

Tximport

NGS_worshop

Data 17th May 2023 – 1st hour

Mirvat Surakhy

This code is for changing the TPM values to counts that can be used for differential gene expression using `deseq2`

The codes are based on the following tutorial

<https://bioconductor.org/packages/release/bioc/vignettes/tximport/inst/doc/tximport.html>

Upload the required libraries

```
library("tximport")
library("dplyr")
```

List all directories containing data

```
samples <- list.files(path = "./salmon_quants", full.names = T, pattern="_sample_quant$")

samples

##[1] "./salmon_quants/5aza_rep1_sample_quant"
##[2] "./salmon_quants/5aza_rep2_sample_quant"
##[3] "./salmon_quants/DMSO_rep1_sample_quant"
##[4] "./salmon_quants/DMSO_rep2_sample_quant"
```

Obtain a vector of all filenames including the path

```
files <- file.path(samples, "quant.sf")
files

## [1] "./salmon_quants/5aza_rep1_sample_quant/quant.sf"
## [2] "./salmon_quants/5aza_rep2_sample_quant/quant.sf"
## [3] "./salmon_quants/DMSO_rep1_sample_quant/quant.sf"
## [4] "./salmon_quants/DMSO_rep2_sample_quant/quant.sf"
```

Assign a shorter name for each element

```
names(files) <- list.files("salmon_quants")
```

Read the annotation file

```
tx2gene <- read.csv("tx2gene_ens109.csv")  
#use the head command to see first few lines of the tx2gene  
head(tx2gene)
```

```
##   X          TXNAME          GENEID  
## 1 1 ENST00000636745 ENSG00000283640  
## 2 2 ENST00000636387 ENSG00000283640  
## 3 3 ENST00000643422 ENSG00000284971  
## 4 4 ENST00000645631 ENSG00000284971  
## 5 5 ENST00000636567 ENSG00000283158  
## 6 6 ENST00000636545 ENSG00000283350
```

Run tximport

```
txi <- tximport(files, type="salmon", tx2gene=tx2gene[,c("TXNAME", "GENEID")]  
, countsFromAbundance="lengthScaledTPM", ignoreTxVersion = TRUE)
```

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

```
## 1 2 3 4  
## summarizing abundance  
## summarizing counts  
## summarizing length
```

```
#view the head of the output
```

```
head(txi[["counts"]])
```

```
##           5aza_rep1_sample_quant 5aza_rep2_sample_quant  
## ENSG000000000003           16.231425           16.5572893  
## ENSG000000000005           0.000000           0.0000000  
## ENSG000000000419           49.226447           53.1368343  
## ENSG000000000457           8.178605           14.7950203  
## ENSG000000000460           7.244632           8.2400971  
## ENSG000000000938           0.000000           0.9825487  
##           DMSO_rep1_sample_quant DMSO_rep2_sample_quant  
## ENSG000000000003           19.218635           11.464534  
## ENSG000000000005           0.000000           0.0000000  
## ENSG000000000419           62.466155           61.953858  
## ENSG000000000457           28.102041           16.140238  
## ENSG000000000460           4.750972           1.127791  
## ENSG000000000938           0.000000           0.000000
```

Extract the counts from the txi list, round the numbers in the counts list and convert it to a data frame

```
counts <- txi$counts %>% round()  
counts <- as.data.frame(counts)  
#change the column names  
colnames(counts) <- c("5aza_rep1", "5aza_rep2", "DMSO_rep1", "DMSO_rep2" )
```

##view the output, can you notice the change

head(counts)

##	5aza_rep1	5aza_rep2	DMSO_rep1	DMSO_rep2
## ENSG00000000003	16	17	19	11
## ENSG00000000005	0	0	0	0
## ENSG000000000419	49	53	62	62
## ENSG000000000457	8	15	28	16
## ENSG000000000460	7	8	5	1
## ENSG000000000938	0	1	0	0

save the results

```
write.table(counts, "counts/Count_fromtximport_Salmon.txt", sep = "\t", quote = F)
```

```
sessionInfo()
```

R version 4.2.1 (2022-06-23)

Platform: x86_64-pc-linux-gnu (64-bit)

Running under: Ubuntu 20.04.6 LTS

Matrix products: default

BLAS: /mnt/service/software/packages/r/R-4.2.1/lib/R/lib/libRblas.so

LAPACK: /mnt/service/software/packages/r/R-4.2.1/lib/R/lib/libRlapack.so

locale:

[1] LC_CTYPE=en_GB.UTF-8	LC_NUMERIC=C
[3] LC_TIME=en_GB.UTF-8	LC_COLLATE=en_GB.UTF-8
[5] LC_MONETARY=en_GB.UTF-8	LC_MESSAGES=en_GB.UTF-8
[7] LC_PAPER=en_GB.UTF-8	LC_NAME=C
[9] LC_ADDRESS=C	LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_GB.UTF-8	LC_IDENTIFICATION=C

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] dplyr_1.0.10 tximport_1.26.1 XML_3.99-0.11 png_0.1-8

loaded via a namespace (and not attached):

[1] pillar_1.8.1	compiler_4.2.1	tools_4.2.1	digest_0.6.30
[5] bit_4.0.4	jsonlite_1.8.3	evaluate_0.17	lifecycle_1.0.3
[9] tibble_3.1.8	pkgconfig_2.0.3	rlang_1.0.6	DBI_1.1.3
[13] cli_3.4.1	rstudioapi_0.14	parallel_4.2.1	yaml_2.3.6
[17] xfun_0.39	fastmap_1.1.0	generics_0.1.3	crayon_1.5.2
[21] vctrs_0.5.0	hms_1.1.2	bit64_4.0.5	tidyselect_1.2.0
[25] glue_1.6.2	R6_2.5.1	fansi_1.0.3	vroom_1.6.0
[29] readr_2.1.3	tzdb_0.3.0	magrittr_2.0.3	htmltools_0.5.3

[33] ellipsis_0.3.2 assertthat_0.2.1 utf8_1.2.2