lab10r

Rachel Kraft

2/19/2022

Section 1

5) Proportion of G/G in population

- download the CSV, read

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")

table(mxl$Genotype..forward.strand.)
```

```
##
## A|A A|G G|A G|G
##  22  21  12   9
```

```
table(mxl$Genotype..forward.strand.) / nrow(mxl) *100
```

```
##
##     A|A     A|G     G|A     G|G
## 34.3750 32.8125 18.7500 14.0625
```

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 ORMDL3 expression.

13) Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes

read dataset

```
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
```

how many samples of each genotype?

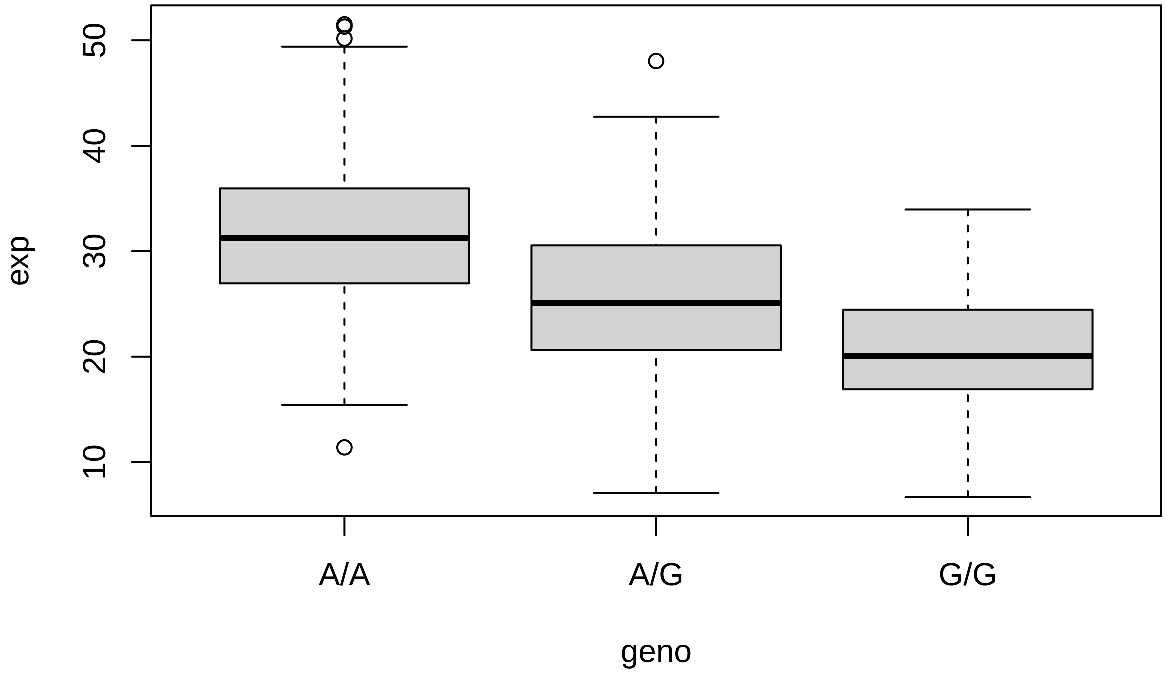
```
table(expr$geno)
```

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

108 samples for AA, 233 for AG, 121 for GG

to find median expression levels, save output of boxplot() to an R object

```
bp <- boxplot(exp~geno, data=expr)
```



```
# stats to give 5 number summary for the boxplot
bp$stats
```

```
##          [,1]     [,2]     [,3]
## [1,] 15.42908  7.07505  6.67482
## [2,] 26.95022 20.62572 16.90256
## [3,] 31.24847 25.06486 20.07363
## [4,] 35.95503 30.55183 24.45672
## [5,] 49.39612 42.75662 33.95602
```

