

## Class 12 pt.2 HW: Q13, Q14

Sion Kang (PID: A17052234)

2025-02-13

### Section 4: Population Scale Analysis

One sample is obviously not enough to know what is happening in a population. 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.

#### Q13

How many samples do we have?

```
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
```

The sample size for each genotype

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

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

**A/A: 108, A/G: 233, G/G: 121**

Finding the median expression levels for each genotype

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      11.40   27.02   31.25   31.82   35.92   51.52
```

```
summary(expr$exp[expr$geno == "A/G"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      7.075  20.626  25.065  25.397  30.552  48.034
```

```
summary(expr$exp[expr$geno == "G/G"])
```

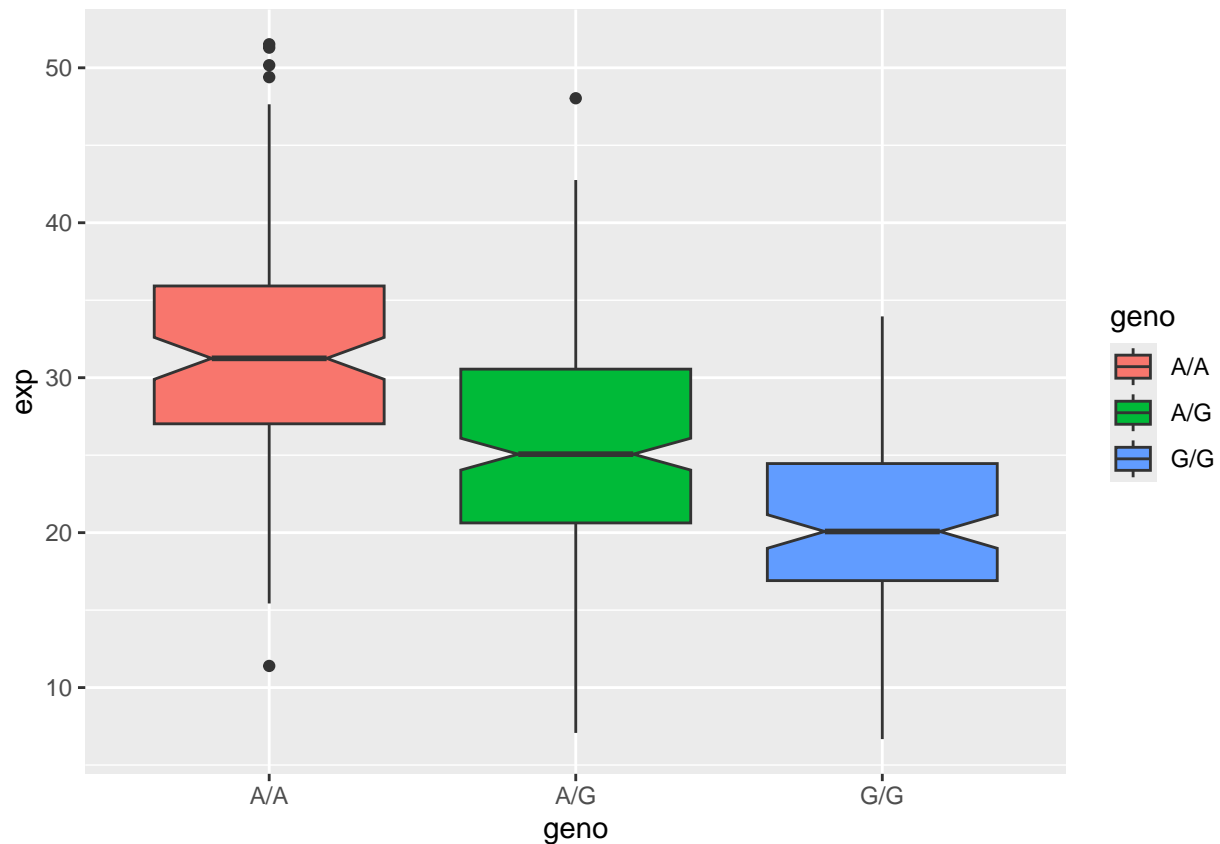
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  6.675  16.903   20.074   20.594  24.457   33.956
```

A/A median: 31.25; A/G median: 25.065; G/G median: 20.074

```
library(ggplot2)
```

Q. 14 Let's make a boxplot

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



From this boxplot, we can infer that the expression value is correlated to the genotypes (A/A, G/G). Thus, we can infer that the SNP effects the expression of ORMDL3.