

Lab 12

Keilyn Duarte (PID 16881868)

Section 4: Population Scale Analysis [HOMEWORK]

Reading the expression genotype results file

```
expresults <- read.table(file="rs8067378_ENSG00000172057.6.txt")  
head(expresults)
```

	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

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.

Find number of G/G genotype samples

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
gg <- expresresults %>%  
  filter(geno == "G/G")  
  
nrow(gg)
```

```
[1] 121
```

Find number of A/G genotype samples

```
ag <- expresresults %>%  
  filter(geno == "A/G")  
  
nrow(ag)
```

```
[1] 233
```

Find number of A/A genotype samples

```
aa <- expresresults %>%  
  filter(geno == "A/A")  
  
nrow(aa)
```

```
[1] 108
```

Find the median expression level for G/G genotype

```
median(gg$exp)
```

```
[1] 20.07363
```

Find the median expression level for the A/G genotype

```
median(ag$exp)
```

```
[1] 25.06486
```

Find the median expression level for the A/A genotype

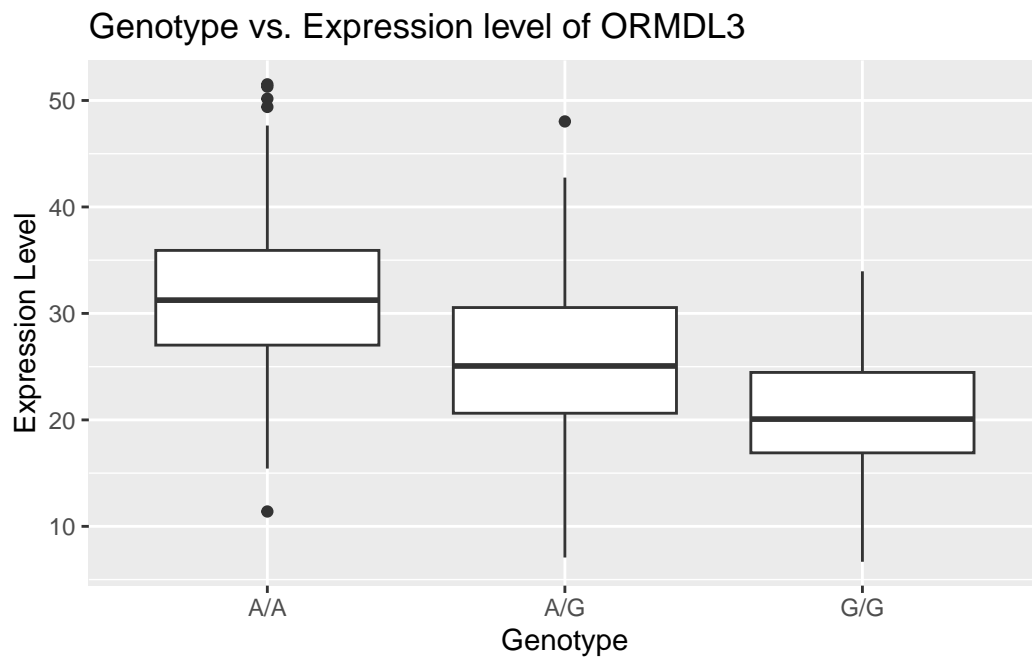
```
median(aa$exp)
```

```
[1] 31.24847
```

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(expresresults, aes(x=geno, y=exp)) +
  geom_boxplot() +
  xlab("Genotype") +
  ylab("Expression Level") +
  labs(title="Genotype vs. Expression level of ORMDL3")
```



The “A/A” genotype has the highest expression level of ORMDL3 and the “G/G” genotype has the lowest. This suggests that the SNP does indeed effect the expression level of ORMDL3.