

# Class 12

Juan Gonzalez (PID: A69036681)

#Section 1: Proportion of G/G in a population

```
mxl<- read.csv("1000_Genomes_Project_Phase_3-SampleGenotypes-Homo_sapiens_Variation_Sample_r  
#head(mx1)
```

```
#table(mx1$Genotype...forward.strand.)/nrow(mx1)
```

#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
```

```
colnames(expression)
```

```
NULL
```

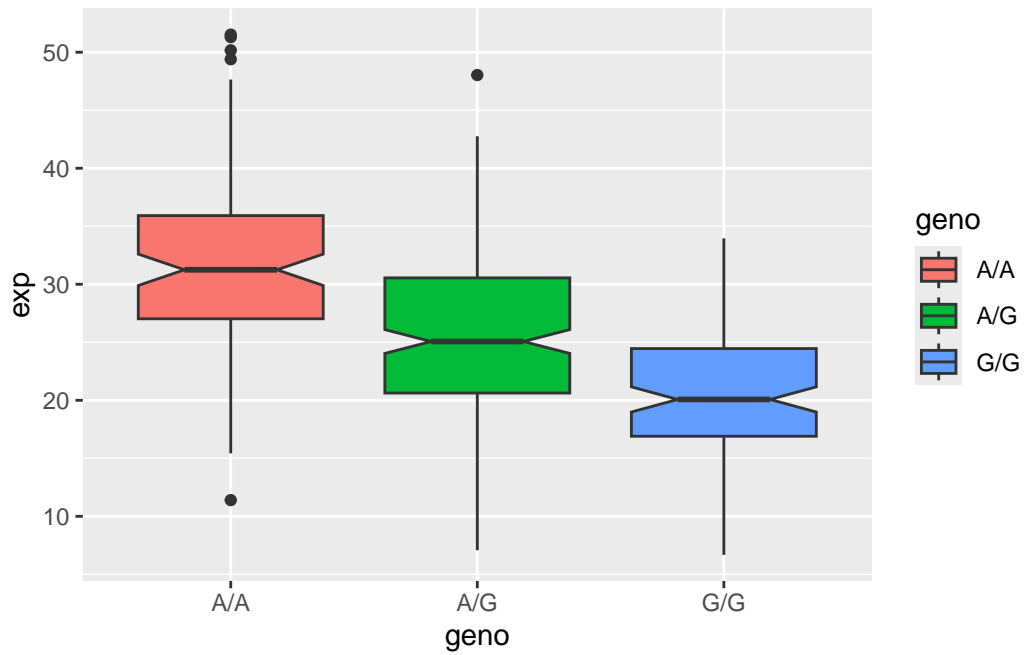
```
median_expression <- tapply(expr$exp, expr$geno, median)  
print(median_expression)
```

```
      A/A      A/G      G/G  
31.24847 25.06486 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?

```
library(ggplot2)
```

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



G/G is expressed down compared to A/A. The SNP will affect ORMDL3 because of the reduced expression of G/G.