Genome, Pathway, and Interaction Bioinformatics: Session Introduction

P. Karp, E. Neumann, A. Hartemink, and P. Romero

Pacific Symposium on Biocomputing 8:101-103(2003)

GENOME, PATHWAY, AND INTERACTION BIOINFORMATICS

PETER KARP, PEDRO R. ROMERO
Bioinformatics Research Group, SRI International
333 Ravenswood Avenue, Menlo Park, CA 94025
{pkarp,promero}@ai.sri.com

ERIC NEUMANN
Beyond Genomics
40 Bear Hill Road, Waltham, MA
eneumann@beyondgenomics.com

ALEXANDER J. HARTEMINK
Department of Computer Science and
Center for Bioinformatics and Computational Biology
Duke University, Box 90129
Durham, NC 27708-0129
amink@cs.duke.edu

The complete sequencing of the genomes of hundreds of organisms has fundamentally transformed both the focus and the practice of modern biology. While these efforts have largely succeeded in revealing the vocabulary of the "Book of Life", the syntax and semantics of the book remain in large measure undiscovered—the book reads more like Lewis Carroll's *Jabberwocky* than Shakespeare's *Hamlet*. Fathoming the meaning of the genomic vocabulary is the critical next step in better understanding biological systems and their complexity.

At the heart of this new challenge lies the need to understand the interplay of genes and their protein products. From the various interactions between genes and proteins, biological pathways and networks arise. These causal pathways and networks are responsible for the development, maintenance, regulation, and responsiveness of all living systems. Only when we go beyond the study of individual biological molecules to the analysis of their relationships does the complexity of biological systems come into full view. This degree of complexity calls for more powerful models and representations, new mathematical and computational frameworks, and possibly even altogether different analytical paradigms. The Genome, Pathway, and Interaction Bioinformatics session at PSB 2003 is dedicated to presenting novel research that attempts to answer this call by addressing fundamental issues in the recovery and analysis of the interactions, pathways, and networks governed by the genomes of all organisms.

This field has been significantly transformed by the emergence of new high-throughput functional genomics technologies that have enhanced our ability to observe the functioning of complex biological systems. Experimental data

elucidating various aspects of bioloigical interactions, pathways, and networks is being generated at an ever-increasing rate. This large ensemble of information contains patterns that reflect pathway dynamics, and thus can be used to deduce causal pathway structures. Progress in this field will depend at least partially on the intelligent analysis and mining of this high-throughput functional genomics data in order to infer pathways and their regulation.

Specific relevant types of high-throughput data include gene expression profiles collected from high-density nucleotide arrays, protein expression profiles collected from 2D gels and newer array-based assays, gene-protein interaction data describing the binding location along the genome of DNA-binding proteins collected using chromatin immunoprecipitation, protein-protein interaction data collected using various hybrid systems, and the immense quantity of both gene and protein sequence information accumulating in publicly available databases. Each of these types of data presents a different perspective on the structure of biological pathways and networks, and the most effective analytical methods will likely draw their conclusions not from just a single type of data, but from the combined evidence of data of many different types.

The papers in the session this year examine a number of different approaches to the problem of network inference and modeling but retain some common themes. The point of widest disagreement among the papers seems to be the appropriate framework with which to model complex biological systems, the choices ranging from S-systems and extensions thereof, to Bayesian and dynamic Bayesian networks, to hybrid Petri nets, to full-blown systems of differential equations. The theme of greatest commonality among the papers seems to be the need for simulation to effectively evaluate the ability of these models to represent and recover properties of complex biological systems.

Beyond the issues addressed in the papers of this session, a number of other challenges in this field remain. Future work in the field will likely address such thorny questions as: How do we compare and align individual pathways or entire networks once uncovered? Can we infer qualitative network properties when the immense number of quantitative parameters that govern their behavior is not known? How can we effectively collect, organize, and integrate pathway information? Are there new experimental techniques that can be developed to reveal even more insights into the functioning of biological pathways? Finally, we note that in addition to the functional aspects of networks of interaction, spatial and temporal factors must also be taken into account when modeling complex biological systems, the difficulty of which is only beginning to be appreciated.

The field of genome, pathway, and interaction bioinformatics still includes many challenges and holds much promise; a better understanding of the interaction networks in complex biological systems will enable numerous advances in biotechnology, not the least of which is an enhanced ability to target therapeutics appropriately in diseased cells.