class19

Groot (PID: A15485151)

11/30/2021

#Proportion of G/G in a population

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
```

```
table(mxl$Genotype..forward.strand.)
```

```
##
## A|A A|G G|A G|G
##  22  21  12   9
```

```
table(mxl$Genotype..forward.strand.)/nrow(mxl) * 100
```

```
##
##      A|A      A|G      G|A      G|G
## 34.3750 32.8125 18.7500 14.0625
```

##Section 4: Population Scale Analysis > 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.

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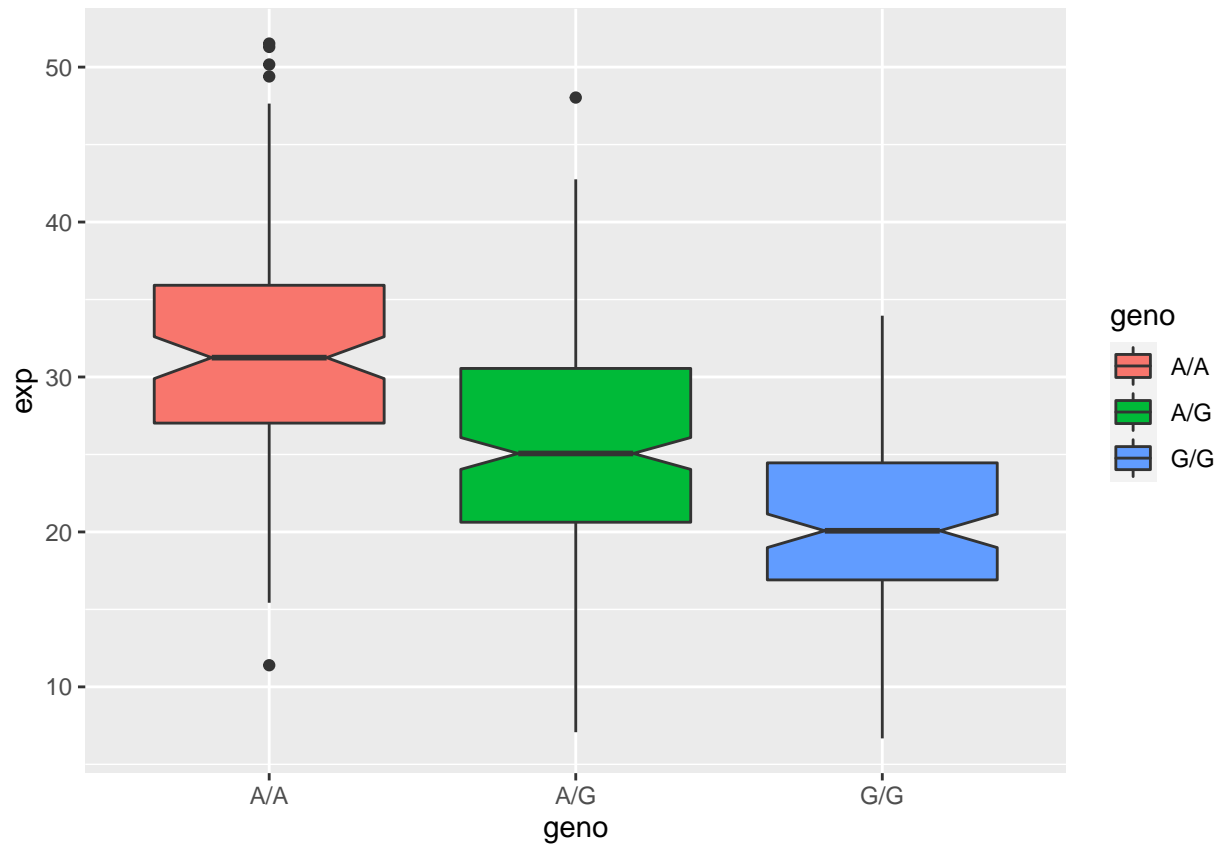
```
table(expr$geno)
```

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

sample size: A/A 108 A/G 233 G/G 121

