EDITORIAL

Integrating scientific cultures

Two communities, two cultures

A key challenge of Systems Biology is that it must integrate several disciplines, each with a very different culture for disseminating results. Within biology, manuscripts describing new work are almost always published in peer-reviewed periodicals. In contrast, within computer science and the engineering fields, new methods and results are typically presented as full-length papers at meetings and workshops. Just as journals have Editorial Boards that handle review of manuscripts, such conferences assemble large and reputable Program Committees, which fulfill the same purpose. Publication in the best conferences, as for the best journals, is highly competitive.

This past December, several hundred scientists convened in La Jolla, California for the Second Annual RECOMB Workshop on Systems Biology (December 1–3, 2006; http://chianti.ucsd.edu/recombsysbio06/). The meeting, which was held jointly with the RECOMB Workshop on Computational Proteomics, took place at the California Institute for Information Technology and Telecommunications in the University of California San Diego campus. RECOMB, which stands for *Research in Computational Biology*, has for a decade sponsored conferences that attract high-quality papers in bioinformatics, primarily from computer science.

In an effort to integrate the computational and experimental biology communities, RECOMB and *Molecular Systems Biology* entered into a partnership by which original, peer-reviewed papers are presented orally at the Workshop on Systems Biology and then appear as full-length manuscripts in the pages of the journal. The precise publication model was formulated after much discussion between the editors of the journal and the organizers of RECOMB. It is original and, we hope, will serve as a case study for future conferences. First, a list of 45 reviewers was approved by both *Molecular Systems Biology* and RECOMB to form the conference Program Committee. Based on three reviews, the RECOMB chairs considered each paper for oral presentation at the conference. Papers chosen for oral presentation were then submitted to the journal, along with the three referee reports. Among these papers, *Molecular Systems Biology* decided to invite a subset of the authors to submit revisions, in which case the same reviewers were asked to check the revisions before publication.

Approximately 50 papers were submitted to the joint workshops, 20 of which were invited for oral presentation. Seven of these 20 were also accepted for publication in *Molecular Systems Biology*. Thus, 40% of the submitted papers were granted oral presentations at the conference, whereas 14% of all submissions were accepted for publication in *Molecular Systems Biology* as written Reports.

Computational methods for data integration

The seven Reports are now published online. Their topics fall into the three major areas covered during the 3 days of the conference: Regulatory Networks (December 1), Protein Interactions (December 2), and Mass Spectrometry/NMR (December 3).

The three papers in the Regulatory Networks theme present methods for integration of gene expression profiles as the main source of data. The study by Shlomi *et al* (2007) combines the measured expression patterns of enzymes together with their corresponding reaction fluxes to evaluate the impact of transcriptional regulation on *Escherichia coli* metabolism. Their analysis reveals that the flux through ∼15% of reactions appears to be controlled by enzyme expression level. In the work by Zhou *et al* (2007), expression profiles of protein-encoding and microRNA genes are combined with promoter analysis to discover microRNA genes induced by UV-B radiation in *Arabidopsis thaliana*. In the third paper focusing on Regulatory Networks, Martha Bulyk and colleagues apply a technology they developed previously called protein binding microarrays (PBMs), which catalog the complete repertoire of DNA sequence motifs recognized by a transcription factor *in vitro*. Here, they describe an elegant bioinformatic approach that integrates gene expression with PBM data to predict condition-specific functions of transcription factors (McCord *et al*, 2007).

Owing to the numerous technologies that have become available for measuring interactions among proteins, protein networks are a central topic of research in Systems Biology. In the study by Lu *et al* (2007), a known network of protein–protein interactions is integrated with gene expression profiles gathered from mice afflicted with a model for human asthma. Intriguingly, highly connected proteins in the network (so-called 'hub' proteins) have gene expression profiles that are less variable than proteins at the network periphery. Ulitsky and Shamir, 2007 strive to integrate interaction networks of different types. Specifically, they describe methods for integrating protein–protein physical interaction networks with genetic networks of synthetic-lethal interactions in yeast. Although protein–protein and synthetic-lethal interactions can coincide (occur between the same proteins), these interaction types are more often orthogonal, with protein–protein interactions occurring within the same pathway and synthetic-lethals spanning two or more pathways with redundant or synergistic functions. In the present work, the authors implement a method for detecting genetic interactions connecting synergistic pathways that are embedded within the physical and genetic networks.

