

# Class 11: Genome Informatics

Anh Tran

```
mxl <- read.csv("MXL.csv", row.names=1)
```

```
View(mxl)
```

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

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

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

```
      A|A      A|G      G|A      G|G
0.343750 0.328125 0.187500 0.140625
```

#Section 4: Population 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.

```
url <- "https://bioboot.github.io/bimm143_F22/class-material/rs8067378_ENSG00000172057.6.t
```

```
file <- read.table(url, row.names=1)
head(file)
```

	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

The sample size for each genotype

```
table(file$geno)
```

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

```
A <- file[file$geno=="A/A",]
B <- file[file$geno=="A/G",]
C <- file[file$geno=="G/G",]
```

The median values for A/A, A/G, and G/G correspondingly are:

```
median(A$exp)
```

```
[1] 31.24847
```

```
median(B$exp)
```

```
[1] 25.06486
```

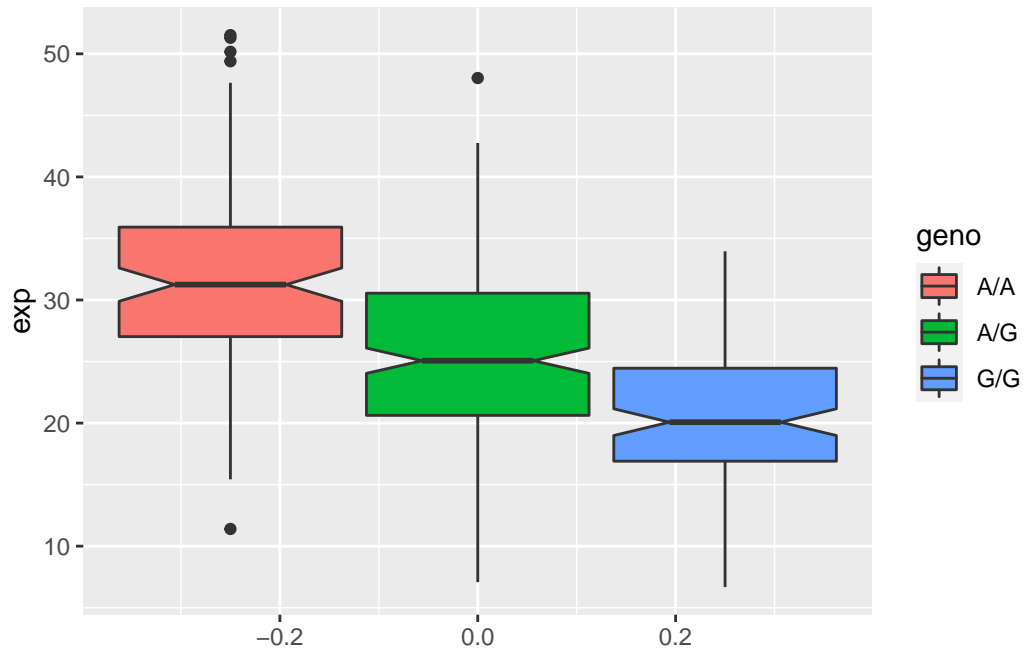
```
median(C$exp)
```

```
[1] 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)
```

```
plot <- ggplot(file) + aes(geno=geno, y=exp, fill=geno) + geom_boxplot(notch=TRUE)  
plot
```



The expression levels are lower for G/G and higher for A/A meaning that having G/G is associated with reduced expression of this gene. Yes, SNP affects the expression of ORMDL3.