

Class 11 HW: Population Analysis

Jessica Diaz-Vigil

2023-05-21

Section 4: Population 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("rs8067378_ENSG00000172057.6.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
```

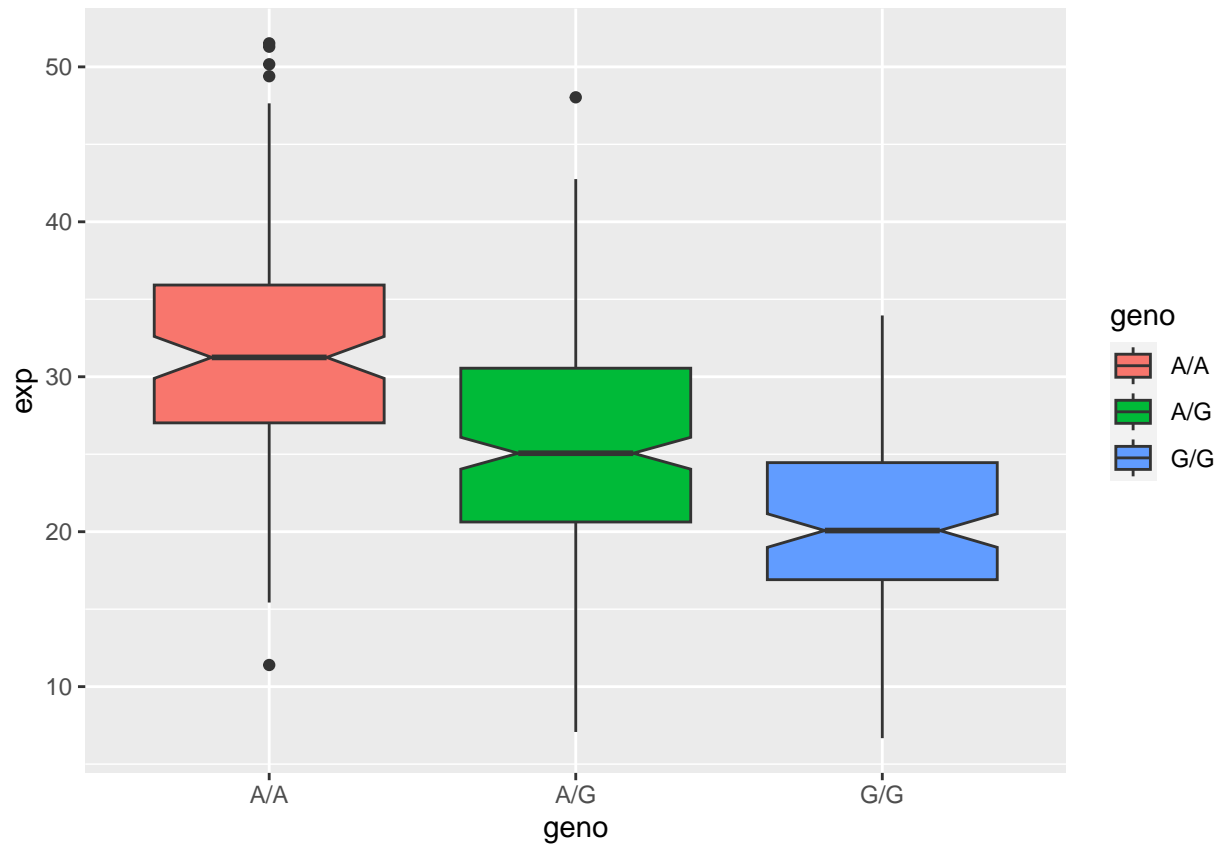
```
table(expr$geno)
```

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

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



The relative difference between A/A and G/G is very large. We can determine this since the notches which represent the median do not overlap. I would think that the SNP does effect the expression of ORMDL3.