Salmon Tutorial DESeq2 Application

Prisma Erika Lopez Jimenez

9/28/2019

```
### Load the quant.sf files as `files` and the sample folders as `folders`
# dir # set the working directory and `dir` as location of quant folders
# "C:/Users/PELJ/Dropbox/bioinformatics/salmon_tutorial/quants"
setwd("C:/Users/PELJ/Dropbox/bioinformatics/software/salmon/salmon_tutorial/quants")
dir=getwd()
folders=list.files(file.path(dir))
files <- file.path(dir,folders, "quant.sf")
names(files)=paste0("sample",1:2)
all(file.exists(files)) # check
```

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

```
### Inspect the biomaRt database: TxDb.Athaliana.BioMart.plantsmart28
### Apparenetly, correct database -- finally!
### Also, shortcut to building a tx2gene

txdb=TxDb.Athaliana.BioMart.plantsmart28 # txdb object
columns(txdb)
```

```
##  [1] "CDSCHROM"   "CDSEND"     "CDSID"      "CDSNAME"    "CDSSTART"
##  [6] "CDSSTRAND"  "EXONCHROM"  "EXONEND"    "EXONID"     "EXONNAME"
## [11] "EXONRANK"   "EXONSTART"  "EXONSTRAND" "GENEID"     "TXCHROM"
## [16] "TXEND"      "TXID"       "TXNAME"     "TXSTART"    "TXSTRAND"
## [21] "TXTYPE"
```

```
keytypes(txdb)
```

```
## [1] "CDSID"    "CDSNAME"  "EXONID"   "EXONNAME" "GENEID"   "TXID"
## [7] "TXNAME"
```

```
k <- keys(txdb, keytype="TXNAME")
tx2gene <- select(txdb, k, "GENEID", "TXNAME")
```

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

```
#tx2gene
#genes=genes(txdb) # GRanges object
#genes$gene_id # access GRances object Ref: https://kasperdanielhansen.github.io/genbioconductor/html/G
```

```
txi <- tximport(files, type="salmon", tx2gene=tx2gene)
```

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

```
## 1 2
## summarizing abundance
## summarizing counts
## summarizing length
```

```
samples <- read.table(file.path("C:/Users/PELJ/Dropbox/bioinformatics/software/salmon/","samples.txt"),
samples
```

```
##    sample        ID Experiment
## 1 sample1 DRR016125       ctrl
## 2 sample2 DRR016126        exp
```

```
dds <- DESeqDataSetFromTximport(txi, samples, ~1) # can't right now because only two samples
```

