

Class 11 - Population Analysis

Hannah Kim

5/10/23

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
# Load the required libraries
library(ggplot2)

# Read the data from the file
data <- read.table("rs8067378_ENSG00000172057.6.txt", header = TRUE)

# Determine sample size for each genotype
sample_sizes <- table(data$Genotype)

# Calculate median expression levels for each genotype
median_expression <- tapply(data$exp, data$geno, median)

# Print sample sizes
cat("Sample Sizes:\n")
```

Sample Sizes:

```
print(sample_sizes)
```

< table of extent 0 >

```
# Print median expression levels
cat("\nMedian Expression Levels:\n")
```

Median Expression Levels:

```
print(median_expression)
```

A/A	A/G	G/G
31.24847	25.06486	20.07363

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

There is some overlap between the expression values of the A/A and G/G genotypes, as seen by the whiskers of the boxplot. This suggests that there are individuals with the A/A genotype who exhibit expression levels similar to or even higher than some individuals with the G/G genotype. It can be inferred that the SNP does affect the expression of ORMDL3. The G/G genotype is associated with higher median expression levels and greater variability in expression compared to the A/A genotype.

```
# Generate a boxplot
# Generate a boxplot
boxplot(exp ~ geno, data = data,
        xlab = "Genotype",
        ylab = "Expression",
        main = "Expression of ORMDL3 by Genotype")
```

Expression of ORMDL3 by Genotype

