Class12

Jessica Gao (PID: 16939806)

Section 1 Proportion of G/G in a population

Downloaded a csv file from Ensemble < https://useast.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=39895595;v=rs8067378;vdb=variation;vf=959672880#373531_tablePanel >

Here we read the csv file

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

```
  Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1                   NA19648 (F)                       A|A ALL, AMR, MXL      -
2                   NA19649 (M)                       G|G ALL, AMR, MXL      -
3                   NA19651 (F)                       A|A ALL, AMR, MXL      -
4                   NA19652 (M)                       G|G ALL, AMR, MXL      -
5                   NA19654 (F)                       G|G ALL, AMR, MXL      -
6                   NA19655 (M)                       A|G ALL, AMR, MXL      -
  Mother
1      -
2      -
3      -
4      -
5      -
6      -
```

```r
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. Population analysis

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.

```
expr <- read.table ("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG00000
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)
```
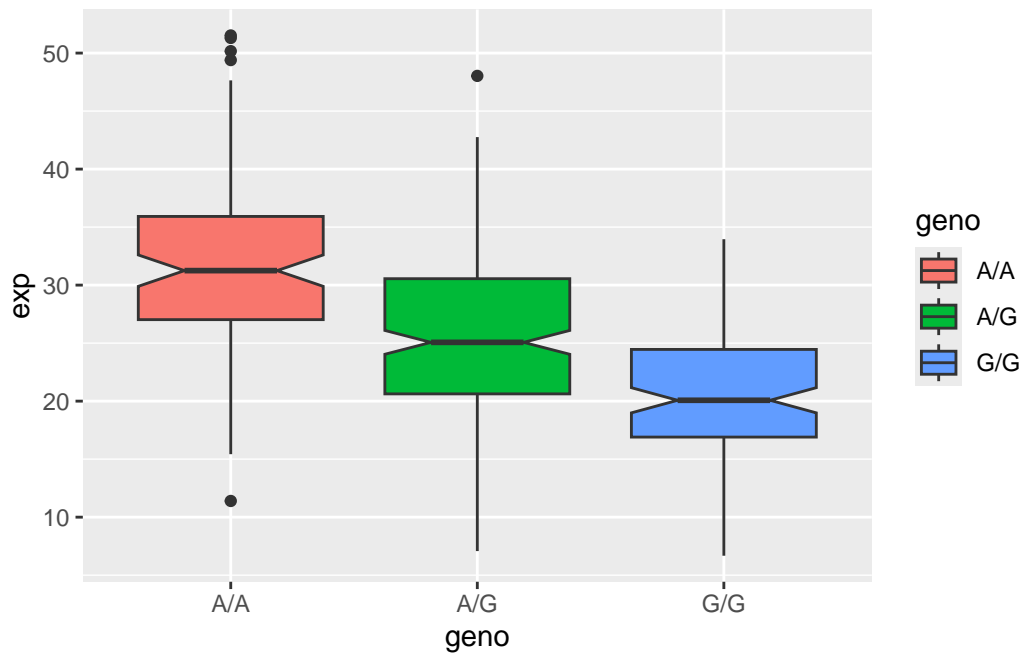
```
A/A A/G G/G
108 233 121
```

```
library(ggplot2)
```

Let's make a boxplot

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?

2

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


A|A seems to be related to an increase in the expression value, and G|G is related to a decrease of the expression value. Yes the SNP effect the expression of ORMDL3.