

Class 19: Genomics

Shitian Li (PID: A13294481)

12/6/2021

```
library(ggplot2)
library(dplyr)

data <- read.table('rs8067378_ENSG00000172057.6.txt')
AA <- filter(data, geno=='A/A')
AG <- filter(data, geno=='A/G')
GG <- filter(data, geno=='G/G')
```

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.

```
nrow(AA)
```

```
## [1] 108
```

```
nrow(AG)
```

```
## [1] 233
```

```
nrow(GG)
```

```
## [1] 121
```

```
median(AA$exp)
```

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

```
median(AG$exp)
```

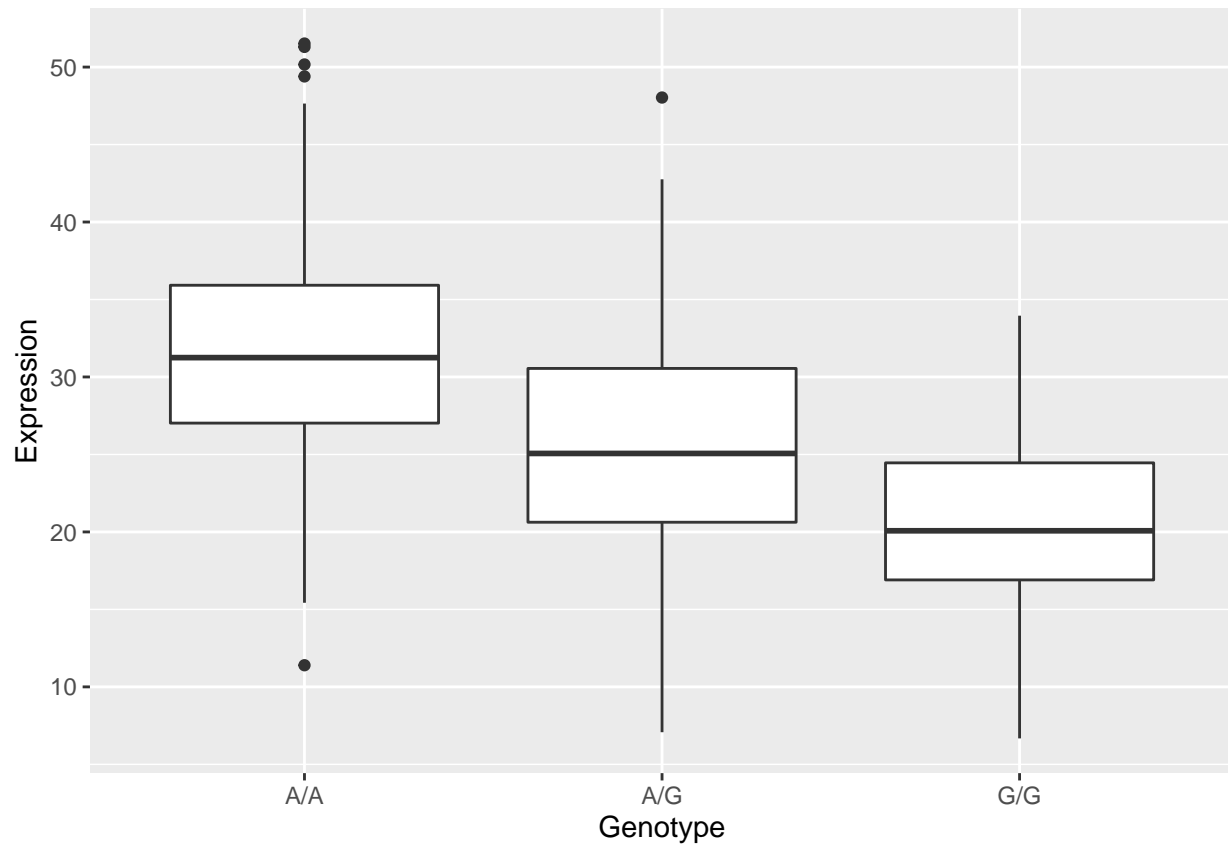
```
## [1] 25.06486
```

```
median(GG$exp)
```

```
## [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?

```
ggplot(data, aes(x=as.factor(geno), y=exp)) +  
  geom_boxplot() +  
  labs(x="Genotype", y="Expression")
```



It seems like expression of G/G is much lower than that of A/A. This SNP, indeed, affects the expression of ORMDL3.