

Lab12

Zainub Darsot (PID: A16294217)

2023-11-11

Section 1

```
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(mxl$Genotype..forward.strand.)
```

```
##
## A|A A|G G|A G|G
## 22 21 12 9
```

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

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

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

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

Section 4: Population Scale Analysis

```
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
```

Total Size:

```
nrow(expr)
```

```
## [1] 462
```

Sample Size:

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

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

Median Expression Levels:

A/A

```
aa <- expr[which(expr$geno == "A/A"),]
head(aa)
```

```
##      sample geno      exp
## 3  HG00361  A/A 31.32628
## 4  HG00135  A/A 34.11169
## 6  NA11993  A/A 32.89721
## 8  NA18498  A/A 47.64556
## 13 NA20585  A/A 30.71355
## 15 HG00235  A/A 25.44983
```

```
median(aa$geno)
```

```
## Warning in mean.default(sort(x, partial = half + 0L:1L)[half + 0L:1L]):
## argument is not numeric or logical: returning NA
```

```
## [1] NA
```

A/G

```
ag <- expr[which(expr$geno == "A/G"),]
head(ag)
```

```
##      sample geno      exp
## 1  HG00367  A/G 28.96038
## 2  NA20768  A/G 20.24449
## 7  HG00256  A/G 31.48736
## 10 HG00115  A/G 33.85374
## 11 NA20806  A/G 16.29854
## 12 HG00278  A/G 19.73450
```

```
median(ag$geno)
```

```
## [1] "A/G"
```

G/G

```
gg <- expr[which(expr$geno == "G/G"),]
head(gg)
```

```
##      sample geno      exp
## 5  NA18870  G/G 18.25141
## 9  HG00327  G/G 17.67473
## 17 NA12546  G/G 18.55622
## 20 NA18488  G/G 23.10383
## 23 NA19214  G/G 30.94554
## 28 HG00112  G/G 21.14387
```

```
median(gg$geno)
```

```
## [1] "G/G"
```

```
library(ggplot2)
```

```
ggplot(expr) + aes(geno, exp, fill=geno) +
  geom_boxplot(notch=T)
```

