

# Expressed RNAs from RNAseq (TPM calculation)

Melina Klostermann

22 September, 2023

## Contents

1	Libraries and settings	1
2	What was done?	1
3	Files	1
4	Calculate tpm	2
5	Session Info	3

## 1 Libraries and settings

```
# -----  
# libraries  
# -----  
library(tidyverse)  
library(GenomicFeatures)  
library(GenomicRanges)  
  
here <- here::here()  
# -----  
# settings  
# -----  
out <- paste0(here, "/Supporting_scripts/TPMs-RNAseq/")
```

## 2 What was done?

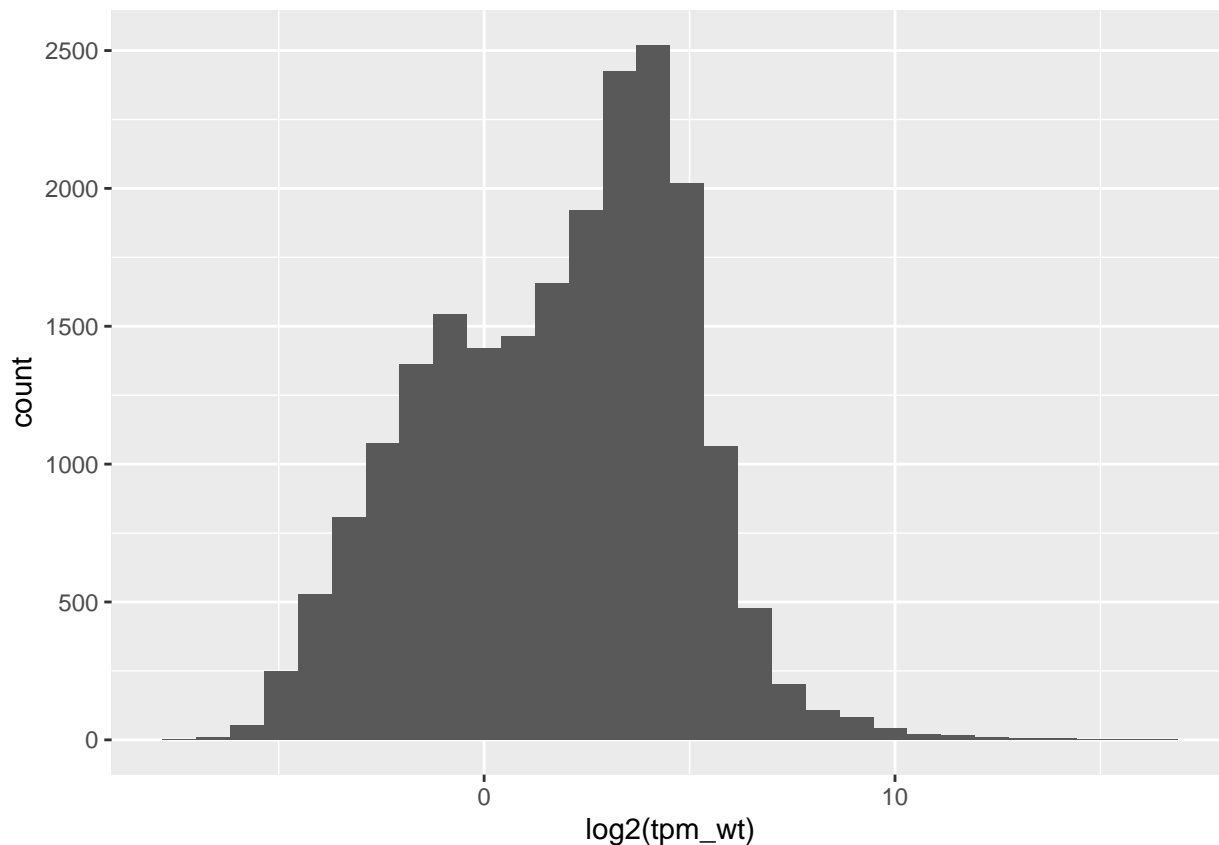
- Calculate TPMs
- Use TPM filter to get a list of expressed RNAs

## 3 Files

```
RNAseq_read_counts <- read_csv(paste0(here, "/Figure3/01_Ribosome_profiling_pipeline/RNAadjusted_countma  
  
#annotation  
anno <- readRDS(paste0(here, "/Supporting_scripts/annotation_preprocessing/annotation.rds"))
```

