

class10

Katie Chau

2/21/2022

Population Analysis

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
```

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

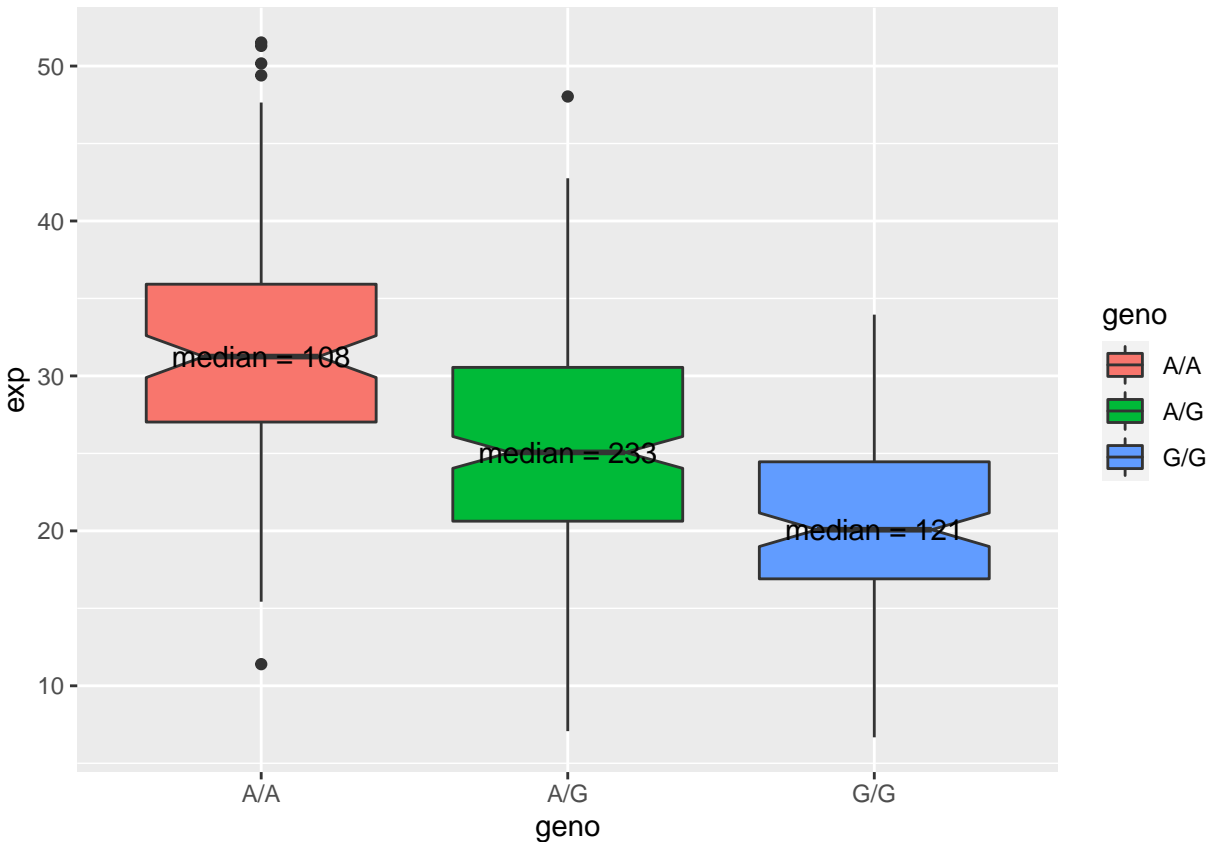
```
##
## A/A A/G G/G
## 108 233 121
```

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.

```
library(ggplot2)
```

Let's make a boxplot

```
n_fun <- function(geno){
  return(data.frame(y = median(geno), label = paste0("median = ",length(geno))))
}
ggplot(expr)+aes(x=geno,y=exp,fill=geno)+
  geom_boxplot(notch=TRUE)+
  stat_summary(fun.data = n_fun, geom = "text")
```



I followed the lab walk through, but I'm still not sure how you get the median() or median values. There is no median function available so I looked online to find help. I also could not find the example R script.

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.

The A/A expression levels are higher than the G/G expression levels by about 10 on average. Yes, I believe the SNP does effect the expression of ORMDL3.