Section 4: Population Scale Analysis

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

```
expr <- read.table("class12_file.txt")
  head(expr)
```

```
##    sample geno      exp
## 1 HG00367  A/G 28.96038
## 2 NA20768  A/G 20.24449
## 3 HG00361  A/A 31.32628
## 4 HG00135  A/A 34.11169
## 5 NA18870  G/G 18.25141
## 6 NA11993  A/A 32.89721
```

```
nrow(expr)
```

```
## [1] 462
```

```
# Sample size for each genotype
table(expr$geno)
```

```
##
## A/A A/G G/G
## 108 233 121
```

```
# Median expression levels for A/A
  median(expr$exp[expr$geno == "A/A"])
```

```
## [1] 31.24847
```

```
# Median expression levels for A/G
  median(expr$exp[expr$geno == "A/G"])
```

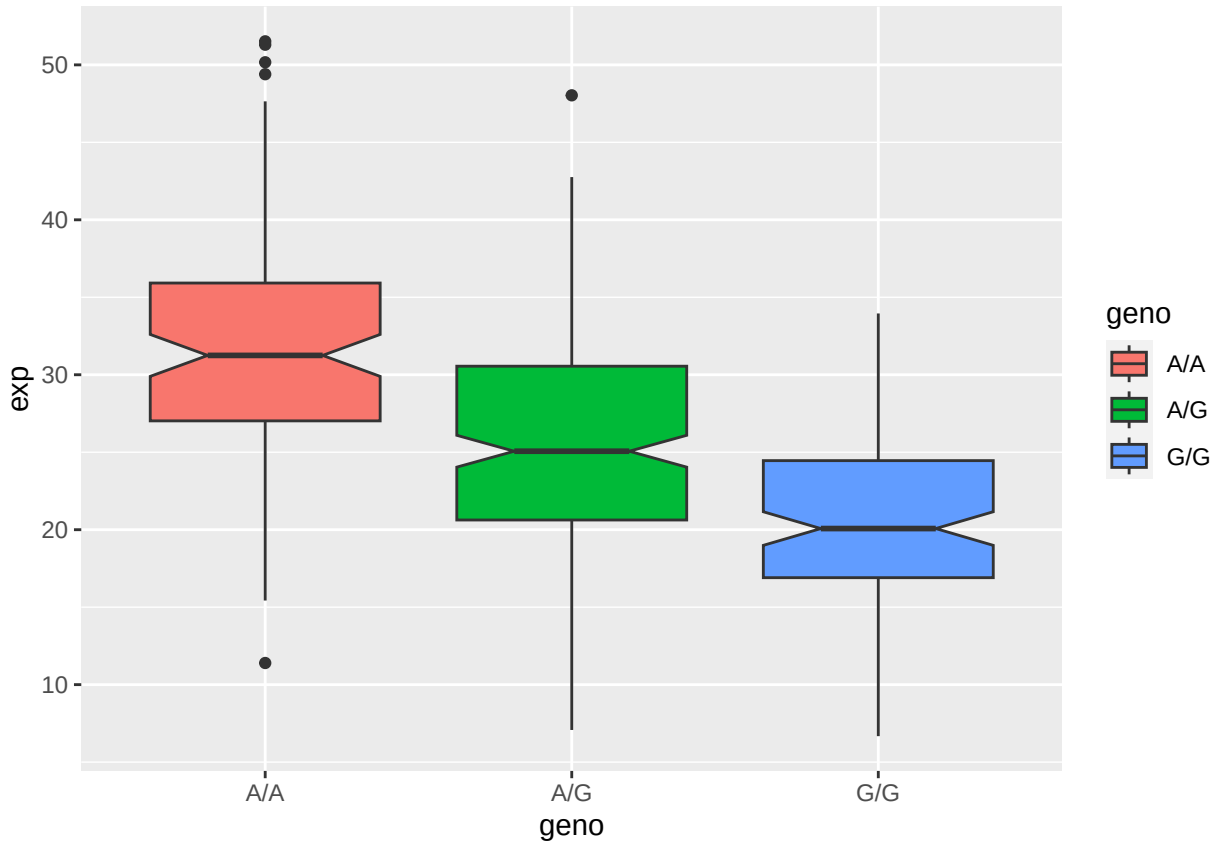
```
## [1] 25.06486
```

```
# Median expression levels for G/G
  median(expr$exp[expr$geno == "G/G"])
```

```
## [1] 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?

```
library(ggplot2)
ggplot(expr) + aes(geno, exp, fill = geno) + geom_boxplot(notch = TRUE)
```



Expression increases with the A allele or decreases with the G allele but the overall effect is not significatn because the expression levels are all overlapping so there is no significant difference