class 11

Jenny Zhou

Identify genetic variants of interest

> Q5: What proportion of the Mexican Ancestry in Los Angeles sample population (MXL) are homozygous for the asthma associated SNP (G|G)?

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

```
  Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1                 NA19649 (M)                       G|G ALL, AMR, MXL      -
2                 NA19652 (M)                       G|G ALL, AMR, MXL      -
3                 NA19654 (F)                       G|G ALL, AMR, MXL      -
4                 NA19676 (M)                       G|G ALL, AMR, MXL      -
5                 NA19719 (F)                       G|G ALL, AMR, MXL      -
6                 NA19720 (M)                       G|G ALL, AMR, MXL      -
  Mother
1      -
2      -
3      -
4      -
5      -
6      -
```

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

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

Now let's look at a different population (British in England and Scotland)

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

```
  Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1                 HG00099 (F)                       G|G ALL, EUR, GBR      -
2                 HG00107 (M)                       G|G ALL, EUR, GBR      -
3                 HG00109 (M)                       G|G ALL, EUR, GBR      -
4                 HG00112 (M)                       G|G ALL, EUR, GBR      -
5                 HG00113 (M)                       G|G ALL, EUR, GBR      -
6                 HG00116 (M)                       G|G ALL, EUR, GBR      -
  Mother
1      -
2      -
3      -
4      -
5      -
6      -
```

```
signif(table(GBR$Genotype..forward.strand.)/ nrow(GBR) *100,3)
```

```
 A|A  A|G  G|A  G|G
25.3 18.7 26.4 29.7
```

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

Lets now dig into this further.

Population Scale Analysis

Determine whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on **ORMDL3** expression.

> 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
```

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

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

```
exp.med <- function(x) {
  # select one specific genotype
  which.geno <- expr$geno == x

  #filter the expression levels for that genotype
  exps <- expr$exp[which.geno]
  #calculate medium of filted expression levels, with 2 decimal points.
  round(median(exps), 2)
}
```

```
exp.med("G/G")
```

```
[1] 20.07
```

```
exp.med("A/A")
```

```
[1] 31.25
```

```
exp.med("A/G")
```

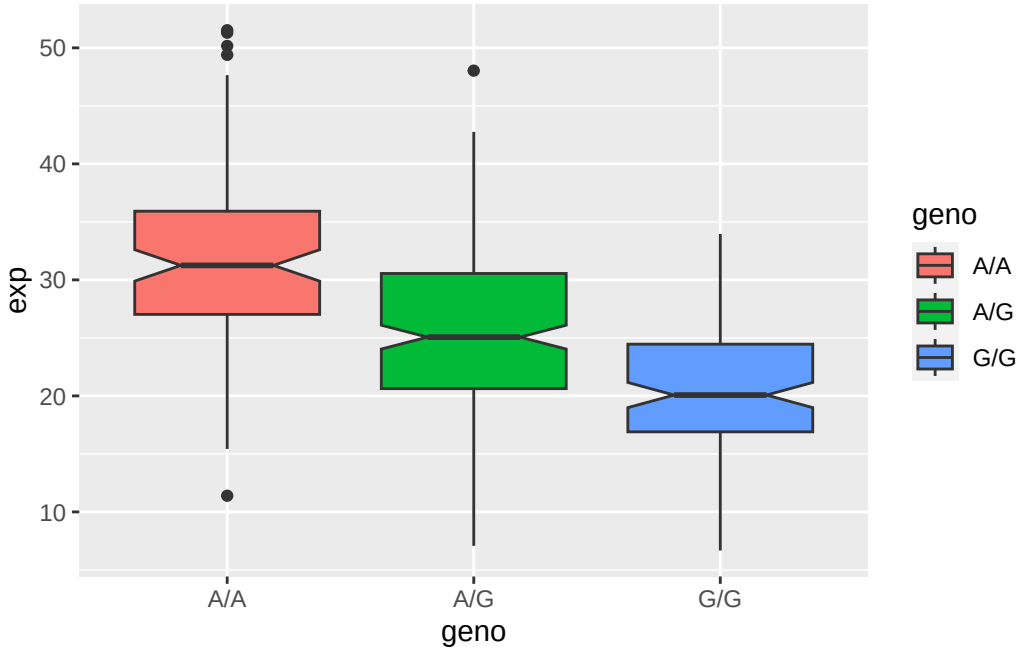
```
[1] 25.06
```

The median expression level for G/G genotype is 20.07. The median expression level for A/A genotype is 31.25. The median expression level for A/G genotype is 25.06.

> 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)
```



SNP effects the expression of ORMDL3. According to the boxplot, comparing to A/A genotype, G/G genotype leads to a lower expression of ORMDL3 by an around 1/3 fold. ORMDL3 expression with a A/G genotype is between that of A/A and G/G.