Editorial

Huiru Zheng

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

Zhongming Zhao

Departments of Biomedical Informatics,
Psychiatry, and Cancer Biology,
Vanderbilt University Medical Center,
Nashville, Tennessee 37232, USA
E-mail: zhongming.zhao@vanderbilt.edu

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
E-mail: ruijiang@tsinghua.edu.cn

Biographical notes: Huiru Zheng (PhD, MSc, BEng) is a 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.

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 Associate Professor in the Departments of Biomedical Informatics, Psychiatry, and Cancer Biology at the Vanderbilt University Medical Center. His research interests include bioinformatics and systems biology approaches to studying complex diseases, genome-wide or large-scale analysis of genetic variation and methylation patterns, next-generation sequencing data analysis, comparative genomics 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.

Systems biology is a field in biology aiming at system-level understanding of biological processes. Systems biology studies biological systems by systematically perturbing them (biologically, genetically, or chemically); monitoring the gene, protein and informational pathway responses; integrating these data and ultimately, formulating mathematical models that describe the structure of the system and its response to individual perturbations. Integrated 'omics' approach has created exciting opportunities for systems biology researchers. To provide a forum for discussion on integrated data analysis approach in systems biology research such as pattern recognition, prediction and data representation and visualisation, we organised a workshop featuring the theme of 'integrated approach' and 'complex biological system'. The workshop was held on 18 December 2011 in Hong Kong, China, in conjunction with the IEEE International Conference on Bioinformatics & Biomedicine (BIBM) 2010. Researchers were invited to submit their original work in the following areas:

- large-scale or cross-species data integration for the reconstruction of networks and pathways
- quantitative understanding of dynamics of regulatory, signalling, interaction and metabolic networks through modelling and simulation techniques
- prediction of protein/RNA structure and biological networks interactions
- data integration and knowledge-driven approach in biomarker identification and drug discovery
- enhancement and enablement of knowledge discovery in functional genomics of disease and other phenotypes through integrated omics approach
- semantic webs and ontology-driven biological data integration methods
- development of integrated systems biology visualisation and analysis tools.

Our call for papers received excellent responses; in total 43 papers were submitted to the workshop. After rigorous reviews by our programme committee, 15 papers were selected to present in our workshop at BIBM 2010. A lunch-time panel discussion was organised and chaired by Dr. Huiru Zheng and Dr. Rui Jiang. The panel discussion topics included: embedding information in DNA; small RNA assembly; progress of visualisation research in systems biology data analysis and multi-scale data generation and modelling.

All workshop authors were invited to expand the scope of their workshop contributions and submit revised versions with new significant results to this special issue. All the papers submitted underwent peer review and six papers were finally included in this special issue.

Glioblastoma is the most common and most lethal brain tumour in humans. Gong et al. investigated the possible cooperative deregulation of microRNAs (miRNAs) and Transcription Factors (TFs) in glioblastoma, under the hypothesis that miRNAs and TFs might have a combinational regulatory effect on glioblastoma genes. They integrated glioblastoma-related miRNAs, TFs and genes in the analysis of glioblastoma-specific regulatory networks and identified 54 Feed-Forward Loops (FFLs). Their results revealed some functions important to carcinogenesis and also some unique functions specific to the FFLs identified.

Incorporating biological meanings in genetic regulatory modules analysis remains a challenge in systems biology. In their study, Tang et al. proposed a structural model-based clustering method for integrative identification of core regulatory modules, which defines several activity measures for network nodes and associates those measures with a weighted clustering algorithm. The method can be applied to analyse both dynamic model systems and expression data, especially together and in consideration of inherent biological meanings. The authors verified the method on diverse model systems.

Moving to the research on the complex gene regulatory network, Hong et al. constructed gene co-expression networks using Gaussian graphic models to identify the ovarian cancer-related genes in two independent ovarian cancer studies. The authors applied decision tree cut algorithm in module identification and presented detail finding on the analysis of the two data sets, including the identification of several common hub genes and four common conserved modules, which were enriched with the genes in four cancer-related, and two ovarian cancer susceptibility genes: *CCEN2* and *BRIC5*.

Hung et al. tackled the research challenge on multi-sequence alignment and proposed a heuristic coding region alignment method, namely CORAL-M, for multiple genome sequences, especially for coding regions. In comparison with previous algorithm developed by the authors, CORAL-M adopts a probabilistic filtration model and the local optimal solution to align genome sequences by the sliding windows to obtain the near-optimal alignment in linear time. In their paper, authors stated that the proposed method can be used to search more potential functional sites.

Wang and Yan introduced a *k*-means-based geometric biclustering (KGBC) algorithm for DNA microarray data analysis, where the Hough transform in column-pair space is used to find sub-biclusters and the *k*-means method is used to optimise the combining process. Authors evaluated the KGBC algorithm on both simulated and real microarray data with different setting of noise levels and overlapping degrees.

The last paper is presented by Tjhi et al. on analysing cell-based screening images for phenotype identification in drug and siRNA study. Authors present a unique analysis method for an exploratory enquiry without presumptions of the expected phenotypes for intricate parameter settings. The authors detailed the analysis steps with the focus on visualisation in the paper. A website was provided to demonstrate the process and case study on two Golgi apparatus screens with 71 drugs and siRNAs being presented.

This workshop and special issue was not possible without the support of BIBM conference organisers or the efforts of many experts who provided valuable advice and critical review comments. We are especially grateful to the anonymous reviewers who helped to improve the quality of the manuscripts.

Workshop: Integrative Data Analysis in Systems Biology

(18 December, 2010, Hong Kong, China, http://rosalind.infj.ulst.ac.uk/idasb/idasb.html)

Organisers: Huiru Zheng, Zhongming Zhao, Rui Jiang

Programme Committee: Huiru Zheng (co-chair, University of Ulster, UK), Zhongming Zhao (co-chair, Vanderbilt University, USA), Rui Jiang (co-chair, Tsinghua University, China), PatrizoArrigo (National Research Council, Italy), JaineBlayney (Queen's University Belfast, UK), Danail Bonchev (Virginia Commonwealth University, USA), Fiona Browne (i-Path Dianostics Ltd, UK), Minghua Deng (Peking University, China), Werner Dublizky (University of Ulster, UK), Xiaodan Fan (Chinese University of Hong Kong, China), Kun Huang (Ohio State University Medical Center, USA),

Kang Li (Queen's University Belfast, UK), Shao Li (Tsinghua University, China), Xiaohui Liu (Brunel University, UK), PiyushOjha (University of Ulster, UK), Antonio Sanfilippo (Pacific Northwest National Laboratory, USA), Youqiang Song (University of Hong Kong, China), Jingchun Sun (Vanderbilt University Medical Center, USA), Momiao Xiong (University of Texas Health Science Center at Houston, USA), Jun Wan (Johns Hopkins University School of Medicine, USA), Haiying Wang (University of Ulster, UK), Yong Wang (Chinese Academy of Sciences, China), Wanling Yang (University of Hong Kong, China), Hong Yue (University of Strathclyde, UK) and Xingming Zhao (Shanghai University, China).