```
library(ggplot2)
```

```
box <- ggplot(expr)+ aes(geno, exp, fill= geno) + geom_boxplot(notch = TRUE)
box
```
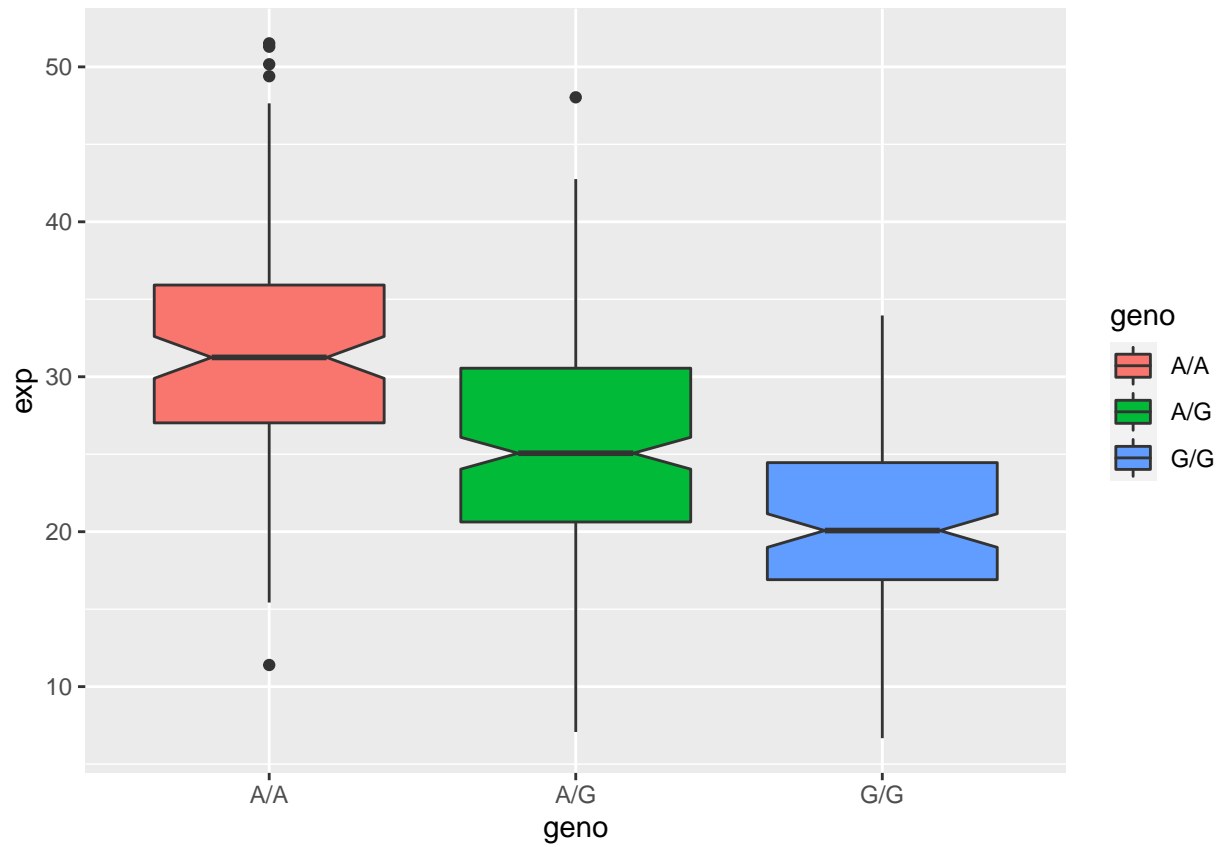

median expression level: A/A 32 A/G 25 G/G 20

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)
```

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

The relative gene expression is reduced from A/A to G/G so we could infer that the G/G genotype variation or the G/G SNP does affect the expression level of ORMDL3.