

# Class 12

Tim

Section 4 Download the file

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.

## Sample size

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

#Median A/A A/G G/G 31.24847 25.06486 20.07363

```
data <- read.table("rs8067378_ENSG00000172057.6.txt")
head(data)
```

	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(data)
```

```
[1] 462
```

```
table(data$geno)
```

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

```
tapply(data$exp, data$geno, median)
```

```
      A/A      A/G      G/G
31.24847 25.06486 20.07363
```

#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

There is clear difference between the med lines of G/G and A/A. A/A is more expressed than G/G shwoing the effect of SNP on ORMDL3

```
library(ggplot2)
```

```
ggplot(data, aes(geno, exp)) +  
  geom_boxplot()
```

