

## class10

Rachel Diao

2/18/2022

#In-class code

Find out how many occurrences of G|G there are in the those of Mexican ancestry in LA.

```
mxl <- read.csv('MXL asthma.csv')
table(mxl$Genotype..forward.strand.)/nrow(mxl)
```

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

Next do the same for British people in England and Scotland.

```
gbr <- read.csv('GBR asthma.csv')
table(gbr$Genotype..forward.strand.)/nrow(gbr)
```

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

```
exp <- read.csv('galaxy results.csv')
which(exp$FPKM > 0)
```

```
## [1] 544 545 547 548 570
```

```
exp$tracking_id[which(exp$FPKM > 0)]
```

```
## [1] "ZPBP2" "GSDMB" "GSDMA" "ORMDL3" "PSMD3"
```

## HOMEWORK

```
df <- read.table('rs8067378_ENSG00000172057.6.txt')

#To find out how many of each genotype
table(df$geno)
```

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

#Find median based on genotype
gg_median <- median(df$exp[which(df$geno == 'G/G')])
aa_median <- median(df$exp[which(df$geno == 'A/A')])
ag_median <- median(df$exp[which(df$geno == 'A/G')])
```

### Question 13

The sample size for each genotype is 108 A/A, 233 A/G, and 121 G/G. The median expression value for A/A is 31.248, G/G is 20.074, A/G is 25.065.

### Question 14

Make a boxplot!

```
library(ggplot2)
ggplot(df, aes(x=geno, y=exp, fill=geno)) +
  geom_boxplot(notch=TRUE, show.legend=FALSE, outlier.shape=NA) +
  geom_point(alpha=0.2, position = position_jitter(width=0.1), show.legend=FALSE) +
  xlab('Genotype') + ylab('Expression')
```

