Editorial

Yunyun Zhou

Department of Data Science,
The University of Mississippi Medical Center,
Jackson, MS 78249, USA
Email: yzhou.umc@gmail.com

Renzhi Cao

Department of Computer Science,
Pacific Lutheran University,
Tacoma, WA 98447, USA
Email: caora@plu.edu

Kai Wang

Department of Pathology & Laboratory Medicine,
Perelman School of Medicine
University of Pennsylvania,
Pennsylvania, PA, 19104, USA
Email: kaichop@gmail.com

Zhongming Zhao

Center for Precision Health,
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: Yunyun Zhou received her PhD in School of Electrical Engineering and Computer Science from Washington State University in 2012. Later she worked as the computational biologist in University of Texas Southwestern Medical Center at Dallas. Currently, she is an Assistant Professor in the Department of Data Science, John D. Bower School of Population Health at the University of Mississippi Medical Center. Her research interests include the development of bioinformatics and biostatistics tools for systems biology data analysis, biomarker discovery for cancer clinical research, next-generation sequencing data analysis, comparative genomics, and quantitative biomedical informatics.

Renzhi Cao received his PhD in Department of Computer Science at the University of Missouri-Columbia in 2016. After that, he became an Assistant Professor in the Department of Computer Science at Pacific Lutheran University in Tacoma. His research interest is mainly focused on developing and applying machine learning and data mining techniques to solve biomedical problems, such as human genome data analysis and protein structure

predictions. In addition, he is interested in promoting early engagement of undergraduate students (especially for women and underrepresented students) in machine learning, bioinformatics, and the data science field by interdisciplinary studies, and inspiring students to pursue advanced STEM education/research careers.

Kai Wang is an Associate Professor at the Raymond G. Perelman Center for Cellular and Molecular Therapeutics of the Children's Hospital of Philadelphia (CHOP) and Department of Pathology & Laboratory Medicine at the University of Pennsylvania (Penn) Perelman School of Medicine. He received a Bachelor's degree from Peking University in China, a Master's degree from Mayo Clinic, and a PhD from the University of Washington, then had postdoctoral training at Penn and CHOP. He was previously an Associate Professor at the University of Southern California (USC) Keck School of Medicine and Columbia University Medical Center (CUMC). His research focuses on the development and application of genomic approaches to study the genetic basis of human diseases and facilitate the implementation of genomic medicine.

Zhongming Zhao received his PhD in Human and Molecular Genetics from The University of Texas MD Anderson Cancer Center UT Health Graduate School of Biomedical Sciences, Houston, Texas in 2000. He holds Chair Professor for Precision Health and is a professor in School of Biomedical Informatics. He is founding director of the Center for Precision Health at The University of Texas Health Science Center at Houston. He is also the founding president of The International Association for Intelligent Biology and Medicine (IAIBM). His research interests include bioinformatics and systems biology approaches to studying complex diseases, deep learning, precision medicine, and pharmacogenomics.

This special issue collects nine papers submitted to *The International Conference on Intelligent Biology and Medicine (ICIBM 2018)*, which was held on 10–12 June, 2018, Los Angeles, CA, USA. The ICIBM 2018, which was built on the success of previous conferences, attracted more than 300 participants from many institutions internationally. ICIBM 2018 provides a leading forum for disseminating the latest research in bioinformatics, systems biology, and intelligent computing. It brings together academic and industrial scientists from computer science, biology, chemistry, medicine, mathematics, and statistics.

The ICIBM program included four keynote speeches, four tutorials, one workshop, nine scientific sessions. The details of all presentations are available on the conference website (http://icibm2018.zhaobioinfo.org/). The four keynote speakers, all world-renowned leaders in Bioinformatics, Systems Biology and Intelligent Computing, are Dr. Joshua C. Denny from Vanderbilt University Medical Center, Dr. Alexander Hoffmann from the University of California Los Angeles, Dr. Jason Moore from the University of Pennsylvania, Dr. Paul D. Thomas from the University of Southern California. Nine scientific sessions included the presentations selected from the rigorous review process handled by a program committee of more than 90 experts in the field based on their scientific merit and technical quality.

The examples of these sessions are NGS Tools, Systems Biology, Bioinformatics, Medical Informatics, Cancer Genomics, and Computational Drug Discovery. These sessions provided a wealth of information on cutting-edge techniques and are well appreciated by the conference participants.

