

Section 4: Population Scale Analysis [HOMEWORK]

Anita Wang (PID: A15567878)

12/5/2021

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. *Hint: The read.table(), summary() and boxplot() functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the boxplot() function to an R object and examining this object. There is also the medium() and summary() function that you can use to check your understanding.*

```
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
```

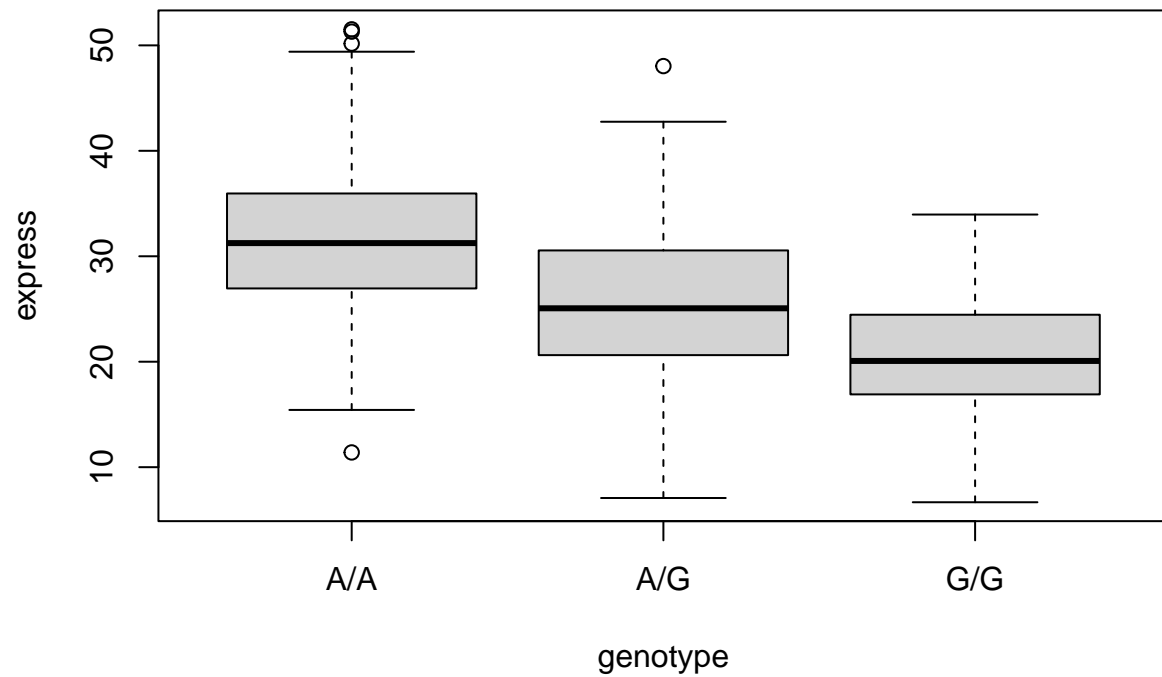
```
table(expr$geno)
```

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

Sample size for each genotype:

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

```
genotype <- expr$geno
express <- expr$exp
med <- boxplot(express~genotype)
```



med

```
## $stats
##      [,1]      [,2]      [,3]
## [1,] 15.42908  7.07505  6.67482
## [2,] 26.95022 20.62572 16.90256
## [3,] 31.24847 25.06486 20.07363
## [4,] 35.95503 30.55183 24.45672
## [5,] 49.39612 42.75662 33.95602
##
## $n
## [1] 108 233 121
##
## $conf
##      [,1]      [,2]      [,3]
## [1,] 29.87942 24.03742 18.98858
## [2,] 32.61753 26.09230 21.15868
##
## $out
## [1] 51.51787 50.16704 51.30170 11.39643 48.03410
##
## $group
## [1] 1 1 1 1 2
##
## $names
## [1] "A/A" "A/G" "G/G"
```

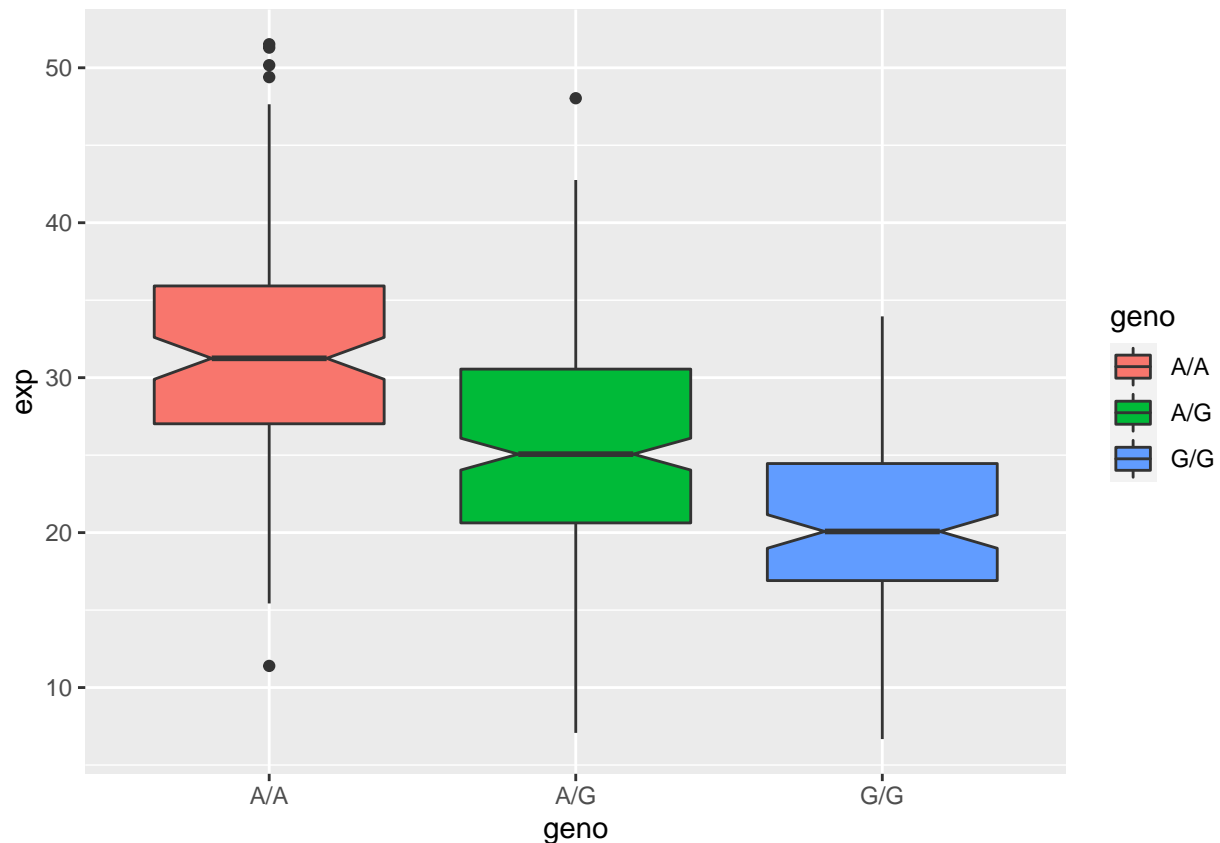
Corresponding median expression levels for each genotype:

A/A A/G G/G 31.24847 25.06486 20.07363

Let's make a boxplot

```
library(ggplot2)
```

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



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? Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

- From the boxplot, you can infer that the SNP does effect the expression of ORMDL3. Having a G/G in this location is associated with having a reduced expression of ORMDL3.