Analysis of ADP-Induced Platelet Aggregation and SNPs

Niramai Arkhom

2024-11-14

## Introduction

This file contains the analysis of ADP-induced platelet aggregation levels in relation to genetic variants (SNPs). The data has been processed to investigate the relationship between SNP genotypes and ADP-induced platelet aggregation.

### Variables in the Data:

* **ADP**: ADP-induced platelet aggregation level.
* **Resistant**: Clopidogrel resistance (coded as 1 for resistance, 0 for non-resistance).
* **Genotypes**: Coded according to an additive model.
  + **rs4244285** (CYP2C19\*2)
  + **rs4986893** (CYP2C19\*3)
  + **rs662** (PON1. 192Q>R)
* **Age**: Age of the individual (in years).
* **Sex**: Gender (0 = male, 1 = female).

## Project Structure

The R project should be organized with the following folder structure:

* **raw data**: Folder containing the original PlateletHW.tsv file.
* **code**: Folder containing R scripts for analysis (including this R Markdown file).
* **clean data**: Folder to save the cleaned data after processing.

## Step 1: Load Data

We begin by loading the PlateletHW.tsv file into R. This dataset contains the raw data for the analysis.

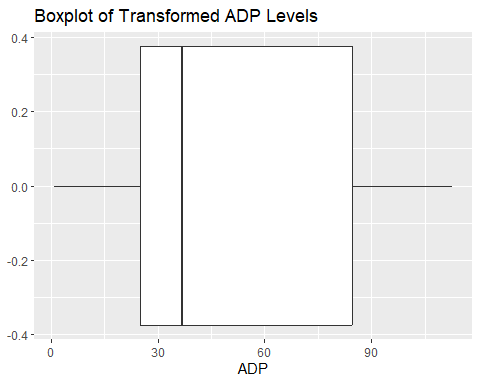
library(dplyr)  
library(ggplot2)  
library(readr)  
  
raw\_data <- read\_tsv("../raw\_data/PlateletHW.tsv")  
print(colnames(raw\_data))

## [1] "IID" "ADP" "Resistance" "rs4244285" "rs4986893"   
## [6] "rs662" "AGE" "SEX" "PON1.192Q>R" "CYP2C19\*2"   
## [11] "CYP2C19\*3"

## Step 2: Check for Outliers in the Data

Next, we check for outliers in the ADP variable. We calculate the minimum ADP value and shift all values if the minimum is less than or equal to zero.

min\_adp <- min(raw\_data$ADP, na.rm = TRUE)  
  
if (min\_adp <= 0) {  
 raw\_data <- raw\_data %>%  
 mutate(ADP = ADP + abs(min\_adp) + 1)  
}  
  
ggplot(raw\_data, aes(x = ADP)) +  
 geom\_boxplot() +  
 ggtitle("Boxplot of Transformed ADP Levels")



## Step 3: Clean the Data

We remove outliers by applying the interquartile range (IQR) method and save the cleaned data in the clean data folder.

Q1 <- quantile(raw\_data$ADP, 0.25)  
Q3 <- quantile(raw\_data$ADP, 0.75)  
IQR <- Q3 - Q1  
outlier\_thresholds <- c(Q1 - 1.5 \* IQR, Q3 + 1.5 \* IQR)  
  
clean\_data <- raw\_data %>%  
 filter(ADP >= outlier\_thresholds[1] & ADP <= outlier\_thresholds[2])  
  
write\_csv(clean\_data, "../clean\_data/PlateletHW\_cleaned.csv")

## Step 4: Test for Association

We now test for an association between the three SNPs (rs4244285, rs4986893, rs662) and ADP-induced platelet aggregation levels using linear regression models. We also account for age and sex as covariates.

results <- list()  
snp\_vars <- c("rs4244285", "rs4986893", "rs662")  
  
for (snp in snp\_vars) {  
 model <- lm(ADP ~ get(snp) + AGE + SEX, data = clean\_data)  
 results[[snp]] <- summary(model)  
 print(paste("Results for SNP:", snp))  
 print(summary(model))  
}

