Editorial

Zhongming Zhao

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

Bing Zhang

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

Yufei Huang

Department of Electrical and Computer Engineering,
The University of Texas at San Antonio,
San Antonio, TX 78249, USA
E-mail: yufei.huang@utsa.edu

Hua Xu

School of Biomedical Informatics,
The University of Texas Health Science Center at Houston,
Houston, TX 77030, USA
E-mail: hua.xu@uth.tmc.edu

Jason E. McDermott

Computational Biology and Bioinformatics Group,
Pacific Northwest National Laboratory,
Richland, WA 99352, USA
E-mail: jason.mcdermott@pnnl.gov

Biographical notes: Zhongming Zhao received his PhD in Human and Molecular Genetics from the graduate school of 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. His research interests include bioinformatics and systems biology approaches to studying complex diseases, personalised medicine, pharmacogenomics, and biomedical informatics.

Bing Zhang received his PhD in Molecular Genetics from the Chinese Academy of Sciences, Shanghai, China in 1999. Currently, he is an Associate Professor in the Departments of Biomedical Informatics and Cancer Biology at the Vanderbilt University Medical Center. His research focuses on the development and application of integrative bioinformatics approaches to the study of complex diseases.

Yufei Huang received his PhD in Electrical Engineering from the State University of New York at Stony Brook in 2001. Since 2002, he has been with the Department of Electrical and Computer Engineering at the University of Texas at San Antonio, where he is now a Professor. His current research focus is on computational biology and computational neuroergonomics, including high throughput biomedical data modelling and integration, genomic regulatory networks, microRNA target and functional prediction, EEG data modelling and characterisation.

Hua Xu received his PhD in Biomedical Informatics from Columbia University at New York City in 2008. Currently he is an Associate Professor in the School of Biomedical Informatics at the University of Texas Health Science Center at Houston. His research interests include clinical natural language processing, biomedical text mining, and healthcare data mining.

Jason E. McDermott received his PhD in Microbiology and Immunology from Oregon Health & Science University in 2000. Currently he is a senior research scientist in the Computational Biology and Bioinformatics Group at the Pacific Northwest National Laboratory in Richland Washington. His current research focuses on network analysis, machine learning, and integrative analysis of high-throughput molecular data.

This special issue collects 10 papers submitted to *The International Conference on Intelligent Biology and Medicine 2013 (ICIBM 2013)*, which was held on 11–13 August, 2013 at the Holiday Inn at Vanderbilt in Nashville, Tennessee, USA. The ICIBM 2013, which was built on the success of ICIBM 2012, attracted more than 110 participants from many institutes in the world. The ICIBM program included four keynote speeches, six scientific presentation sessions, and two poster sessions. Additionally, for the education of next-generation researchers in interdisciplinary fields, we organised one workshop and two tutorials: “Workshop on Next-Generation Sequencing”, “Introduction to Proteome Informatics”, and “Pathway and Network Analysis Tutorial”. The six regular scientific sessions were:

- Next Generation Sequencing (NGS): Analysis and Tools
- Network Analysis
- Genomics
- Systems Biology
- Computational Medicine
- Intelligent Computing.

Forty manuscripts were presented in these six regular sessions, covering recent advances in the areas of bioinformatics, systems biology, medical informatics, and intelligent computing. As we did in ICIBM 2012, all the presenters were selected through a rigorous review of the novelty and significance of their work. The details are available on the conference website (http://bioinfo.mc.vanderbilt.edu/icibm2013/).

Year 2013 officially ushered in the big data era in the biomedical and biological research fields, after exponential growth of data for more than two decades. Among numerous big data events in 2013, we have witnessed an appointment of the first associate director for data science in National Institutes of Health (NIH) and rapid development of the Big Data to Knowledge (BD2K) program. The 10 original papers selected in this special issue describe recent advances in genomics, systems biology, and intelligent computing, reflecting the rapid advances in big data science. The other selected papers were included in other special issues including two supplemental issues to BMC Genomics (http://www.biomedcentral.com/bmcgenomics/supplements/14/S8) and BMC Systems Biology (http://www.biomedcentral.com/bmcsystbiol/supplements/7/S5). Below, we briefly summarise the 10 papers in this special issue.

Next-generation sequencing (NGS) is rapidly emerging as one of the main approaches for biological and biomedical research. Among the NGS platforms, microRNA sequencing (miRNAseq) is a relatively new form of transcriptome sequencing and there are many technical challenges to be solved. In the first paper of this issue, Guo et al. tackled the reproducibility issue of miRNAseq. Through comparing two popular microRNA isolation methods, mirVana and TRIzol, the authors found excellent repeatability of miRNAseq, and suggested mirVana method had better performance measured by the useful reads sequenced, number of miRNA identified, and reproducibility.

