class19_bimm143

Hope (PID: A15652616)

12/5/2021

Section 4: Population Scale Analysis

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
```

We have 462 samples in the data.

What if we want to know the number of each type?

```
table(expr$geno)
```
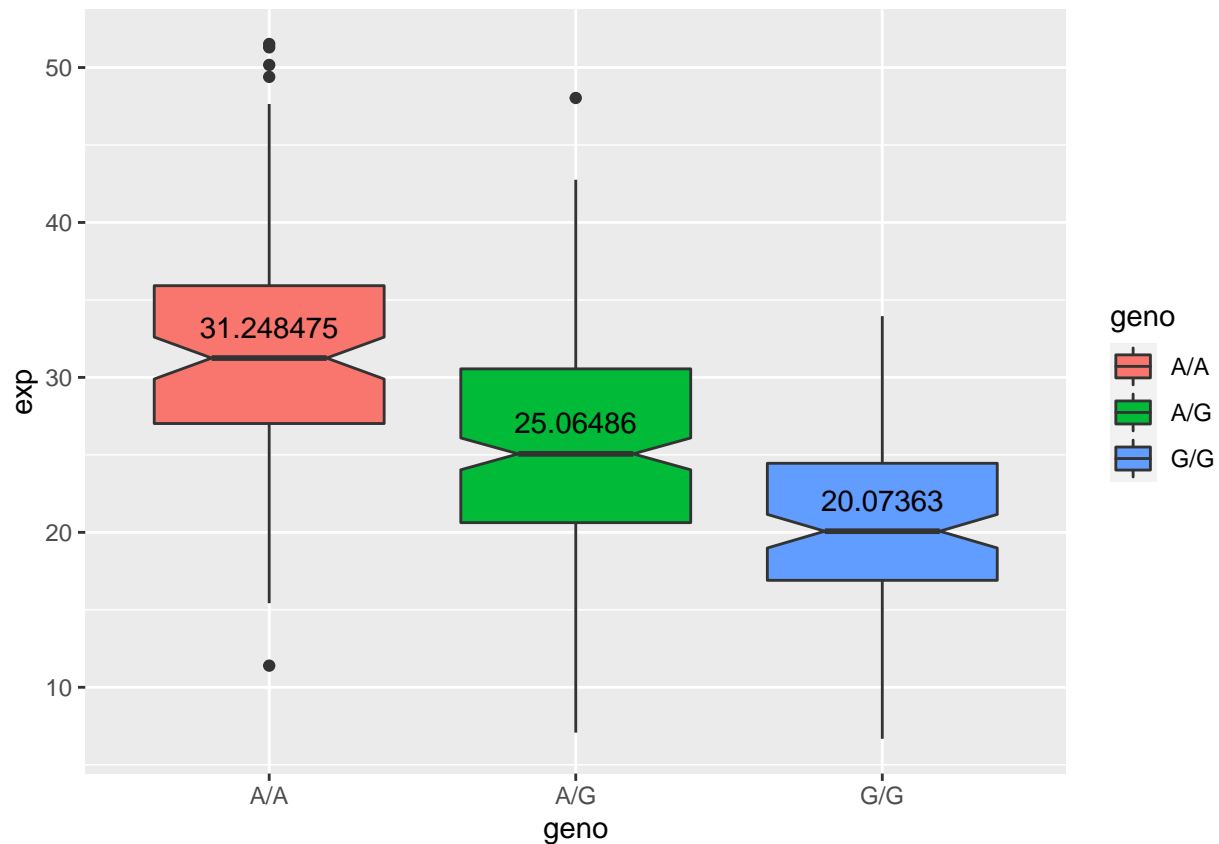
```
##
## A/A A/G G/G
## 108 233 121
```

Now we can see there are 108 A/A, 233 A/G, and 121 G/G

Now we want to make a summary figure of the data using ggplot:

```
library(ggplot2)
```

We are making a boxplot:

```
med <- aggregate(exp ~ geno, expr, median)
ggplot(expr) + aes(geno, exp, fill=geno) + geom_boxplot(notch = TRUE) + geom_text(data = med, aes(label
```


I made a table for the median values to show in the box plot. I am going to show the table:

```
med
```

```
##   geno      exp
## 1  A/A 31.24847
## 2  A/G 25.06486
## 3  G/G 20.07363
```

The medians were found above.

Question 14: 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 seems to be greater expression of ORMDL3 in A/A genotypes than G/G genotypes based on the median alone but the data overall in the boxplot mostly overlaps with each other so it's hard to say for sure.