Editorial

Victor X. Jin

Departments of Molecular Medicine and Biostatistics,
University of Texas Health Science Center at San Antonio,
San Antonio, Texas 78229, USA
Email: jinv@uthscsa.edu

Yufei Huang

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

Jianhua Ruan

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

Zhongming Zhao

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

Biographical notes: Victor X. Jin received his PhD in Biological/ Macromolecular Chemistry in 2001 at Queen's University in Kinston, Ontario, Canada. He is currently an Associate Professor in the Department of Molecular Medicine at University of Health Science Center at San Antonio. He has an impressive level of productivity with more than 75 peer-reviewed papers in some of the most prestigious and highly cited journals such as *Nature*, *Nature Cell Biology*, *Cancer Cell*, *Molecular Cell*, *Genome Research*, *Genome Biology*, *Cancer Research* and *PNAS*. The major focus of his group is to perform functional genomics and develop statistical methods and machine learning algorithms for different high-throughput 'omics data and to apply these experimental and computational tools to understand epi-/genetic regulatory mechanisms in cancers. He is involved in organising numerous conferences and meetings. He is currently an Associate Editor for BMC Medical Genetics and on Editorial Board of several journals.

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 focuses 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.

Jianhua Ruan received his PhD in Computer Science and Engineering from Washington University in St Louis in 2007. He is currently an Associate Professor in the Department of Computer Science and Leader of Computational Systems Biology Core at The University of Texas at San Antonio. His research interests lie in broad areas of bioinformatics, computational systems biology, and data mining. He has published more than 40 papers, many of which are in top tier journals such as *Bioinformatics*, *PLoS Computational Biology*, and *Genome Biology*. He serves on the editorial board of three journals, as well as the review boards of more than 15 journals and program committees of over 10 scientific conferences. He received a Best Performer award in the Third Annual DREAM Reverse Engineering Challenges in 2008. His research has been sponsored by the United States National Institutes of Health, United States National Science Foundation, and the San Antonio Life Sciences Institute.

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 Ingram Professor of Cancer Research and is 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, precision medicine, pharmacogenomics, and biomedical informatics.

This special issue collects seven papers, including five submitted to *The International Conference on Intelligent Biology and Medicine 2014 (ICIBM 2014)*, which was held on 4–6 December, 2014 at the Sheraton Gunter Hotel in San Antonio, Texas, USA. The ICIBM 2014, which was built on the success of ICIBM 2012 and 2013, attracted more than 115 participants from many institutions around the country and the world. The ICIBM program included four keynote speeches, six scientific sessions, four tutorials, nine highlight talks, and a poster session. The details of all presentations are available on the conference website (http://compgenomics.utsa.edu/icibm2014/). Four keynote speakers who are world-renowned leaders in bioinformatics, genomics, systems biology and computational medicine include Dr. Lynda Chin from The University of Texas MD Anderson Cancer Center, Dr. Tim Huang from The University of Texas Health Science Center at San Antonio, Dr. Josh Stuart from University of California, Santa Cruz, and Dr. Jasmine Zhou from The University of Southern California. Six 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*: Gene Regulation and Protein Interaction Networks I
- *Session II*: Gene Regulation and Protein Interaction Networks II
- *Session III*: Algorithms in NGS Data Analysis

- *Session IV*: Metagenomics and Bioinformatics Methods
- *Session V*: Epigenetics and Systems Biology
- *Session VI*: Biomarker Identification Methods.

Four tutorial sessions cover frontier research topics such as proteomics, metabolomics, metagenomics, single cell analysis, and next-generation sequencing and data analysis. These tutorials provided a wealth of information on these cutting-edge techniques and are well appreciated by the conference participants.

Computational biology and bioinformatics are increasingly becoming major disciplines in the biomedical and biological research fields. Among many topics, genomics, systems biology, bio-imaging, and intelligent computing have seen exponentially increases in terms of technologies, data volumes as well as publications. The seven original papers selected in this special issue describe recent advances in genomics, systems biology, and intelligent computing, reflecting the rapid advances in these topics. The other selected papers were included in other special issues including two supplemental issues to BMC Genomics (http://www.biomedcentral.com/bmcgenomics/supplements/16/S7) and Biomed Research International (http://www.hindawi.com/journals/bmri/). Below, we briefly summarise the seven papers in this special issue.

The algorithmic and analytical approaches for inferring a gene regulatory network (GRN) are still challenging tasks. The features of such approaches may affect the performance and the accuracy of interpretation of biological meanings. We have two papers in this special issue to tackle on this important topic.

The study from Mizeranschi et al. is to compare the performance of the conventional single-stage and the modern two-stage algorithms. The authors explored data generated from artificial as well as real GRN systems under different experimental conditions and regulatory structure constraints. The basic model inference algorithm is based on the multi-scale modelling and simulation tool called MultiGrain/MAPPER, which implements a parallel/distributed multi-swarm particle swarm optimisation process to infer the GRN model parameters. They found that the 2-stage approach outperforms the single-stage methods by far in terms of model inference speed without losing the accuracy.

