

Genomics Homework

Brad Hunter (PID: A69038089)

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.

```
data <- read.table("rs8067378_ENSG00000172057.6.txt")  
  
x <- boxplot(exp ~ geno, data = data)
```

```
stats_table <- data.frame(  
  Genotype = x$names,  
  Sample_Size = x$n,  
  Median = round(x$stats[3, ], 2)  
)
```

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.

```
flextable::flextable(stats_table)
```

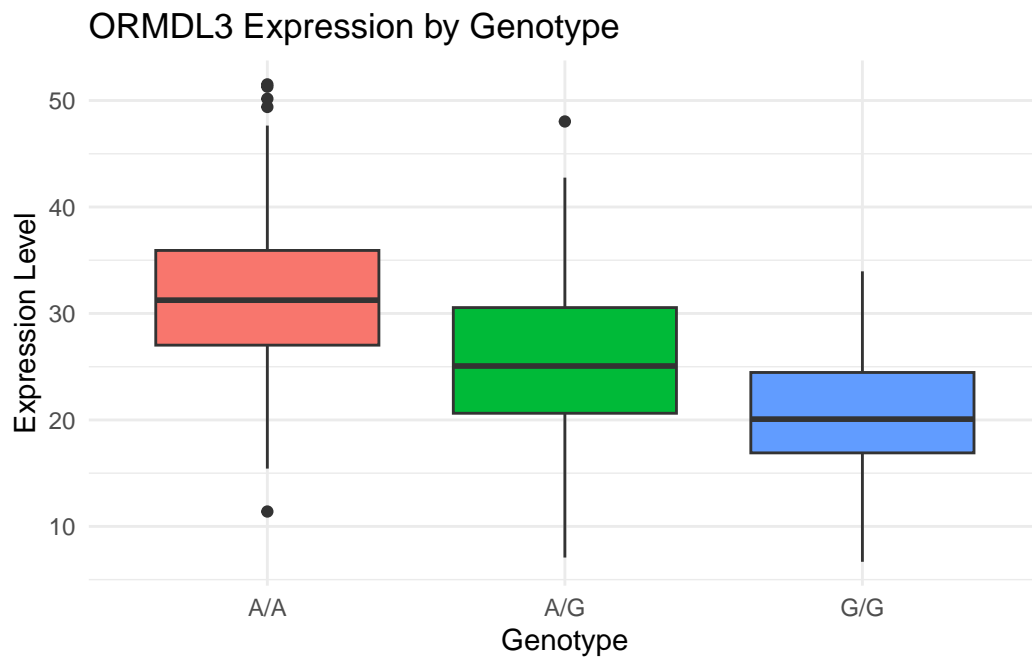
Genotype	Sample_Size	Median
A/A	108	31.25
A/G	233	25.06
G/G	121	20.07

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)
```

Warning: package 'ggplot2' was built under R version 4.5.2

```
ggplot(data, aes(x = geno, y = exp, fill = geno)) +  
  geom_boxplot() +  
  guides(fill = "none", color = "none") +  
  labs(  
    title = "ORMDL3 Expression by Genotype",  
    x = "Genotype",  
    y = "Expression Level") +  
  theme_minimal()
```



It does seem that SNP effects the expression of ORM DL3, the A|A genotype has a higher median expression level compared to the G|G genotype.