Analysis of ADP-Induced Platelet Aggregation and SNPs

Niramai Arkhom

2024-11-15

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"
```

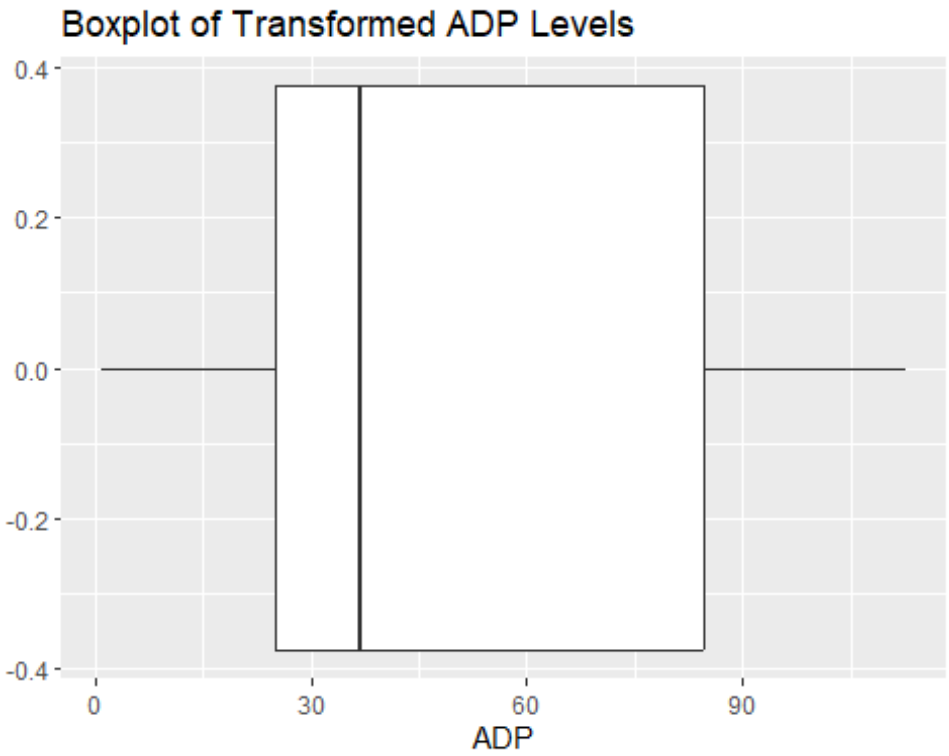
Step 2: Check for Outliers in the Data

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.

```r
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.

```r
Q1 <- quantile(raw_data$ADP, 0.25)
Q3 <- quantile(raw_data$ADP, 0.75)
```

```r
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.

```r
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

The results from the analysis of ADP-induced platelet aggregation levels show distinct associations between specific SNPs and platelet response to ADP, suggesting potential biological implications related to clopidogrel resistance and individual variability in platelet reactivity. SNP rs4244285 (CYP2C192*): Significant association with ADP-induced platelet aggregation (p-value = 0.0001). The CYP2C192 allele is associated with reduced enzyme function, which can lower the conversion of clopidogrel to its active form. Individuals with this SNP variant may experience higher platelet aggregation levels, indicating a potential resistance to clopidogrel treatment. This aligns with known pharmacogenetic impacts of CYP2C192, as reduced enzyme activity affects drug efficacy. SNP rs4986893 (CYP2C193*): Also shows a significant association with ADP levels (p-value = 0.0023). Similar to CYP2C192, the CYP2C193 allele reduces enzyme function and is often linked to higher platelet reactivity. This SNP suggests that carriers may also have a diminished response to clopidogrel due to limited conversion to its active form,

impacting therapeutic outcomes for patients requiring antiplatelet therapy. SNP rs662 (PON1 192Q>R): No significant association with ADP levels (p-value = 0.7521). Unlike CYP2C19 variants, the PON1 192Q>R SNP does not appear to impact ADP-induced platelet aggregation significantly. This suggests that while PON1 plays a role in clopidogrel metabolism, its genetic variation may not directly influence platelet response to ADP in this population. The data underscore the importance of genetic testing for SNPs in the CYP2C19 gene, as certain variants (like CYP2C192 and CYP2C193) are linked to higher ADP-induced platelet aggregation and possible clopidogrel resistance. This can guide clinicians in personalized medicine approaches, identifying individuals who may benefit from alternative therapies or adjusted clopidogrel dosages to optimize antiplatelet efficacy and reduce the risk of adverse cardiovascular events. In contrast, the lack of a significant association for PON1 suggests that not all enzymes involved in clopidogrel metabolism have equal impact on platelet aggregation levels, highlighting the specificity of genetic markers in guiding treatment decisions.

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.