Class 11 Introduction to Genome Informatics Lab

Sindy

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.

```r
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
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
```

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

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

```r
nrow(expr)
```

```
[1] 462
```

```r
table(expr$geno)/nrow(expr)
```
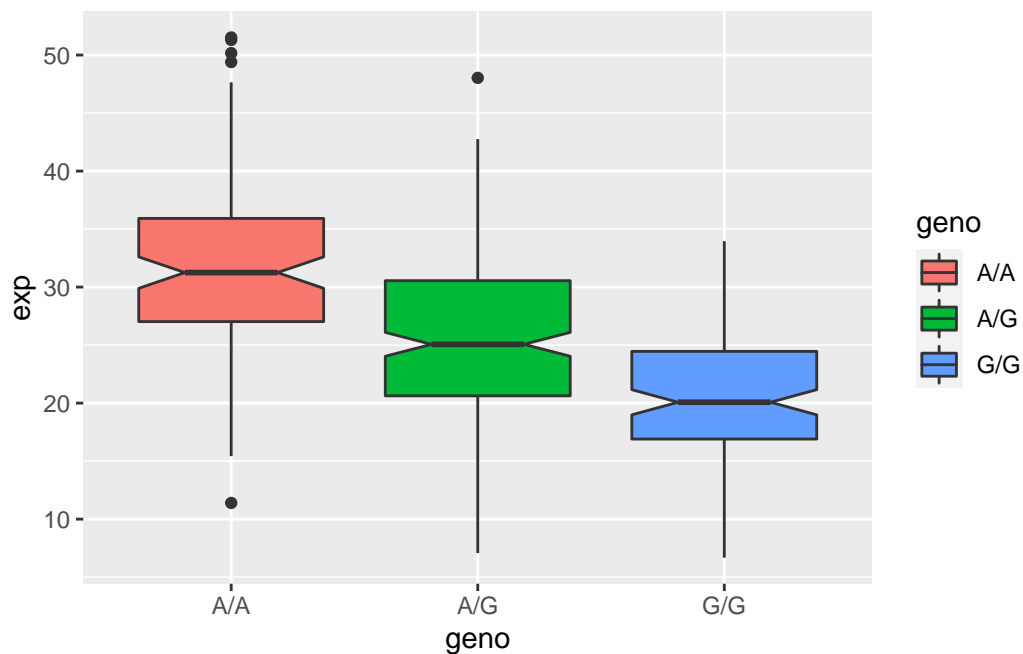
```
       A/A       A/G       G/G
0.2337662 0.5043290 0.2619048
```

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?

Boxplot

```
library(ggplot2)
```

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


what could you infer from the relative expression value between A/A and G/G displayed in this plot?

There is a statistically significant difference between the expression levels of A/A and G/G. A/A is more abundant in the population.

Does the SNP effect the expression of ORMDL3?

Yes. The OMIM entry for '% 611403 ASTHMA-RELATED TRAITS, SUSCEPTIBILITY TO, 6' states that being homozygous GG increases risk of asthma when exposed to environmental factors.

Question 6 Math

```
MXL <- read.csv("MXL_asthma.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
```

Proportion of G homozygous

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

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