tpm_pca

2022-12-1

```r
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.3.6      ✔ purrr   0.3.5
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.1      ✔ stringr 1.4.1
## ✔ readr   2.1.3      ✔ forcats 0.5.2
## ── Conflicts ──────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```r
in_f <- ("../output/count_tpm.tsv")
tpm <- read_table(in_f)
```

```
##
## ── Column specification ───────────────────────────────────────────
## cols(
##   target_id = col_character(),
##   wt_1 = col_double(),
##   wt_2 = col_double(),
##   wt_3 = col_double(),
##   mutant_1 = col_double(),
##   mutant_2 = col_double(),
##   mutant_3 = col_double()
## )
```

```r
tpm_t <- data.table::transpose(tpm, make.names = 1)

pca <-prcomp(tpm_t)
summary(pca)
```

```
## Importance of components:
##                            PC1       PC2       PC3       PC4       PC5
## Standard deviation     7675.5957 3875.0885 2.929e+03 2.077e+03 1.738e+03
## Proportion of Variance    0.6558    0.1671 9.547e-02 4.802e-02 3.362e-02
## Cumulative Proportion     0.6558    0.8229 9.184e-01 9.664e-01 1.000e+00
##                            PC6
## Standard deviation     4.736e-11
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```r
plotdf <- tibble(pc1 = pca$x[,1],
                 pc2 = pca$x[,2],
                 condition = c(1,1,1,2,2,2))

ggplot(plotdf,aes(x=pc1,y=pc2,color=condition))+
  geom_point() +
  theme(legend.position = "none") +
  geom_text(aes(x = pc1 + 200, label=colnames(tpm)[2:7]),hjust =0)+
  xlim(-7000, 14000)
```