

class19

Jibin (PID: A53300326)

2021/12/1

Section 1. Proportion of G/G in a population

```
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 -
```

```
table(mx1$Genotype..forward.strand.)/nrow(mx1)*100
```

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

The proportion is 14.0625%

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
```

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

```
##
## A|A A|G G|A G|G
## 25.27473 18.68132 26.37363 29.67033
```

This variant that is associated with childhood asthma is more frequent in the GBR population than MKL population.

Section 4: Population Scale 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("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
```

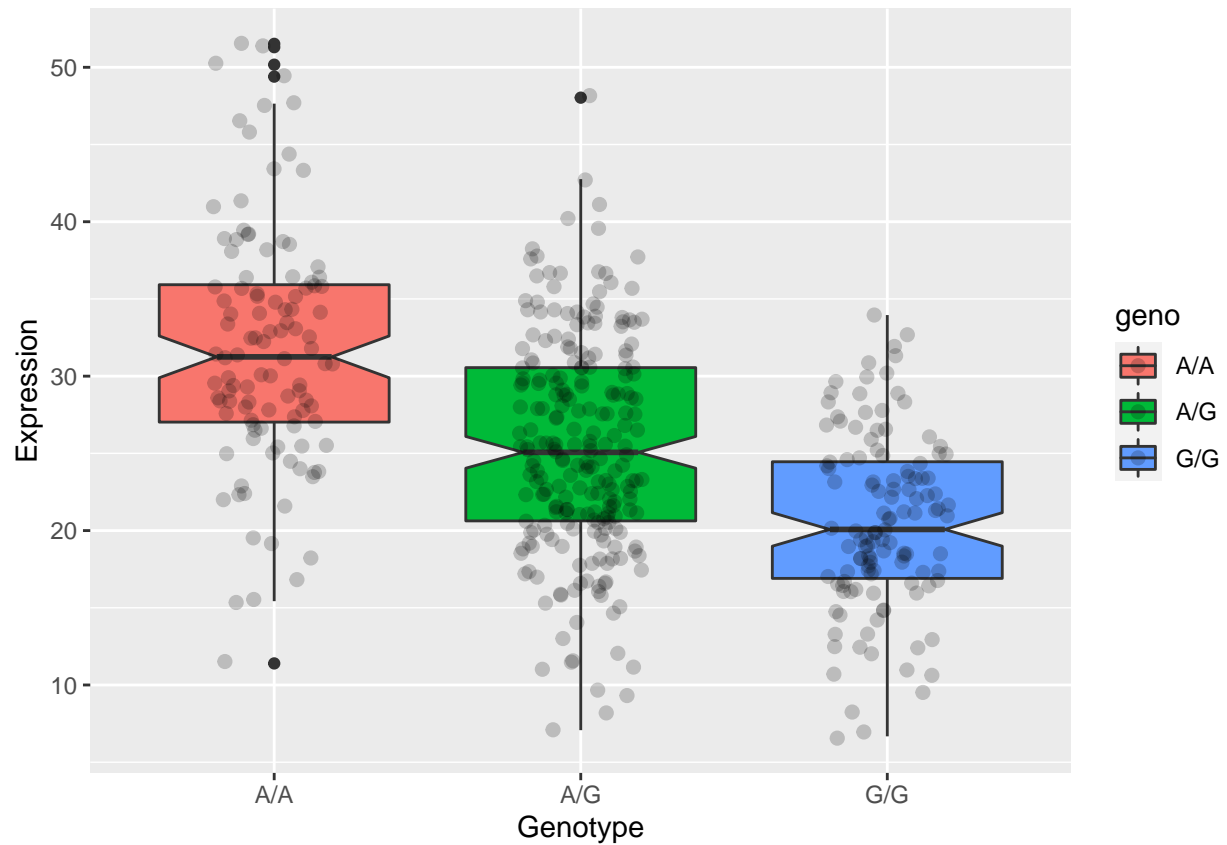
```
summary(expr)
```

```
##      sample      geno      exp
## Length:462      Length:462      Min.   : 6.675
## Class :character Class :character 1st Qu.:20.004
## Mode  :character Mode  :character Median  :25.116
##                                     Mean   :25.640
##                                     3rd Qu.:30.779
##                                     Max.   :51.518
```

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

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

```
library(ggplot2)
ggplot(expr)+ aes(geno, exp, fill=geno)+
  geom_boxplot(notch=T)+
  geom_point(position=position_jitter(h=0.15,w=0.2), alpha=0.2, size=2)+
  labs(x="Genotype", y= "Expression")
```



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

Yes. The G/G SNP dose show a lower expression of ORMDL3 than A/A phenotype.