

Class 19: Boxplot

Tom Quach (A15549142)

12/5/2021

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.

Reading our file

```
expr <- read.table("Expression genotype results.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
```

Determining the sample size for each genotype

```
table(expr$geno)
```

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

Finding the median expression levels for each of these genotypes

```
summary(expr)
```

```
##      sample      geno      exp
## Length:462      Length:462      Min.   : 6.675
## Class :character Class :character 1st Qu.:20.004
## Mode  :character Mode  :character Median  :25.116
##                                     Mean    :25.640
##                                     3rd Qu.:30.779
##                                     Max.    :51.518
```

```
median(expr$exp)
```

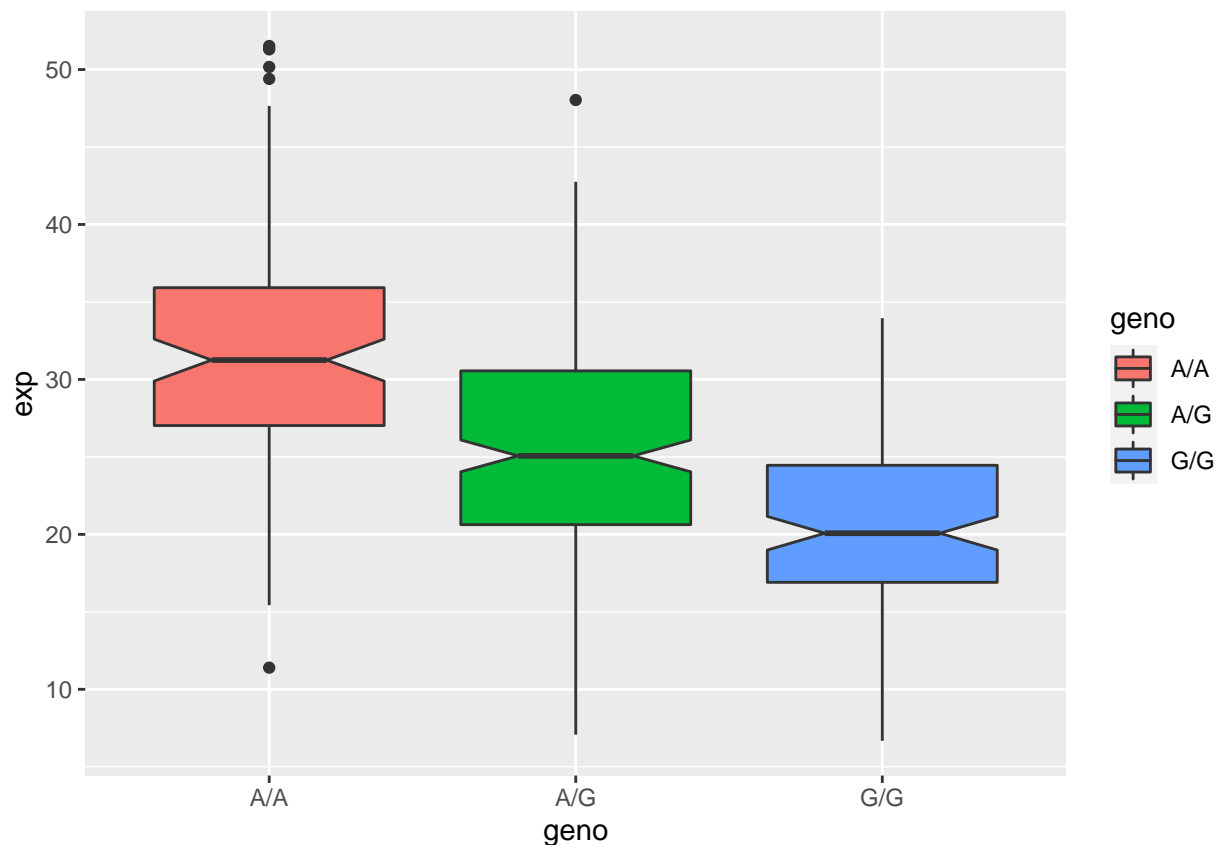
```
## [1] 25.11561
```

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)
```

```
## Lets make the boxplot
```

```
ggplot(expr) + aes(x=geno, y=exp, fill=geno) + geom_boxplot(notch=TRUE)
```



From the boxplot, we can see that the expression value between AA and GG are pretty different, with G/G expression being lower compared to A/A expression. It also looks like having a G/G in this location is definitely associated with the expression of ORMDL3. In sums, the SNP does effect the expression of ORMDL3.