Meeting report

# All systems are go
## Claire E Eyers and Onrapak Reamtong

Address: Michael Barber Centre for Mass Spectrometry, School of Chemistry, Manchester Interdisciplinary Biocentre, Princess Street, Manchester, M1 7DN, UK.

Correspondence: Claire E Eyers. Email: Claire.Eyers@manchester.ac.uk

---

A report on 'Genomes to Systems', the Fourth Conference of the Consortium for Post-Genome Science, Manchester, UK, 17-19 March 2008.

---

The Consortium for Post-Genome Science promotes the application of research advances in genomics, transcriptomics and proteomics to advance our understanding of biological systems. For this information to be useful to the wider scientific community, informatics tools are required for the assimilation and modeling of these systems. The latest Genomes to Systems (G2S) conference [http://www.genomestosystems.org] highlighted recent advances in these areas, with a focus on contemporary technologies in biotechnology, biomedicine and their applications to understanding integrated systems in both normal and disease states.

## The search for disease markers

Early diagnosis of cancer dramatically increases survival rate and because of this, biomarker discovery studies seek to identify early-onset markers. Samir Hanash (Fred Hutchinson Cancer Research Center, Seattle, USA) and Ruedi Aebersold (ETH, Zurich, Switzerland and Institute for Systems Biology, Seattle, USA) highlighted the importance of rigorous large-scale quantitative data acquisition to facilitate the search for disease biomarkers. This was a contrast to previous meetings, which tended to put the emphasis on improving proteomics strategies and data-analysis tools. The primary stumbling block in biomarker discovery appears to be the amount and quality of the data being analyzed. Specifically, Hanash described techniques for discovering biological indicators of cancer in plasma samples, using both human and murine models. Of particular interest was a study aimed at identifying biomarkers for breast cancer using a biobank of plasma samples obtained from women over a ten year period. The Biobank contains samples from over 160,000 women and is being used specifically to look for biomarkers in 1,000 women in samples taken a year prior to diagnosis of breast cancer.

Aebersold also highlighted the challenges associated with defining disease biomarkers and performing hypothesis-driven research. He noted that disease markers are often different for disease subtypes and dependent on associated risk factors, and that these must be identified amid all the biological 'noise'. Different diseases may also perturb overlapping regions of networks and disease-specific signatures of these deregulated networks need to be identified and used for biomarker discovery. Andrey Rzhetsky (University of Chicago, USA) presented methods of text mining that should go some way to help distinguish when a single factor is contributing to multiple diseases. Specifically, he gave examples of gene targets contributing to autism, bipolar disorder and schizophrenia, predicting gene candidates that are both specific to these diseases and shared among them. The topic of multiple disease factors for a given disease was taken further by John Griffiths (Cancer Research UK, Cambridge, UK), whose findings in tumor cells indicate that a multi-target approach (using combinations of two or more drugs) will in all likelihood be necessary to control progression of diseases such as cancer and diabetes. He reported that inhibition of tumor growth was found to be markedly improved when the histone deacetylase inhibitors SAHA and LAQ824 were used in combination.

Several compelling cases of disease-specific genomic markers involved in responses to drugs were presented. Epilepsy can present in multiple forms, with different patients exhibiting a different pattern of resistance and response to therapeutic agents. Sanjay Sisodiya (University College London, UK) presented data showing that mutation of the gene *SCN1A*, which encodes a sodium channel, is associated with drug-resistant forms of epilepsy. On the same theme, Caroline Lee

(National University of Singapore, Singapore) reported a combination of three single-nucleotide polymorphisms in the blood-brain barrier transporter MDR1 that can be used as a marker of Parkinson's disease among ethnic Chinese, and Ann Daly (Newcastle University, UK) reported the characterization of specific genetic polymorphisms in genes encoding the enzymes UGT2B7, CYP2C8 and ABCC2, associated with hepatotoxicity induced by the non-steroidal anti-inflammatory drug Diclofenac.

