Class 11: Genome Informatics

Andrew Sue

We are reading in the Mexican American Ancestry data for the SNPs associated with childhood asthma.

Q1. What are those 4 candidate SNPs?

rs12936231, rs8067378, rs9303277, and rs7216389

Q2. What three genes do these variants overlap or effect?

ZPBP2, IKZF3, GSDMB

Q3. What is the location of rs8067378 and what are the different alleles for rs8067378?

Location: Chromosome 17:39895095

Q4: Name at least 3 downstream genes for rs8067378?

Genes GSDMB, GSDMA and PSMD3.

```
mxn_data <- read.csv("Asthma_genotypes.csv", row.names = 1)
mxn_data
```

```
           Genotype..forward.strand. Population.s. Father Mother
NA19648 (F)                      A|A ALL, AMR, MXL      -      -
NA19649 (M)                      G|G ALL, AMR, MXL      -      -
NA19651 (F)                      A|A ALL, AMR, MXL      -      -
NA19652 (M)                      G|G ALL, AMR, MXL      -      -
NA19654 (F)                      G|G ALL, AMR, MXL      -      -
NA19655 (M)                      A|G ALL, AMR, MXL      -      -
NA19657 (F)                      A|G ALL, AMR, MXL      -      -
NA19658 (M)                      A|A ALL, AMR, MXL      -      -
NA19661 (M)                      A|G ALL, AMR, MXL      -      -
```

```
NA19663 (F)                    A|A ALL, AMR, MXL        -      -
NA19664 (M)                    G|A ALL, AMR, MXL        -      -
NA19669 (F)                    A|A ALL, AMR, MXL        -      -
NA19670 (M)                    A|A ALL, AMR, MXL        -      -
NA19676 (M)                    G|G ALL, AMR, MXL        -      -
NA19678 (F)                    A|A ALL, AMR, MXL        -      -
NA19679 (M)                    A|G ALL, AMR, MXL        -      -
NA19681 (F)                    A|G ALL, AMR, MXL        -      -
NA19682 (M)                    A|G ALL, AMR, MXL        -      -
NA19684 (F)                    A|G ALL, AMR, MXL        -      -
NA19716 (F)                    G|A ALL, AMR, MXL        -      -
NA19717 (M)                    A|G ALL, AMR, MXL        -      -
NA19719 (F)                    G|G ALL, AMR, MXL        -      -
NA19720 (M)                    G|G ALL, AMR, MXL        -      -
NA19722 (F)                    G|A ALL, AMR, MXL        -      -
NA19723 (M)                    G|G ALL, AMR, MXL        -      -
NA19725 (F)                    A|G ALL, AMR, MXL        -      -
NA19726 (M)                    A|A ALL, AMR, MXL        -      -
NA19728 (F)                    A|A ALL, AMR, MXL        -      -
NA19729 (M)                    A|G ALL, AMR, MXL        -      -
NA19731 (F)                    A|A ALL, AMR, MXL        -      -
NA19732 (M)                    A|G ALL, AMR, MXL        -      -
NA19734 (F)                    G|A ALL, AMR, MXL        -      -
NA19735 (M)                    G|G ALL, AMR, MXL        -      -
NA19740 (F)                    A|A ALL, AMR, MXL        -      -
NA19741 (M)                    A|A ALL, AMR, MXL        -      -
NA19746 (F)                    A|A ALL, AMR, MXL        -      -
NA19747 (M)                    G|A ALL, AMR, MXL        -      -
NA19749 (F)                    A|G ALL, AMR, MXL        -      -
NA19750 (M)                    A|G ALL, AMR, MXL        -      -
NA19752 (F)                    A|G ALL, AMR, MXL        -      -
NA19755 (F)                    A|A ALL, AMR, MXL        -      -
NA19756 (M)                    G|A ALL, AMR, MXL        -      -
NA19758 (F)                    A|G ALL, AMR, MXL        -      -
NA19759 (M)                    G|A ALL, AMR, MXL        -      -
NA19761 (F)                    G|A ALL, AMR, MXL        -      -
NA19762 (M)                    A|A ALL, AMR, MXL        -      -
NA19764 (F)                    A|A ALL, AMR, MXL        -      -
NA19770 (F)                    A|G ALL, AMR, MXL        -      -
NA19771 (M)                    A|A ALL, AMR, MXL        -      -
NA19773 (F)                    A|A ALL, AMR, MXL        -      -
NA19774 (M)                    A|G ALL, AMR, MXL        -      -
NA19776 (F)                    A|G ALL, AMR, MXL        -      -
```

```
NA19777 (M)                          A|A ALL, AMR, MXL      -      -
NA19779 (F)                          G|A ALL, AMR, MXL      -      -
NA19780 (M)                          A|A ALL, AMR, MXL      -      -
NA19782 (F)                          G|A ALL, AMR, MXL      -      -
NA19783 (M)                          A|G ALL, AMR, MXL      -      -
NA19785 (F)                          A|A ALL, AMR, MXL      -      -
NA19786 (M)                          G|A ALL, AMR, MXL      -      -
NA19788 (F)                          A|G ALL, AMR, MXL      -      -
NA19789 (M)                          G|G ALL, AMR, MXL      -      -
NA19792 (M)                          A|A ALL, AMR, MXL      -      -
NA19794 (F)                          G|A ALL, AMR, MXL      -      -
NA19795 (M)                          A|G ALL, AMR, MXL      -      -
```

Q5. What proportion of the Mexican Ancestry in Los Angeles sample population (MXL) are homozygous for the asthma associated SNP (G|G)?

14% G|G phenotype.

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

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

Q6. Back on the ENSEMBLE page, use the "search for a sample" field above to find the particular sample HG00109. This is a male from the GBR population group. What is the genotype for this sample?

The genotype of this sample is G/G

We are reading in now the GBR dataset and finding the HG00109 sample.

```
gbr <- read.csv("Asthma_genotypes_gbr.csv", row.names = 1)

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

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

To compare the gene expression of the entire population, we will take the population data.

3

```r
expr <- read.table("gene_expression.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
```

```r
nrow(expr)
```

```
[1] 462
```

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

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

Lets make a summary plot (boxplot) of the data

```r
library(ggplot2)

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