

Class 19 HW

Matt Hashimoto

12/6/2021

Population Scale 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.”

```
# Read file from url
finaldata <- read.table(url("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG00000177170.txt"))

# Check summary of the overall data
summary(finaldata)
```

```
##      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
```

```
# Create new tables based on genotype
AAtable <- finaldata[-which(finaldata$geno != "A/A"),]
AGtable <- finaldata[-which(finaldata$geno != "A/G"),]
GGtable <- finaldata[-which(finaldata$geno != "G/G"),]

# Check new tables
summary(AAtable)
```

```
##      sample      geno      exp
## Length:108      Length:108      Min.   :11.40
## Class :character Class :character 1st Qu.:27.02
## Mode  :character Mode  :character Median :31.25
##                                     Mean  :31.82
##                                     3rd Qu.:35.92
##                                     Max.   :51.52
```

```
summary(AGtable)
```

```
##      sample      geno      exp
## Length:233      Length:233      Min.   : 7.075
```

```
## Class :character   Class :character   1st Qu.:20.626
## Mode  :character   Mode  :character   Median :25.065
##                                           Mean  :25.397
##                                           3rd Qu.:30.552
##                                           Max.   :48.034
```

```
summary(GGtable)
```

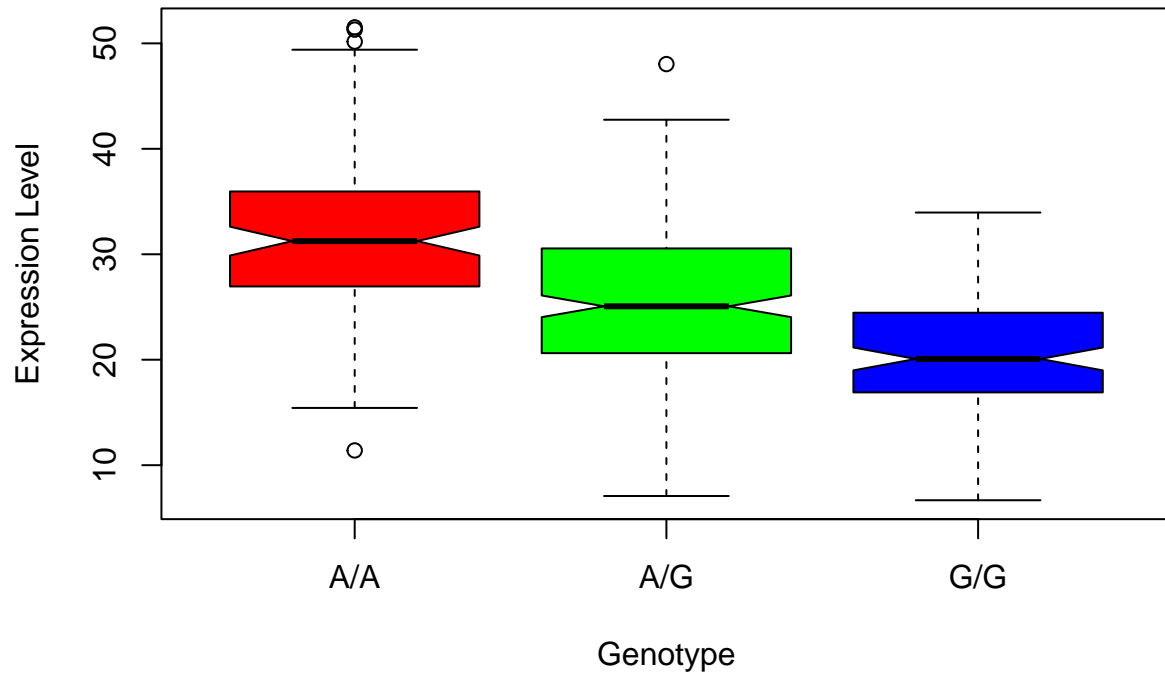
```
##      sample          geno          exp
## Length:121      Length:121      Min.   : 6.675
## Class :character   Class :character   1st Qu.:16.903
## Mode  :character   Mode  :character   Median :20.074
##                                           Mean  :20.594
##                                           3rd Qu.:24.457
##                                           Max.   :33.956
```

The sample size of A/A is 108, A/G is 233, and G/G is 121. The median expression level for A/A is 31.25, A/G is 25.065, G/G is 20.074.

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?”

```
boxlist <- list("A/A" = AAtable$exp, "A/G" = AGtable$exp, "G/G" = GGtable$exp)
boxplot(boxlist,
        main = "Affect of Genotype on Expression Level in ORMDL3",
        xlab = "Genotype",
        ylab = "Expression Level",
        notch = TRUE,
        col = c("red", "green", "blue"))
```

Affect of Genotype on Expression Level in ORMDL3



From this box plot, it is reasonable to infer that genotype does significantly affect expression levels of the ORMDL3 gene. The median value for the G/G data falls well outside the IQR of the A/A data, indicating significance. Thus, there is strong evidence that the associated SNP affects the expression of ORMDL3.