## Systems biology and proteomics

One of the most important aspects of systems biology is translating the biological information into models that can be manipulated and used in simulations. The Systems Biology Markup Language (SBML) [http://sbml.org], a computer-readable format for representing biological models, has been evolving over the past eight years with the input of users and software developers. Mike Hucka (California Institute of Technology, Pasadena, USA) described the current status of SBML and developments soon to be implemented. Rather than further complicating an already over-extended system, SBML is being modularized, so that biochemical species can themselves be annotated, with localization, substructures and post-translational modifications defined, instead of having to specify different biological states of the same protein as separate entities.

Having defined the language of computational modeling, Nicolas Le Novère (European Bioinformatics Institute, Hinxton, UK) defined the 'minimum information requested in the annotation of biochemical models' (MIRIAM) [http://www.ebi.ac.uk/compneur-srv/miriam] - guidelines for the curation of quantitative models that can be used by the systems biology community. Le Novere continued by defining the 'minimum information about a simulation experiment' (MIASE), guidelines enabling the research community to repeat and make use of previous simulations of a given model. An overview of COPASI (a COmplex PAthway SImulator) software for the modeling, simulation and analysis of biological systems was given by Ursula Kummer (University of Heidelberg, Germany). COPASI [http://www.copasi.org/tiki-index.php] has been developed by Kummer and colleagues in collaboration with Pedro Mendes (University of Manchester, UK), as a user-friendly, platform-independent tool for the analysis of biochemical networks.

Hiroaki Kitano (Systems Biology Institute, Tokyo, Japan) gave an entertaining presentation on the graphical representation of biological networks. Unlike engineers, who have specific graphical expressions for representing defined functions and/or species, the bioscience community uses the same graphical notation for different biological processes. Kitano described the use of the standardized systems biology graphical notation (SBGN) [http://www.sbgn.org] for biological pathways, which essentially generates a biological circuit diagram. While the widespread use of a standard notation of this sort would be extremely useful for the bioscience community, making biological network diagrams visually transparent, only time will tell how quickly it will filter into general use.

Several presentations focused on evaluating and developing novel techniques for X-ray crystallography and NMR to help improve analysis of protein structure and function. Using X-ray crystallography to understand protein function can lead to the discovery of new therapeutic agents, as demonstrated by Larry DeLucas (University of Alabama, Birmingham, USA) and Stephen Cusack (EMBL, Grenoble, France). DeLucas presented novel crystallographic techniques for studying drug design based on analyzing the structures of the proteins the drugs are intended to target. He described the application of self-interaction chromatography (SIC) to measure protein-protein interactions, using this technique to facilitate structure-based drug design. He also demonstrated that certain additives, such as polysaccharides and amino acids, can lead to the preferential precipitation of certain protein species without prior separation, due to changes in hydrophobicity. A reduction in growth rate was also found to improve the quality of crystals for structural analysis. Cusack described how structural elucidation of the influenza virus polymerase can help to understand viral transcription mechanisms. Using a library-based screening technique called 'expression of soluble proteins by random incremental truncation' (ESPRIT), he has elucidated the structure of the polymerase subunit PB2 and defined specific regions of the protein involved in nuclear import and mRNA cap-binding. Elucidation of these motifs has helped define the mechanism of transcription of viral mRNAs. In addition, modeling studies in the presence and absence of viral mRNAs were used to define areas of interest that may be targeted for antiviral drug design.

