Class11HW

Laura Biggs

2022-11-01

Read in expression genotype results

```
genotype_data <- read.table("rs8067378_ENSG00000172057.6.txt")
View(genotype_data)
```

Q13: Sample size/median for each genotype AG median: 25.065 AA median: 31.25 GG median: 20.074

```
table(genotype_data$geno)
```

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

```
AG <- genotype_data[grepl("A/G", genotype_data$geno), ]
summary(AG)
```

```
##     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
```

```
AA <- (genotype_data[grepl("A/A", genotype_data$geno), ])
summary(AA)
```

```
##     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
```

```
GG <- (genotype_data[grepl("G/G", genotype_data$geno), ])
summary(GG)
```

```
##     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
```

Q14: Boxplot The SNP is G|G which reduces ORMDL3 expression relative to the A|A genotype. A|A is expressed in approximately 30% of the sample while G|G is expressed in 20% of the samples.

```
boxplot(genotype_data$exp ~ genotype_data$geno)
```