

Correlation Analysis between Gene Expression and SED (Question 1)

Shiyi Wang

12/10/2020

Subject: Transcriptome data of 41 breast tumors with available three-dimensional tumor diameters.

Questions: What are the gens significantly correlated with SED of breast cancers?

Object: To reveal the genes significantly correlated with SED

Filter criteria:

1. p-Value < 0.01
2. means > 10
3. CV > 1

Note: This is the reproduction of the results in the paper.

```
library(tidyverse)

## -- Attaching packages -----
## v ggplot2 3.3.2      v purrr  0.3.3
## v tibble  3.0.3      v dplyr  0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(readr)

# import data
data <- read_csv("data/SED_vs_Transcriptome-Profile/Correlation_and_CV.csv") %>%
  print()

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   correlation = col_double(),
##   pvalue = col_double(),
```

```
## Means = col_double(),
## SD = col_double(),
## CV = col_double()
## )

## # A tibble: 30,195 x 6
##   X1          correlation pvalue Means    SD    CV
##   <chr>          <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 AMMECR1L    -0.00906 0.956   925. 176. 0.190
## 2 EPC1         0.175   0.286   913. 176. 0.193
## 3 REV1         0.233   0.153 1047. 205. 0.196
## 4 PRDM2       -0.152   0.356 1470. 289. 0.197
## 5 PKNX1       -0.142   0.388   436.  86.9 0.199
## 6 NRF1         0.198   0.228   426.  87.4 0.205
## 7 YTHDC1       0.152   0.355 2819. 579. 0.205
## 8 WDR33       -0.186   0.256 1505. 309. 0.206
## 9 ILKAP        0.380   0.0171  486. 101. 0.208
## 10 ELF2        0.0858  0.603 1352. 285. 0.211
## # ... with 30,185 more rows
```

```
# filter the data using the filter criteria:
```

```
data_filtered <- data %>%
  filter(pvalue < 0.01) %>%
  filter(Means > 10) %>%
  filter(CV > 1) %>%
  print()
```

```
## # A tibble: 54 x 6
##   X1          correlation pvalue Means    SD    CV
##   <chr>          <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 ADAMTS12    -0.536 0.000436 278.  278.  1.00
## 2 POU2F3     -0.533 0.000475 133.  159.  1.19
## 3 FAIM2      -0.512 0.000871  91.1 150.  1.65
## 4 RFFL       -0.507 0.000986 107.  107.  1.00
## 5 C9orf172   -0.498 0.00124  33.7  36.5  1.08
## 6 MMP13      -0.476 0.00218 1364. 2245.  1.65
## 7 PDE10A     -0.474 0.00228  274.  379.  1.38
## 8 LOC101930481 0.471 0.00248  11.4  19.0  1.67
## 9 GRIN1      -0.463 0.00303  93.5 149.  1.59
## 10 PPP1R1A   -0.461 0.00312  145.  237.  1.64
## # ... with 44 more rows
```

```
# How many genes significantly correlated with SED?
```

```
# (i.e., How many genes pass the filter?)
```

```
dim(data_filtered)
```

```
## [1] 54  6
```

```
# sort the data ascendingly by the "correlation" value
```

```
data_sorted <- data_filtered %>%
  arrange(correlation) %>%
  print()
```

```
## # A tibble: 54 x 6
##   X1          correlation pvalue Means    SD    CV
##   <chr>          <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 ADAMTS12    -0.536 0.000436 278.  278.  1.00
```

```
## 2 POU2F3      -0.533 0.000475 133. 159. 1.19
## 3 FAIM2       -0.512 0.000871 91.1 150. 1.65
## 4 RFFL        -0.507 0.000986 107. 107. 1.00
## 5 C9orf172    -0.498 0.00124 33.7 36.5 1.08
## 6 MMP13       -0.476 0.00218 1364. 2245. 1.65
## 7 PDE10A      -0.474 0.00228 274. 379. 1.38
## 8 GRIN1       -0.463 0.00303 93.5 149. 1.59
## 9 PPP1R1A     -0.461 0.00312 145. 237. 1.64
## 10 WNT7B      -0.460 0.00323 249. 280. 1.13
## # ... with 44 more rows
```

```
write.csv(data_sorted, 'figure/Correlation_Analysis_results.csv')
```

```
# How many genes show significant negative correlation with SED?
# (i.e., How many genes have a correlation value < 0?)
data_sorted %>%
  filter(correlation < 0) %>%
  nrow() %>%
  print()
```

```
## [1] 39
```

```
# How many genes show significant positive correlation with SED?
# (i.e., How many genes have a correlation value > 0?)
data_sorted %>%
  filter(correlation > 0) %>%
  nrow() %>%
  print()
```

```
## [1] 15
```