

Guest Editors Introduction to the Special Section on ISBRA 2014

Mitra Basu, Yi Pan, and Jianxin Wang



THIS special section consists of a selection of papers presented at the 10th International Symposium on Bioinformatics Research and Applications (ISBRA 2014), which was held at Zhangjiajie, China, June 28-30, 2014. ISBRA 2014 received 119 full paper submissions from over 20 countries and every paper went through a very rigorous review process. It provided a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications. After a careful consideration, 48 papers were accepted as regular papers. A number of authors were invited to submit extended versions of their symposium papers to this special section. Following a rigorous review process, seven papers were selected for publication.

CRYO-ELECTRON microscopy (Cryo-EM), in combination with three-dimensional (3D) image processing techniques, is progressively becoming a more reliable and efficient tool for determining the structural characterization of biological macromolecular complexes. In "A Two-Phase Improved Correlation Method for Automatic Particle Selection in Cryo-EM," Fa Zhang, Yu Chen, Fei Ren, Xuan Wang, Zhiyong Liu, and Xiaohua Wan presented a two-phase improved correlation method for automatic, fast particle selection by combining the advantages of feature-based methods and template-matching methods. Several optimization strategies such as a modified ADAboost algorithm, divide and conquer technique, cascade strategy, and GPU parallel technique are employed to improve feature recognition ability and reduce processing time. Experimental results on the benchmark of Cryo-EM images show that their method can improve the accuracy and processing speed of particle selection significantly.

During the past two decades, mass spectrometry has gradually become a standard technique for the high-throughput characterization of large biomolecules, including peptides

and proteins. Mixture spectra occur quite frequently in a typical wet-lab mass spectrometry experiment, which result from the concurrent fragmentation of multiple precursors. The ability to efficiently and confidently identify mixture spectra is essential to alleviate the existent bottleneck of low mass spectra identification rate. In "An Approach for Peptide Identification by *De Novo* Sequencing of Mixture Spectra," Yi Liu, Bin Ma, Kaizhong Zhang, and Gilles Lajoie formulated the mixture spectra *de novo* sequencing problem mathematically, and proposed a dynamic programming algorithm for the problem. Both simulated and real mixture spectra datasets are used to demonstrate the meritorious aspects of their proposed algorithm.

De novo peptide sequencing using tandem mass spectrometry (MS/MS) data has become a major computational method for sequence identification in recent years. With the development of new instruments and technology, novel computational methods have emerged with enhanced performance. In "NovoExD: *De Novo* Peptide Sequencing for ETD/ECD Spectra," Yan Yan, Anthony J. Kusalik, and Fang-Xiang Wu presented a *de novo* sequencing method, called NovoExD, for ECD/ETD spectra. NovoExD applies a new form of spectrum graph with multiple edge types (called a GMET), considers multiple peptide tags, and integrates amino acid combination (AAC) and fragment ion charge information. Experiments conducted on three different datasets show that the average full length peptide identification accuracy of NovoExD is as high as 88.70 percent, and that NovoExDs average accuracy is more than 20 percent greater on all datasets than that of the competing method pNovo+.

In recent years, a remarkable amount of protein-protein interaction (PPI) data are being available owing to the advance made in experimental high-throughput technologies. However, the experimentally detected PPI data usually contain a large amount of spurious links, which poses new challenges to both computational and biological scientists. In "Identifying Spurious Interactions in the Protein-Protein Interaction Networks Using Local Similarity Preserving Embedding," Lin Zhu, Su-Ping Deng, Zhu-Hong You, and De-Shuang Huang developed a new embedding algorithm called Local Similarity Preserving Embedding (LSPE) to rank the interaction possibility of protein links. Their experimental results on benchmark PPI networks demonstrated that LSPE outperformed the state-of-the-art methods for eliminating topological false links.

In recent years, the cost of high-throughput sequencing technology including metagenomics sequencing and rRNA

- M. Basu is with the Computer & Information Science & Engineering Directorate, National Science Foundation, 4201 Wilson Blvd., Arlington, VA 22230. E-mail: mbasu@nsf.gov.
- Y. Pan is with the Department of Computer Science, Georgia State University, Atlanta, GA 30302. E-mail: yipan@gsu.edu.
- J. Wang is with the School of Information Science and Engineering, Central South University, Changsha 410083, China. E-mail: jxwang@mail.csu.edu.cn.

