class 11: RNAseq and Galaxy

Jimmi Nguyen

RNAseq and Galaxy

Q1: What are those 4 candidate SNPs?

The 4 candidate SNPs rs12936231, rs8067378, rs9303277, and rs7216389

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

These variants overlap with GSDMB, ZPBP2, IKZF3.

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

The location/alleles are Chromosome 17: 39,895,045-39,895,145

Q4: Name at least 3 downstream genes for rs8067378?

Three down stream genes are PSMD3, NR1D1, CDC6.

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

9/64

```
[1] 0.140625
```

The portion of MXL homozygous for asthma associated SNP is 0.141

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 for this sample is G|G.

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 first file. The file size is 741.9 KB with the format of fastqsanger.

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

The GC content was 54% and the sequence lengths range between 50 minimum -75 maximum

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

mean per base sequence qualities are 33 at the lowest and 40 at the highest. None of the bases have a mean quality score of below 20.

Q10: Where are most the accepted hits located?

The most accepted hits are located on chromosome 17.

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

There are interesting genes of ZPBP2, GSDMB, LRRC3C.

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 gene is 136853. The genes with zero FPKM are ZPBP2, GSDMA, GSDMB.