Wang et al. instead focused on understanding the GRN for TCF7L2, a transcription factor (TF) downstream of Wnt signalling pathway. The rationale for this work is that this TF may influence transcription of several genes by binding to distinct regulatory regions, and consequently, plays an important role in a wide range of cell functions and is thus implicated in several diseases like type 2 diabetes and cancer. From the identified thousands of TCF7L2 binding sites and their associated TF partners, they analysed ChIP-seq data by searching for motifs in the enriched peak regions based on TF-specific position weight matrix (PWM). The authors found an association of TFs FOXO1 and CAD with the up-regulated genes, and an association of TFs AP2α, PBF and AP1 with the down-regulated genes. TCF7L2 and GATA3 were found to be associated with both up and down-regulated genes. That study uncovered new TCF7L2 associated regulatory networks, which may contribute to further study of the mechanisms related to Wnt pathway in breast cancer cells or other human diseases.

Next-generation sequencing (NGS) is becoming a routine high throughput tool in studying various biological and biomedical questions. Among the NGS platforms,

Ion Torrent personal genome medicine (PGM) is a relatively new sequencing platform, offering fast and real-time measurement. Budd et al. used this sequencing machine to systematically genotype marker genes *VKORC1*, *CYP2C9*, *CYP2C19*, and *CYP2D6*. Their study showed that there exists a large range of genetic variation within and between various ethnic groups residing in the USA. With exception of *CYP2C9*, the wild type genotype is not the most common variant. Each gene sequenced showed a unique pattern of distribution that significantly differed within and across ethnic groups. These findings call into question the concept of a 'normal' patient. Their results highlight the need for and applicability of pharmacogenomic testing. Genetic determination of patient response groups can assist physicians tailor therapeutic regimens and increase the likelihood of medication success, which is an important topic in precision medicine.

More and more new computational approaches have been shown promising in predicting side-effect drug designing and reducing cost and time. Jahid and Ruan has designed an ensemble approach that combines the results from different classification models. They applied their approach to 1385 side-effects in the SIDER database for 888 drugs. Results showed that the approach outperformed the previously published approaches and standard classifiers. Furthermore, they applied our method to a number of uncharacterised drug molecules in the DrugBank database and predicted their side-effect profile for future usage. Results from various sources confirmed that their method could predict the side-effects for uncharacterised drugs, and more importantly, could predict rare side-effects, which are often ignored by other computational approaches.

Computational imaging is becoming a new important bioinformatics domain. The research topics range from processing images in cell biology, pathology, to radiology. This special issue includes a paper dealing with this aspect of informatics. Liang et al. developed a framework applicable to 3D pathology analytical imaging, with an application to whole slide images of sequential liver slices for 3D vessel structure analysis. The complete analysis workflow consists of image registration, segmentation, vessel cross-section association, interpolation, and volumetric rendering. To identify biologically-meaningful correspondence across adjacent slides, the authors formulated a similarity function. The optimal solution to this multi-object grouping problem was then obtained by constrained Integer Programming. They quantitatively compared the machine-generated vessel reconstruction results with those produced by human annotations on six primary vessels. They found a satisfactory concordance as measured both by region-based and distance-based metrics, demonstrating a promising 3D vessel analysis framework for whole slide images of liver tissue sections.

In another study, Banwait et al. developed a computational framework to identify mRNAs with enriched structural localisation features in their 3'-untranslated regions (UTRs). They analysed all nuclear encoded mitochondrial proteins and reported a potential alternative to peptide signal based localisation of mitochondrial proteins. They were able to further extend the mitochondrial proteome approach to identify seven new proteins that were not previously labelled to be localised to mitochondria by Uniprot but with literature support of their involvement in mitochondria-related functions.

In the last paper, Al-Shammari et al. aimed to investigate whether germ-line gene mutations are a significant biological factor in 29 major primary human cancers. Using data obtained from multiple databases, they identified 424 genes from 8879 cancer mutation records. By integrating these gene mutation records, the authors constructed a Human Cancer Map (HCM). The features include the observations that missense, nonsense, and regulatory mutations might play a central role in connecting cancers with

the associated genes, and these mutations are distributed in almost all chromosomes. The analytic results suggest that, of all mutation classes missense, nonsense and regulatory mutation classes are over-expressed in the human genome and so are likely to have a significant impact on human cancer aetiology and pathological mechanisms. They also observed that several chromosomes (chromosomes 1, 2, 3, 15, and 17) tended to contribute to cancer genes disproportionally compared with other chromosomes, whereas chromosomes 21 and Y did not show any contribution to any of the cancer-associated genes. This sheds new insights into how the distribution and interconnections of gene mutations have the influence on the development of cancers.

We thank all the reviewers for judging the scientific merits of the manuscripts submitted to ICIBM 2014 including this special issue. We are grateful to the local organisation committee members and volunteers for making ICIBM 2014 a success. We thank the National Science Foundation (NSF grant IIS-1451135) and The University of Texas Health Science Center at San Antonio (UTHSCSA) for financial support and The University of Texas at San Antonio (UTSA) and Vanderbilt University (VU) for the support of ICIBM 2014.