

# Class 11 Introduction to Genome Informatics Lab

Justin Robinson (PID: A16307501)

2024-05-07

## Population Scale Analysis Homework

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.

```
geno_table <- read.table(url("https://bioboot.github.io/bgg213_W19/class-material/rs8067378_ENSG000001"))
```

```
sum(geno_table$geno == "A/G")
```

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

```
sum(geno_table$geno == "A/A")
```

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

```
sum(geno_table$geno == "G/G")
```

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

Stats for the A/A group:

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
AA_geno <- geno_table %>%  
  filter(geno == "A/A")
```

```
summary(AA_geno)
```

```
##      sample      geno      exp
## Length:108      Length:108      Min.   :11.40
## Class :character Class :character 1st Qu.:27.02
## Mode  :character Mode  :character Median  :31.25
##                                     Mean   :31.82
##                                     3rd Qu.:35.92
##                                     Max.   :51.52
```

The median expression of ORMDL3 for the A/A genotype is 31.25

Stats for the A/G group:

```
AG_genotype <- geno_table %>%
  filter(genotype == "A/G")
summary(AG_genotype)
```

```
##      sample      geno      exp
## Length:233      Length:233      Min.   : 7.075
## Class :character Class :character 1st Qu.:20.626
## Mode  :character Mode  :character Median  :25.065
##                                     Mean   :25.397
##                                     3rd Qu.:30.552
##                                     Max.   :48.034
```

For the A/G genotype it is 25.065

Stats for the G/G group:

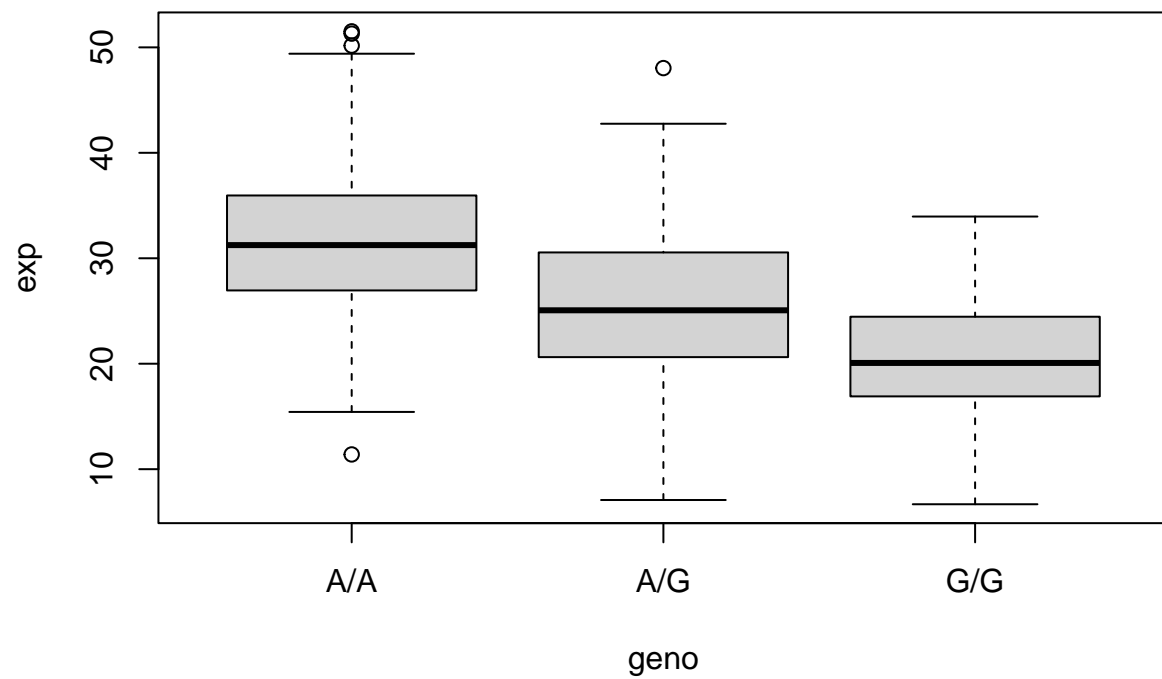
```
GG_genotype <- geno_table %>%
  filter(genotype == "G/G")
summary(GG_genotype)
```

```
##      sample      geno      exp
## Length:121      Length:121      Min.   : 6.675
## Class :character Class :character 1st Qu.:16.903
## Mode  :character Mode  :character Median  :20.074
##                                     Mean   :20.594
##                                     3rd Qu.:24.457
##                                     Max.   :33.956
```

For the G/G genotype it is 20.074

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?

```
boxplot(exp ~ genotype, data = geno_table)
```



Based on the above boxplot, it would seem that the A/A genotype has a higher expression of the ORMDL3 gene. A/G is second and G/G is last in terms of ORMDL3 expression.