Guest Editors' Introduction: Selected Papers from ACM-BCB 2013

Srinivas Aluru and Donna K. Slonim

In recent years, computational methods have increasingly become an integral part of the toolkit for performing biological and biomedical research. The fourth ACM Conference on Bioinformatics, Computational Biology, and Biomedical Informatics, held in Washington D.C. in September 2013, provided a valuable forum for an exchange of information about such methods. We are pleased to introduce here a special section devoted to the top papers selected from this meeting.

The conference featured talks on papers submitted to the proceedings track; additional "highlights" talks on previously-published work; excellent keynotes by Steven Salzberg, Ruth Nussinov, and Isaac Kohane; tutorials; workshops; and poster sessions. The reviewing process for the papers in the Proceedings Track was organized by 18 area chairs who oversaw the reviewing of submissions in nine thematic areas. We extend our special thanks to the area chairs and reviewers for their substantial efforts on behalf of this meeting.

Of the 122 manuscripts submitted to the proceedings track, 44 were chosen for full-length talks at the meeting. This special section contains eight papers selected from among those 44, including the winners of both the Best Paper and Best Student Paper awards. The conference versions of the invited papers were revised with a significant addition of new content and re-reviewed before being accepted for publication in the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (*TCBB*). Although the accepted papers, like the conference itself, span a wide range of areas, common themes include biological networks and translational applications.

In "MRFy: Remote Homology Detection for Beta-Structural Proteins Using Markov Random Fields and Stochastic Search," the winner of the Best Student Paper award, Noah M. Daniels, Andrew Gallant, Norman Ramsey, and Lenore J. Cowen tackle the challenge of finding new members of protein structural families when their sequences have sufficiently diverged that homology-based methods fail. Profile Hidden Markov Models have limited success in these cases because they only characterize interactions between residues that are relatively proximate in the underlying sequence. By incorporating Markov random fields, Daniels et al. enable detection of interactions between residues that are more distant in the underlying sequence, though proximate in the folded protein. Their methods lead to improved recognition of several protein families with beta-structural motifs.

Recent work has led to an increase in our understanding of the degree to which post-translational modifications of proteins influence protein expression and function. Knowledge bases representing current information about post-translational modifications are being compiled, but such expert curation requires computational assistance to keep pace with the growing collection of biomedical knowledge. In "RLIMS-P 2.0: A Generalizable Rule-Based Information Extraction System for Literature Mining of Protein Phosphorylation Information," Manabu Torii, Cecilia N. Arighi, Gang Li, Qinghua Wang, Cathy H. Wu, and K. Vijay-Shanker describe a novel biocuration tool for extracting protein phosphorylation information from the literature. Their method proves both accurate and scalable to large corpora of either abstracts or full-text articles.

"Phenotype-Dependent Coexpression Gene Clusters: Application to Normal and Premature Ageing," by Kun Wang, Avinash Das, Zheng-Mei Xiong, Kan Cao, and Sridhar Hannenhalli is one of two papers in this special section that use computational modeling to study dynamic aging processes. In this paper, Wang and colleagues model transcriptional changes with age in normal fibroblast cell culture and in cells derived from patients with Hutchinson Gilford progeria syndrome (HPGS), a genetic disorder that causes severely accelerated aging. Through clustering and regression analyses of RNA-seq data from these cells, the team identifies functional pathways relevant to normal aging and disease. Their results, which are consistent with prior work on cellular aging, are the first describing how aging processes change in HPGS.

"Global Network Alignment in the Context of Aging," by Fazle Elahi Faisal, Han Zhao, and Tijana Milenković, takes a comparative genomic approach to the dynamics of aging. They evaluate the algorithmic and scoring parts of existing network alignment methods, propose a new global network alignment approach that combines the top performing algorithm with the best scoring function, and demonstrate its impact in deriving new human protein-protein interaction network information from model organisms. This work allows for novel inferences about aging-related network dynamics that are difficult to study directly in human subjects.

