

Class11

Hetian Su

Q13

```
# read in the file
pop <- read.table('rs8067378_ENSG000000172057.6.txt')
```

```
summary(pop[pop$geno=='A/G',])
```

sample	geno	exp
Length:233	Length:233	Min. : 7.075
Class :character	Class :character	1st Qu.:20.626
Mode :character	Mode :character	Median :25.065
		Mean :25.397
		3rd Qu.:30.552
		Max. :48.034

The genotype A/G has sample size 233 and median 25.065.

```
summary(pop[pop$geno=='A/A',])
```

sample	geno	exp
Length:108	Length:108	Min. :11.40
Class :character	Class :character	1st Qu.:27.02
Mode :character	Mode :character	Median :31.25
		Mean :31.82
		3rd Qu.:35.92
		Max. :51.52

The genotype A/A has sample size 108 and median 31.25

```
summary(pop[pop$geno=='G/G',])
```

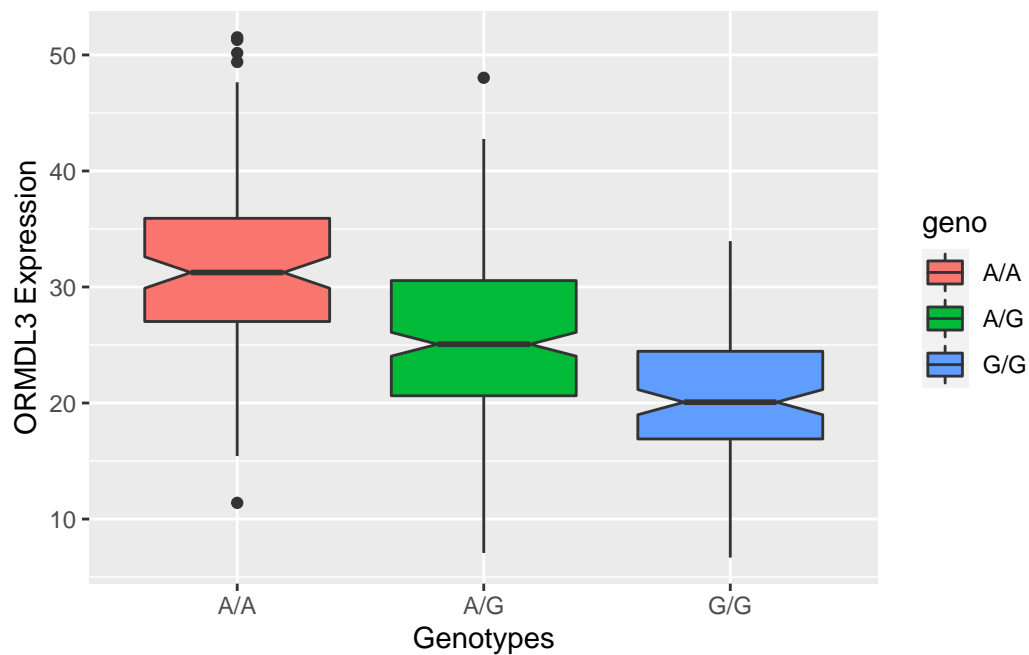
sample	geno	exp
Length:121	Length:121	Min. : 6.675
Class :character	Class :character	1st Qu.:16.903
Mode :character	Mode :character	Median :20.074
		Mean :20.594
		3rd Qu.:24.457
		Max. :33.956

The genotype G/G has sample size 121 and median 20.074.

Q14

```
library(ggplot2)
```

```
ggplot(pop)+
  aes(geno, exp, fill=geno)+
  geom_boxplot(notch = T)+
  labs(x='Genotypes', y='ORMDL3 Expression')
```



Comparing A/A and G/G, we can infer that homozygous A/A is correlated with a significantly higher ORMDL3 expression level. Combining with the fact that one allele changing to G results in a intermediate expression level, we can propose that allele A is a “stronger” allele correlated with high ORMDL3 expression, whereas G is a “weaker” allele.