

Class 11 hw

Jennifer

Section 4: Population Scale Analysis

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.

We will read our file into R using 'read.table()'.

```
expr <- read.table("https://bioboot.github.io/bimm143_F22/class-material/rs8067378_ENSG000  
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

The 'nrow()' function will give us the sample size.

```
nrow(expr)
```

```
[1] 462
```

The 'table()' function can help give us the sample size for each genotype.

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

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

108 for A/A, 233 for A/G, 121 for G/G

We can use the ‘aggregate()’, ‘list()’, and ‘median()’ function to help us find the median expression levels for each genotype.

```
aggregate(expr$exp, list(expr$geno), median)
```

	Group.1	x
1	A/A	31.24847
2	A/G	25.06486
3	G/G	20.07363

The table above shows the median expression levels for each genotype.

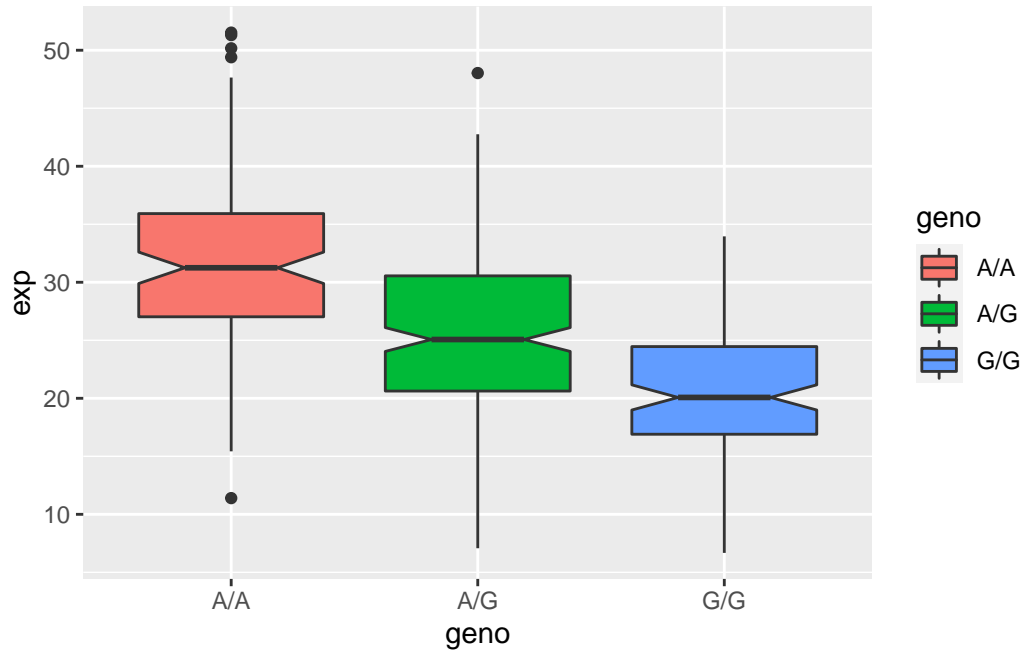
Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

We can read ggplot2 to make a boxplot.

```
library(ggplot2)
```

Let’s make a boxplot.

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



From the expression in this plot we can see that the expression value between A/G and G/G show that G/G is correlated with reducing the expression of the gene. SNP does effect the expression of ORMDL3.