Lab 19: Galaxy

Nicholas Do (PID: 15053002)

12/5/2021

#Section 1. Proportion of G/G in a population

> Q1: What are those 4 candidate SNPs?

rs12936231, rs8067378, rs9303277, and rs7216389

> Q2: What three genes do these variants overlap or effect?

ZPBP2, IKZF3, and GSDMB

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

Chromosome 17: 39,894,595-39,895,595

> Q4: Name at least 3 downstream genes for rs8067378?

GSDMB, CSF3, PSMD3

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

```
data <- read.csv("373530-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378 (2).csv")
data
```

```
##    Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
## 1                  HG03052 (F)                       A|G AFR, ALL, MSL      -
## 2                  HG03054 (M)                       A|G AFR, ALL, MSL      -
## 3                  HG03055 (F)                       A|G AFR, ALL, MSL      -
## 4                  HG03057 (M)                       A|A AFR, ALL, MSL      -
## 5                  HG03058 (F)                       G|A AFR, ALL, MSL      -
## 6                  HG03060 (M)                       A|A AFR, ALL, MSL      -
## 7                  HG03061 (F)                       G|A AFR, ALL, MSL      -
## 8                  HG03063 (M)                       G|A AFR, ALL, MSL      -
## 9                  HG03064 (F)                       A|A AFR, ALL, MSL      -
## 10                 HG03066 (M)                       A|G AFR, ALL, MSL      -
## 11                 HG03069 (M)                       A|G AFR, ALL, MSL      -
## 12                 HG03072 (M)                       A|G AFR, ALL, MSL      -
## 13                 HG03073 (F)                       G|A AFR, ALL, MSL      -
## 14                 HG03074 (M)                       G|G AFR, ALL, MSL      -
## 15                  HG03077 (M)                       G|A AFR, ALL, MSL      -
## 16                  HG03078 (M)                       G|G AFR, ALL, MSL      -
## 17                  HG03079 (F)                       G|G AFR, ALL, MSL      -
## 18                  HG03081 (M)                       A|G AFR, ALL, MSL      -
## 19                  HG03082 (F)                       G|G AFR, ALL, MSL      -
## 20                  HG03084 (M)                       G|G AFR, ALL, MSL      -
## 21                  HG03085 (F)                       G|G AFR, ALL, MSL      -
## 22                  HG03086 (F)                       G|G AFR, ALL, MSL      -
## 23                  HG03088 (F)                       G|G AFR, ALL, MSL      -
## 24                  HG03091 (F)                       G|A AFR, ALL, MSL      -
## 25                  HG03095 (F)                       G|A AFR, ALL, MSL      -
## 26                  HG03096 (M)                       G|G AFR, ALL, MSL      -
## 27                  HG03097 (F)                       G|G AFR, ALL, MSL      -
## 28                  HG03209 (M)                       A|A AFR, ALL, MSL      -
## 29                  HG03212 (F)                       A|G AFR, ALL, MSL      -
## 30                  HG03224 (M)                       G|G AFR, ALL, MSL      -
## 31                  HG03225 (M)                       G|G AFR, ALL, MSL      -
## 32                  HG03376 (M)                       G|G AFR, ALL, MSL      -
## 33                  HG03378 (F)                       G|A AFR, ALL, MSL      -
## 34                  HG03380 (F)                       G|A AFR, ALL, MSL      -
## 35                  HG03382 (M)                       G|A AFR, ALL, MSL      -
## 36                  HG03385 (M)                       A|A AFR, ALL, MSL      -
## 37                  HG03388 (M)                       G|A AFR, ALL, MSL      -
## 38                  HG03391 (M)                       A|A AFR, ALL, MSL      -
## 39                  HG03394 (M)                       A|G AFR, ALL, MSL      -
## 40                  HG03397 (M)                       G|G AFR, ALL, MSL      -
## 41                  HG03401 (F)                       G|A AFR, ALL, MSL      -
## 42                  HG03410 (F)                       A|G AFR, ALL, MSL      -
## 43                  HG03419 (F)                       G|G AFR, ALL, MSL      -
## 44                  HG03428 (F)                       G|G AFR, ALL, MSL      -
## 45                  HG03432 (M)                       G|G AFR, ALL, MSL      -
## 46                  HG03433 (M)                       G|G AFR, ALL, MSL      -
## 47                  HG03436 (M)                       G|G AFR, ALL, MSL      -
## 48                  HG03437 (F)                       A|A AFR, ALL, MSL      -
## 49                  HG03439 (M)                       A|G AFR, ALL, MSL      -
## 50                  HG03442 (M)                       G|G AFR, ALL, MSL      -
## 51                  HG03445 (M)                       G|G AFR, ALL, MSL      -
## 52                  HG03446 (F)                       G|A AFR, ALL, MSL      -
## 53                  HG03449 (F)                       A|A AFR, ALL, MSL      -
## 54                  HG03451 (M)                       A|G AFR, ALL, MSL      -
## 55                  HG03452 (F)                       A|G AFR, ALL, MSL      -
## 56                  HG03455 (F)                       A|A AFR, ALL, MSL      -
## 57                  HG03457 (M)                       G|G AFR, ALL, MSL      -
## 58                  HG03458 (F)                       A|G AFR, ALL, MSL      -
## 59                  HG03460 (M)                       A|A AFR, ALL, MSL      -
## 60                  HG03461 (F)                       A|G AFR, ALL, MSL      -
## 61                  HG03464 (F)                       A|G AFR, ALL, MSL      -
## 62                  HG03469 (M)                       A|A AFR, ALL, MSL      -
## 63                  HG03470 (F)                       G|G AFR, ALL, MSL      -
## 64                  HG03472 (M)                       A|G AFR, ALL, MSL      -
## 65                  HG03473 (F)                       G|A AFR, ALL, MSL      -
## 66                  HG03476 (F)                       G|A AFR, ALL, MSL      -
## 67                  HG03478 (M)                       A|G AFR, ALL, MSL      -
## 68                  HG03479 (F)                       A|A AFR, ALL, MSL      -
## 69                  HG03484 (M)                       G|G AFR, ALL, MSL      -
## 70                  HG03485 (F)                       G|G AFR, ALL, MSL      -
## 71                  HG03547 (M)                       G|G AFR, ALL, MSL      -
## 72                  HG03548 (F)                       G|A AFR, ALL, MSL      -
## 73                  HG03556 (M)                       G|A AFR, ALL, MSL      -
## 74                  HG03557 (F)                       A|A AFR, ALL, MSL      -
## 75                  HG03558 (F)                       G|A AFR, ALL, MSL      -
## 76                  HG03559 (M)                       G|G AFR, ALL, MSL      -
## 77                  HG03563 (F)                       G|A AFR, ALL, MSL      -
## 78                  HG03565 (M)                       A|A AFR, ALL, MSL      -
## 79                  HG03567 (F)                       G|G AFR, ALL, MSL      -
## 80                  HG03571 (M)                       A|A AFR, ALL, MSL      -
## 81                  HG03572 (F)                       G|A AFR, ALL, MSL      -
## 82                  HG03575 (F)                       G|A AFR, ALL, MSL      -
## 83                  HG03577 (M)                       G|A AFR, ALL, MSL      -
## 84                  HG03578 (F)                       A|A AFR, ALL, MSL      -
## 85                  HG03583 (F)                       G|G AFR, ALL, MSL      -
##    Mother
## 1       -
## 2       -
## 3       -
## 4       -
## 5       -
## 6       -
## 7       -
## 8       -
## 9       -
## 10      -
## 11      -
## 12      -
## 13      -
## 14      -
## 15      -
## 16      -
## 17      -
## 18      -
## 19      -
## 20      -
## 21      -
## 22      -
## 23      -
## 24      -
## 25      -
## 26      -
## 27      -
## 28      -
## 29      -
## 30      -
## 31      -
## 32      -
## 33      -
## 34      -
## 35      -
## 36      -
## 37       -
## 38       -
## 39       -
## 40       -
## 41       -
## 42       -
## 43       -
## 44       -
## 45       -
## 46       -
## 47       -
## 48       -
## 49       -
## 50       -
## 51       -
## 52       -
## 53       -
## 54       -
## 55       -
## 56       -
## 57       -
## 58       -
## 59       -
## 60       -
## 61       -
## 62       -
## 63       -
## 64       -
## 65       -
## 66       -
## 67       -
## 68       -
## 69       -
## 70       -
## 71       -
## 72       -
## 73       -
## 74       -
## 75       -
## 76       -
## 77       -
## 78       -
## 79       -
## 80       -
## 81       -
## 82       -
## 83       -
## 84       -
## 85       -
```

