ENGINEERING TRIPOS PART IIB

Thursday 30 April 2009 9 to 10.30

Module 4G1

SYSTEMS BIOLOGY

*Answer **all** three questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

There are no attachments.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
Engineering Data Book
CUED approved calculator allowed

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

1 (a) Discuss, with one example, the algorithmic complexity of global sequence alignment and describe how we can trade space with time. [50%]

(b) Describe the following sequence alignment algorithms BLAST, PSI-BLAST and PatternHunter, and discuss the differences between them. [50%]

2 (a) Explain the need for pre-processing of microarray data before the analysis of differential expression, and describe its key steps. How can graphical tools used for exploratory data analysis help you with each of the steps? Discuss alternative methods of normalisation and the assumptions associated with their application. [40%]

(b) Describe, for a two-colour microarray experiment, how the experimental design can be specified to enable the computational modelling and analysis of differential expression. How can this specification be used to quantify the precision associated with each of the contrasts the experiment aims to estimate? Explain the caveats of these precision estimates by discussing the concept of effective replication. [30%]

(c) Explain the difficulties of judging statistical significance and the merits of a moderated t-statistic, in particular the empirical Bayesian log odds, for assessing differential expression in microarray experiments. [30%]

(TURN OVER

3 (a) Consider a birth-death process described by $x \xrightarrow{f(x)} x+1, \quad x \xrightarrow{g(x)} x-1$.

(i) Write down the master equation for the probability of x taking value k at time t. [15%]

(ii) Derive using the master equation a differential equation for the mean of x. Hence show that at equilibrium $\langle f(x)\rangle = \langle g(x)\rangle$. [25%]

(iii) A differential equation for $\langle x^2\rangle$ is given by

$$\frac{d\langle x^2\rangle}{dt} = 2\langle x(f(x)-g(x))\rangle + \langle f(x)+g(x)\rangle$$

By linearising $f(x)$ and $g(x)$ about the mean of x i.e.

$$f(x) \approx f(\langle x\rangle) + f'(\langle x\rangle)(x - \langle x\rangle),$$
$$g(x) \approx g(\langle x\rangle) + g'(\langle x\rangle)(x - \langle x\rangle)$$

derive an expression for the variance of x at equilibrium. When is this approximation for $f(x)$ and $g(x)$ valid?
(Note: $f'(\langle x\rangle)$ and $g'(\langle x\rangle)$ denote respectively the derivatives of $f(x)$ and $g(x)$ with respect to x and evaluated at $x = \langle x\rangle$). [30%]

(b) Consider now the inhibition mechanism

$$y \xrightarrow{\frac{K}{K+y^h}} y+1, \quad y \xrightarrow{\beta y} y-1$$

Using the approximation in part (a)(iii) derive an expression for the variance of y at equilibrium in terms of its mean and the constants K, β, h. Hence discuss how the variance of y changes with an increasing Hill coefficient h, for a given mean of y. [30%]

END OF PAPER