**BIM3007 2022 Term 1**

**Assignment #2**

**Deadline: 23:59:59 13th Nov.**, 2022 (Delayed submission is not allowed for any reason)

**Percentage:** 10%

**Purpose:** to enhance the learning outcomes for the topics in “Genome Project”, “Noninvasive prenatal testing (NIPT)” and “Whole exome sequencing (WES)”.

**1 [Genome Project] (40 points)**

Please perform the ‘Genome Assembly, Genome Polishing and Genome Evaluation’ on the *Escherichia coli* genome (Genome size = 4.8 Mb) based on Nanopore reads (<https://figshare.com/ndownloader/files/21623145>) and Illumina pair-end reads (<ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR987/006/SRR9873306/SRR9873306_1.fastq.gz> and <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR987/006/SRR9873306/SRR9873306_2.fastq.gz>; you can download them by using the command ‘**wget**’ n

**2 [Noninvasive prenatal testing (NIPT)] (30 points)**

Please compute and give the results of the standard z-scores, NCV scores and regression based z-scores (RBZ) on chromosome 13, 18 and 21 based on the test dataset (<https://github.com/molgenis/NIPTeR/raw/master/test_samples/Trisomy21.rds>) and the control dataset (<https://github.com/molgenis/NIPTeR/raw/master/test_samples/NIPTeR_cleaned_87_controls.rds>). Next, you need to justify whether this test sample has the disease (if it suffers the disease, please provide the type of disease), and give the reason.

3 **[Whole exome sequencing (WES)] (30 points)**

Given pair-end reads (test\_1.fastq and test\_2.fastq) and structure variation reference databases (Mills\_and\_1000G\_gold\_standard.indels.hg38.vcf, dbSNP\_138.hg19.vcf and Homo\_sapiens\_assembly38.known\_indels.vcf) in [Server IP:10.21.53.251; Account: BIM3007; Password: BIM3007], please perform variant calling for chromosome 17 and provide the analysis process.