## class11 HW

Ayanna Kerr (PID: A17143404)

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.

```
file <- read.table("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG000
summary(file)</pre>
```

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

table(file\$geno)

A/A A/G G/G 108 233 121

```
df <- as.data.frame(file)
head(df)</pre>
```

```
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
```

```
median <- aggregate(df$exp ~ df$geno, FUN = stats::median)
median</pre>
```

```
df$geno df$exp
1 A/A 31.24847
2 A/G 25.06486
3 G/G 20.07363
```

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(file, aes(x = geno, y = exp))+
  geom_boxplot()+
  labs(x = "Genotype", y = "Expression")+
  geom_jitter(width = 0.2, alpha = 0.4)+
  theme_bw()
```

