

class14 RNA-seq

Pham Vo

2/21/2022

Section 4: One sample is obviously not enough to know what is happening in a population. You are interested in assessing genetic differences on a population scale. So, you processed about ~230 samples and did the normalization on a genome level. Now, you want to find whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on ORM DL3 expression.

Q13. How many samples do we have?

```
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
```

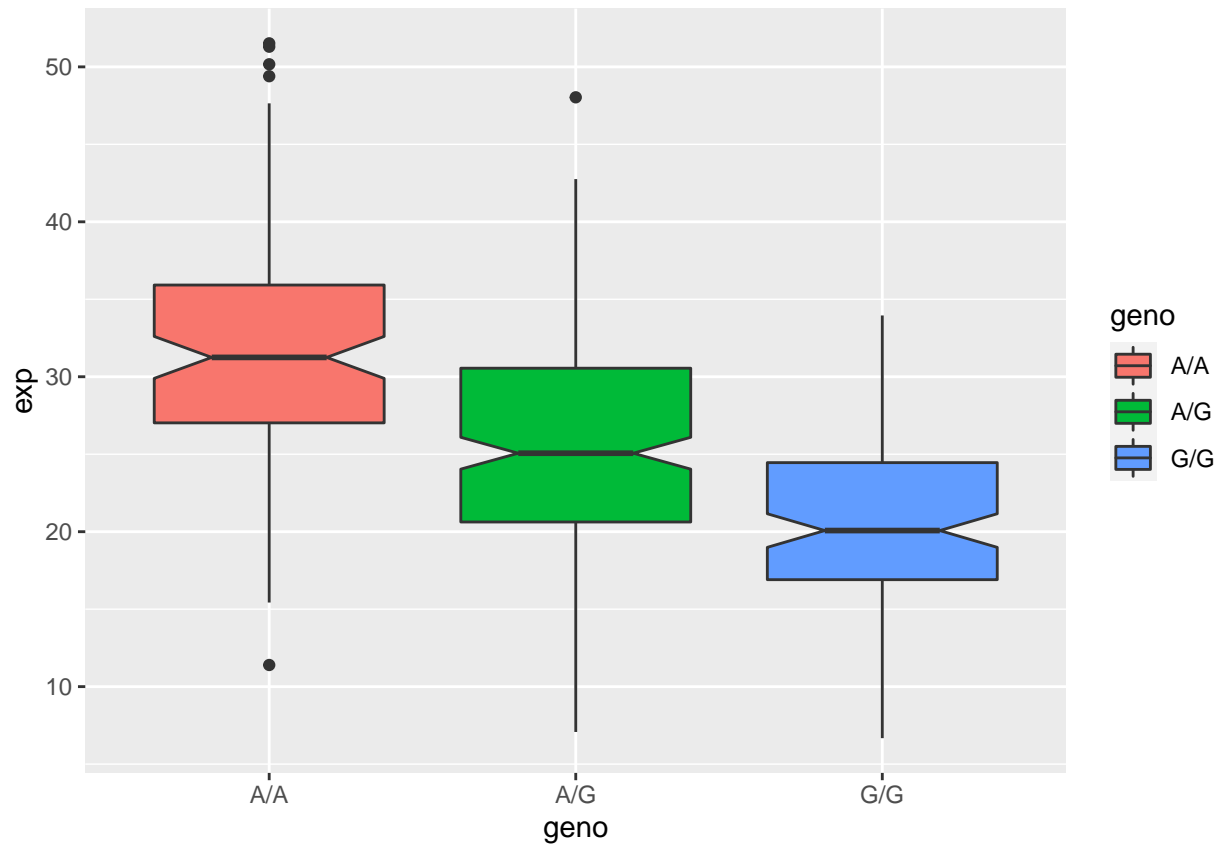
```
table(expr$geno)
```

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

```
library(ggplot2)
```

Lets make a box plot

```
library(ggplot2)
plot.expr <- ggplot(expr) + aes(x=geno, y=exp, fill=geno) +
  geom_boxplot(notch=TRUE)
plot.expr
```



Q14. 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 ORM DL3? Yes, G/G SNP does effect the expression of ORM DL3.