

Class 12: Q13 and Q14

Renny Ng (A98061553)

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

```
snp <- read.table("rs8067378_ENSG000000172057.6.txt")  
table(snp$geno)
```

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

There are 108 A/AA, 233 A/G, and 121 G/G genotypes.

```
AA <- snp$geno=="A/A"  
AG <- snp$geno=="A/G"  
GG <- snp$geno=="G/G"
```

```
AAmedian <- median(snp$exp[snp$geno=="A/A"])  
AAmedian
```

```
[1] 31.24847
```

```
AGmedian <- median(snp$exp[snp$geno=="A/G"])  
AGmedian
```

```
[1] 25.06486
```

```
GGmedian <- median(snp$exp[snp$geno=="G/G"])  
GGmedian
```

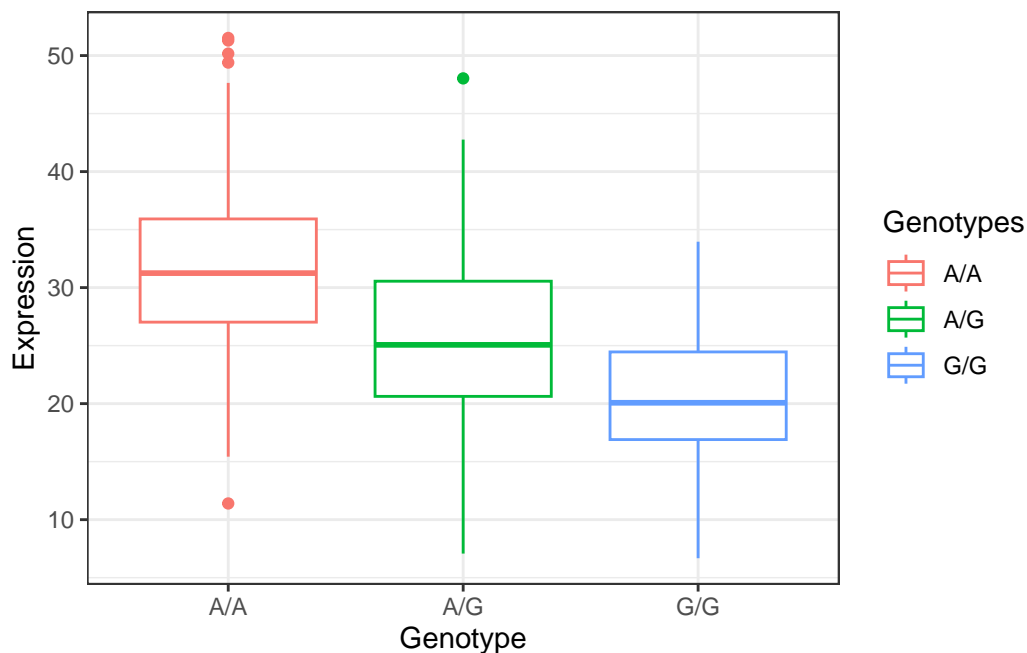
[1] 20.07363

The median expression levels for each genotype are as follows: A/A: 31.24847 A/G: 25.06486 G/G: 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(snp) +
  aes(geno, exp, col=geno) +
  geom_boxplot() +
  xlab("Genotype") +
  ylab("Expression") +
  theme_bw() +
  guides(color=guide_legend("Genotypes"))
```



From this data, I infer that the relative expression of G/G is reduced compared to A/A. It appears that this SNP does indeed affect the expression of ORMDL3.