Class 11: Genome Informatics

Kira

Section 1. Proportion of G/G in a Population

We can now read a CSV file that we downloaded from Ensemble.

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(mxl)
```

```
  Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1                  NA19648 (F)                       A|A ALL, AMR, MXL      -
2                  NA19649 (M)                       G|G ALL, AMR, MXL      -
3                  NA19651 (F)                       A|A ALL, AMR, MXL      -
4                  NA19652 (M)                       G|G ALL, AMR, MXL      -
5                  NA19654 (F)                       G|G ALL, AMR, MXL      -
6                  NA19655 (M)                       A|G ALL, AMR, MXL      -
  Mother
1      -
2      -
3      -
4      -
5      -
6      -
```

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

```

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

```
# How many of each genotype based on total number of individuals (rows)
table(mxl$Genotype..forward.strand.) / nrow(mxl) * 100
```

```
    A|A     A|G     G|A     G|G 
34.3750 32.8125 18.7500 14.0625 
```

We might want to compare the proportion of SNPs to another population (GBR).

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378 (1)(gbr).c
head(gbr)
```

```
  Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1                  HG00096 (M)                       A|A ALL, EUR, GBR      -
2                  HG00097 (F)                       G|A ALL, EUR, GBR      -
3                  HG00099 (F)                       G|G ALL, EUR, GBR      -
4                  HG00100 (F)                       A|A ALL, EUR, GBR      -
5                  HG00101 (M)                       A|A ALL, EUR, GBR      -
6                  HG00102 (F)                       A|A ALL, EUR, GBR      -
  Mother
1      -
2      -
3      -
4      -
5      -
6      -
```

```
round(table(gbr$Genotype..forward.strand.) / nrow(gbr) * 100,2)
```

```
  A|A   A|G   G|A   G|G 
25.27 18.68 26.37 29.67 
```

The proportion of individuals that are G|G is higher in the GBR population than in the MXL population.

Section 4: Homework Questions

> 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
```

```
# The total number of samples
nrow(expr)
```

```
[1] 462
```

```
# To find the sample size for each genotype
table(expr$geno)
```

```

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

There are 108 individuals with genotype A/A, 233 individuals with genotype A/G, and 121 individuals with genotype G/G.

```
# Putting the data into a data frame and pulling out relevant genotype G/G
expr.df <- data.frame(expr)
gg <- expr.df[expr.df$geno == "G/G",]
```

```
# To find the median expression value for the G/G genotype
round(mean(gg$exp),2)
```

```
[1] 20.59
```

```
# Repeating to select values for for A/A and A/G
ag <- expr.df[expr.df$geno == "A/G",]
aa <- expr.df[expr.df$geno == "A/A",]
```

```
# Repeating to find the median expression value for A/A and A/G
round(mean(ag$exp),2)
```

```
[1] 25.4
```

```
round(mean(aa$exp),2)
```

```
[1] 31.82
```

The median expression for each genotype is as follows:

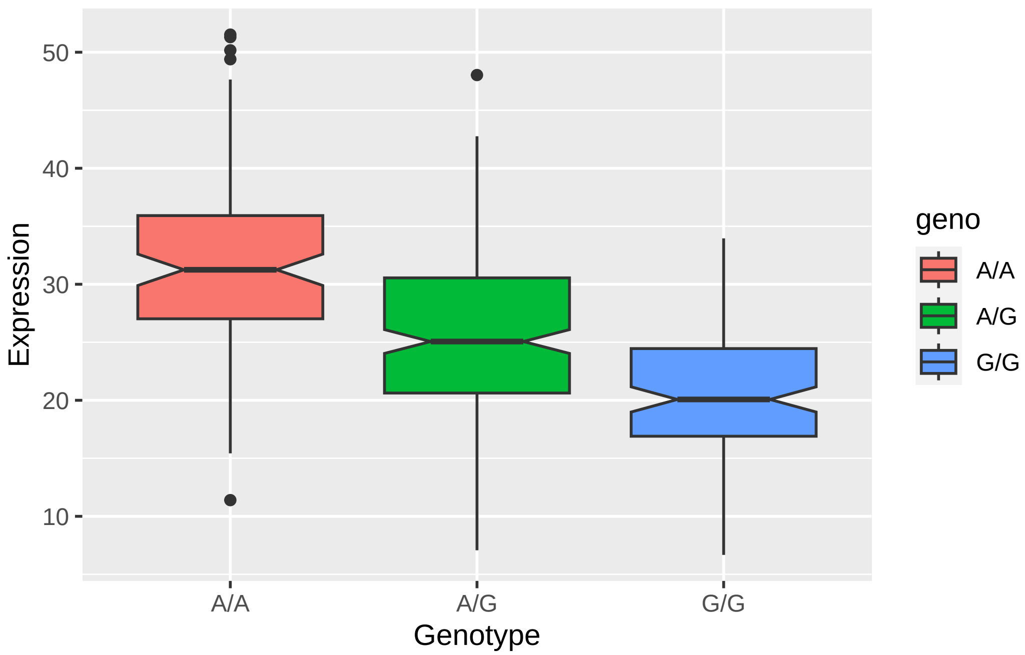
G/G : 20.59

A/G : 25.4

A/A : 31.82

> 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(x=geno,y=exp,fill=geno) + geom_boxplot(notch=TRUE) + xlab("Genotype") +
```



The expression of ORMDL3 decreases in the G/G genotype compared with the A/A genotype. We could hypothesize that the SNP influences expression but would need more detail/data to confirm this.