

# Quiz 3

Monikrishna Roy

2021-10-17

## Question 1

Load the file GSE9782.RData into your RStudio environment. Calculate log base 2 of the expression. Compute the difference in mean base 2 expression between the two groups (named logFC). Note that you have to use the apply(). Do not use for loop.

```
filePath = "Project-1/"
fileName = "GSE9782.RData"
load(paste0(filePath, fileName))

treated <- which(group == "treated")
control <- which(group != "treated")

# log base 2 of the expression
log2Data <- log(x = dataGSE9782, base = 2)

# a function to calculate the difference in mean
cal_mean_diff <- function(x, treated, control) {
  mean(x[treated]) - mean(x[control])
}

# used apply to call the function
logFC <- apply(log2Data, 1, cal_mean_diff, treated, control)
```

## Question 2

Perform t-test using the t.test() function to compare the control and treated groups. Output the p-values and the t-scores of the test (named PValue and TScore, respectively). Note that you have to use apply().

```
# Function to calculate p-value
cal_p_value <- function(x, treated, control) {
  t.test(x[treated], x[control])$p.value
}

# Function to calculate t-score
cal_t_score <- function(x, treated, control) {
  t.test(x[treated], x[control])$statistic
}

PValue <- apply(log2Data, MARGIN = 1, FUN = cal_p_value, treated, control)
TScore <- apply(log2Data, MARGIN = 1, FUN = cal_t_score, treated, control)
```

## Question 3

Output a data frame that contains the following columns: 1. gene ids (row names of expression data matrix), 2. t-test's p-values, 3. t-scores, and 4. logFC.

```
# rownames used as gene ids
geneIds <- rownames(log2Data)

df <- data.frame(
  row.names = NULL,
  "GeneID" = geneIds,
  "PValue" = PValue,
  "TScore" = TScore,
  "LogFC" = logFC
)
```

#### Question 4

Plot the volcano plot of, in which the x-axis is logFC and y-axis is  $-\log_{10}$  of p-values. Use the “col” parameter to color the gene with absolute logFC > 1 and p-value < 0.05 in red and the remaining ones in black in the volcano plot.

```
plot(
  x = df$LogFC,
  y = -log10(df$PValue),
  xlab = 'logFC',
  ylab = '-log10(p-value)',
  main = "Volcano plot",
  col = ifelse(abs(df$LogFC) > 1 & df$PValue < 0.05, 'red', 'black'),
  xlim = c(-2, 2)
)
abline(h = -log10(0.05), col = "red")
abline(v = -1, col = "blue")
abline(v = 1, col = "blue ")
```

**Volcano plot**