Many presentations focused on the adaptation and development of established techniques to understand more of the proteome. Simon Gaskell (University of Manchester, UK) described how complementary methods of tandem mass spectrometry (MS/MS) can be exploited to gain more raw data from proteomics samples. Fragmentation of a given peptide ion was induced using both collision-induced dissociation and electron transfer dissociation, techniques that differ significantly in their mechanisms of fragmentation and thus generate different types of fragment ions. Sequential fragmentation in this manner increases the dimensionality of the acquired data and enables deeper mining of the proteome. He echoed the need for quantitative proteomics data for biological systems, and discussed recent developments for the absolute quantification of sites of protein phosphorylation using a modified QconCAT-based strategy. QconCAT permits the

absolute quantification of proteins by MS in a multiplexed manner, and was developed in a collaboration between Gaskell and Rob Beynon (University of Liverpool, UK). Coding sequences for the peptides being used as quantification standards are concatenated for expression as a single artificial protein in *Escherichia coli*. The QconCAT protein can then be isotopically labeled, purified and quantified as a single entity, prior to proteolysis and the generation of more than 50 purified stoichiometric reference peptides in a single step. The advantages of QconCAT over synthetic peptide production for the absolute quantification of proteins in complex mixtures, namely the ease and cost-effectiveness, were also discussed by Beynon. Also raised was the need to examine the extent of analyte digestion in order to accurately determine absolute amounts of proteins.

Simon Hubbard (University of Manchester, UK) discussed the selection of peptides for use as standards for absolute protein quantification in a semi-automated fashion. A combinatorial approach employing three independent machine-learning methods has been developed that predicts those tryptic methods most likely to 'fly' (that is, ionize and be detected) during liquid chromatography-mass spectrometry (LC/MS) analysis, with a positive predictive value of greater than 79%. Hubbard also discussed new bioinformatics tools under development, including open-source software for extracting quantitative data from a variety of proteomics platforms, currently called the SILACanalyzer and described the updated peptide identification database PepSeekerGOLD [http://www.ispider.manchester.ac.uk/pepseeker]. Pepseeker has so far been used to help elucidate peptide-fragmentation mechanisms in different types of mass spectrometer and as a tool to predict sites of trypsin hydrolysis. Beynon presented a strategy for simplifying analyses using positional, rather than shotgun, proteomics, which involve enriching for and analyzing amino-terminal peptides. Using this strategy, free amines at the protein amino terminus and lysine side chains are chemically blocked, the sample is trypsinized and peptides containing free amine groups (present at the newly generated peptide amino termini) are removed. Quantification using only the amino-terminal peptide was shown to give similar data to those obtained with multiple peptides, arguing against the requirement for large-scale peptide analysis.

Anne Dell (Imperial College London, UK) discussed the analysis of glycosylated proteins, drawing attention to the Consortium for Functional Glycomics, which has recently developed a database of glycan structures and glycan-binding proteins [http://www.functionalglycomics.org] as a tool to interpret MS data in a semi-automated fashion. She also explained how knowledge of the biosynthetic pathways of glycans is critical for interpreting MS data and gave examples of how glycan signatures might be used as

diagnostic markers in cancer. She also explained how knowledge of the biosynthetic pathways of glycans is critical for interpreting MS data and used the altered glycosylation status of haptoglobin in the sera of patients with prostate cancer as an example of how specific glycan signatures might be used as diagnostic markers in cancer.

The conference culminated in a plenary lecture by Hans Westerhoff (University of Manchester, UK and Vrije Universiteit, Amsterdam), discussing how the study of biological systems as networks could be used to identify therapeutic drug targets. The multifactorial nature of diseases means that their treatment should be considered in the context of the networks that are perturbed. Importantly, therefore, drugs that target the most 'fragile' part of the perturbed network might be the most physiologically relevant. As an extension to this, he also suggested that drugs for parasitic infections are likely to have reduced host toxicity if they are targeted to factors where the parasite and host have different fragility coefficients. Furthermore, when considering drug targets and toxicity, it is important to realize that points of robustness change in normal versus tumor cells, and that it is the difference of these fragility coefficients that is likely to lead to the development of successful novel therapeutics. The 2008 Genomes to Systems conference provided a stimulating environment to discuss the (multiple) current, and future, directions of systems biology research and we look forward to the next meeting in this series.

## Acknowledgements