

# lab 11

Chan-yu Kuo

Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
expr <- read.table("rs8067378_ENSG00000172057.6.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

```
nrow(expr)
```

```
[1] 462
```

sample size for each genotype . the median expression levels for each genotypes (A/A, A/G, and G/G)

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

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

```
AA_median<-median(expr$exp[expr$geno=="A/A"])
AG_median<-median(expr$exp[expr$geno=="A/G"])
GG_median<-median(expr$exp[expr$geno=="G/G"])
AA_median
```

```
[1] 31.24847
```

```
AG_median
```

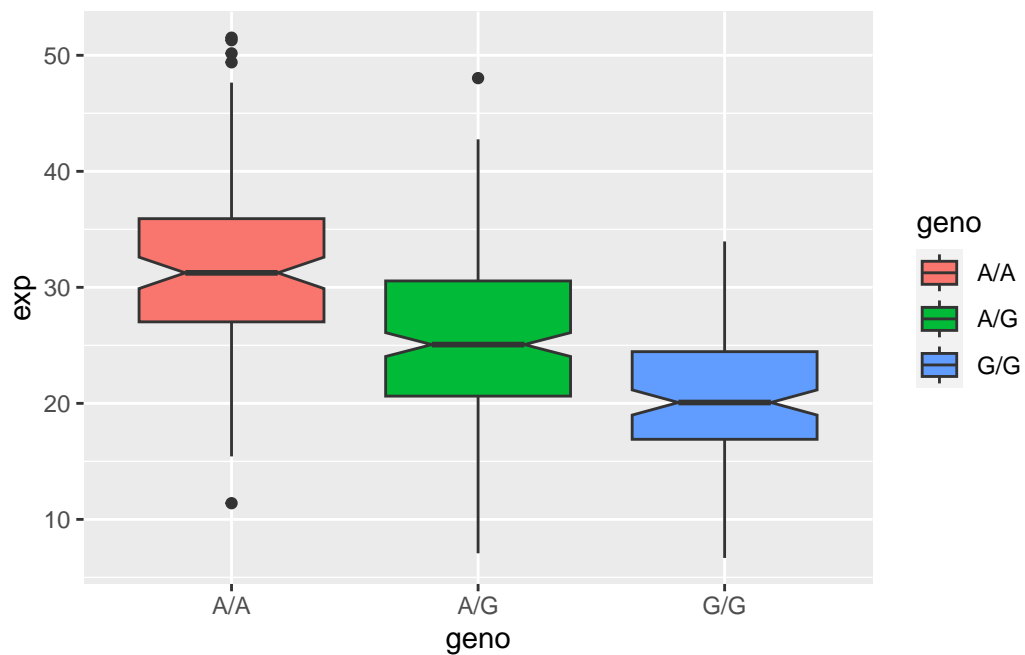
```
[1] 25.06486
```

```
GG_median
```

```
[1] 20.07363
```

q14

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



q14: What can you infer?

The expression value on average is higher for A/A than G/G. anomalies is also more common in A/A than on G/G. Based on the median expression value, as well as the general distribution of the expression level, the SNP does effect the expression fo ORM DL3. When having an SNP (From A to G), the expression value goes down