

class10

Vince (PID: A15422556)

2/17/2022

MXL Population

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
table(mxl$Genotype..forward.strand.) / nrow(mxl)
```

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

GBR Population

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
table(gbr$Genotype..forward.strand.) / nrow(gbr)
```

```
##
##      A|A      A|G      G|A      G|G
## 0.2527473 0.1868132 0.2637363 0.2967033
```

Expression by Genotype Analysis

Q13. Read file into R, determine sample size and median expression levels. Sample size for A/A is 108, for A/G is 233, and for G/G is 121. Median expression levels for A/A is 31.25, for A/G is 25.065, and for G/G is 20.074.

```
x <- read.table("rs8067378_ENSG000000172057.6.txt")
table(x$geno)
```

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

```
print("G/G")
```

```
## [1] "G/G"
```

```
summary(x[x$geno == "G/G", 3])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  6.675  16.903  20.074  20.594  24.457  33.956
```

```
print("A/G")
```

```
## [1] "A/G"
```

```
summary(x[x$geno == "A/G", 3])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.075  20.626  25.065  25.397  30.552  48.034
```

```
print("A/A")
```

```
## [1] "A/A"
```

```
summary(x[x$geno == "A/A", 3])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.40   27.02   31.25   31.82   35.92   51.52
```

Q14. Is the mean expression different based on genotype? Yes, the SNP seems to cause a decrease in expression of ORMDL3.

Summary Overview Figure

```
library(ggplot2)
```

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

