Week 6 RefSeq Tuesday Lab and Homework Questions

David Alvarez

Section 1. Proportion of G/G in a population

From url "https://useast.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=17:39894399-39895800;v=rs8067378;vdb=variation;vf=105535077#373531_tablePanel"

```
mxl <- read.csv("SampleGenotypes.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      -
```

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

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

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

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

Homework

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.

```r
expr <- read.table("Homework_genotypes.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
```

```r
# Number of total sample

nrow(expr)
```

```
[1] 462
```

```r
# Sample size by genotype
sample_size <- table(expr$geno)
sample_size
```

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

```
# Corresponding median expression levels for each of these genotypes

# By looking at the boxplot below, the median expression levels for A/A is around 31%, A/G
```

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?

```
library("ggplot2")
ggplot(expr) + aes(geno, exp, fill=geno) + geom_boxplot(notch=TRUE) + labs(x="Genotype", y
```
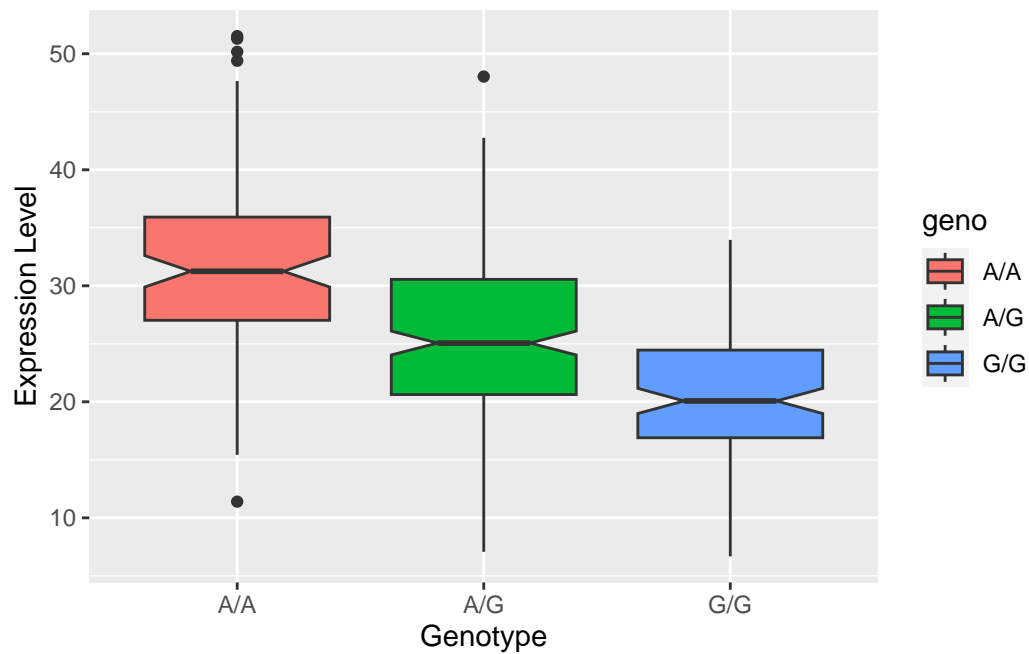


```
# From looking at the boxplot, it can be inferred that subjects with the G/G genotype have
```