## 4 Calculate tpm

```
#####  
# get gene transcript length  
#####  
anno <- anno[anno$type == "exon"]  
anno <- split(anno, anno$geneID)  
anno <- reduce(anno)  
  
gene_length <- lapply(anno, function(x) sum(width(x))) %>% unlist()  
  
# get count matrix  
RNAseq_read_counts$...1 <- sub("\\\\.\\.*", "", RNAseq_read_counts$...1)  
RNAseq_read_counts <- as.data.frame(RNAseq_read_counts)  
rownames(RNAseq_read_counts) <- RNAseq_read_counts$...1  
RNAseq_read_counts$...1 <- NULL  
RNAseq_read_counts <- as.matrix(RNAseq_read_counts)  
  
# order gene length  
gene_length <- gene_length[rownames(RNAseq_read_counts)]  
  
# calculate tpm  
x <- RNAseq_read_counts / gene_length  
tpm <- t( t(x) * 1e6 / colSums(x, na.rm = T) )  
  
# tpm of all wt  
tpm <- as.data.frame(tpm) %>%  
  rownames_to_column(var = "gene_id") %>%  
  rowwise() %>%  
  mutate(tpm_wt = median(c(WT_1411, WT_1601, WT_1710), na.rm = T))  
  
# plot tpm of wt  
ggplot(tpm, aes(x = log2(tpm_wt)))+  
  geom_histogram()
```



```
# expressed genes
expressed_genes <- tpm %>% subset(tpm_wt >= 1) %>% pull(gene_id)

saveRDS(expressed_genes, paste0(out, "expressed_genes.rds"))
```

## 5 Session Info

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## 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] stats4      stats      graphics  grDevices  utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] GenomicFeatures_1.50.4 AnnotationDbi_1.60.2  Biobase_2.58.0
```

```

## [4] GenomicRanges_1.50.2   GenomeInfoDb_1.34.9   IRanges_2.32.0
## [7] S4Vectors_0.36.2       BiocGenerics_0.44.0   lubridate_1.9.2
## [10] forcats_1.0.0          stringr_1.5.0         dplyr_1.1.2
## [13] purrr_1.0.1            readr_2.1.4           tidyr_1.3.0
## [16] tibble_3.2.1           ggplot2_3.4.2         tidyverse_2.0.0
## [19] knitr_1.43
##
## loaded via a namespace (and not attached):
## [1] bitops_1.0-7            matrixStats_1.0.0
## [3] bit64_4.0.5            filelock_1.0.2
## [5] progress_1.2.2         httr_1.4.6
## [7] rprojroot_2.0.3        tools_4.2.2
## [9] utf8_1.2.3             R6_2.5.1
## [11] DBI_1.1.3              colorspace_2.1-0
## [13] withr_2.5.0            tidyselect_1.2.0
## [15] prettyunits_1.1.1      bit_4.0.5
## [17] curl_5.0.1             compiler_4.2.2
## [19] cli_3.6.1              xml2_1.3.5
## [21] DelayedArray_0.24.0    labeling_0.4.2
## [23] rtracklayer_1.58.0     scales_1.2.1
## [25] rappdirs_0.3.3         digest_0.6.33
## [27] Rsamtools_2.14.0       rmarkdown_2.23
## [29] XVector_0.38.0         pkgconfig_2.0.3
## [31] htmltools_0.5.5        MatrixGenerics_1.10.0
## [33] highr_0.10             dbplyr_2.3.3
## [35] fastmap_1.1.1          rlang_1.1.1
## [37] rstudioapi_0.15.0      RSQLite_2.3.1
## [39] farver_2.1.1           BiocIO_1.8.0
## [41] generics_0.1.3         BiocParallel_1.32.6
## [43] vroom_1.6.3            RCurl_1.98-1.12
## [45] magrittr_2.0.3         GenomeInfoDbData_1.2.9
## [47] Matrix_1.5-4.1         Rcpp_1.0.11
## [49] munsell_0.5.0          fansi_1.0.4
## [51] lifecycle_1.0.3        stringi_1.7.12
## [53] yaml_2.3.7             SummarizedExperiment_1.28.0
## [55] zlibbioc_1.44.0        BiocFileCache_2.6.1
## [57] grid_4.2.2            blob_1.2.4
## [59] parallel_4.2.2         crayon_1.5.2
## [61] lattice_0.21-8         Biostrings_2.66.0
## [63] hms_1.1.3             KEGGREST_1.38.0
## [65] pillar_1.9.0           rjson_0.2.21
## [67] codetools_0.2-19       biomaRt_2.54.1
## [69] XML_3.99-0.14          glue_1.6.2
## [71] evaluate_0.21          png_0.1-8
## [73] vctrs_0.6.3           tzdb_0.4.0
## [75] gtable_0.3.3          cachem_1.0.8
## [77] xfun_0.39             restfulr_0.0.15
## [79] GenomicAlignments_1.34.1 memoise_2.0.1
## [81] timechange_0.2.0       here_1.0.1

```