UNIVERSITY OF CAMBRIDGE

M. PHIL. IN COMPUTATIONAL BIOLOGY

Friday, 12 May, 2017 2:00 pm to 4:00 pm

COMPUTATIONAL BIOLOGY

*Attempt **ALL** questions.*

*There are **THREE** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS

Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS

Calculator - students are permitted to bring an approved calculator.

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Computational Neuroscience

(a) Describe briefly the following three approaches to modelling neurons:

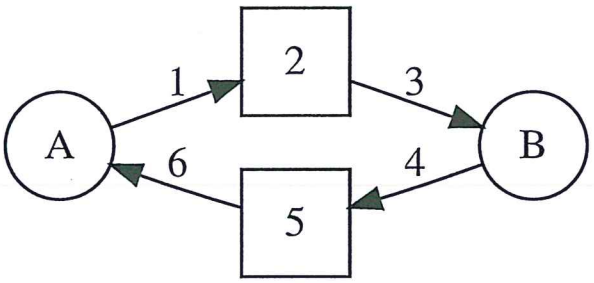
(i) Hodgkin-Huxley conductance-based model.

(ii) Integrate and fire model.

(iii) Firing-rate based model.

Name one advantage of each style of model. [60%]

(b) Explain each of the six numbered elements of the diagram below in the dynamic clamp. Explain how the dynamic clamp can be used to make virtual connections between neurons. Why is this useful? [25%]



(c) Name three differences between using a perceptron and a multiple-layer perceptron for classification problems. [15%]

2 Genome Informatics

(a) Outline the Illumina method of sequencing and explain what limits read length. [25%]

(b) What are the advantages/disadvantages of Illumina sequencing versus other methods? [25%]

(c) Compare and contrast the approach you would take for a small prokaryotic vs a large eukaryotic genome for:

(i) sequencing and assembly

(ii) gene identification [50%]

3 Functional Genomics

(a) Describe why we obtain continuous values from a microarray experiment, and counts (discrete values) from a sequencing experiment if in both techniques we are measuring intensity from an image. Briefly describe the typical contents of the following file-types that are commonly used in Genomics analyses:

(i) Fastq

(ii) Bam [20%]

(b) Suppose you want to investigate the effect of the knockout of a given gene in the regulation of different pathways. Discuss the steps in the analysis needed to go:

(i) from bam files to a list of differentially expressed genes

(ii) from a list of differentially expressed genes to the detection of the pathways affected by the knockout [20%]

(c) In the context of an RNA-SEQ experiment, suppose that we want to compare the expression of different genes amongst several samples. Is it a good normalisation strategy to divide the read counts by the total read depth for each sample? Justify your answer. Describe two different methods for sample to sample normalisation. [20%]

(d) In the context of a ChIP-SEQ experiment describe:

(i) the difference between a peak overlap analysis (occupancy analysis) and a differential binding analysis (binding affinity analysis)

(ii) the goals and the basic steps of a motif analysis [20%]

(e) How can you use genotype information to improve copy number calling? [20%]

END OF PAPER