Image analysis, normalisation and gene interaction modelling

C.A. Glasbey, C-D. Mayer & D. Husmeier
Biomathematics and Statistics Scotland
King's Buildings, Edinburgh, EH9 3JZ, Scotland

chris@bioss.ac.uk

This talk will cover aspects of three stages in the analysis of microarray gene expression data: image analysis of scanned arrays, normalisation and modelling of gene interactions.

The standard operating procedure of the Genomic Technology and Informatics Centre at Edinburgh University, is to laser scan cDNA microarrays several times using increasing settings of photo-multiplier tube gain, and possibly also laser power. The single set of scan results are chosen for subsequent analysis that have the highest settings without any pixel values having been censored at 65535, the saturation level of the scanner. We show that noise levels in the data can be reduced by combining pixel values from all scans, taking into account the censoring.

Least trimmed squares (LTS) is one of several methods that has been used to estimate the normalisation between channels or arrays, on the assumption that a proportion of genes are not differentially expressed (Huber et al, 2002). We have applied LTS to the simultaneous normalisation of a series of one-channel experiments. This leads to a robust form of principal components analysis.

Bayesian networks have been proposed as models for genetic regulatory interactions (Friedman et al, 2000). However, it is hard to infer a network from microarray data, because interactions between many genes have to be learned from small data sets, typically containing only a few dozen time points during a cell cycle. We test the viability of the Bayesian network paradigm in a simulation study. First, gene expression data are simulated from a realistic biological network involving DNAs, mRNAs, inactive protein monomers, and active protein dimers. Then, interaction networks are inferred from these data in a reverse engineering approach.

Friedman, N., Linial, M., Nachman, I. and Pe'er D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, **7**, 601-620.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**, S96-S104.