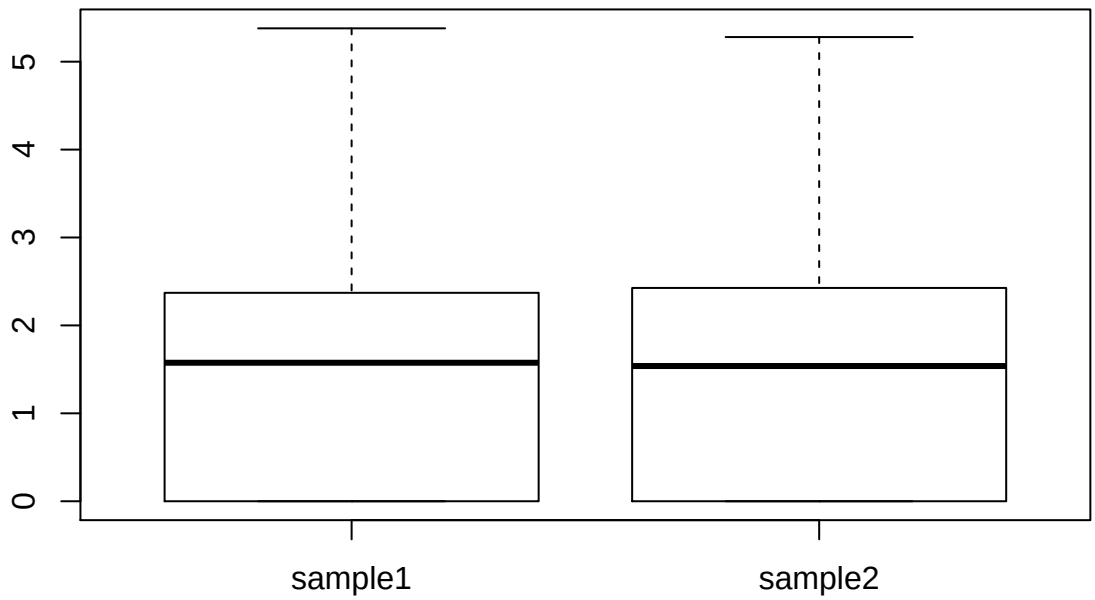
```
## using counts and average transcript lengths from tximport
```

```
keep <- rowSums(counts(dds) >= 5) >= 4

# option:
#table(keep)
#boxplot(log10(counts(dds)+1))
# Alt:
dds <- estimateSizeFactors(dds)
```

```
## using 'avgTxLength' from assays(dds), correcting for library size
```

```
boxplot(log10(counts(dds,normalized=TRUE)+1))
```




```
dds.de = DESeq(dds)
```

```
## Warning in DESeq(dds): the design is ~ 1 (just an intercept). is this
## intended?
```

```
## using pre-existing normalization factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
```

```
results(dds.de)
```

```
## log2 fold change (MLE): Intercept
## Wald test p-value: Intercept
## DataFrame with 32753 rows and 6 columns
##                   baseMean       log2FoldChange             lfcSE
##                  <numeric>            <numeric>         <numeric>
## AT1G01010 62.3511991981287     5.96234540250536 0.610138345220897
## AT1G01020 99.8393688017216     6.64153690819906 0.564065044873806
## AT1G01030 29.6186272043248     4.88843286949759   0.7475824563292
## AT1G01040 379.408034062211     8.56760641828424 0.482025440706227
## AT1G01050 796.569125487938     9.63765575261937  0.47168861951817
```

```
## ...                    ...                  ...                ...
## ATMG01350 5.07112370878421     2.34230546901558  1.55811883263992
## ATMG01360 68.6141478402139     6.10043417732706 0.598028426298321
## ATMG01370 17.5110880776415     4.13019682545762 0.910865296597283
## ATMG01400 1.00001934421169 2.79075283494943e-05  3.10207250262414
## ATMG01410               0                   NA                NA
##                          stat               pvalue                padj
##                     <numeric>            <numeric>           <numeric>
## AT1G01010     9.77212045302074 1.48314861200816e-22  2.3827700474694e-22
## AT1G01020     11.7744167424608 5.28815960853722e-32 9.77294831555547e-32
## AT1G01030      6.5389882120842 6.19364061708629e-11 8.52393572880534e-11
## AT1G01040      17.774178901702 1.12025897189913e-70 5.00690241937198e-70
## AT1G01050     20.4322414275423 8.64414434008747e-93 8.26744834955924e-92
## ...                        ...                  ...                ...
## ATMG01350     1.50329064763759    0.132764110768059    0.151526094103512
## ATMG01360     10.2009100388213  1.9642328967704e-24 3.24674422792175e-24
## ATMG01370        4.53436621297  5.7776709656623e-06 7.39882176076353e-06
## ATMG01400 8.99641395418271e-06    0.999992821900203    0.999999974989234
## ATMG01410                   NA                   NA                   NA
```

```
vsd <- vst(dds)
assay(vsd)[1:2,1:2]
```

```
##            sample1  sample2
## AT1G01010 6.474874 6.517420
## AT1G01020 7.210520 6.766467
```

```
### References:
## Refer to .Rmd file: Intro to transcript/genome annotations access
## introductory chunks
# https://combine-lab.github.io/salmon/getting_started/ # main salmon page
# https://bioconductor.github.io/BiocWorkshops/rna-seq-data-analysis-with-deseq2.html #
# http://127.0.0.1:31884/library/tximport/doc/tximport.html #importing quant.sf files reference
# https://bioconductor.org/packages/devel/bioc/vignettes/GenomicFeatures/inst/doc/GenomicFeatures.pdf

# https://bioconductor.riken.jp/packages/3.0/data/annotation/ # package list names for access to transc
# https://ropensci.github.io/biomartr/articles/Functional_Annotation.html # Guide to access plants (ath


## Database access options using AnnotationHub (Bioconductor Forums)
# https://support.bioconductor.org/p/115371/
# https://support.bioconductor.org/p/109092/
# https://support.bioconductor.org/p/111536/
# https://davetang.org/muse/2017/08/08/getting-started-arabidopsis-thaliana-genomics/ # get contents of

## Database access options using Ensembl
# http://127.0.0.1:23132/library/ensembldb/doc/ensembldb.html
# https://support.bioconductor.org/p/104194/
# https://support.bioconductor.org/t/ensembldb/ # bioconductor post forums
```