

QBS 103 Final Project V1

Elizabeth Chin

2023-07-23

Gene: ABCA1

This gene encodes a protein located in the cell membrane, and it belongs to a group of proteins called ATP-binding cassette (ABC) transporters. These proteins are responsible for moving various molecules in and out of cells. There are seven different subgroups of ABC genes, and this particular protein belongs to the ABC1 subgroup, which is unique to multicellular organisms. This protein plays a role in removing cholesterol from cells by acting as a pump to push cholesterol out of the cell. When both copies of this gene have mutations, it leads to conditions known as Tangier disease and familial high-density lipoprotein (HDL) deficiency, affecting how cholesterol is processed in the body.

```
#set working directory
setwd("/Users/li_li/desktop/103_Final_Project/Data")

#read in csv files
gene.exp <- read.csv("QBS103_finalProject_geneExpression.csv", row.names = 1)

meta.data <- read.csv("QBS103_finalProject_metadata.csv", row.names = 1)

#subset gene expression ABCA1
gene.ABCA1 <- gene.exp['ABCA1', ]

#transform rows to columns
gene.ABCA1 <- as.data.frame(t(gene.ABCA1))
head(gene.ABCA1)
```

```
##                ABCA1
## COVID_01_39y_male_NonICU 32.30
## COVID_02_63y_male_NonICU 15.84
## COVID_03_33y_male_NonICU 34.38
## COVID_04_49y_male_NonICU 14.24
```

```
## COVID_05_49y_male_NonICU 18.39
## COVID_06_.y_male_NonICU 3.64
```

```
#column bind two data frames
```

```
ABCA1.gene.exp.meta <- cbind(gene.ABCA1,meta.data)
head(ABCA1.gene.exp.meta, 1)
```

```
##                ABCA1 geo_accession                status
## COVID_01_39y_male_NonICU 32.3      GSM4753021 Public on Aug 29 2020
##                X.Sample_submission_date last_update_date type
## COVID_01_39y_male_NonICU          Aug 28 2020          Aug 29 2020  SRA
##                channel_count                source_name_ch1 organism_ch1
## COVID_01_39y_male_NonICU          1 Leukocytes from whole blood Homo sapiens
##                disease_status age    sex icu_status apacheii
## COVID_01_39y_male_NonICU disease state: COVID-19 39 male          no          15
##                charlson_score mechanical_ventilation
## COVID_01_39y_male_NonICU          0                      yes
##                ventilator.free_days
## COVID_01_39y_male_NonICU          0
##                hospital.free_days_post_45_day_followup
## COVID_01_39y_male_NonICU          0
##                ferritin.ng.ml. crp.mg.l. ddimer.mg.l_feu.
## COVID_01_39y_male_NonICU          946          73.1          1.3
##                procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen sofa
## COVID_01_39y_male_NonICU          36          0.9          513      8
```

