# Editorial

Stephen Kwok-Wing Tsui

THE special section of *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* is a collection of papers presented at the 18th Asia Pacific Bioinformatics Conference (APBC2020), which was a virtual conference held in Seoul, Korea, from August 18 to 20 in 2020. This conference brought together researchers, academics and industrial practitioners to explore cutting-edge research, development and novel applications in the field of bioinformatics. We are grateful and proud to present the 3-day program that includes three outstanding keynote speeches, 61 excellent oral presentations and two special lectures. Selected high-quality conference papers related to bioinformatics and computational biology were invited to submit to this special issue and all these papers have been rigorously peer-reviewed. Finally, eight papers were selected for publication in *TCBB*. We would like to express our sincere thanks to the program committee members and researchers who reviewed the selected papers. Special thanks also go to the editorial staff of *TCBB*. The publication of this special section would not be possible without their commitment and dedication. Year 2020 marked the emergence of the global COVID-19 pandemic, which has totally changed the living style of people around the world. In the last two years, genomics and bioinformatics played an important role in the infection control, tracing of viral spread and vaccine development. We believe bioinformatics research will significantly contribute to the scientific community as well as our society in the future.

In the paper titled "protein2vec: Predicting Protein-Protein Interactions Based on LSTM " by Jiongmin Zhang, Man Zhu, and Ying Qian, the authors proposed a new method to characterize a protein with a vector based on the Gene Ontology (GO) terms annotated to it and combines the information of both the GO and known protein-protein interactions. The network embedding algorithm was applied on the GO network to generate feature vectors for each GO term and then Long Short-Time Memory was used to encode the feature vectors of the GO terms annotated with a protein into another vector. Finally, two protein vectors are forwarded into a feedforward neural network to predict the interaction between the two corresponding proteins. Experimental results showed that protein2vec outperforms many traditional semantic similarity methods.

In the paper titled "A New Approach to Deriving Prognostic Gene Pairs from Cancer Patient-Specific Gene Correlation Networks" by Byungkyu Park, Wook Lee and Kyungsook Han, the authors proposed a new method for inferring prognostic gene pairs from patient-specific gene correlation networks. Evaluation of this method using data of three cancer types showed that gene pairs can serve as more reliable prognostic signatures for cancer when compare with genes. Analysis of patient-specific gene networks suggests that prognosis of individual cancer patients is affected by the existence of prognostic gene pairs in the patient-specific network and by the size of the patient-specific network. This approach will be useful for finding gene pairs to predict survival time of patients and to tailor treatments to individual candidates.

In the paper titled "DCHap: A Divide-and-Conquer Haplotype Phasing Algorithm for Third-Generation Sequences" by Yanbo Li and Yu Lin, the authors developed a divide-and-conquer algorithm to phase haplotypes using third-generation sequencing reads. When DCHap was benchmarked against three state-of-the-art phasing tools on both PacBio SMRT data and ONT Nanopore data, results showed that DCHap generates more accurate or comparable results while being scalable for higher coverage and longer reads. Thus, DCHap is a fast and accurate algorithm for haplotype phasing using third-generation sequencing data.

In the paper titled "CIR-Net: Automatic Classification of Human Chromosome Based on Inception-ResNet Architecture" by Chengchuang Lin, Gansen Zhao, Zhirong Yang, Aihua Yin, Xinming Wang, Li Guo, Hanbiao Chen, Zhaohui Ma, Lei Zhao, Haoyu Luo, Tianxing Wang, Bichao Ding, Xiongwen Pang, and Qiren Chen, the authors developed an automatic chromosome classification approach based on Inception-ResNet. Moreover, a simple but effective augmentation method called CDA was proposed for improving the performance of CIR-Net. Results showed that this method achieved a very high classification accuracy on the clinical G-band chromosome dataset whose training dataset is insufficient. Moreover, the proposed augmentation method significantly improves classification accuracy comparing to other methods.