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.2017.2676859

(16S or 18S) target sequencing has decreased dramatically, which made the studies of microbiome a promising field. Microbiome datasets are often comprised of different representations or views which provide complementary information to understand microbial communities, such as metabolic pathways, taxonomic assignments and gene families. In "Microbiome Data Representation by Joint Nonnegative Matrix Factorization with Laplacian Regularization," Xingpeng Jiang, Xiaohua Hu, and Weiwei Xu proposed a LJ-NMF method to integrate functional and phylogenetic profiles from HMP for a comprehensive view of a given microbiome study. The experimental results indicate that the proposed method offers a more efficient framework for microbiome data analysis, compared to other variants of NMF methods.

Although various biological resources are available, due to function diversity of proteins, it is very challenging to design effective methods that make good use of those various biological resources for annotating functions of unknown proteins. In "Predicting Protein Functions by Using Unbalanced Random Walk Algorithm on Three Biological Networks," Wei Peng, Min Li, Lu Chen, and Lusheng Wang proposed a new method named ThrRW, which takes several steps of random walking on three different biological networks: protein interaction network (PIN), domain co-occurrence network (DCN), and functional interrelationship network (FIN), respectively, so as to infer functional information from neighbors in the corresponding networks. The results of experiment on *S. cerevisiae* data shown that their method achieves better prediction performance not only than the methods that consider both PIN data and GO term similarities, but also than the methods using both PIN data and protein domain information.

Essential proteins are indispensable for the survival or reproduction of an organism. With the development of high-throughput techniques, a large number of protein-protein interaction data are available, which promotes the studies of essential proteins from the network level. In "United Complex Centrality for Identification of Essential Proteins from PPI Networks," Min Li, Yu Lu, Zhibei Niu, and Fang-Xiang Wu proposed a new method, United complex Centrality (UC), to identify essential proteins by integrating the protein complexes with the topological features of protein-protein interaction networks. The experimental results shown that the proposed method UC performs obviously better than the eight previously proposed methods (DC, IC, EC, SC, BC, CC, NC, and LAC) for identifying essential proteins.

As the guest editors of this special section, we would like to thank the Program Committee members and external reviewers for volunteering their time to review and discuss symposium papers. We would like to extend special thanks to the Steering and General Chairs of the symposium for their leadership, and to the Finance, Publication, Publicity, and Local Organization Chairs for their hard work in making ISBRA 2014 a successful event. Furthermore, we would like to thank the Editor-in-Chief, Dr. Ying Xu and Associate Editor-in-Chief, Dr. Dong Xu for providing us with the opportunity to showcase some of the exciting research presented at ISBRA 2014 in the *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Last but not the

least, we would like to thank all of the ISBRA 2014 authors. The symposium could not continue to thrive without their high quality contributions.

Mitra Basu
Yi Pan
Jianxin Wang
Guest Editors



Mitra Basu received the PhD degree in electrical engineering from Purdue University. She joined the US National Science Foundation (NSF)'s Directorate for Computer, Information Science and Engineering (CISE), in 2008 as the lead program director for the Expeditions in Computing program, one of the flagship programs in CISE. During her tenure at the US National Science Foundation, she has served as acting deputy division director in the Divisions of Computer and Network Systems and Intelligent Information Systems. Prior to her US National Science Foundation appointment, she was a tenured faculty member at the City University of New York (CUNY). In 2006, she joined the United States Naval Academy as a visiting chaired professor of information technology. Currently, she holds a visiting professors position in the Computer Science Department, Johns Hopkins University. She has published multiple research articles and book chapters in image processing/pattern recognition, neural networks/learning theory, and computational biology/bioinformatics, and has served on numerous conference program committees in these areas. She received the NSF Service and Director's awards. She is a senior member of the IEEE.



Yi Pan received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, in 1991. He is a Regents' professor of computer science and an interim associate dean and chair of biology with Georgia State University. He joined Georgia State University, in 2000 and was promoted to full professor in 2004, named a distinguished university professor in 2013, and designated a Regents' professor (the highest recognition given to a faculty member by the University System of Georgia), in 2015. He served as the chair of Computer Science Department from 2005-2013. He is also a visiting Changjiang chair professor with Central South University, China. His profile has been featured as a distinguished alumnus in both the Tsinghua Alumni Newsletter and the University of Pittsburgh CS Alumni Newsletter. His research interests include parallel and cloud computing, wireless networks, and bioinformatics. He has published more than 330 papers including more than 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 8800 times. He has served as an editor-in-chief or editorial board member of 15 journals including seven IEEE Transactions. He received many awards including IEEE Transactions Best Paper Award, four other international conference or journal Best Paper Awards, four IBM Faculty Awards, two JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at more than 50 international conferences around the world.



Jianxin Wang received the BS and MS degrees in computer science from the Central South University of Technology, P. R. China, and the PhD degree in computer science from Central South University. Currently, he is the vice dean and a professor in the School of Information Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network. He has published more than 200 papers in various international journals and refereed conferences. He is a senior member of the IEEE.