Kelemen et al. applied a unique combination of methods, including clustering, statistical methods, and a support vector machine, to test the hypothesis that an immune cell's movement pattern can convey key information about the cell's function, antigen specificity, and environment. Specifically, they applied those methods to classify different datasets of imaged fluorescently labelled T cells in mouse liver. The authors concluded that functional differences and environmental conditions could be predicted based on chosen movement attributes of the cells. Future work was proposed to study the T cell movement classification.

Koparde et al. presented the Bayesian-like Operational Taxonomic Unit eXaminer (BOTUX), a new tool for classifying 16S rRNA gene sequences into Operational Taxonomic Units (OTUs) that addresses the problem of overestimation caused by errors introduced during PCR amplification and DNA sequencing steps. Through the benchmarking tests using both the real and simulated datasets of 16S rDNA sequence, the authors demonstrated that BOTUX could accurately identify OTUs with comparable or better clustering efficiency and low execution times than other OTU algorithms.

Protein–protein interaction (PPI) data has been widely used in omic data analyses, especially for exploring the molecular mechanisms of cancer proteins' interactions. In this issue, Xiong et al. addressed an important issue of topological centrality of cancer proteins in the human interactome. The authors compared network properties (degree centrality, clustering coefficient, and betweenness centrality) of four sets of human proteins (cancer proteins, non-cancer disease proteins, non-cancer essential proteins, and control proteins) in yeast PPI network. Based on this study, the authors suggested that topological characteristics of cancer proteins in PPI network and quality of PPI data are

important, enhancing the previous studies of the features of cancer proteins in PPI network.

Localisation of a targeting embryo in an embryonic image is the first step in the analysis of spatial and temporal gene expression and gene–gene interactions. In this issue, Li and Ananta introduced an active contour based scheme to localise *Drosophila* embryos in RGB images. They proposed an algorithmic strategy to detect and restore open active contours and demonstrated the effectiveness of the proposed localisation scheme in experiments.

Prabhakaran et al. explored whether tRNA encoded in the genome of *Aeromonas* phages could facilitate the translation of phage proteins. They revealed considerable differences between synonymous codon preferences of GC-rich and GC-poor *Aeromonas* phages with respect to their host. Their phylogenetic analyses suggested that the presence of tRNAs in *Aeromonas* phages is a derived trait. They also found that tRNAs encoded in the phage genome correspond to synonymous codons overused in the phage genes but not in the host genes.

Differential expression analysis has been a standard approach for detecting differences in case-control transcriptomic data. Wang et al. studied a pair of alternate techniques for identifying genes of interest in case-control microarray experiments: differential normalised Shannon entropy and differential coefficient of variation. Through a systematic analysis of 16 human disease datasets, the authors demonstrated that these techniques could identify genes not found by differential expression, and those genes are likely to be disease-relevant. They suggested that these two techniques could serve as viable alternatives or complements to standard differential expression in an array analysis pipeline.

High throughput bacterial transcriptome experiments using NGS technology (RNASeq) generate data with extremely high and imbalanced sequencing coverage, which may lead to over- or under-estimation of gene expression levels in standard differential expression analysis. To tackle this problem, Zhang et al. evaluated strategies to identify expression differences of genes with high coverage in bacterial transcriptome data using either raw sequence reads or unique reads with duplicate fragments removed. They also proposed a generalised linear model (GLM) based approach to identify imbalance in read coverage based on sequence compositions. They showed that analysis using raw reads could identify more differentially expressed genes with more accurate fold change than using unique reads.

Chen et al. introduced a supervised learning-based method to mine the relationship between microbes and periodontitis with 16S rRNA sequences. They clustered the microbial sequence reads into OTUs and constructed the OTU-profile matrix. Then, the sample-taxa profile was used as the input feature to train a model for predicting the periodontal disease. They demonstrated the method could effectively predict disease status. This study helps us to better understand the association between microbes and diseases and the method is useful for disease diagnosis and forensics.

In the last paper, Andersen et al. introduced a generic framework to specify and execute strategies for the systematic exploration of spaces of graphs. Specifically, they used the Double Pushout formalism to derive new graphs. They applied the framework to the complex systems of chemical reactions. The strategies framework could serve not only as a convenient tool for exploration but also allow a detailed modelling of the constraints in chemical networks.

We thank all the reviewers for judging the scientific merits of the manuscripts submitted to ICIBM 2013 including this special issue. We are grateful to the local organisation committee members and volunteers for making ICIBM 2013 a success. We would like to acknowledge Drs. Kun Huang, Dongxiao Zhu, David Tabb, and Alexander Pico for organising the workshop and tutorials in ICIBM 2013. We thank the National Science Foundation (NSF grant IIS-1329380) and Vanderbilt Center for Quantitative Sciences for financial support of ICIBM 2013.