BIM3007 Computational Genomics and Proteomics (2022-2023 Term 1)

Assignment #1

Deadline: Please submit your assignment through the blackboard e-learning system before **23:59:59 16th Oct**, 2022 (Delayed submission is not allowed for any reasons)

Percentage: 10%

Purpose: to enhance the learning outcomes for the topics in "Human genome project", "Evolution of Sequencing technology", "Exome-sequencing data analysis" and "Genome-wide association study".

1. [**Human genome project**] (**20 points**)
    1.1 Please describe the size of human genome, the updated number of protein-coding genes identified in the human genome, as well as the average number of genetic variants each Chinese people could have. (10 points)
    1.2 Please illustrate a gene structure, describe potential impact of genetic variants on a gene product. (10 points)

2. [**Evolution of Sequencing technology**] (**20 points**)
    2.1 Please describe the principle of 1st, 2nd and 3rd generation sequencing technologies, respectively. (10 points)
    2.2 Please describe the pros and cons of these sequencing technologies. (10 points)

3. [**Exome-sequencing data analysis**] (**15 points**)
    3.1 Please describe the principle and main goal of Exome-sequencing technology. (5 points)
    3.2 Please describe the process of Exome-sequencing data analysis and the related software (10 points)

4. [**Exome-sequencing data analysis**] (**25 points**)

    Given the following pedigree,
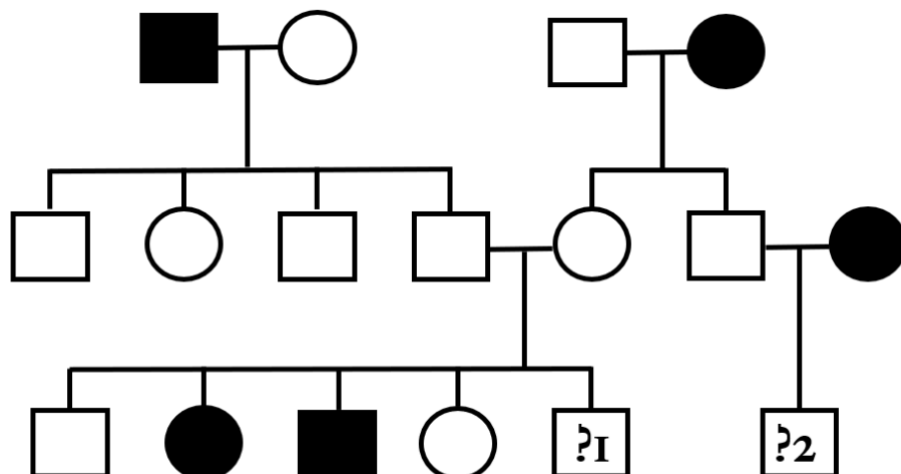
    4.1 What type of inheritance pattern would be the most likely explanation for the above pedigree? (5 points)
    4.2 Based on the answer to 4.1, what is the probability that the two new-born males marked

with ?1 and ?2 will be affected by the condition? (5 points)

4.3 Please design a study to identify the pathogenic variant for the condition and describe the criteria you used to define the pathogenic variant? (15 points)

5. [**Genome-wide association study**] (**20 points**)

Given the genotype frequency of SNP1 (A/a) and SNP2 (B/b) in a population as shown below:

SNP2 (B/b)

SNP1 (A/a)	BB	Bb	bb
AA	23	51	34
Aa	37	48	35
aa	13	19	11

5.1 Please estimate the frequencies of the 4 haplotypes: AB, Ab, aB, ab, using EM algorithm (5 iterations). (10 points)

5.2 Please describe whether the two SNPs are in linkage disequilibrium and why? (10 points)