While genome-wide measurements of DNA sequences and mRNA expression levels are now routine, techniques such as mass spectrometry and nuclear magnetic resonance (NMR) are bringing genome-scale analysis to the two other major classes of biological molecules: proteins and metabolites. Tandem mass spectrometry is the method of choice for protein identification, with the most popular protocols involving a search of available peptides to find one that best explains the spectra. However, a direct search of translated cDNA databases, while inclusive, suffers from high redundancy. Nathan Edwards (2007) provides an interesting approach to massively compress EST databases, thereby reducing the otherwise prohibitive computational cost of such extensive searches and enabling the identification of novel peptides. The work by Feala *et al* (2007) uses ^{1}H NMR spectroscopy to study the metabolic response to hypoxia in the fruitfly *Drosophila melanogaster*. Surprisingly, the authors observe an accumulation of lactate, alanine, and acetate by-products, which was not predicted based on existing knowledge of fly metabolism. Using an expanded model, the authors suggest that growth tolerance to hypoxia may rely on the ability to convert pyruvate to acetate and alanine by altering the ATP/H $+$ ratio.

Systems Biology as an integrative discipline

This series of papers clearly illustrates the increasing importance and sophistication of data integration methods. Early efforts in bioinformatics concentrated on finding the internal structure of individual genome-wide data sets. With the explosion of the 'omics' technologies, comprehensive coverage of the multiple aspects of cellular physiology is progressing rapidly, generating vast amounts of data on mRNA profiles, microRNAs, protein and metabolic abundances, and protein interactions. A systems-level understanding of living organisms requires taking into account, systematically, all of their mRNA, protein, and small molecule components. Importantly, it will also require integrating all of the known properties of a given class of components (e.g., protein abundance, enzymatic activity, localization, physical interactions, etc.). Computational techniques able to combine these large and heterogenous sets of data will be essential for the generation of unified mechanistic models of cellular processes, one of the main goals of Systems Biology.

To generate these ambitious models, scientists from different disciplines must be able to communicate with each other in the first place. The challenge of different publication models in different disciplines is that, if one is interested in following the latest results or methods in Systems Biology, it is sometimes confusing where to look. In particular, the biological community is often unaware of relevant progress in bioinformatics, simply because it appears first at a conference, even one that is well attended and highly regarded. Both RECOMB and ISCB (International Society for Computational Biology) have collaborated with bioinformatics journals in the past. Here, we have adopted a journal/conference co-presentation model for Systems Biology. By partnering with a journal well known to biologists, additional exposure is provided to relevant computational methods, while attracting a body of excellent work that might have otherwise been dispersed. It is our sincere wish that this initial experiment, conducted in partnership between RECOMB and *Molecular Systems Biology*, will continue to grow as a means of integrating not only systems-level measurements, but also the scientific communities that analyze them.

Acknowledgements

Trey Ideker[1,2], Vineet Bafna[2] and Thomas Lemberger[3]
[1]Department of Bioengineering, University of California San Diego, La Jolla, CA, USA;
[2]Department of Computer science, University of California San Diego, La Jolla, California, USA and
[3]European Molecular Biology Organization, Heidelberg, Germany

References

Edwards NJ (2007) Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. *Mol Syst Biol* **3**: 102

Feala JD, Coquin L, McCulloch AD, Paternostro G (2007) Flexibility in energy metabolism supports hypoxia tolerance in *Drosophila* flight muscle: metabolomic and computational systems analysis. *Mol Syst Biol* **3**: 99

Lu X, Jain VV, Finn PW, Perkins DL (2007) Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *Mol Syst Biol* **3**: 98

McCord RP, Berger MF, Philippakis AA, Bulyk ML (2007) Inferring condition-specific transcription factor function from DNA binding and gene expression data. *Mol Syst Biol* **3**: 100

Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* **3**: 101

Ulitsky I, Shamir R (2007) Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol Syst Biol* **3**: 104

Zhou X, Wang G, Zhang W (2007) UV-B responsive microRNA genes in *Arabidopsis thaliana*. *Mol Syst Biol* **3**: 103