Meeting Report

Unravelling Nature's Networks: From Microarray and Proteomic Analysis to Systems Biology

A Biochemical Society Focused Meeting held in association with the University of Sheffield Centre for Bioinformatics and Computational Biology, University of Sheffield, 21–22 July 2003.

Nicholas A. M. Monk[1] and Neil Lawrence[2]

1. Centre for Bioinformatics and Computational Biology, and Division of Genomic Medicine, University of Sheffield, Royal Hallamshire Hospital, Sheffield, S10 2JF, UK
2. Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK

The robust and adaptable behaviours of cells and tissues depend on the operation of complex regulatory biochemical networks. The elucidation of the structure and functioning of such networks poses many daunting challenges. Recently developed experimental techniques, such as large-scale profiling of gene expression and protein interactions, provide unprecedented amounts of information on the molecular composition of cells. The size (and often variable quality) of the resulting data sets necessitates the use of sophisticated computational schemes for the analysis, mining, and integration of the data. In all but the simplest cases, the complexity of the networks is such that it is impossible to provide an intuitive picture of the principles governing their dynamic behaviour without synthesising the experimental data into a coherent mathematical model of the underlying system.

How techniques from many different disciplines can be brought together to explore biochemical networks was the topic of a lively and successful two day focused meeting "Unravelling Nature's Networks: From Microarray and Proteomic Analysis to Systems Biology", held at the University of Sheffield in July. The aim of this meeting was to explore the progress made in this multi-disciplinary process, together with the major outstanding problems that remain to be solved. The speakers and participants covered the spectrum of disciplines involved in acquiring, analysing, and integrating large-scale biological data.

In the first session, two speakers illustrated the types of techniques that are available for gathering primary data on cells and tissues. Steve Dower (Sheffield, UK) described how fluorescence imaging of signal transduction has been used to show that even under identical conditions, there can be significant quantitative differences between cellular responses. These results show that data obtained from homogenised tissues may not always capture the full complexity of cellular processes. Rick Livesey (Cambridge, UK) explained how expression profiling using cDNA microarrays could be combined with mutagenesis and in situ hybridisation to map out transcriptional networks underlying the development of the mouse forebrain. The combination of high throughput profiling with more traditional techniques of developmental biology allows developmental pathways to be mapped out rapidly and with high confidence.

Data from high throughput techniques is currently of quite variable quality, often containing significant levels of noise, false positives and false negatives. Three talks addressed different aspects of the critical problem of how best to extract useful information from such data sets. Mahesan Niranjan (Sheffield, UK) used the specific example of microarray expression analysis to review basic techniques for classification and how ROC curves can be used in their evaluation. Terry Speed (Parkville, Australia) showed how a statistical approach could be used to improve the quality of reconstructed spectra for the SEQUEST algorithm for protein identification through tandem mass spectrometry. Patrick Kemmeren (Utrecht, Netherlands) addressed the problem of variable data quality by showing how the integration of data obtained from disparate techniques could be used to improve the reliability and usability of data; integration allows features such as interactions to be identified and classified with increased

confidence, with a concomitant (and sometimes dramatic) reduction in the scale of the data.

The second day of the workshop focussed on the reconstruction and modelling of interaction networks. Lev Soinov (Hinxton, UK) showed how this could be achieved through rule based classification systems that assign simple functional relationships between genes and their products. Such approaches are particularly appealing to experimental biologists as they are very easy to interpret and validate. Preliminary analyses of the structure of entire high throughput data sets have suggested that the underlying interaction networks have a characteristic overall structure, characterised by scale-free and small world properties. Current data sets contain only a small sample of the entire set of interactions, however. Alun Thomas (Utah, USA) used the particular case of protein-protein interaction networks to highlight the influence that sampling techniques can have on perceived network structure; in particular, samples of (non-scale-free) random networks can exhibit scale-free structure. Local network structures that have been derived from intensive investigation can be used to give complementary information on basic network structure, and suggest specific models that can be tested.

Once an interaction network has been reconstructed from data, it is necessary to develop mathematical models of the network to study the dynamical behaviour of the network. Jaroslav Stark (London, UK) and Kwang-Hyun Cho (Uslan, Korea) discussed how systems of nonlinear differential equations could be used to model the dynamics of restricted networks of interactions operating in specific biological contexts, such as apoptosis or signal transduction. Both talks highlighted the fact that it is often difficult to deduce the behaviour of even quite simple networks without the use of a model. Both also illustrated the problems faced when constructing models, such as the need for time-series data of high spatial and temporal resolution.

One biological process for which detailed information is available is the yeast cell cycle, and Béla Novák (Budapest, Hungary) described a detailed model based on modularity of the cell cycle machinery. The insights into cell cycle dynamics that can be gained through analysis of this model provide an excellent illustration of the potential power of mathematical modelling when combined with high quality data. The meeting concluded with a presentation from Eric Mjolsness (Irvine, USA), who gave an impressive demonstration of how the "Cellerator" package can be used to integrate biochemical interactions into the spatio-temporal context of a biological system. This represents a significant step for the modelling of systems such as developing tissues in which cell division and movement are critical.

The invited presentations were complemented by a number selected from submitted abstracts. The poster prize “Microarray Analysis” by Mark Schena, kindly donated by John Wiley & Sons, was awarded to Keira Curtis for her work on “Control Analysis of Microarray Data”.