Introduction

Jianhua Ruan

Department of Computer Science,
The University of Texas at San Antonio,
San Antonio, Texas 78249, USA
Email: jianhua.ruan@utsa.edu

Wenjin Zheng

School of Biomedical Informatics,
The University of Texas Health Science Center at Houston,
Houston, TX 77030, USA
Email: wenjin.j.zheng@uth.tmc.edu

Zhandong Liu

Department of Pediatrics,
Baylor College of Medicine,
Houston, TX 77030, USA
Email: Zhandong.liu@bcm.edu

Zhongming Zhao

School of Biomedical Informatics,
The University of Texas Health Science Center at Houston,
Houston, TX 77030, USA
Email: zhongming.zhao@uth.tmc.edu

Biographical notes: Jianhua Ruan received his PhD in Computer Science and Engineering from Washington University in St Louis in 2007. He is currently Associate Professor in the Department of Computer Science at The University of Texas at San Antonio. His research interests lie in broad areas of bioinformatics, computational systems biology, and data mining.

Wenjin Zheng received his PhD in Biochemistry and Molecular Biology from UT Southwestern Medical Center in 1997 and an MS in Computer Science from UT Dallas in 2000. He is currently Associate Professor in the School of Biomedical Informatics and the Director of the Data Science and Informatics Core for Cancer Research, UT Health Science Center at Houston. His research interests include large scale genome information integration and visualisation, and mining literature to discover novel gene-gene relationships.

Zhandong Liu received his PhD in Genomics and Computational Biology from University of Pennsylvania in 2010 and an MS in Computer Science from Wayne State University in 2004. He is currently an Assistant Professor in the Department of Pediatrics at Baylor College of Medicine. His research interests include topics in genomics, transcription regulation, disease genes

prioritisation, and machine learning. He often examines problems in these areas using methods and models from computer science, statistics and related disciplines.

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 holds Chair Professor for Precision Health, School of Biomedical Informatics and is the founding director of the Center for Precision Health at the University of Texas Health Science Center at Houston (UTHealth). Before coming to UTHealth, he was a Professor in the Departments of Biomedical Informatics, Psychiatry and Cancer Biology at Vanderbilt University School of Medicine and the chief bioinformatics officer of the Vanderbilt-Ingram Cancer Center. His research interests include bioinformatics and systems biology approaches to studying complex diseases, precision medicine, pharmacogenomics, and biomedical informatics.

This special issue collects 10 papers submitted to *The International Conference on Intelligent Biology and Medicine 2016 (ICIBM 2016)*, which was held on December 8–10, 2016 in Houston, Texas, USA. The ICIBM 2016, which was built on the success of previous ICIBM conferences, attracted more than 150 participants from many institutions around the world. The ICIBM program included four keynote speeches, four conference invited speakers, eight scientific sessions, four workshops/tutorials, and an interactive poster session. The details of presentations are available on the conference website (https://www.uth.edu/cph/icibm/). Four keynote speakers are world-renowned leaders in bioinformatics, genomics, systems biology and computational medicine; they are Dr. Trey Ideker from the University of California at San Diego, Dr. John Weinstein from The University of Texas MD Anderson Cancer Center, Dr. Edward Marcotte from The University of Texas at Austin, and Dr. Yi Xing from The University of California at Los Angeles. Eight scientific sessions included the presentations selected from rigorous review process handled by a program committee of more than 80 experts in the field based on their scientific merit and technical quality. These sessions are the following: Session I: Genome Structure Analysis and Visualisation; Session II: NGS Analysis and Tools I; Session III: Computational Drug Discovery; Session IV: Medical Informatics and Big Data; Session V: NGS Analysis and Tools II; Session VI: Cancer Genomics; Session VII: Systems Biology; and Session VIII: Big Data and Machine Learning. Four tutorial sessions cover frontier research topics such as pathway and network analysis, multi-omics clustering analysis, microbiome sequencing and genotyping, and ENCODE data analysis. These tutorials provided a wealth of information on these cutting-edge techniques and are well appreciated by the conference participants.

Here, we provide an editorial report of the supplement to the *International Journal of Computational Biology and Drug Design* that includes 10 research articles selected from 77 manuscripts submitted to ICIBM 2016. Each manuscript was reviewed by at least two reviewers (most by three reviewers) and was substantially revised by taking care of the reviewers' critiques before further review and final acceptance. We summarise these papers in this special issue as below.

Next-generation sequencing technology such as RNA-sequencing (RNA-seq) continues to revolutionise many frontier biomedical research areas. During the past several years, single cell RNA-sequencing has gained widespread popularity, allowing biologists to study cell-specific transcriptomic changes. In this supplement, Gao et al. studied the technical variability of single cell RNAseq data with a thorough comparative analysis of RNA-seq data using low-input RNA and more than 35 single neuron cells. They found that the quantity of input RNA is negatively correlated with the variation of gene expression from technical replicates and the variation contributed from bioinformatics pipeline was generally minor when compared to the quantity of input RNA. They also demonstrated that t-SNE was an effective dimension reduction approach to reduce noise from very low-input RNA-Seq.

