

Homework 11

Bomin Xie A16144147

Homework 11

You are interested in assessing genetic differences on a population scale. So, you processed about ~230 samples and did the normalization on a genome level. Now, you want to find whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on ORMDL3 expression.

```
expression = read.table("https://marcos-diazg.github.io/BIMM143_SP23/class-material/class1")
summary(expression)
```

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(expression$exp[expression$geno == "A/A"])
```

```
[1] 31.24847
```

```
median(expression$exp[expression$geno == "A/G"])
```

```
[1] 25.06486
```

```
median(expression$exp[expression$geno == "G/G"])
```

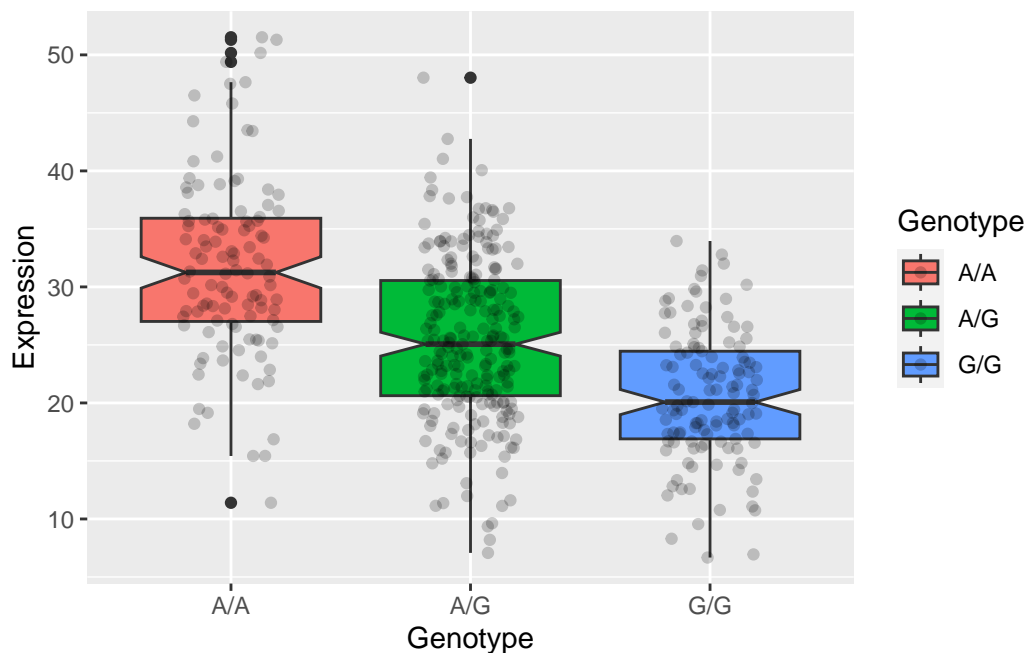
[1] 20.07363

```
table(expression$geno)
```

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

Q13: The sample size for each genotype is: 108 samples for the A/A genotype, 233 for A/G, and 121 for G/G. The median of each genotype is 31.2 (A/A), 25.1 (A/G), and 20.1 (G/G).

```
library(ggplot2)  
ggplot(expression) + aes(geno, exp, fill=geno) +  
  geom_boxplot(notch = TRUE) +  
  geom_jitter(color="black", alpha=0.2, width = 0.2) +  
  xlab("Genotype") + ylab("Expression") + labs(fill = "Genotype")
```



Q14: From the graph, A/A has higher expression on ORMDL3 than G/G genotype. The genotype difference is correlated to the expression differences, which indicates the SNP might affect the ORMDL3 expression.