

Editorial

NEW ANALYTICAL TECHNIQUES FOR THE INTERPRETATION OF MICROARRAY DATA

The recent development of array-based techniques is rapidly accelerating progress in many areas of biomedical research. These technologies include DNA, protein or combinatorial chemistry arrays though, in the last few years, DNA microarrays have largely been the main focus in this area. For the first time DNA microarrays have given a global view of the expression levels of thousands of genes and this has already led to significant findings. Cluster analysis has led to the discovery of sub-types in gene expression patterns for certain cancers pointing towards new molecular taxonomies for these diseases. Using microarray data, classifiers have been constructed capable of predicting the metastatic potential of a primary tumor. Statistical scoring of features has led to the discovery of new genes associated with cancer and other diseases in addition to indicating their function and dependencies.

Clearly the advent of these technologies will revolutionize biology and medicine but their full utilization will depend heavily on accurate data processing and analysis techniques. The central role of data analysis will become even more critical in the future as array data is successfully integrated with other massive datasets such as genome sequence data or data from high-throughput drug screening techniques.

Many methods are applicable to analysis of array-based datasets. However, the most relevant are algorithmic techniques drawn from machine learning and statistical methods. As indicated above, analysis of DNA microarray data has frequently utilized machine learning techniques including cluster analysis, classification and feature selection, outlier detection and the discovery of learning curves, for example. Similarly statistical methods have been used to isolate anomalous gene expression, abnormal expression in chromosomal regions and to give accurate pre-processing techniques to correct for experimental artifacts during data acquisition.

This Section on the Analysis of Microarray Data presented in this issue, followed from the NIPS Workshop on Machine Learning Techniques for Bioinformatics held at Whistler Resort, British Columbia, Canada in December, 2001. A central theme of the Workshop was the development and application of new machine learning techniques for functional interpretation of gene expression data from microarray experiments. In the call for papers we encouraged submissions from authors from the broader data analysis community, particularly statistics, in addition to encouraging submissions from authors presenting papers at the Workshop.

As expected new developments in cluster analysis and classification feature prominently among the accepted papers. Other authors analyse inference of regulatory structures and gene interaction or propose statistical adjustments for signal censoring of microarray data. It is reasonable to expect that the future development of array-based technologies will utilize these methods but demand considerable further innovation in data analysis.

I would like to thank all those who have provided direct or indirect help in the preparation of this Section. In particular I would like to thank Sayan Mukherjee from MIT who co-organised the original Workshop, the NIPS Workshop Co-Chairs Barak Pearlmutter and Virginia de Sa and the NIPS Foundation for providing an excellent venue and facilities for the Workshop. Thanks also to the UK EPSRC for a grant (GR/R96255) in partial support of the Workshop and the Special Section. In addition I would also like to thank Martin Bishop and the *Bioinformatics* editorial office for all their patience and support in the preparation of this volume. Last, but not least, I would like to thank the many referees for their tireless efforts in reviewing the papers - in many cases they distinctly improved the clarity and presentation of accepted papers.

Colin Campbell
Bristol University
E-mail: c.campbell@bristol.ac.uk