Expressed RNAs from RNAseq (TPM calculation)

Melina Klostermann

04 July, 2023

Contents

1 Libraries and settings

```
# ----------------------------------------
# libraries
# ----------------------------------------
library(tidyverse)
library(GenomicFeatures)
library(GenomicRanges)
# ----------------------------------------
# settings
# ----------------------------------------
out <- "/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/Figure5/TPMs-RN
```

2 What was done?

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

3 Files

```
RNAseq_read_counts <- read_csv("/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR

#annotation
anno <- readRDS("/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/Metho
```

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()
```

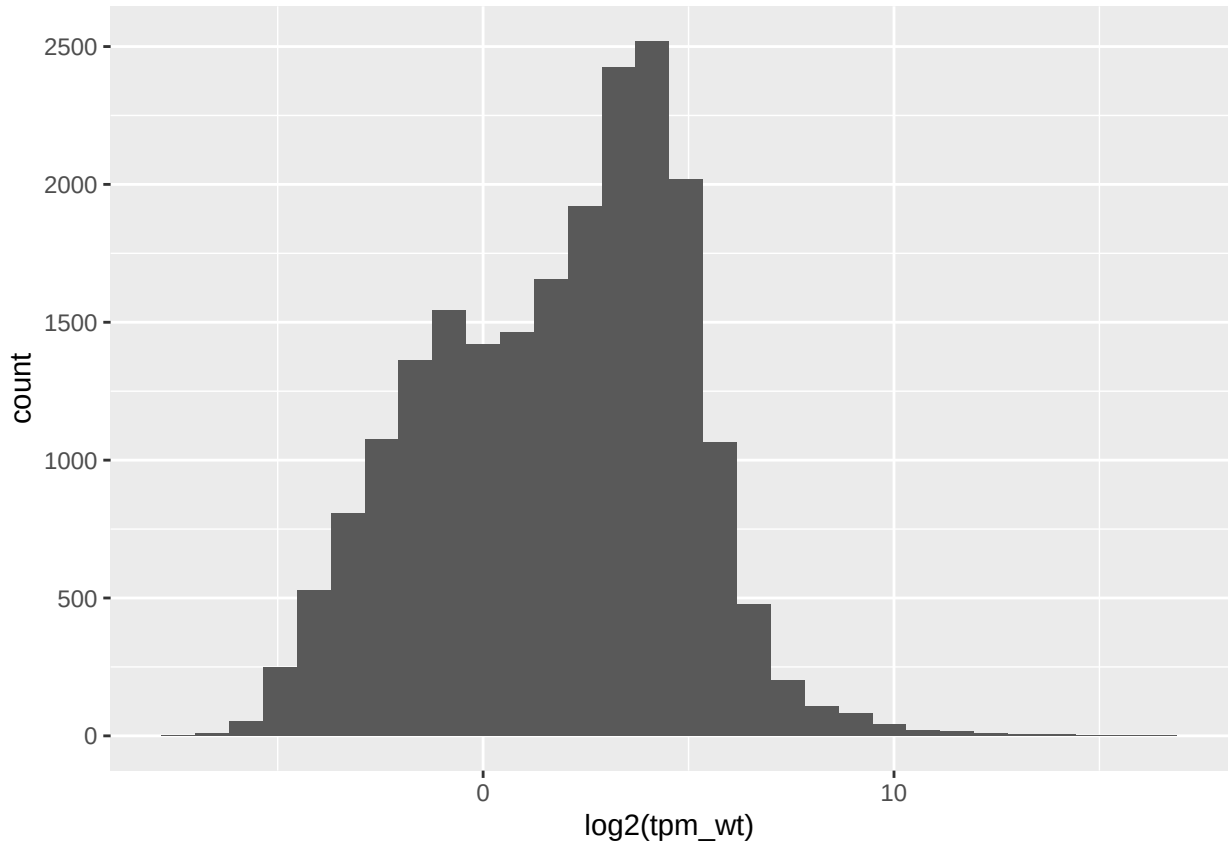
```
# get count matrix
RNAseq_read_counts$...1 <- sub("\\..*", "", RNAseq_read_counts$...1)
RNAseq_read_counts <- as.data.frame(RNAseq_read_counts)
rownames(RNAseq_read_counts) <- RNAseq_read_counts$...1
RNAseq_read_counts$...1 <- NULL
RNAseq_read_counts <- as.matrix(RNAseq_read_counts)

# order gene length
gene_length <- gene_length[rownames(RNAseq_read_counts)]

# calucalte 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"))
```