

# W8\_Genomics\_Lab

Vera Sophia Beliaev

2/17/2022

## Section 4: Population 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")
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
```

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

```
## [1] 462
```

```
#Make a table counting each genotype
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
```

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? The A/A genotype has much higher expression than the G/G genotype as evidenced by the boxplot. Having an A/A genotype corresponds to high expression of ORMDL3 while having a G/G genotype corresponds to low expression of ORMDL3.

```
library(ggplot2)
```

Make a boxplot

```
ggplot(expr) + aes(geno, exp, col= geno) + geom_boxplot(notch=TRUE)
```

