

Great Lakes Bioinformatics Conference (GLBIO) 2015 Special Section Editorial

Tijana Milenković¹ and Sarath Chandra Janga²

THE GLBIO conference is organized by the Great Lakes Bioinformatics Consortium to provide an interdisciplinary forum for the discussion of research findings and methods in bioinformatics. An important goal for the conference is to foster long term collaborative relationships and networking opportunities within the domain of computational approaches to biology. In 2015, GLBIO was held during May 18-20, 2015 at Purdue University in West Lafayette, IN, USA. With over 250 attendees, outstanding education, the celebration of GLBIO's 10th anniversary, and 118 poster presentations, GLBIO 2015 was a rousing success. A call for submissions on a wide range of topics was made, and those full length papers that were accepted by the conference committee for oral talks at the conference and whose topics fell under the scope of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* were invited to submit to this GLBIO 2015 special section. We thank all of the reviewers of the papers, whose invaluable efforts contributed to the high quality of the conference program as well as the GLBIO 2015 special section. The following papers are a part of the special section.

The paper "Encoding Data Using Biological Principles: The Multisample Variant Format for Phylogenomics and Population Genomics" by James B. Pease and Benjamin K. Rosenzweig won the GLBIO 2015 Best Paper Award. The rapid progress of phylogenomics and population genomics has driven increase in the size of multi-genomic datasets and thus the complexity of genome-wide analyses. So, the authors introduce the Multisample Variant Format (MVF), specifically designed to store multiple sequence alignments for phylogenomics and population genomic analyses. MVF allows for encoding of aligned sites with specific biological information, rapid filtering and quality control of data, speed-ups of many computational analyses, effective data sharing, fast complex genome-wide visualizations, etc.

- T. Milenković is with Department of Computer Science and Engineering, Eck Institute for Global Health, and Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: tmilenko@nd.edu.
- S. C. Janga is with Department of Biohealth Informatics, School of Informatics and Computing, Indiana University-Purdue University (IUPUI) and Center for Computational Biology and Bioinformatics, Indiana University's School of Medicine, Indianapolis, IN 46202, USA. E-mail: scjanga@iupui.edu.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2018.2849800

In the paper titled "Towards Extracting Supporting Information About Predicted Protein-Protein Interactions," Adam Roth, Sandeep Subramanian, and Madhavi K. Ganapathiraju deal with the problem of identifying information about protein-protein interactions (PPIs) from biomedical literature. Specifically, the authors introduce UPSITE, a text mining tool for extracting from PubMed abstracts evidence in support of a query PPI. With the development of computational methods for PPI prediction, it is hard to validate the predicted PPIs. UPSITE can help with this, by using each predicted PPI as its query, thus relieving the burden of validating the predicted PPIs via biological experiments.

Satwica Yerneni, Ishita K. Khan, Qing Wei, and Daisuke Kihara introduce a new computational method for predicting PPIs in their paper titled "IAS: Interaction Specific GO Term Associations for Predicting Protein-Protein Interaction Networks." The new method can supplement or verify PPIs identified experimentally. The authors develop a novel scoring method for predicting PPIs from Gene Ontology (GO) annotations of proteins, which unlike existing methods that consider functional similarity as an indication of interaction between proteins, considers GO term associations of known interacting protein pairs in multiple organisms.

In the paper "A Framework for Identifying Genotypic Information from Clinical Records: Exploiting Integrated Ontology Structures to Transfer Annotations between ICD Codes and Gene Ontologies," Seyedsasan Hashemikhabir, Ran Xia, Yang Xiang, and Sarath Chandra Janga present a generic framework for integrating diverse ontologies to facilitate annotation transfer and apply the developed approach to merge GO and International Classification of Disease 9 codes (ICD9). Mapping the diseases to GO terms, the authors extend disease terms to human gene annotations and demonstrate using currently available annotations from Malacards and disease ontology annotations for human genes that their mappings are significantly enriched with known annotations and can also uncover novel associations. *Disclaimer: since this paper is co-authored by Guest Editor Sarath Chandra Janga, to avoid any conflict of interest, its review process was handled solely by Guest Editor Tijana Milenković, through her independent TCBB Associate Editor account rather than through the shared GLBIO 2015 Guest Editor account.*

In the paper "REPA: Applying Pathway Analysis to Genome-Wide Transcription Factor Binding Data," Pranjal Patra, Tatsuo Izawa, and Lourdes Peña-Castillo study the problem of linking the biological phenomenon in question

to the underlying molecular pathways. Genome-wide transcription factor binding data have become increasingly available. Hence, the authors develop REPA (Regulatory Enrichment Pathway Analysis) to apply gene set enrichment analysis to genome-wide transcription factor binding data, in order to infer associations between transcription factors and biological pathways.

Rodrigo F. Ramalho, Sujun Li, Predrag Radivojac, and Matthew W. Hahn in their paper “Proteomic Evidence for In-Frame and Out-of-Frame Alternatively Spliced Isoforms in Human and Mouse” seek evidence for translation of alternatively spliced transcripts, especially those that result in a change in reading frame. With this goal in mind, the authors collect exon-skipping cases previously found by RNA-Seq and apply a computational approach to screen millions of mass spectra coming from seven human and six mouse tissues. This work suggests that both in-frame and out-of-frame translation may be used to regulate protein activity or localization.

In the paper titled “Discovering Gene Regulatory Elements Using Coverage-Based Heuristics,” Rami Al-Ouran, Robert Schmidt, Ashwini Naik, Jeffrey Jones, Frank Drews, David Juedes, Laura Elnitski, and Lonnie Welch recognize an issue with existing genomic regulatory motif discovery algorithms, which often produce very long lists of putative transcription factor binding sites, thus making it difficult to select a manageable set of candidate motifs for experimental validation. To address this issue, the authors introduce the motif selection problem and provide coverage-based search heuristics for its solution. The new algorithms reveal new biological insights about the regulatory code of the human genome and provide new insights into the biology of hepatocellular carcinoma and multiple sclerosis.



Tijana Milenkovic received the PhD degree in computer science from UC Irvine, in 2010. She is an associate professor of computer science and engineering. She has been a Notre Dame faculty since 2010. She leads the Complex Networks (CoNe) Lab, solving challenging problems in network science, data mining, algorithms, computational biology, social networks, etc. She won the 2016 AFOSR YIP and 2015 NSF CAREER awards, and has been the lead PI on three other NSF or NIH grants. She has been an associate editor of *IEEE/ACM Transactions on Computational Biology and Bioinformatics* since 2014. She is committed to increasing participation of women in computer science.



Sarath Chandra Janga received the PhD degree from the MRC Laboratory of Molecular Biology and the University of Cambridge, in 2010. He is an associate professor in the Department of Bio-health Informatics, School of Informatics and Computing, Indiana University-Purdue University, Indianapolis (IUPUI). He has been a faculty at IUPUI since 2012. He leads a Regulatory Genomics and Systems Biology Group. His research interests include understanding the design principles and inferring post-transcriptional gene regulatory networks within the broader field of genomics and systems biology his lab works on. Research and community service initiated by his lab is supported by multi-year NIH, NSF, and foundation grants.

▷ **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**