GUEST EDITOR'S INTRODUCTION



Computationally Driven Experimental Biology

T.M. Murali, Virginia Tech

The complexity, diversity, and richness of experimental data on cellular systems are inspiring the development of computational analysis techniques that can directly prioritize and suggest new experiments.

dvances in high-throughput experiments have created a data-rich environment for biology. This phenomenon is especially prominent in molecular biology, where a dizzying variety of data on genomic sequences, molecular levels and activities, and interaction pathways and networks is becoming available at increasing rates.¹

In principle, the potential now exists for most new biological experiments to be founded on an analysis of high-throughput, large-scale data. Since each data type provides insight into a specific facet of cellular activity, integrated analysis is necessary to construct a complete model of cellular systems. Accordingly, researchers have developed numerous computational techniques for analyzing and interpreting large-scale molecular datasets. Many of these approaches are successful in finding patterns of biological activity that are of interest to biologists. However, for the most part, analyzing these results to develop hypotheses and propose experiments to validate them remains challenging.

CLOSING THE GAP

Many factors contribute to this gap between computational analysis and experimental follow-up.

First, automated analyses can yield a high volume of results. The required manual analysis of these results can overwhelm an experimenter who desires to develop and validate a small number of hypotheses.

Second, experimentalists sometimes design experiments and generate new data without fully taking into account how their computational collaborators can analyze this data.

Third, analysis is often performed by quantitative scientists who rely on publicly available datasets. Therefore, the results of their work do not reach the experimentalists who generated the data and who might benefit from novel or improved analysis.

Fourth, many bioinformatic approaches do not directly address the issue of how a life scientist might use the results of computational analyses to drive the next round of experiments.

IN THIS ISSUE

An emerging set of algorithms is systematically tackling these challenges. This special issue highlights the latest advances, opportunities, and challenges in this area.

The authors were encouraged to write articles that provide their own personal perspectives on these topics, in some cases by summarizing multiple papers they have published with their collaborators. For this reason, the research described in these articles might not comprehensively cover the literature on methods that have the explicit goal of prioritizing and directing new experiments.¹⁻⁵

The four cover features in this special issue differ in their focus, including the description of a scientific collaboration that transcends disciplinary boundaries, discussions of challenging computational issues that arise in the analysis of genomic, molecular, and network data and the algorithmic approaches for tackling them, and an article highlighting numerous computational questions that arise in the new field of synthetic biology in the context of redesigning viral genomes.

In "Close Encounters of the Collaborative Kind," Michael Mayhew and his Duke University colleagues describe a successful scientific collaboration between an experimental group and a computational group to study the fundamental process by which all cells grow and divide. The authors describe the genesis of the collaboration, the challenges in communicating with scientists who speak a different language, strategies they used for improving the interpretability of the computational models, and how continued collaboration improved the validation of the computational models and the biological data.

The interplay between computation and biology will continue to yield improved quantitative models, novel and powerful algorithms, and sophisticated biological insights.

In "Using Protein Interaction Networks to Understand Complex Diseases," Mehmet Koyutürk focuses on using such networks to obtain insights into mechanisms that underlie diseases. He surveys current computational approaches that integrate networks with disease-associated differences (dysregulation) in the abundance of individual molecules. His article stresses that improvements in methods to model and measure combinatorial dysregulation of genes can lead to more refined algorithms to tease out the mechanisms that underlie complex diseases from molecular biology datasets.

Cancers are among the most complex and deadliest diseases that afflict human beings. Recent advances in sequencing have enabled the discovery of genetic mutations in the genomes of patients diagnosed with cancer. In "Algorithms and Genome Sequencing: Identifying Driver Pathways in Cancer," Fabio Vandin, Eli Upfal, and Benjamin Raphael discuss methods they have developed to identify cellular processes that contain genes harboring mutations that might cause cancer. Distinguishing such driver mutations from so-called passenger mutations is necessary to identify those that are a priority for experimental study.

Synthetic biology is an exciting new area that aims to design new genetic sequences with specified functions. In "Redesigning Viral Genomes," Steven Skiena describes multiple applications for his work, including the design of antiviral vaccines and refactoring viral genomes to make them more tractable for experimental manipulation.

he informative and thought-provoking articles included in this special issue are intended to give *Computer*'s readers a glimpse of the exciting developments taking place in computationally driven experimental biology. In a discussion of progress in the reverse direction, Saket Navlakha and Ziv Bar-Joseph provide a review of how computational modeling of biological processes can improve the design of algorithms.⁶ Researchers in this field anticipate that the interplay between computation and biology will continue to yield improved quantitative models, novel and powerful algorithms, and sophisticated biological insights.

References

- 1. A.R. Joyce and B.O. Palsson, "The Model Organism as a System: Integrating 'Omics' Data Sets," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, 2006, pp. 198-210.
- 2. Y. Guan et al., "Systematic Planning of Genome-Scale Experiments in Poorly Studied Species," *PLoS Computational Biology*, vol. 6, no. 3.
- 3. R.D. King et al., "The Automation of Science," *Science*, Apr. 2009, vol. 324, no. 5923, pp. 85-89.
- E. Szczurek et al., "Elucidating Regulatory Mechanisms Downstream of a Signaling Pathway Using Informative Experiments," *Molecular Systems Biology*, vol. 5, article no. 287.
- 5. C.H. Yeang et al., "Validation and Refinement of Gene-Regulatory Pathways on a Network of Physical Interactions," *Genome Biology*, vol. 6, no. 7, 2005, R62+.
- 6. S. Navlakha and Z. Bar-Joseph, "Algorithms in Nature: The Convergence of Systems Biology and Computational Thinking," *Molecular Systems Biology*, vol. 7, article no. 546.

T.M. Murali is an associate professor in the Department of Computer Science at Virginia Tech, where he codirects the Center for Systems Biology of Engineered Tissues. The goal of Murali's research in computational systems biology is to build phenomenological and predictive models of the intricate interaction networks that govern the functioning of a living cell. Murali is a member of ACM and the International Society for Computational Biology. Contact him at murali@cs.vt.edu.

CN Selected CS articles and columns are available for free at http://ComputingNow.computer.org.