Class 12

Adam Bisharat

Section 4: Population Scale Analysis [HOMEWORK] One sample is obviously not enough to know what is happening in a population. You are interested in assessing genetic differences on a population scale. So, you processed about $\sim\!230$ samples and did the normalization on a genome level. Now, you want to find whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on ORMDL3 expression.

```
expr <- read.table("Genome.txt")
head(expr)</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
```

```
nrow(expr)
```

[1] 462

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

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

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(x=geno, y=exp, fill=geno) + geom_boxplot(notch=T)
```