## [1] "Results for SNP: rs4244285"  
##   
## Call:  
## lm(formula = ADP ~ get(snp) + AGE + SEX, data = clean\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -56.093 -23.649 -9.168 30.273 61.761   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 61.1741 12.8821 4.749 3.82e-06 \*\*\*  
## get(snp) 12.9533 3.3261 3.894 0.000133 \*\*\*  
## AGE -0.2565 0.1963 -1.307 0.192731   
## SEX -1.8824 4.6734 -0.403 0.687526   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 30.17 on 207 degrees of freedom  
## Multiple R-squared: 0.07678, Adjusted R-squared: 0.0634   
## F-statistic: 5.738 on 3 and 207 DF, p-value: 0.0008638  
##   
## [1] "Results for SNP: rs4986893"  
##   
## Call:  
## lm(formula = ADP ~ get(snp) + AGE + SEX, data = clean\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -56.63 -22.36 -11.56 28.13 63.23   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 66.2939 12.9576 5.116 7.09e-07 \*\*\*  
## get(snp) 26.1477 8.4722 3.086 0.0023 \*\*   
## AGE -0.2618 0.1989 -1.316 0.1895   
## SEX -0.3504 4.7425 -0.074 0.9412   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 30.56 on 207 degrees of freedom  
## Multiple R-squared: 0.05272, Adjusted R-squared: 0.03899   
## F-statistic: 3.84 on 3 and 207 DF, p-value: 0.0105  
##   
## [1] "Results for SNP: rs662"  
##   
## Call:  
## lm(formula = ADP ~ get(snp) + AGE + SEX, data = clean\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -53.59 -23.80 -12.99 33.41 60.64   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 66.0935 13.7039 4.823 2.74e-06 \*\*\*  
## get(snp) 1.0252 3.2415 0.316 0.752   
## AGE -0.2501 0.2040 -1.226 0.222   
## SEX -1.1655 4.8677 -0.239 0.811   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31.25 on 207 degrees of freedom  
## Multiple R-squared: 0.009613, Adjusted R-squared: -0.004741   
## F-statistic: 0.6697 on 3 and 207 DF, p-value: 0.5715

## Step 5: Write Summary of Results

After performing the association tests, we compile the results for each SNP and write a summary to a text file.

summary\_text <- "Association analysis between SNPs and ADP-induced platelet aggregation levels:\n"  
for (snp in snp\_vars) {  
 snp\_results <- results[[snp]]  
 summary\_text <- paste0(summary\_text,   
 "SNP: ", snp, "\n",  
 " Estimate: ", snp\_results$coefficients[2, 1], "\n",  
 " Std. Error: ", snp\_results$coefficients[2, 2], "\n",  
 " t-value: ", snp\_results$coefficients[2, 3], "\n",  
 " p-value: ", snp\_results$coefficients[2, 4], "\n\n")  
}  
  
write(summary\_text, file = "../clean\_data/association\_summary.txt")

## Summary

Association analysis between SNPs and ADP-induced platelet aggregation levels:

* **SNP: rs4244285** (CYP2C19\*2)
  + **Estimate**: 12.9533
  + **Standard Error**: 3.3261
  + **t-value**: 3.8945
  + **p-value**: 0.0001
* **SNP: rs4986893** (CYP2C19\*3)
  + **Estimate**: 26.1477
  + **Standard Error**: 8.4722
  + **t-value**: 3.0863
  + **p-value**: 0.0023
* **SNP: rs662** (PON1 192Q>R)
  + **Estimate**: 1.0252
  + **Standard Error**: 3.2415
  + **t-value**: 0.3163
  + **p-value**: 0.7521

## Conclusion

This analysis provides insights into the relationship between specific SNPs and ADP-induced platelet aggregation levels, considering age and sex as covariates. The summary results have been written to a text file for further review.

## Instructions

Create an R project with the following structure: raw data code clean data Load the raw data (PlateletHW.tsv) into R and clean it by removing outliers. Test for associations between the SNPs and ADP-induced platelet aggregation levels. Save the cleaned data and a summary of the association analysis.