Class 12- Genome Informatics

Emily Ignatoff (A16732102)

Q5. What proportion of the Mexican Ancestry in Los Angeles sample population (MXL) are homozygous for the asthma associated SNP (G|G)?

```
mxl <- read.csv('mexican_la.csv')

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

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

Homework: Population Scale Analysis

I will be looking at 230 samples of RNASeq for the asthma associated SNPs

```
data1 <- read.csv("popdata.csv")
```

Q13. How many of each genotype and what are their corresponding median expression levels?

```
table(data1$geno)
```

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

```
median_by_geno <- tapply(data1$exp, data1$geno, median)
print(median_by_geno)
```

```
      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?

```r
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
by_geno <- ggplot(data1) + aes(x=geno, y=exp, fill=geno) + geom_boxplot() +
  labs(xlab="Genotype", ylab="Expression level", title="Expression level by Genotype")
by_geno
```
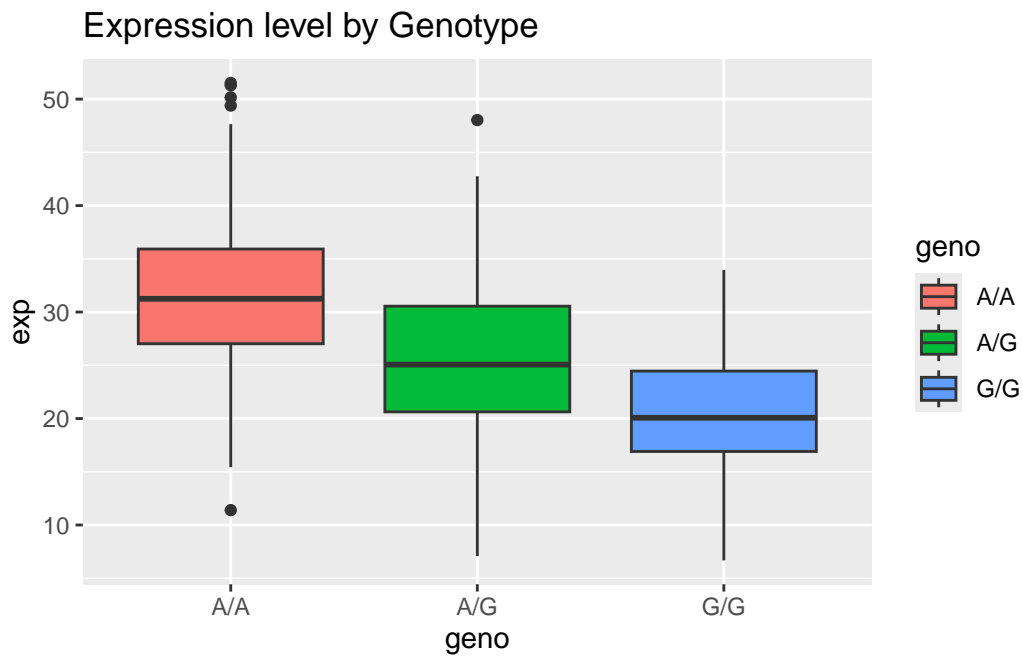

Expression of A|A is much higher than expression of G|G. This SNP likely reduces the expression of ORDML3.