Week 11: Genome informatics

Bryn Baxter (PID A69038039)

Section 1: Porportion on G/G in a population

Download CSV file from esemble

We need to read this CSV file

```
mxl <-read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378 (2).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
```

Now lets look at different population. I picked the GBR.

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
```

Find proportion of G|G

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

```
     A|A      A|G      G|A      G|G
25.27473 18.68132 26.37363 29.67033
```

This variant that is associated with childhood asthma is more frequent in the GBR population than MKL population.

Now lets dig into this further.

Section 4: Population Scale Analysis (HOMEWORK)

How many samples do we have?

```
expr <- read.table("Expression genotype results.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
```

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

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

```
library(ggplot2)
```

Lets make a box plot:

```
ggplot(expr)+ aes(geno,exp, fill=geno)+
  geom_boxplot(notch=T) +
  geom_jitter(width=0.2, alpha=0.2)+
  labs(x="Genotype", y="Expression")
```