

Genomics Lab

Justin Lu (A16318305)

2024-05-09

Section 4

```
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
```

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.

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

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

```
summary(expr$exp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.675  20.004   25.116   25.640   30.779   51.518
```

The sample sizes are listed above, and the median expression is listed. However, I wasn't sure how to calculate medians by genotype, but it's around 32 for A/A, 25 for A/G, and 20 for G/G based on the boxplots.

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?

A/A does seem to have a higher expression value compared to G/G, although the difference is not statistically significant. I would say based on the distribution and location of the data points for each genotype though, that the SNP does affect the expression of ORMDL3.

```
library(ggplot2)
```

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

