Systems biology approaches in biological and biomedical research: opportunities and challenges

Zhongming Zhao*

Departments of Biomedical Informatics, Psychiatry, and Cancer Biology, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA Email: zhongming.zhao@vanderbilt.edu *Corresponding author

Rui Jiang

Ministry of Education Key Laboratory of Bioinformatics and Bioinformatics Division, Department of Automation and Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China Email: ruijiang@tsinghua.edu.cn

Huiru Zheng

Computer Science Research Institute, School of Computing and Mathematics, University of Ulster, BT37 0QB, UK Email: h.zheng@ulster.ac.uk

Biographical notes: Zhongming Zhao received his PhD in Human and Molecular Genetics from the University of Texas Health Science Center at Houston and MD Anderson Cancer Center, Houston, Texas in 2000. Currently, he is an Ingram Associate Professor of Cancer Research and an Associate Professor in the Departments of Biomedical Informatics, Psychiatry, and Cancer Biology at the Vanderbilt University Medical Center. He has published more than 130 papers. His research interests include bioinformatics and systems biology approaches to studying complex diseases, personalised medicine, pharmacogenomics, and biomedical informatics.

Rui Jiang received his PhD in Automatic Control Engineering from Tsinghua University, China in 2002. He is currently an Associate Professor in the Department of Automation at Tsinghua University, China. His research interests include pattern recognition, machine learning, statistical genetics, bioinformatics, and systems biology.

182 Z. Zhao, R. Jiang and H. Zheng

Huiru Zheng (PhD, MSc, BEng) is a Senior Lecturer in Computer Science at the University of Ulster, UK. Her research interests include healthcare informatics, bioinformatics and systems biology, particularly the application of data mining and machine learning in intelligent data analysis. She has over 100 publications in journals and conferences in these areas.

The last decade has witnessed the exponential growth of omics data, thanks to the unprecedented advancement of biotechnologies, especially microarray, mass spectrometry, and next generation sequencing technologies. We have also seen the release of numerous knowledge based databases such as Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases. Accordingly, integrative and systems biology approaches have been demonstrated powerful in the analysis of all kinds of omic datasets. Demands on systems biology approaches will be even stronger in the near future, as the ultimate goal in biological and biomedical research is to understand how the biological system works at the cellular and organism levels. To provide a forum for promoting integrative data analysis approaches in systems biology, we organised a workshop featuring the theme of 'integrated approach' and 'complex biological system'. The first workshop, entitled 'Integrative Data Analysis in Systems Biology', was held on 18 December 2011 in Hong Kong, China, in conjunction with the IEEE International Conference on Bioinformatics & Biomedicine (BIBM) 2010. In November 2011, we organised the second workshop in Atlanta, Georgia, USA, in conjunction with the BIBM 2011. Our call for papers and attendance has been well received by the communities, with a total 43 papers being submitted to the first workshop and 19 papers to the second workshop. We had a full day workshop on 12 November 2012, including four sessions (Integration, Tool and Proteomics, Genomics, and Network). In this special issue, we included nine papers that were selected from the original submissions and were accepted after rigorous peer reviews of their substantially expanded version of the manuscripts. Below, we briefly summarise these papers in this special issue.

Wnt signalling pathway is one of the most important signalling pathways, which regulates embryogenesis and differentiation. While the core members are well defined, its broader network has not been yet. Saha et al. attempted to identify novel members of the Wnt regulatory network and to identify sub-networks of the larger Wnt signalling network that are active in different biological contexts using complementary computational approaches. Their approaches might be useful to identify promising novel Wnt signalling proteins and define Wnt sub-networks in specific conditions or diseases.

Identification of disease related single nucleotide polymorphisms (SNPs) is very important for understanding the disease mechanisms and developing better clinical therapies. Jiang et al. developed an integrative method based on sequence conservation and domain-domain interaction network to prioritise candidate non-synonymous SNPs (nsSNPs) and to test the common disease rare variant hypothesis (CD-RV). Their work showed the power to detect the disease related nsSNPs.

Dempsey et al. presented a network model for analysing experimental data aiming to identify clusters and nodes of high relevance using GO annotations. Their results indicated that their method may help detect causative relationships based on GO annotations. The proposed approach, which is different from the traditional ones, is useful for uncovering true functional relationships in the experimental data.

