

class10: Genome Informatics

Ari_Fon (PID: A15390446)

2/17/2022

#Examine 1000 Genome Data

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 Ensemble

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
table(mx1$Genotype..forward.strand.)/nrow(mx1)
```

```
##
##      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.

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

```
##      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
```

First try at this question. Is the mean expression different based in gneotype?

```
x$geno== "G/G"
```

```
## [1] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE
## [25] FALSE FALSE FALSE TRUE TRUE FALSE TRUE TRUE FALSE FALSE TRUE FALSE
## [37] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
## [49] TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE
## [61] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
## [73] TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
## [85] TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE FALSE
## [109] TRUE TRUE TRUE FALSE FALSE TRUE TRUE FALSE TRUE TRUE TRUE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE
## [133] FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE
## [157] FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
## [169] FALSE TRUE TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
## [193] TRUE TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
## [205] FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE
## [229] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## [241] TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
## [253] TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
## [265] FALSE FALSE TRUE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [277] FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE TRUE TRUE
## [289] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## [301] FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [313] FALSE TRUE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
## [325] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## [337] FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
## [349] FALSE FALSE TRUE FALSE FALSE FALSE TRUE TRUE TRUE FALSE FALSE FALSE
## [361] TRUE TRUE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE
## [373] TRUE FALSE TRUE TRUE FALSE TRUE TRUE TRUE TRUE FALSE TRUE FALSE
## [385] TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [421] TRUE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
## [433] FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [445] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
## [457] TRUE TRUE FALSE FALSE FALSE FALSE
```

```
summary(x[x$geno== "G/G", "exp"])
```

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

```
table(x$geno)
```

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

```
summary(x[x$geno== "A/G", "exp"])
```

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

```
summary(x[x$geno== "A/A", "exp"])
```

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

```
summary(x[x$geno== "G/G", "exp"])
```

```
##      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
```

```
library(ggplot2)
```

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

```
ggplot(x) + aes(geno,exp, fill=geno) + geom_boxplot(notch = TRUE)
```