```
#convert age to an integer
```

```
ABCA1.gene.exp.meta$age <- as.integer(ABCA1.gene.exp.meta$age)
```

```
## Warning: NAs introduced by coercion
```

```
head(ABCA1.gene.exp.meta, 1)
```

```
##                ABCA1 geo_accession                status
## COVID_01_39y_male_NonICU 32.3      GSM4753021 Public on Aug 29 2020
##                X.Sample_submission_date last_update_date type
## COVID_01_39y_male_NonICU          Aug 28 2020          Aug 29 2020  SRA
##                channel_count                source_name_ch1 organism_ch1
## COVID_01_39y_male_NonICU          1 Leukocytes from whole blood Homo sapiens
##                disease_status age    sex icu_status apacheii
## COVID_01_39y_male_NonICU disease state: COVID-19 39 male          no          15
##                charlson_score mechanical_ventilation
## COVID_01_39y_male_NonICU          0                      yes
##                ventilator.free_days
## COVID_01_39y_male_NonICU          0
##                hospital.free_days_post_45_day_followup
## COVID_01_39y_male_NonICU          0
##                ferritin.ng.ml. crp.mg.l. ddimer.mg.l_feu.
## COVID_01_39y_male_NonICU          946          73.1          1.3
##                procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen sofa
## COVID_01_39y_male_NonICU          36          0.9          513      8
```

```

#creating histogram for gene expression

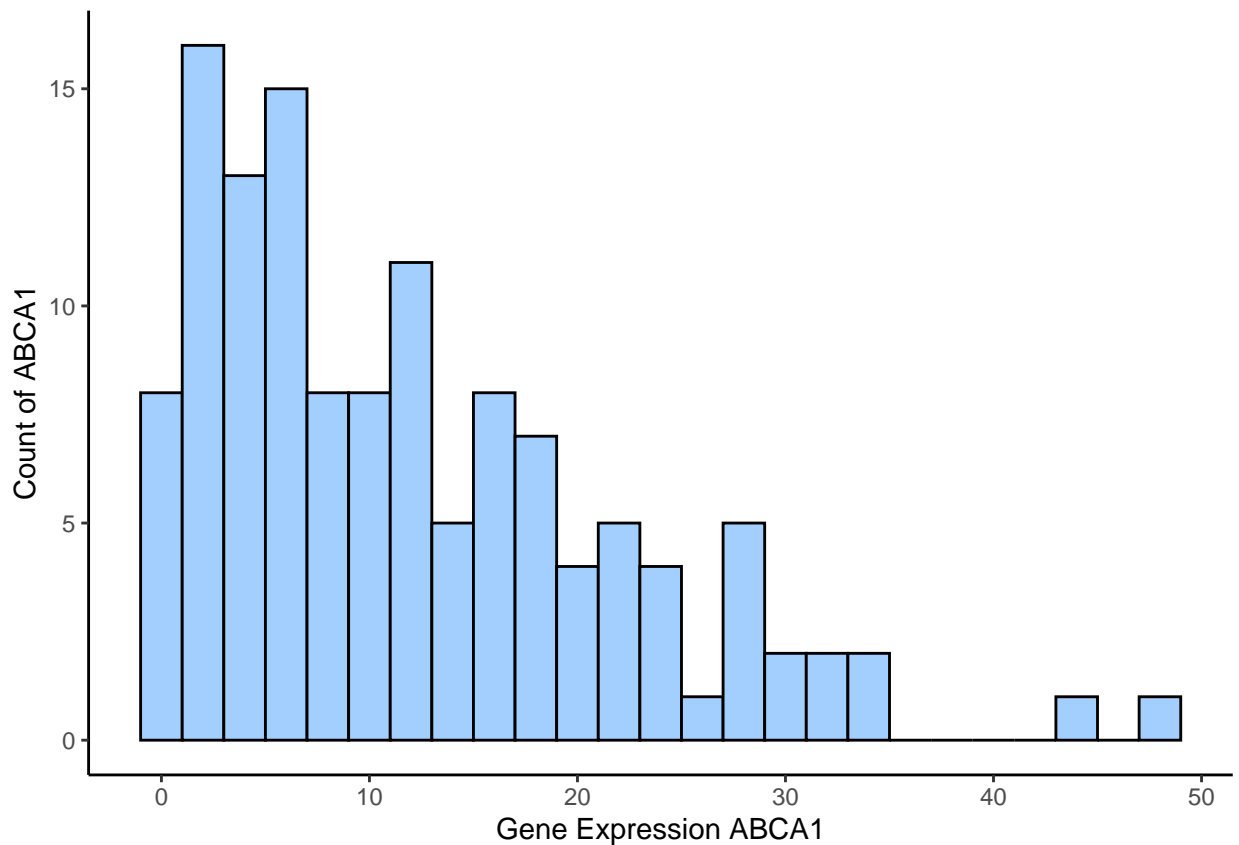
#load the required package
library(ggplot2)

#define color palette
my_palette1 <- c("#A2CFFE") # Choose any colors you like

#create theme
newBlankTheme <- theme(# Remove all the extra borders and grid lines
  panel.border = element_blank(), panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  # Define my axis
  axis.line = element_line(colour = "black", linewidth = rel(1)),
  # Set plot background
  plot.background = element_rect(fill = "white"),
  panel.background = element_blank(),
  legend.key = element_rect(fill = 'white'),
  # Move legend
  legend.position = 'top')

#create histogram
ggplot(ABCA1.gene.exp.meta, aes(x=ABCA1.gene.exp.meta$ABCA1)) +
  geom_histogram(binwidth = 2, fill = my_palette1, col="black") +
  labs(x = 'Gene Expression ABCA1', y = 'Count of ABCA1') +
  newBlankTheme