Introduction

In another GO related work, Devignes et al. evaluated functional fuzzy classification of genes using *IntelliGO* semantic distance with the help of reference sets, which consist of genes from KEGG pathways and Pfam clans in two species (human and yeast). The authors systematically compared three classification approaches (hierarchical ascending clustering, fuzzy C-means clustering, and popular tool DAVID – Database for Annotation Visualisation and Integrated Discovery).

Yousri and Elkaffash proposed a combined t-test score and contingency based measures for identifying association functional groups to multiple clinical types. They applied this approach to breast cancer genes to investigate association between functional group and clinical status.

Jaromczyk et al. tackled the fundamental protein-coding sequence search by introducing a linear-time, deterministic algorithm for finding a longest open reading frame (ORF) in an alternatively spliced gene represented by a splice graph. By pruning the search space, the authors could find the longest ORF among all splicing variants using their algorithm.

Kong et al. presented an integrated computational tracking algorithm for quantitative analysis of microtubule dynamics in live-cell microscopy images. The authors also proposed a web interface that replicates their current manual analysis work flow of microtubule images, and demonstrated the added functionalities of the interface including how to share image data and analysis results. The system has the capability to integrate other automated MT analysis algorithms.

Post-translational modifications (PTMs) play most important roles in the accomplishment of biological processes and molecular functions. Currently, it is challenging to identify two PTMs from a tandem mass spectrum. Li et al. tackled this issue by proposing a new algorithm based on a Pair of Peak Set (PPS) and a scoring function to detect two PTMs with unknown types. The authors demonstrated their method by applying it to a large tandem mass spectrometry (MS/MS) dataset.

In the last paper, Li and Liu developed an algorithm based on graph search to generate short peptide sequence tags by taking advantage of pairs of peaks with high intensities in tandem mass spectrometry. Specifically, the authors applied the PPS proposed in the previous paper and used the pair peak values of highest intensities as the root of a tree in the algorithm. In their evaluation using the 2420 experimental MS/MS spectra with two PTMs, the authors showed that their algorithm could achieve better accuracy than the PepNovo approach as well as with higher efficiency.

We thank the BIBM conference organisers, especially Dr. Tony Hu, for the continuous and strong support of our IDASB workshops during the past two years. We also would like to thank numerous experts who provided us valuable and timely advice and critical review comments, which helped improve the quality of the manuscripts. Finally, we thank Ms. Liz Harris, Mr. Jim Corlett, and other editorial staff for their support and encouragement.

The Second Workshop on Integrative Data Analysis in Systems Biology (12 November 2011, Atlanta, Georgia, USA, http://bioinfo.mc.vanderbilt.edu/idasb/)

Organisers: Zhongming Zhao, Huiru Zheng, Rui Jiang

Program Committee: Chairs: Zhongming Zhao (Vanderbilt University, USA), Huiru Zheng (University of Ulster, UK), Rui Jiang (Tsinghua University, China). Members: Patrizo Arrigo (National Research Council, Italy), Michael Berthold (University of

184 Z. Zhao, R. Jiang and H. Zheng

Konstanz, Germany), Jaine Blayney (Queen's University Belfast, UK), Minou Bina (Purdue University, USA), Danail Bonchev (Virginia Commonwealth University, USA), Fiona Browne (Queen's University Belfast, UK), Jack Chen (Simon Fraser University, Canada), Minghua Deng (Peking University, China), Werner Dublizky (University of Ulster, UK), Xiaodan Fan (Chinese University of Hong Kong, China), Jan Freudenberg (Feinstein Institute for Medical Research, Northshore-LIJ Health System, USA), Chittibabu (Babu) Guda (University of Nebraska Medical Center, USA), Yufei Huang (University of Texas at San Antonio, USA), Che-lun Hung (Providence University, Taiwan), Peilin Jia (Vanderbilt University, USA), Hyunju Lee (Gwangju Institute of Science and Technology, Korea), Kang Li (Queen's University Belfast, UK), Shao Li (Tsinghua University, China), Xiaohui Liu (Brunel University, UK), Piyush Ojha (University of Ulster, UK), Antonio Sanfilippo (Pacific Northwest National Laboratory, USA), Youqiang Song (University of Hong Kong, China), Qingguo Wang (Vanderbilt University, USA), Jun Wan (Johns Hopkins University, USA), Haiying Wang (University of Ulster, UK), Yong Wang (Chinese Academy of Sciences, China), Wanling Yang (University of Hong Kong, China), Ying Wang (Xiamen University, China), Hong Yue (University of Strathclyde, UK), Xingming Zhao (Shanghai University, China), and Shihua Zhang (Chinese Academy of Sciences, China).