

Class 12: Genome Informatics

Liana Melikian (A16675734)

#Section 1. Proportion of G/G in a population Downloaded a CSV files from Ensemble Here
we read this CSV file

Q5.

```
mxl=read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")  
head(mx1)
```

	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(mx1$Genotype..forward.strand.)
```

A A	A G	G A	G G
22	21	12	9

```
table(mx1$Genotype..forward.strand.)/nrow(mx1)*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. I picked the TSI population.

```
tsi=read.csv("373537-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(tsi)
```

```

Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
1              NA20502 (F)              A|G ALL, EUR, TSI      -
2              NA20503 (F)              G|A ALL, EUR, TSI      -
3              NA20504 (F)              A|G ALL, EUR, TSI      -
4              NA20505 (F)              G|A ALL, EUR, TSI      -
5              NA20506 (F)              A|G ALL, EUR, TSI      -
6              NA20507 (F)              A|G ALL, EUR, TSI      -
Mother
1      -
2      -
3      -
4      -
5      -
6      -

```

```
table(tsi$Genotype..forward.strand.)
```

```

A|A A|G G|A G|G
26 27 31 23

```

```
round(table(tsi$Genotype..forward.strand.)/nrow(tsi)*100,2)
```

```

      A|A      A|G      G|A      G|G
24.30 25.23 28.97 21.50

```

This variant that is associated with childhood asthma is more frequent in the TSI population than the MXL population.

##Section 4: Population Scale Analysis

How many samples do we have?

```
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
```

```
nrow(expr)
```

```
[1] 462
```

Q13.

```
counts=table(expr$geno)
counts
```

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

```
medians=tapply(expr$exp, expr$geno,median)
medians
```

```
      A/A      A/G      G/G
31.24847 25.06486 20.07363
```

```
library(ggplot2)
```

Q14.

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