In the paper titled "A Scalable Embedding Based Neural Network Method for Discovering Knowledge From Biomedical Literature" by Shengtian Sang, Xiaoxia Liu, Xiaoyu Chen, and Di Zhao, the authors presented a model which incorporates knowledge graph, graph embedding and deep learning methods for both open and closed Literature-based discovery (LBD). In this work, how deep learning combining with graph embedding techniques can be applied to LBD tasks, including discovering new knowledge from unrelated literature and providing logical explanations for the relations between entities, has been demonstrated. The experimental results suggest that incorporating knowledge graph and deep learning methods is an effective way for capturing the underlying complex associations between entities hidden in the literature.

• *The author is with the School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E-mail: kwtsui@cuhk.edu.hk.*

*Digital Object Identifier no. 10.1109/TCBB.2022.3155845*

In the paper titled "Hierarchical Structured Component Analysis for Microbiome Data Using Taxonomy Assignments" by Sun Kim, Nayeon Kang, and Taesung Park, the authors proposed a hierarchical structural component model for microbiome data using taxonomy information as well as operational taxonomic units (OTUs) table data. They then calculated simultaneously coefficient estimates of OTUs and taxa of the two layers inserted in the hierarchical model. Through their analysis, the association between taxa or OTUs and disease status were inferred. Using simulation study and real microbiome data analysis, this model has shown to reveal the relations between each taxon and disease status and identify the key OTUs of the disease at the same time.

In the paper titled "Evaluation of Experimental Protocols for Shotgun Whole-Genome Metagenomic Discovery of Antibiotic Resistance Genes" by Ken Hung-On Yu, Xiunan Fang, Haobin Yao, Bond Ng, Tak Kwan Leung, Ling-Ling Wang, Chi Ho Lin, Agnes Sze Wah Chan, Wai Keung Leung, Suet Yi Leung, and Joshua Wing Kei Ho, the authors evaluated the seven protocols in terms of robust detection of antibiotic resistance genes (ARGs) and microbial abundance estimation at various sequencing depths. Results showed that the data generated by the seven protocols are largely similar and the inter-protocol variability is significantly smaller than the variability between samples or sequencing depths. This systematic benchmarking study sheds light on the impact of sequencing depth, experimental protocol and DNA input amount on ARG detection in human stool samples.

In the paper titled "Seq-BEL: Sequence-Based Ensemble Learning for Predicting Virus-Human Protein-Protein Interaction" by Yingjun Ma, Tingting He, Yuting Tan, and Xingpeng Jiang, the authors proposed a new method based on projection neighborhood non-negative matrix factorization to predict the potential virus-human protein-protein interactions (PPIs). Based on the amino acid sequence of proteins and the currently known virus-human PPI network, Seq-BEL calculates various features and similarities of human proteins and viral proteins, and then combines these similarities and features to score the potential of virus-human PPIs. Results showed that Seq-BEL achieved success in predicting potential virus-human PPIs and outperforms other state-of-the-art methods. Moreover, Seq-BEL has good predictive performance for new human proteins and new viral proteins.

Stephen Kwok-Wing Tsui
*Guest Editor*

**Stephen Kwok-Wing Tsui** received the PhD degree in biochemistry from the Chinese University in Hong Kong (CUHK), in 1995. He is currently a professor and the associate director (research) with the School of Biomedical Sciences, as well as director of the Hong Kong Bioinformatics Centre in CUHK. He was also a former member of the International HapMap Consortium and worked on the single nucleotide polymorphisms of human chromosome 3p. During the SARS outbreak in 2003, his team was one of the earliest teams that cracked the complete genome of the SARS-coronavirus. He has published more than 240 scientific papers in referred journals, including *Nature*, *Nature Machine Intelligence*, *New England Journal of Medicine*, *Lancet*, *PNAS, Genome Biology*, *Nucleic Acids Research* and *Bioinformatics*. He has more than 15000 paper citations and his *H*-index is 43. He is also an editor of several international journals including *Frontiers in Genetics*, *Gene* and *International Journal of Oncology*. He was the recipient of an Excellent Research Award, Food and Health Bureau, Hong Kong in 2019. His research interests include bioinformatics, comparative genomics, and multi-omics of human diseases.