

2 Bioinformatics (pl219)

- (a) You are a data scientist working at a hospital. A former in-patient claims to have been infected with HIV during their previous stay in the hospital. You have access to blood samples of a number of patients who were hospitalised at the same time as the claimant.
- (i) Describe how you would investigate the claim. [5 marks]
- (ii) Discuss how to evaluate the robustness of your finding. [5 marks]
- (b) You are given a table of gene expression data for different patients and controls (healthy people). The rows represent different genes, columns represent a group of patients or controls. Discuss an algorithm to identify the most important genes involved in the disease, its complexity and limitations. [5 marks]
- (c) A new deep-sea animal species is captured by a British expedition in the Mariana Trench, the deepest trench in the world, and a DNA sample is successfully sequenced. Species in extreme environment have usually adapted by accumulating large numbers of genomic rearrangements and mutations with respect to ancestor species. Discuss DNA coverage and reads length for robust genome assembly in the case the genome of the new species contains large number of repeated genomic regions. [5 marks]