

Class11: Genomics

Alice (PID:A16799081)

Section 1. Proportion of G/G in a population

```
mxl <- read.csv("MXL.csv")
```

```
table(mx1$Genotype..forward.strand.)
```

A A	A G	G A	G G
22	21	12	9

```
table(mx1$Genotype..forward.strand.) / nrow(mx1) * 100
```

A A	A G	G A	G G
34.3750	32.8125	18.7500	14.0625

Now let's look at GBR:

```
gbr <- read.csv("GBR.csv")
```

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

A A	A G	G A	G G
23	17	24	27

```
table(gbr$Genotype..forward.strand.) / nrow(gbr) * 100
```

```
      A|A      A|G      G|A      G|G
25.27473 18.68132 26.37363 29.67033
```

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")
```

```
nrow(expr)
```

```
[1] 462
```

The size of each genotypes are as follows:

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

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

```
summary(expr)
```

sample	geno	exp
Length:462	Length:462	Min. : 6.675
Class :character	Class :character	1st Qu.:20.004
Mode :character	Mode :character	Median :25.116
		Mean :25.640
		3rd Qu.:30.779
		Max. :51.518

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

```
medians
```

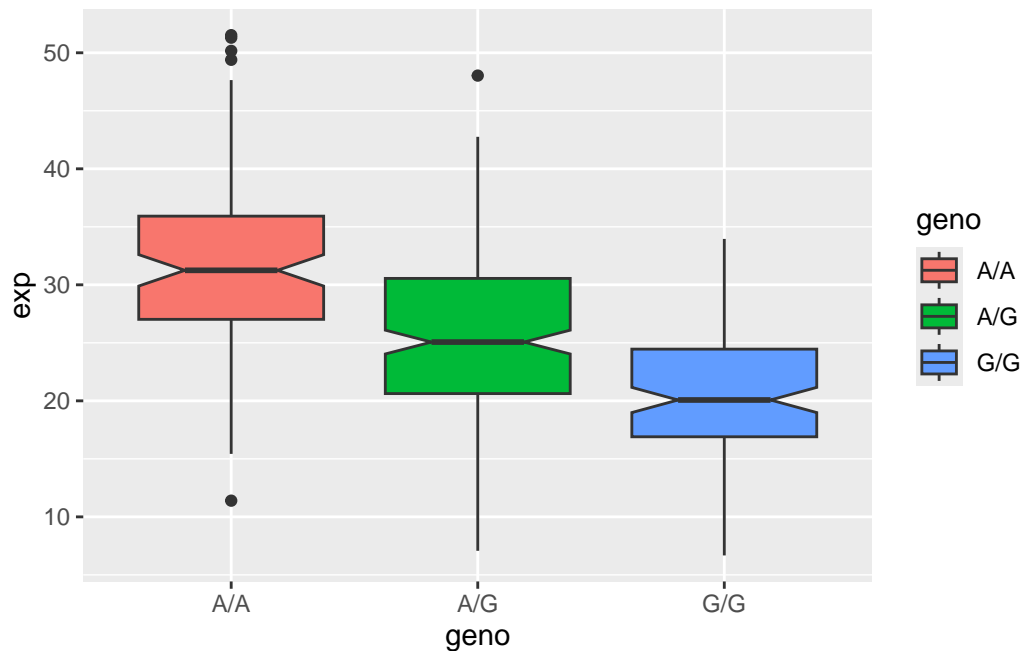
A/A	A/G	G/G
31.24847	25.06486	20.07363

A/A has 108 samples, A/G has 233 samples, and G/G has 121 samples. The corresponding median expression levels are 31.25, 25.06, and 20.07.

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 = T)
```



A/A have a higher expression value than G/G as displayed in the plot. This infers that the SNP does affect the expression of ORMDL3.