

class11ws

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

##Identify Genetic Variants of Interest

Q1: What are those 4 candidate SNPs?

rs 12936231, rs8067378, rs9303277, rs7216389

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

ZPBP2 >Q3: What is the location of rs8067378 and what are the different alleles for rs8067378?

A/G, Ancestral:G, MAF:0.43 (G). Chromosome 17;39895095

Q4: Name at least 3 downstream genes for rs8067378?

zpbp2-201, ccds11353.2, ccds11352.1

Proportion of G?G in a population

read in csv

```
mx1 <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378 (2).csv")
head(mx1)
```

	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(mx1$Genotype..forward.strand.)
```

A A	A G	G A	G G
22	21	12	9

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

```
table(mx1$Genotype..forward.strand.)/nrow(mx1)*100
```

A A	A G	G A	G G
34.3750	32.8125	18.7500	14.0625

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?

homozygous

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!

3863 sequences,

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

53% > Q9: How about per base sequence quality? Does any base have a mean quality score below 20?

20 is your threshold, no

Q10: Where are most the accepted hits located?

PSMD3, ORMDL3 > Q11: Following Q10, is there any interesting gene around that area?

no > 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, GSDMB, ZBPB2