
## ENSMUSG00000103161    0    0    0    0      1      0      0      0
## ENSMUSG00000102331    0    0    0    0      0      0      1      3
## ENSMUSG00000102948    2    0    1    1      0      0      0      0
```

```
head ( read_counts , n = 3)
```

```
##                    ctl1 ctl2 ctl3 ctl4 treat1 treat2 treat3 treat4
## ENSMUSG00000064842    0    0    0    0      1      0      0      0
## ENSMUSG00000051951    2    0    1    1     11      2      1     24
```

```
## ENSMUSG00000103147    0    1    0    0      0      0      0      0
```

```
# make a data . frame with meta - data where row. names should match the individual sample names

sample_info <- data.frame (condition = gsub("_ [0 -9]+ " , " " , names(read_counts) ) , row.names = name

sample_info
```

```
##        condition
## ctl1        ctl1
## ctl2        ctl2
## ctl3        ctl3
## ctl4        ctl4
## treat1    treat1
## treat2    treat2
## treat3    treat3
## treat4    treat4
```

```
# generate the DESeqDataSet
DESeq.ds <- DESeqDataSetFromMatrix(countData = read_counts , colData = sample_info , design = ~conditio

# you can check the result using the accessors described above :
colData (DESeq.ds) %>% head
```

```
## DataFrame with 6 rows and 1 column
##        condition
##         <factor>
## ctl1        ctl1
## ctl2        ctl2
## ctl3        ctl3
## ctl4        ctl4
## treat1    treat1
## treat2    treat2
```

```
assay (DESeq.ds,"counts" ) %>% head
```

```
##                    ctl1 ctl2 ctl3 ctl4 treat1 treat2 treat3 treat4
## ENSMUSG00000064842    0    0    0    0      1      0      0      0
## ENSMUSG00000051951    2    0    1    1     11      2      1     24
## ENSMUSG00000103147    0    1    0    0      0      0      0      0
## ENSMUSG00000103161    0    0    0    0      1      0      0      0
## ENSMUSG00000102331    0    0    0    0      0      0      1      3
## ENSMUSG00000102948    2    0    1    1      0      0      0      0
```

```
rowData (DESeq.ds) %>% head
```

```
## DataFrame with 6 rows and 0 columns
```

```
# test what counts () returns
counts(DESeq.ds) %>% str
```

```
##  int [1:28474, 1:8] 0 2 0 0 0 2 0 1 1 5 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:28474] "ENSMUSG00000064842" "ENSMUSG00000051951" "ENSMUSG00000103147" "ENSMUSG0000010
##   ..$ : chr [1:8] "ctl1" "ctl2" "ctl3" "ctl4" ...
```

```
# remove genes without any counts
DESeq.ds <- DESeq.ds [ rowSums ( counts ( DESeq.ds ) ) > 0 , ]
```

```r
colSums(counts(DESeq.ds)) # should be the same as colSums ( readcounts )
```

```
##      ctl1      ctl2      ctl3      ctl4    treat1    treat2    treat3
##  41375067  34613772  35482626  37119570  40026818  36909672  40303811
##    treat4
## 173631720
```

```r
colSums(read_counts)#GOOD
```

```
##      ctl1      ctl2      ctl3      ctl4    treat1    treat2    treat3
##  41375067  34613772  35482626  37119570  40026818  36909672  40303811
##    treat4
## 173631720
```

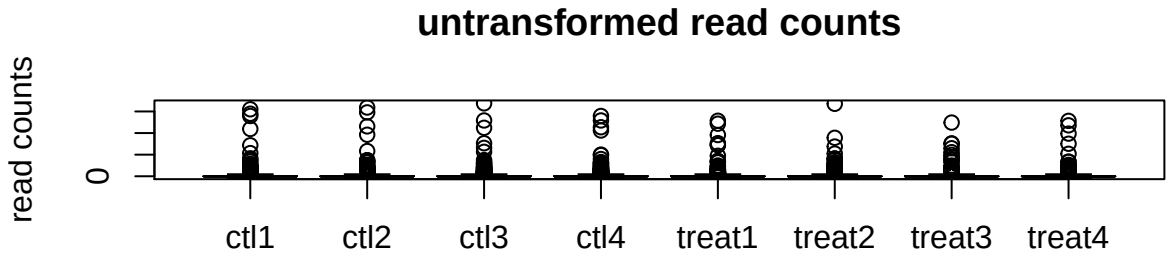
```r
# calculate the size factor and add it to the data set
DESeq.ds <- estimateSizeFactors(DESeq.ds)
sizeFactors(DESeq.ds)
```

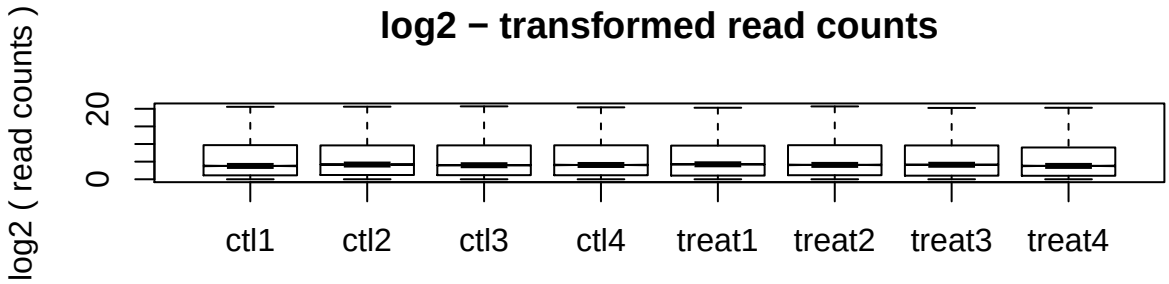
```
##      ctl1      ctl2      ctl3      ctl4    treat1    treat2    treat3
## 0.8613852 0.7601755 0.8089857 0.8337696 0.9487156 0.8192834 0.9787590
##    treat4
## 4.0539751
```

```r
# if you check colData () again , you see that this now contains the sizeFactors
colData(DESeq.ds)
```

```
## DataFrame with 8 rows and 2 columns
##        condition        sizeFactor
##         <factor>         <numeric>
## ctl1        ctl1 0.861385212241083
## ctl2        ctl2 0.760175549092362
## ctl3        ctl3 0.808985675298092
## ctl4        ctl4 0.833769556412077
## treat1    treat1  0.94871557100813
## treat2    treat2 0.819283353107582
## treat3    treat3 0.978758960142215
## treat4    treat4  4.05397505356715
```

```r
# counts () allows you to immediately retrieve the _ normalized _ read counts
counts_normalized <- counts(DESeq.ds , normalized = TRUE )

# transform size - factor normalized read counts to log2 scale using a pseudocount of 1
counts_lognorm <- log2(counts_normalized + 1)


par( mfrow =c(2 ,1) ) # to plot the following two images underneath each other

boxplot ( counts_normalized , notch = TRUE , main = " untransformed read counts " , ylab = " read counts

# box plots of log2 - transformed read counts
boxplot (counts_lognorm , notch = TRUE ,main = " log2 - transformed read counts " , ylab = " log2 ( read
```


untransformed read counts
read counts
0
ctl1
ctl2
ctl3
ctl4
treat1
treat2
treat3
treat4


log2 − transformed read counts
log2 (read counts)
0
20
ctl1
ctl2
ctl3
ctl4
treat1
treat2
treat3
treat4