

# HW Class 12

Samson A16867000

## 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

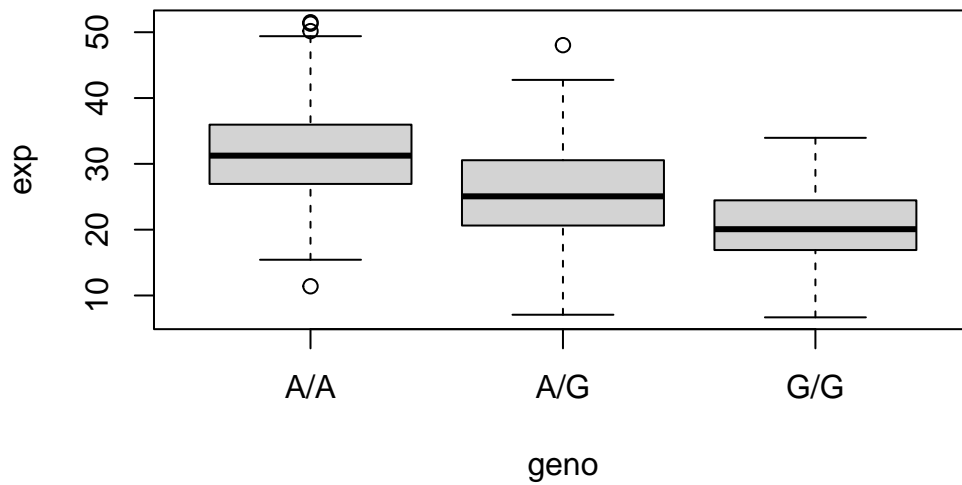
```
nrow(expr)
```

```
[1] 462
```

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

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

```
geno.exp <- boxplot(exp ~ geno, data = expr)
```



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

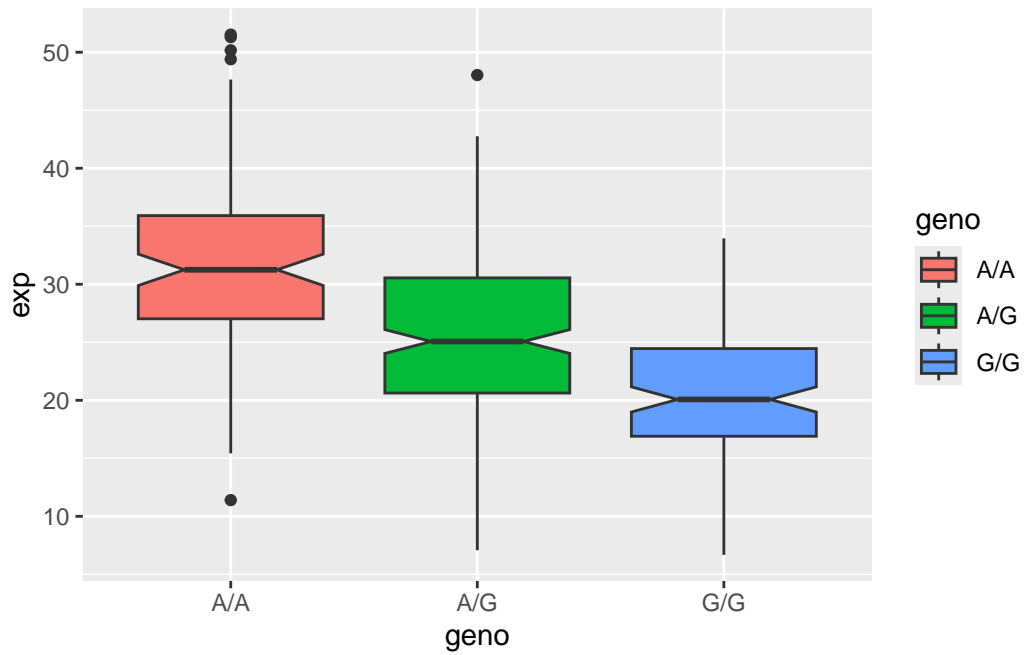
```
  geno    exp
1  A/A 31.24847
2  A/G 25.06486
3  G/G 20.07363
```

median expression levels for each of the genotypes

```
library(ggplot2)
```

. 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 ORM DL3?

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



Based on the boxplot it seems that having G/G is associated with having less expression compared to A/A. Yes the SNP effects the expression of ORMDL3