```



```

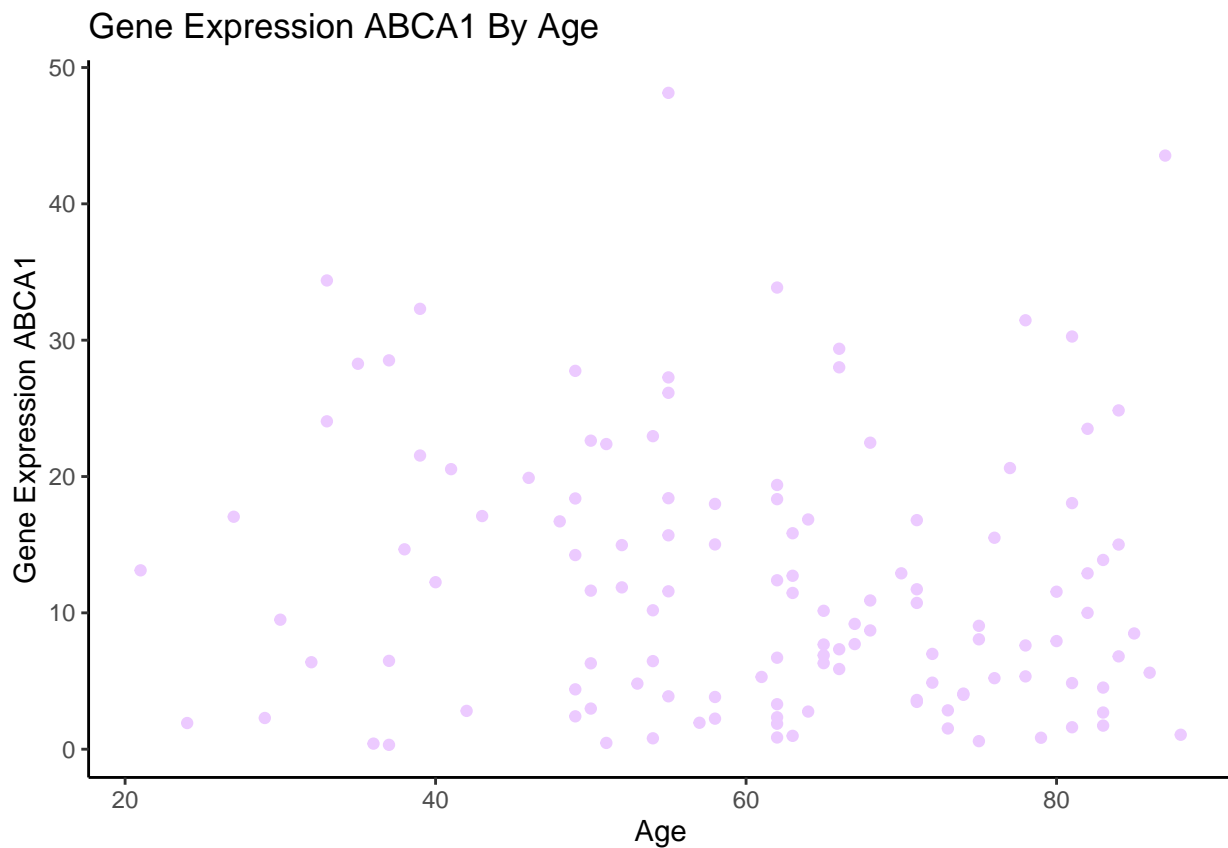
#create scatterplot for gene expression and age

#define color palette
my_palette2 <- c("#ECCAFF") # Choose any colors you like

#create scatterplot
ggplot(ABCA1.gene.exp.meta,aes(x=ABCA1.gene.exp.meta$age,y=ABCA1.gene.exp.meta$ABCA1)) +
  geom_point(color=my_palette2) +
  labs(x = "Age", y = "Gene Expression ABCA1", title = "Gene Expression ABCA1 By Age") +
  newBlankTheme

```

```
## Warning: Removed 3 rows containing missing values ('geom_point()').
```



```

#create boxplot for gene expression separated by sex and icu status

#load the required package
library(dplyr)

```

```

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

#define color palette
my_palette3 <- c("#FFEEC4", "#B1E3C7") # Choose any colors you like

#create a new data frame to filter the unknown in sex
filtered_sex <- ABCA1.gene.exp.meta %>%
  filter(!grepl('unknown', sex))

#plot new data frame to create box plot
ggplot(filtered_sex, aes(x=filtered_sex$sex, y=filtered_sex$ABCA1, fill=filtered_sex$icu_status)) +
  geom_boxplot() +
  scale_fill_manual(values = my_palette3) +
  labs(x = "Sex",
       y = "Gene Expression ABCA1",
       fill = "ICU Status",
       title = "Gene Expression ABCA1 By Sex & ICU Status") +
  newBlankTheme
```

