Class 12

Natalie Ogg A91030809

Section 4: Population Analysis

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 ~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.

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.

How many samples do we have? ^

```
expr <- read.table("expression data.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
```

Sample sizes:
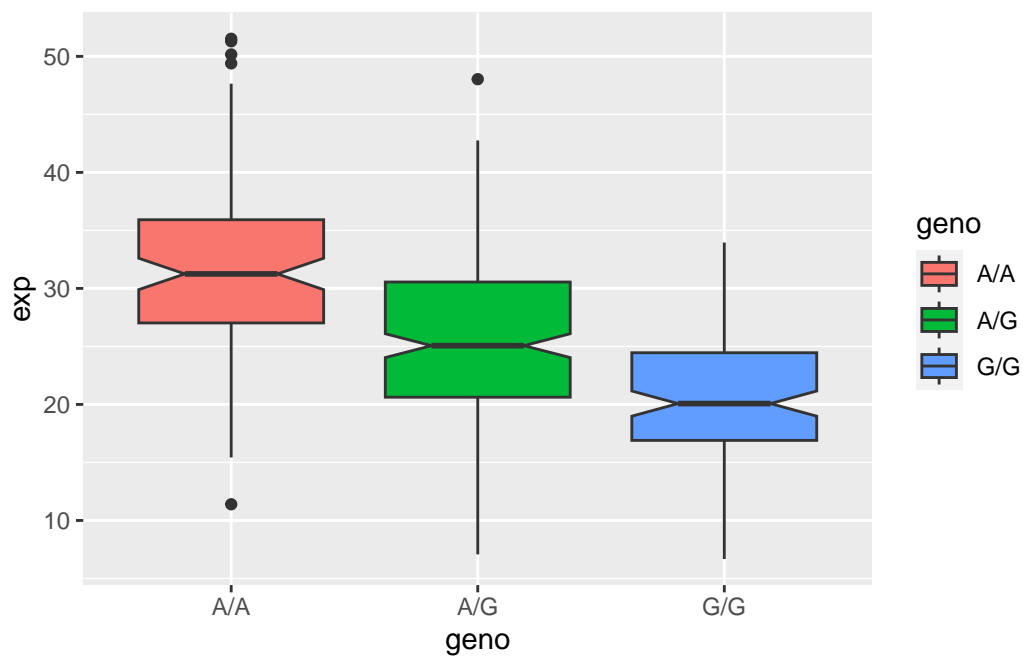
```
nrow(expr)
```
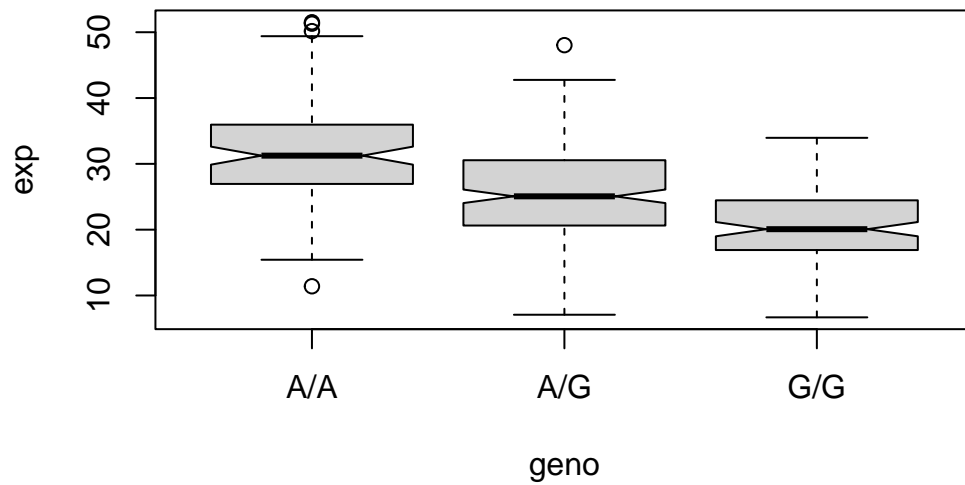
```
[1] 462
```

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

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

Let's make a box plot!

```
library(ggplot2)
ggplot(expr, aes(geno, exp, fill=geno)) + geom_boxplot(notch=TRUE)
```



```
box <- boxplot(exp ~ geno, data = expr, notch = TRUE)
```

```
box
```

```
$stats
         [,1]     [,2]     [,3]
[1,] 15.42908  7.07505  6.67482
[2,] 26.95022 20.62572 16.90256
[3,] 31.24847 25.06486 20.07363
[4,] 35.95503 30.55183 24.45672
[5,] 49.39612 42.75662 33.95602

$n
[1] 108 233 121

$conf
         [,1]     [,2]     [,3]
[1,] 29.87942 24.03742 18.98858
[2,] 32.61753 26.09230 21.15868

$out
[1] 51.51787 50.16704 51.30170 11.39643 48.03410

$group
```

```
[1] 1 1 1 1 2

$names
[1] "A/A" "A/G" "G/G"
```

```
  median(box$stats[,1])
```

```
[1] 31.24847
```

```
  median(box$stats[,2])
```

```
[1] 25.06486
```

```
  median(box$stats[,3])
```

```
[1] 20.07363
```

Median expression levels:

A/A: 31.2

A/G: 25.1

G/G: 20.1

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?

Based on the boxplot and median values, it appears that the A allele has a positive correlation with ORMDL3 gene expression. A/A genotypes result in higher gene expression than G/G genotypes, with heterozygosity resulting in medium levels of expression.