In another paper, Wijayawardena et al. used single cell RNA-seq data to study tumour cell heterogeneity in breast cancer and its correlation with chemotherapeutic resistance. They found a decrease of heterogeneity during the transition to Paclitaxel tolerance and high mutation rates in stressed single cells, suggesting genetic instability of cancer cells that ultimately results in development of drug resistance. This study may have significant implications for personalised and efficient treatment against breast cancer.

Chromosome spatial organisation in the nucleus plays critical roles in gene expression and has significant downstream effect on proper cellular functions such as DNA replication, DNA repair and chromosome translocation. Hi-C uses high-throughput sequencing to survey chromatin interactions in a genome scale and has paved the path to understand chromosome structure. Here, Zhang et al. proposed a genetic algorithm to reconstruct 3D model of chromosome structure from HiC data and showed that their method is more flexible and more accurate than the existing algorithms.

With the massive amount of data easily generated from next-generation sequencing, simply aligning the short reads to genome reference sequence is still a bottleneck in many applications. Al Kawam et al. proposed a read alignment algorithm based on a heterogeneous system composed of graphics processing units (GPUs) and multicore central processing units (CPUs). The design exploits the unique features of GPU and CPU simultaneously to maximise throughput. Their comparison with the BWA-mem alignment tool indicated a significant speedup.

It is well known that consistency between mRNA expression level and protein expression level is often very low. Yazdanparast et al. compared the protein expression level measured with reverse-phase protein arrays (RPPA) and RNAseq data from more than 400 primary breast tumours and 33 cell lines from The Cancer Genome Atlas (TCGA). While the overall correlation is low, however, they found a very high correlation between mRNA expression and protein expression for 207 cancer-related genes, suggesting that mRNA expression for these genes can serve as biomarkers for their protein products in drug target selections and cancer subtype classification. In addition, they found a high correlation between cell lines and primary tumours, confirming that breast cancer cell lines can serve as reasonably good models for primary breast cancer tumours.

Protein structure prediction is a long-standing problem in computational biology. This supplement includes a study by Devaurs et al. that analysed data generated from hydrogen/deuterium exchange detected by mass spectrometry (HDX-MS) for the complement protein c3d, and found that the c3d conformation generated by conformational sampling from HDX-MS was highly similar to the available crystal

structures but allowed the refining of the model. This finding may impact many HDX-based applications, from structural analyses to ligand-interaction studies.

In another study, Dhusia et al. modelled the protein structure of ornithine decarboxylase (ODC) enzyme, which plays an important role in growth as well as stress and disease resistance in plants, and then screened a large database of herbal products for natural ingredients that potentially can bind to the active site of ODC and inhibit the function of ODC. Using a tool called Autodock Vina, they predicted several molecules, including conessine which is a steroid alkaloid, to be the most effective inhibitor of ODC. Further experimental validation is needed to validate their prediction.

Systems biology approaches to understanding complex diseases is one of the greatest challenges in modern biology. While gene regulatory network-based methods have been shown promising, the space of all possible network configurations is very large and the amount of experimental data needed to fully reconstruct a gene regulatory network is often prohibitive. In this supplement, Song et al. proposed a sophisticated algorithm to reconstruct gene regulatory network using high-dimensional differential equations after dimensional reduction, and applied the algorithm to model genetic mechanisms of dynamic gene regulation during latent HIV reactivation. By combining clustering and other dimensional reduction techniques, they are able to build a dynamic regulatory network that included 95 co-expressed gene modules, which offered new insights for understanding the biological process underlying latent HIV reactivation, and may lead to novel strategies for eradication of HIV.

An alternative approach to network-based understanding of complex diseases is based on known pathways, which relies on pre-existing knowledge of gene-gene relationships but is less dependent on sample size and the results are often easier to evaluate and validate. Here, Mitchel et al. presented a method to detect associations between known pathways and Alzheimer's disease (AD) using genome-wide association studies (GWAS) data. Using this algorithm, they identified 133 AD-associated pathways, five of which are common in most AD patients. Their method can help with a comprehensive understanding of the molecular mechanisms underlying complex diseases such as AD and can lead to novel markers for disease prediction.

Electroencephalogram (EEG) recording of brain activities has enabled data-driven approaches to research in cognitive neuroscience. This supplement includes an article by Chikara et al. that reported a study to explore human brain activity changes in responses to stop-signal tasks, a well-known design for evaluating the so-called response inhibition, which is a hallmark of executive control. Their study revealed interesting brain activity change patterns in different brain regions for both left- and right-hand responses, which may lead to better understanding of neurological disorders such as schizophrenia disorder and attention deficit hyperactivity disorder.

We thank all the reviewers for assessing the scientific merits of, as well as valuable comments on these manuscripts included in this special issue. We thank the National Science Foundation (NSF grant IIS-1645823) for financial support and the University of Texas Health Science Center at Houston for hosting ICIBM 2016.