Networks and network-based methods are a mainstay of systems biology research. Nevertheless, such research does

- *S. Aluru is with the School of Computational Science and Engineering, Georgia Institute of Technology, 266 Ferst Drive, Atlanta, GA 30332. E-mail: aluru@cc.gatech.edu.*
- *D. Slonim is with the Department of Computer Science, Tufts University, 161 College Avenue, Medford, MA 02155. E-mail: Donna.Slonim@tufts.edu.*

For information on obtaining reprints of this article, please send e-mail to: reprints.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2015.2389551

not always sufficiently account for the fact that biological networks are not known perfectly, especially those computationally inferred from data. The paper "Reachability Analysis in Probabilistic Biological Networks" by Haitham Gabr, Andrei Todor, Alin Dobra, and Tamer Kahveci addresses this issue in the context of studying the reachability problem in signaling networks. The problem is to determine whether a signal travels from a set of receptor molecules to a set of reporter molecules in a signaling network, where each edge is associated with a probability indicating likelihood of the underlying interaction. While this problem is deceptively simple on the surface, analyzing a probabilistic network with n edges is equivalent to analyzing 2^n deterministic networks! Indeed, the authors prove that the problem is #P-complete. As a key contribution, the authors develop an elegant polynomial representation for the edges, allowing the reachability problem to be solved by multiplying a polynomial-sized subset of terms. They also provide a space-efficient means for computing it. The technique is likely to be of broader utility.

Multiple sequence alignment is a widely studied problem in bioinformatics. In their paper "GLProbs: Aligning Multiple Sequences Adaptively", authors Yongtao Ye, David W.-L. Cheung, Yadong Wang, Siu-Ming Yiu, Qing Zhang, Tak-Wah Lam, and Hing-Fung Ting propose a new twist to this old problem. Their approach analyzes sequences to first determine whether they should be aligned globally or locally, and then adaptively utilizes a progressive alignment approach to construct the multiple sequence alignment. Using MSAProbs with pair-HMMs adapted in this way, the authors show superior results to many widely used multiple sequence aligners, particularly for sequences with significant divergence.

Mining electronic health records (EHRs) to develop predictive diagnoses and model disease progression is a bourgeoning field in data-driven health informatics. Hui Li, Xiaoyi Li, Murali Ramanathan, and Aidong Zhang develop risk models for bone diseases in "Prediction and Informative Risk Factor Selection of Bone Diseases". They developed a generalized risk factor learning framework using restricted Boltzmann machines and multi-layer perceptrons to learn the combinatorial risk factors that have the best predictive value. They also develop the concept of disease memory models to capture risk factors for different patient cohorts.

The paper "Capturing Uncertainty by Modeling Local Transposon Insertion Frequencies Improves Discrimination of Essential Genes" by Michael A. DeJesus and Thomas R. Ioerger received the Best Paper Award at the conference. The authors developed a Bayesian model to more accurately predict the essentiality of individual genes in a microbial organism based on transposon mutagenesis data. This is an important step in countering pathogenic organisms by targeting genes that are essential for their survival. Transposon mutagenesis is used to randomly insert transposon sequences within the genome, and test if any potential gene interruption caused by the insertions is harmful to the survival of the organism. Deep sequencing is then used to study insertion locations. Prior models assumed uniform probability of insertion throughout the genome. The authors' work estimates local frequencies based on experimental data, and uses this information to develop a hierarchical Bayesian method to better identify genes essential for survival.

Overall, the papers published in this special section contain an eclectic mix of topics and provide a snapshot of work being carried out at the leading edge of the field. We learned a great deal by reading the works in this collection, and we hope that the readers find them to be equally valuable.

Srinivas Aluru, *Guest Editor*
Donna K. Slonim, *Guest Editor*

Srinivas Aluru is a professor in the School of Computational Science and Engineering within the College of Computing at the Georgia Institute of Technology. Earlier, he held faculty positions at Iowa State University, the Indian Institute of Technology Bombay, New Mexico State University, and Syracuse University. He conducts research in high-performance computing, bioinformatics and systems biology, combinatorial scientific computing, and applied algorithms. He pioneered the development of parallel methods in computational biology, and contributed to the assembly and analysis of complex plant genomes. He is a fellow of the American Association for the Advancement of Science (AAAS) and the IEEE.

Donna K. Slonim received the BS degree in computer science from Yale University, the MS degree in computer science from the University of California at Berkeley, and the PhD degree in computer science from the Massachusetts Institute of Technology. She is on the faculty at Tufts University in Medford, Massachusetts, where she is an associate professor of computer science in the School of Engineering and a member of the Genetics faculty at the Sackler School of Graduate Biomedical Sciences at Tufts University School of Medicine. She previously worked in the pharmaceutical industry on computational pharmacogenomics. Her current research focuses on computational methods for translational medicine, including network dynamics, transcription, and disease gene discovery and characterization.