class 11Structural Bioinformatics

> 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.

How many samples do we have?

```
expr <- read.table("textfile.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
```
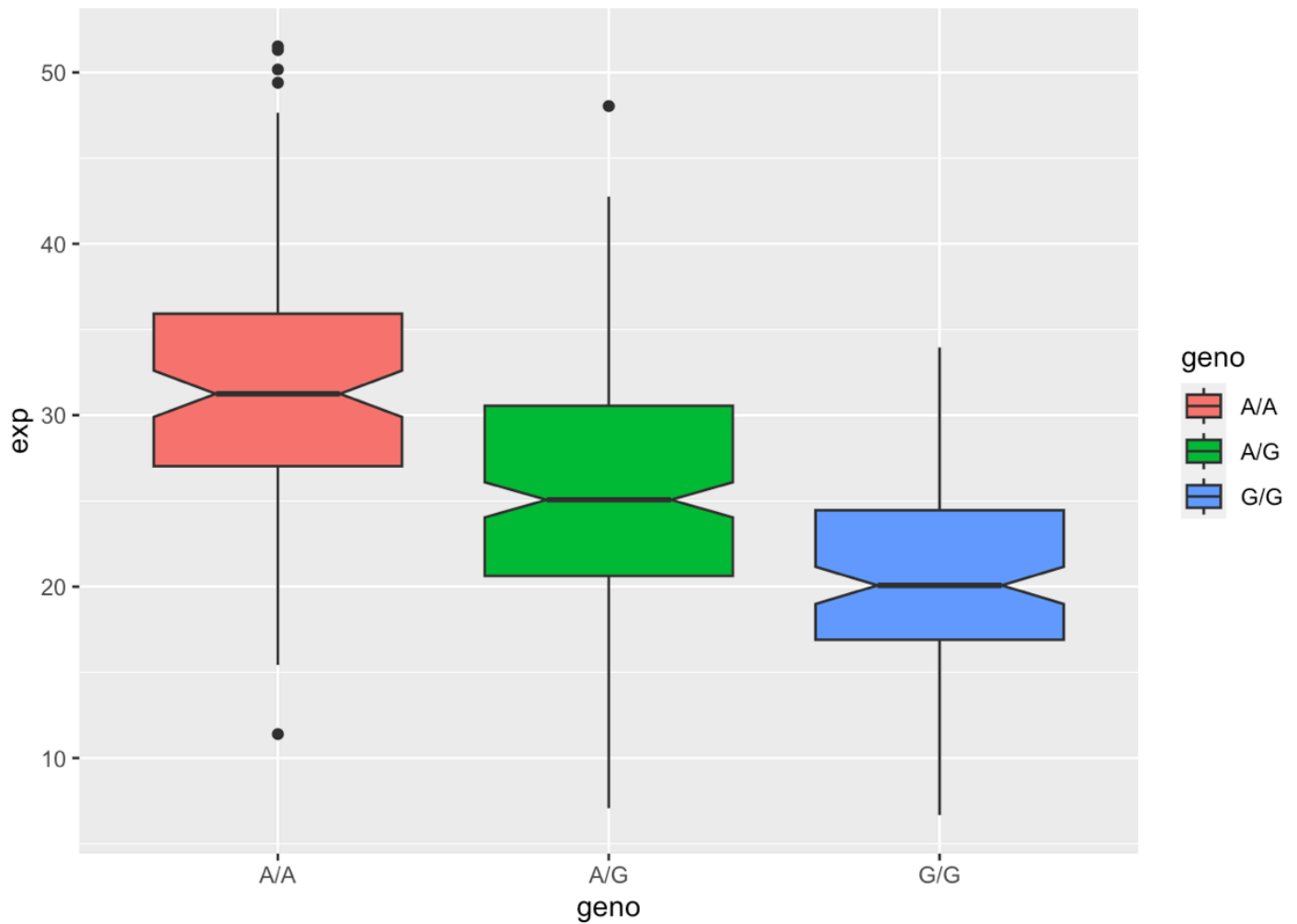
```
table(expr$geno)
```

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

```
library(ggplot2)
```

> 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? Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

lets make a boxplot

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


I could infer that A|A has a higher expression than G|G. The reduced expression of G|G could show that SNP does effect the expression of ORMLDL3 since it altered the gene expression level, potentially influencing the disease function and immune response.