Class 12: Introduction to Genome Informatics Lab

Noel Lim (PID: A17652474)

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?

Located on Chromosome 17:39895095(forward strand) Alleles: A/C/G // Ancestral: G // MAF: 0.43 (G)

Q4. Name at least 3 downstream genes for rs8067378?

RARA, CASC3, WIPF3

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

14.06%

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?

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!

3,863 sequences 741.9 KB

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

Sequence length is 3863 GC content is 54%

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

The per base sequence quality is high (all are in the green regions), and there is nothing below 20.

Q10. Where are most the accepted hits located?

Chromosome 17, q arm

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

PSMD3

Q12. Cufflinks again produces multiple output files that you can inspect from your right-hand-side 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?

128189 for FPKM of ORMDL3 Other genes are: LRRC3C, GSDMB