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

Assignment #3

Note: Please write down your name and student ID when you submit the assignment to the blackboard e-learning system.

Deadline: Please submit your assignment through the system before **23:59:59 4th Dec**, 2022 (Delayed submission is not allowed for any reasons)

Percentage: 10%

Purpose: to enhance the learning outcomes for the topics in "Association tests", "Quality control" and "Genome-wide association study".

Please download the data for this assignment through:

https://connecthkuhk-my.sharepoint.com/:f:/g/personal/yfwangbm_connect_hku_hk/ElBy4U1nQl1JpGHn0J9wvjoBtppjv8e45sxBJIuOmKkY3Q

(1) How many samples are included in the dataset? (10 points)
(2) How many variants are included in the dataset? (10 points)
(3) Please perform quality control for samples and record the details in each step. (20 points)
(4) Please perform quality control for genetic variants and record the details in each step. (20 points)
(5) Please perform association tests for the QCed data and illustrate the results using Manhattan plot. (30 points)
(6) Please show the QQ plot for the association results and calculate the inflation factor. (10 points)