Thanks to the development of high-throughput NGS technologies generating a huge amount of whole genome data, biologists now can observe and measure thousands of molecules in cells simultaneously. This wealth of data provides an opportunity to interpret the complex molecular system, although unprecedented challenge. Machine learning and statistical modelling play critical roles in this field of systems biology, and the demand for more efficient and accurate algorithms is still urgent. The nine original papers selected in this special issue describe some state-of-the-art machine learning and statistical methods in bioinformatics, systems biology, and intelligent computing, reflecting the rapid advances in these topics. Among these manuscripts, most of them are related to novel machine learning frameworks to improve the integrative analysis for –omics data.

Double minute chromosomes (DMs) are important circular fragments of extrachromosomal DNA that are highly related to malignant tumour cell behaviour and drug resistance. Mardugalliamov et al. developed a novel Hidden Markov Model-based approach AmpliconFinder to detect DM amplicons, which would help to locate DMs and inform precision therapy to cancer treatment. Their approach substantially improved the ability to discover DMs in simulated genomic data and their experiment has shown that the sensitivity is increased by an average of 20%.

Drug design is a fundamental research field for new drug discovery. Alam et al. developed a new computational method to design small molecule stabilisers targeting mutant (V210I) human prion protein (HuPrP) against familial Creutzfeldt-Jakob disease. They collected target protein structure from protein data bank (PDB) Database and minimised the energy to identify drug binding pockets in the target protein. The de-novo ligand design is then applied and generated a pharmacophore for further virtual screening of small molecules. The final molecules are selected by docking with the target pocket of the protein. Their experiment demonstrated that the final five molecules are ideal candidates for further drug development.

Another hot topic for drug discovery is the prediction of whether a drug combination of arbitrary orders is likely to induce adverse drug reactions. Chiang et al. proposed a new machine learning method that is based on novel kernels within support vector machines (SVM) to measure the similarities among drug combinations. Their algorithm is evaluated by ROC curve and demonstrated a high performance with AUC = 0.921.

Genome-wide association study (GWAS) is significant for identifying disease causal genes, however much remains unknown about the structural connectivity of these genes to a specific disease. For example, Alzheimer's disease (AD) is a very common form of brain dementia with more than 20 risk genes identified and verified previously. However, it is still unknown how are these genes correlated to the brain-wide breakdown of structural connectivity in AD patients. Yan et al. used data from the Alzheimer's disease neuroimaging initiative (ADNI) database and constructed brain networks. They applied a targeted genetic association analysis with the help of regression model to discover abnormal brain-wide network alterations under the control of 34 AD risk SNPs. Their research has a lot of interesting findings, such as the significant association of anisotropy and rs10498633 in SLC24A4.

Genome alignment programs are usually used for identifying high scoring fragments between a query and target sequence. It is still challenging to do that for not evolutionary similar genomes. Aljouie et al. utilised the graphics processing unit (GPU) to speed up this alignment process for aligning short reads to a genome sequence and then extended the fragment to obtain a complete alignment. Their method gives at least 20% higher accuracy than existed method.

Understanding the complicated interactions among microbial communities is very important. However, the existing models are biased for different microbial communities. Apaydin et al. proposed a flexible framework to model microbial communities with uncertainty. They proposed a new algorithm, P-OptCom, which relied on the coordinated decision making between community level and many microorganisms decision makes to support robust solutions. They analysed the trade-offs among the members of microbial communities and closely approximate the actual experimental measurements.

Unsupervised learning, or clustering, methods are essential in grouping co-expressed genes from the same communities. Although there are many traditional methods can perform clustering analysis for gene expression data, improving the cluster prediction accuracy is a key step for interpreting biological meanings. Cui et al. proposed a new approach for gene expression clustering analysis by integrating deep learning architecture with a network-based method. Their deep learning approach, Robust Autoencoder, provided a more accurate high-level representation of the feature sets and outperformed traditional clustering methods such as hierarchical clustering and K-means.

Identification of differentially expressed genes in time course experiments which are associated certain disease phenotypes is critically important for precision medicine. Soltanalizadeh et al. developed a dynamic network-based algorithm which can capture time-dependent information to identify differentially expressed genes in three cancer cell lines. Their results outperform than several currently available methods. Importantly, they reported several novel genes, such as BMP6 and ARSJ, in response to hypoxia function in breast cancer cells, which had not been identified previously.

The final paper described a novel deep learning method to analyse the quality of predicted protein, which is crucial for computational predicting protein structures and designing drugs. Smith et al. analysed the topology of the predicted protein structure and used their findings to train a deep learning model based on convolutional neural network (CNN). Their method achieved an overall correlation of 0.41 on testing dataset CASP12.

We thank all the reviewers for evaluating the scientific merits of the manuscripts submitted to ICIBM 2018 including this special issue. We are grateful to the local organising committee members and volunteers for making ICIBM 2018 an excellent venue for exchanging research results and ideas for fostering collaboration, and for training next- generation bioinformaticians in biological and biomedical fields.