

Homework 11: Population Scale Analysis

Arshiya Zarmahd (PID: A16247996)

Section 4

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 |

```
nrow(expr)
```

```
[1] 462
```

462 individuals in this data file.

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

| | | |
|-----|-----|-----|
| A/A | A/G | G/G |
| 108 | 233 | 121 |

This is the breakdown of how many individuals there are with each genotype.

```
summary(expr)
```

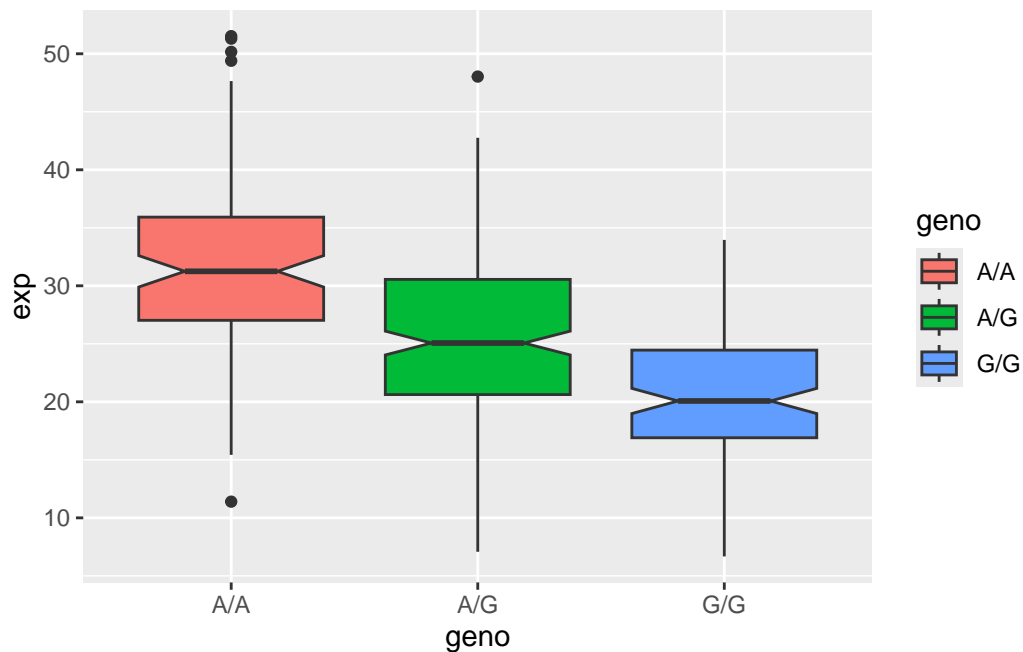
| sample | geno | exp |
|------------------|------------------|----------------|
| Length:462 | Length:462 | Min. : 6.675 |
| Class :character | Class :character | 1st Qu.:20.004 |
| Mode :character | Mode :character | Median :25.116 |
| | | Mean :25.640 |
| | | 3rd Qu.:30.779 |
| | | Max. :51.518 |

Here is a quick summary of the data file.

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?

```
library(ggplot2)
```

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



Here is a boxplot of the data file. A|A is the highest expressed genotype, because it has the highest median, with G|G being the lowest expressed genotype.