

class12

Laurie Chang A16891192

##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

```
nrow(expr)
```

```
[1] 462
```

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

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

There are 108 individuals who have the A/A genotype, 233 individuals with the A/G genotype, and 121 individuals with the G/G genotype.

```
median_expression <- aggregate(expr$exp ~ expr$geno, data = expr, FUN = median)
median_expression
```

```
expr$geno expr$exp
1      A/A 31.24847
2      A/G 25.06486
3      G/G 20.07363
```

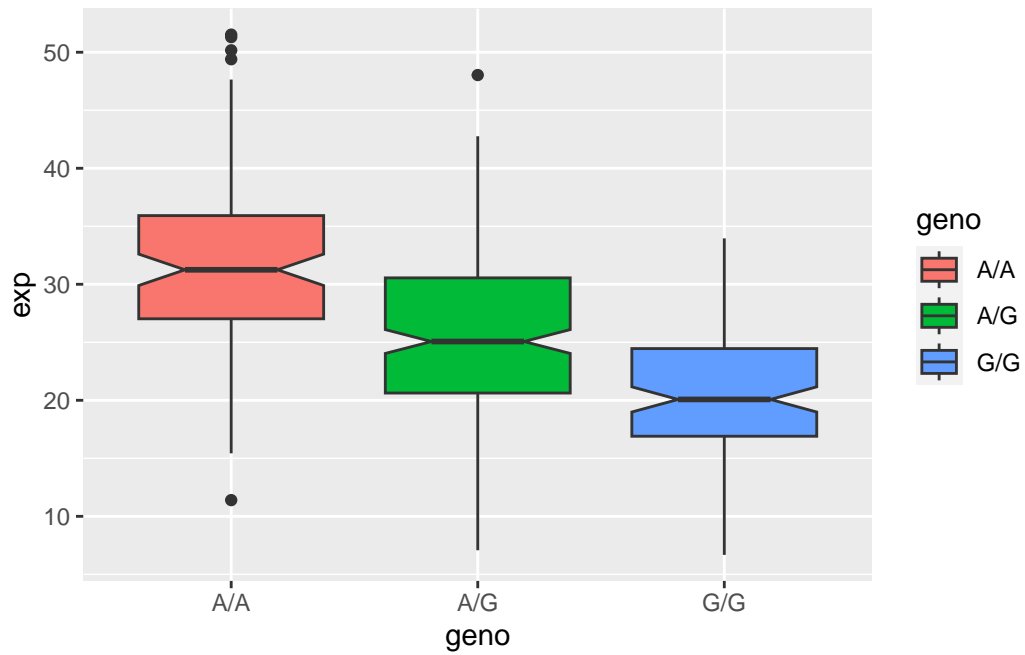
The median expression level for the A/A genotype is 31.24847. The median expression level for the A/G genotype is 25.06486. The median expression level for the G/G genotype is 20.07363.

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?

Let's make a boxplot!

```
library(ggplot2)

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



I can infer that there is a significant decrease in expression in the G/G phenotype in comparison to the A/A phenotype. It appears the SNP does effect the expression of ORM DL3.