```
# medians are in row 3
bp$stats[3,]
```

```
## [1] 31.24847 25.06486 20.07363
```

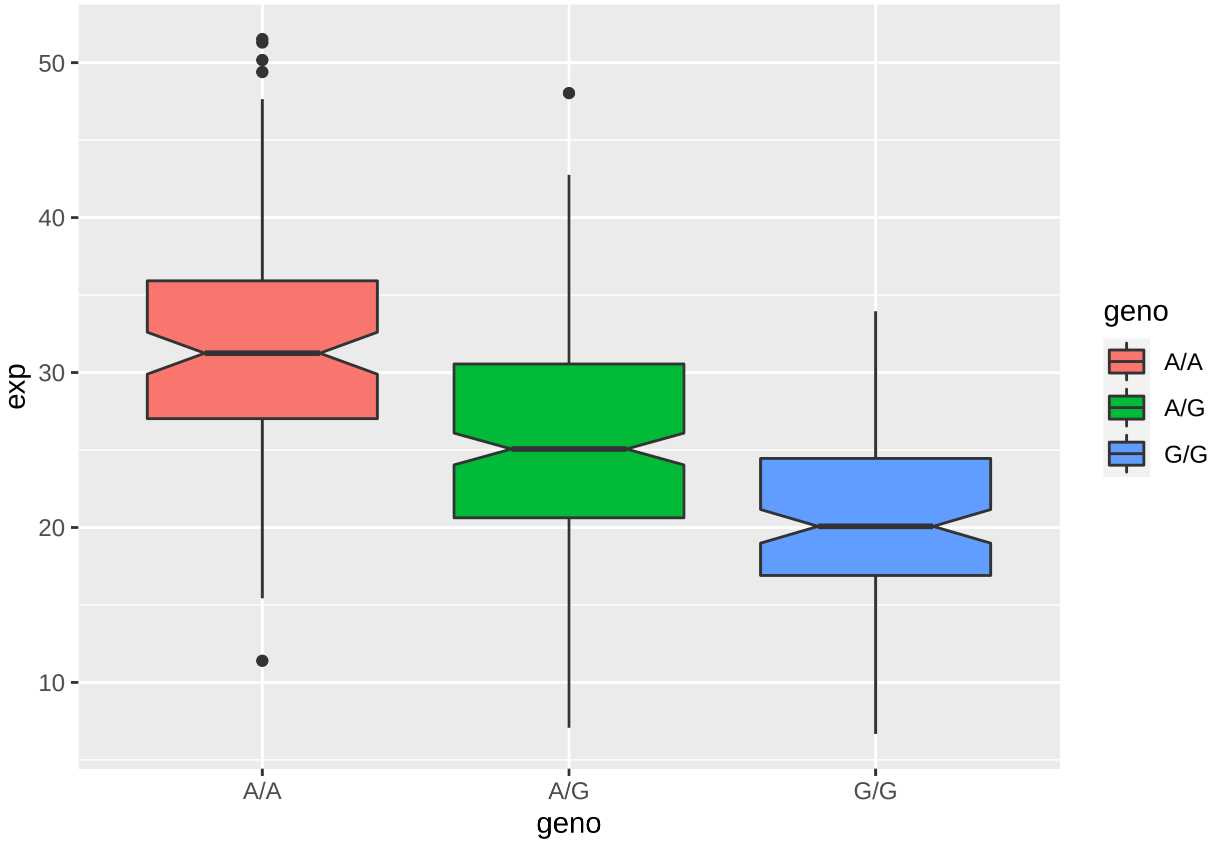
for AA (column 1), median expression level is 31.24847 for AG (column 2), median expression level is 25.06486 for GG (column 3), median expresison level is 20.07363

14) 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?

```
library(ggplot2)
```

boxplot for this data, genotype vs expression level to show the median expression level for each genotype

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



From this plot, there is definitely less expression of ORMDL3 in individuals with the homozygous GG genotype opposed to AA. Yes, the SNP decreases expression of ORMDL3.