

Class 10: Genome Informatics

Soobin (PID:A15201229)

2/17/2022

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

```
# Read genotype file from Ensembl  
  
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")  
  
table(mxl$Genotype..forward.strand.) / nrow(mxl)
```

```
##  
##      A|A      A|G      G|A      G|G  
## 0.343750 0.328125 0.187500 0.140625
```

What about a different population? Here we take the British in England and Scotland (GBR)

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")  
  
table(gbr$Genotype..forward.strand.) / nrow(gbr)
```

```
##  
##      A|A      A|G      G|A      G|G  
## 0.2527473 0.1868132 0.2637363 0.2967033
```

Expression by Genotype Analysis

I want to read my RNA-Seq Expression results into R. This file is not a CSV but rather has fields separated by space.

```
expression <- read.table("rs8067378_ENSG00000172057.6.txt")  
head(expression)
```

```
##      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
```

Now we will look at the other genotypes.

```
table(expression$geno)
```

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

```
summary( expression[ expression$geno == "A/A", 3 ] )
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.  
##    11.40   27.02   31.25   31.82   35.92   51.52
```

```
summary( expression[ expression$geno == "A/G", 3 ] )
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.  
##     7.075  20.626  25.065  25.397  30.552  48.034
```

```
summary( expression[ expression$geno == "G/G", 3 ] )
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.  
##     6.675  16.903  20.074  20.594  24.457  33.956
```

Make a summary overview figure

Make a boxplot figure

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