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

```
##
## A|A A|G G|A G|G
##  16  18  22  29
```

There are 29 out of the 85 total that are homozygous for the asthma associated SNP, for a proportion of 34.12%

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

```
##
##      A|A      A|G      G|A      G|G
## 18.82353 21.17647 25.88235 34.11765
```

> 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?

This sample's genotype is G|G

#Section 2: Initial RNA-Seq analysis

> Q7: How many sequences are there in the first file? What is the file size and format of the data? Make sure the format is fastqsanger here!

There are 3,863 sequences in the HG00109_1.fastq file.

> Q8: What is the GC content and sequence length of the second fastq file?

HG00109_2 has a %GC of 54%

> Q9: How about per base sequence quality? Does any base have a mean quality score below 20?

The per base sequence quality remains within the green region, in the range of 30 - 40, which is very good.

#Section 3: Mapping RNA-Seq reads to genome

> Q10: Where are most the accepted hits located?

A lot of the accepted hits are located around the 38,120,000 to 38,160,000 region.

> Q11: Following Q10, is there any interesting gene around that area?

There is another large group of accepted htis around the 38,060,000 to 38,080,000 region as well.

> Q12: Cufflinks again produces multiple output files that you can inspect from your right-handside galaxy history. From the "gene expression" output, what is the FPKM for the ORMDL3 gene? What are the other genes with above zero FPKM values?

The FPKM for ORMDL3 is 136853.

The other genes are:

ZPBP2, 4613.49 GSDMB, 26366.3 GSDMA, 133.634 PSMD3, 299021