(a) Explain how a microar: n your
answer, include brief def:

2007 IIB 4G1 Dr G Vinnicombe
gv@eng.cam.ac.uk

- reverse transcription
- probe
- target
- hybridisation
- control spots
- two-colour vs single-channel arrays

[40%]

(b) Describe, with a simple example, the rank products technique for detecting differentially expressed genes. How is significance of the ranking assessed? [30%]

(c) What is hierachical clustering? Draw a sketch to illustrate your answer. [10%]

(d) Describe, with examples, two ways how hierachical clustering can be applied to study gene expression data, depending on how samples are taken from the gene expression matrix. [20%]

---

2)

Short crib:

(a)

See lecture notes, including cartoon guide.

- reverse transcription: converting mRNA back into cDNA, suitable for use as target material.

- probe: short (~25 base) sequence of DNA that is attached to the array in a known location.

- target: material that has been extracted from the cells under study, and fluorescently labelled.

- hybridisation: if the probe and target are complementary to each other, they will bind together. This process is known as hybridisation.

- control spots: In addition to target cDNA from the cells, control cDNA of known concentation can be added into the target material. We then expect to see hybridisation of this CDNA with the corresponding control spots on the array. This acts as one type of control.

- two-colour vs single-channel arrays: two-colour arrays are where the experimenter uses two fluorescent tags to label two different samples, and mix those samples onto the array. The fluorescence levels must be measured twice, to detect the intensity of the two different fluorescences. With a single-channel array, only one type of sample cDNA is tested, and so only one fluorescence tag is used.

(b) Rank products: see lecture notes/ Breitling et al for details. Answer should include definition of rank product (RP), simple example of ranking (geometric means may not be exact). Ranking assessed by using bootstrap to simulate null distribution of RPs, and estimate of false discovery rate.

(c) Hierachical clustering allows us to iteratively group together samples. In aggloremative clustering, input vectors are repeatedly merged by finding e.g. the closest pair of inputs. [Show sketch of input samples and tree, e.g. 2-d example from lecture notes.]

[d] Each sample is a vector in N-d space. If our gene expression data is regarded as a matrix, with data from one experiment in a column, and data from one gene in a row, we can cluster the data in two ways: by row (gene) or by column (experiment). Clustering by gene allows us to see which genes have similar expression profiles (e.g. cell cycle example in lecture notes); clustering by experiment allows us to see if there is a molecular signature that discriminate e.g. the two types of cell analysed in different experiments (AML/ALL example).

3 a) ~~Bk~~ Bookwork

30%

b) $k_1 \sim \frac{1}{20}$ mins if repressor only knocked off by DNA polymerase.

$$k_1 (1 - P_{on}) = f P_{on}$$

$$\Rightarrow P_{on} = \frac{k_1}{k_1 + f}$$

20%

c) $N = 1$ ON

$N = 0$ OFF

linear, so means satisfy

$$\dot{N} = k_1(1-N) - fN$$
$$\dot{M} = \beta N - \gamma M$$
$$\Rightarrow A = \begin{bmatrix} -(k_1+f) & 0 \\ \beta & -\gamma \end{bmatrix}$$

$$D = \begin{pmatrix} 2f\hat{N} & 0 \\ 0 & 2\beta\hat{N} \end{pmatrix}$$
$$\hat{N} = P_{on} = \frac{k_1}{k_1+f}$$

FDT

$$A\Sigma + \Sigma A^T + D = 0, \quad A\Sigma = \begin{pmatrix} -(k_1+f) & 0 \\ \beta & -\gamma \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$= \begin{pmatrix} -(k_1+f)\Sigma_{11} & -(k_1+f)\Sigma_{12} \\ \beta\Sigma_{11} - \gamma\Sigma_{12} & \beta\Sigma_{12} - \gamma\Sigma_{22} \end{pmatrix}$$

$$\Rightarrow (k_1+f)\Sigma_{11} = f\hat{N} \Rightarrow \Sigma_{11} = \frac{f\hat{N}}{k_1+f}$$

$$\beta\Sigma_{11} = \gamma\Sigma_{12} + (k_1+f)\Sigma_{12}$$
$$\Rightarrow \Sigma_{12} = \frac{\beta}{\gamma + k_1 + f} \frac{f\hat{N}}{k_1+f}$$
$$\gamma\Sigma_{22} = \beta\Sigma_{12} + \beta\hat{N} = \beta\hat{N}\left(\frac{\beta}{\gamma+k_1+f} \cdot \frac{1}{k_1+f} f + 1\right)$$

50%

← should show it decreases with k_1, f - k_1/f fixed. - Need to check