

# Guest Editorial for the 12th Asia Pacific Bioinformatics Conference

Shuigeng Zhou and Yi-Ping Phoebe Chen

**T**HIS special section includes eight papers that were selected from the papers presented at the Twelfth Asia Pacific Bioinformatics Conference (APBC2014), which was held in Shanghai, China, 17-19 January, 2014. The eight papers cover diverse topics, including protein complex prediction, nucleosome positioning, biological data classification and clustering, sequence comparison, genome reconstruction, and enhancers etc.

In "From Function to Interaction: A New Paradigm for Accurately Predicting Protein Complexes Based on Protein-to-Protein Interaction Networks," Bin Xu and Jihong Guan propose a novel paradigm for protein complex detection. They first cluster proteins using Biology Process (BP) annotations from Gene Ontology (GO). Then, they map the resulting proteins clusters onto a PPI network, extract connected subgraphs consisting of clustered proteins as nodes from the PPI network and expand each connected subgraph with protein nodes that have rich links to the proteins in the subgraph. The expanded subgraphs are taken as predicted complexes. The proposed method Cpredictor is applied to two PPI data sets of *S. cerevisiae* for predicting protein complexes. Experimental results show that Cpredictor outperforms the existing methods.

In "Beyond Fixed-Resolution Alignment-Free Measures for Mammalian Enhancers Sequence Comparison," Matteo Comin and Davide Verzotto present a parameter-free alignment-free statistic based on variable-length words. Different from traditional alignment-free methods that are based on fixed length patterns or, in other words, tied to a fixed resolution, the proposed statistic is built upon variable-length words, and thus multiple resolutions are allowed, which thus can capture the great variability of lengths of CRMs. They evaluate several alignment-free statistics on simulated data and real ChIP-seq sequences, and find that the new statistic is highly successful in discriminating functionally related enhancers and, in almost all experiments, it outperforms the fixed-resolution methods.

In "A Novel Wavelet-Based Approach for Predicting Nucleosome Positions Using Structural Feature," Yanglan Gan, Guobing Zou, Jihong Guan, and Guangwei Xu analyze the positioning of well-positioned and fuzzy nucleosomes from a novel structural perspective, and introduce a new

computational approach WaveNuc for predicting the positions of well-positioned and fuzzy nucleosomes based on continuous wavelet transformation. Their systematic comparative analysis demonstrates that the two kinds of nucleosomes exhibit different propeller twist characteristics. Well-positioned nucleosomes tend to locate at sharp peaks of propeller twist profile, while the positions of fuzzy nucleosomes correspond to broader peaks. By using this findings, they apply WaveNuc to detecting the two kinds of peaks of propeller twist structural profile along the genome. The new method is evaluated and compared with existing methods on real data sets, and experimental results show that the new method can resolve accurately complex configurations of fuzzy nucleosomes, which leads to better performance of nucleosome positioning prediction on the whole genome.

In "SeeSite: Characterizing Relationships between Splice Junctions and Splicing Enhancers," Christine Lo, Boyko Kakaradov, Daniel Lokshtanov, and Christina Boucher present a novel method to characterize co-occurring relationships between splice site motifs and splicing enhancers. As the authors claimed, this is the first formulation and computational attempt to detect co-occurring sequence elements in RNA sequence data. They demonstrate that certain ESEs are preferentially associated with weaker splice sites, and that there exists a co-occurrence relationship with splice site motifs.

In "An Improved Ensemble Learning Method for Classifying High-Dimensional and Imbalanced Biomedicine Data," to classify high-dimensional and skewed biomedicine data, Hualong Yu and Jun Ni combine the asymmetric bagging ensemble classifier (asBagging) and an improved random subspace (RS) generation strategy called feature subspace (FSS). FSS is a novel method to promote the balance level between accuracy and diversity of base classifiers in asBagging. They conduct extensive experiments on four benchmark biomedicine data sets, and the experimental results show that the proposed ensemble learning method outperforms many baseline approaches in terms of Accuracy, F-measure, G-mean and AUC evaluation criterions.

In "Probabilistic Reconstruction of Ancestral Genomes with Gene Insertions and Deletions," Fei Hu, Jun Zhou, Lingxi Zhou, and Jijun Tang propose a new adjacency-based method PMAG+ to infer ancestral genomes under a more general model of gene evolution involving gene insertions and deletions (indels), in addition to gene rearrangements. The new method is an improvement on its previous version PMAG by developing a new way to infer ancestral gene contents and reducing the adjacency assembly problem to an instance of TSP. A series of experiments are carried out to extensively validate

- S. Zhou is with the Shanghai Key Lab of Intelligent Information Processing, and the School of Computer Science, Fudan University, 220 Handan Road, Shanghai 200433, China. E-mail: sgzhou@fudan.edu.cn.
- Y.-P.P. Chen is with the Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia. E-mail: phoebe.chen@latrobe.edu.au.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.  
Digital Object Identifier no. 10.1109/TCBB.2014.2327487

PMAG+ and compare it with the state of the art method GapAdj. Experimental results validate the advantage of the proposed method.

In "Maximum Likelihood Estimation of GEVD: Applications in Bioinformatics," Minta Thomas, Anneleen Daemen, and Bart De Moor develop a method called maximum likelihood estimation of generalized eigenvalue decomposition (MLGEVD) to improve the existing generalized eigenvalue decomposition (GEVD) and demonstrate the new method's applications to Bioinformatics by three case studies for which microarray data sets with corresponding clinical/literature information are available. The case studies show that MLGEVD outperforms GEVD and significantly improves diagnosis, prognosis and prediction of therapy response.

In "A Simple but Powerful Heuristic Method for Accelerating k-Means Clustering of Large-scale Data in Life Science," Kazuki Ichikawa and Shinichi Morishita study the properties of the Pearson correlation distance and propose a simple but powerful heuristic method to prune unnecessary computation of Pearson correlation distance, which thus boosts the efficiency of k-means clustering. They also conduct extensive experiments to validate the proposed method.

We would like to express our gratitude to the authors for contributing their works to this special section, and to the reviewers for their hard work in reviewing the manuscripts and giving insightful and constructive feedback comments. Without their support and cooperation, this special section could not come out.

Shuigeng Zhou  
Yi-Ping Phoebe Chen  
*Guest Editors*



**Shuigeng Zhou** (M'07) received the bachelor's degree from the Huazhong University of Science and Technology (HUST) in 1988, the master's degree from the University of Electronic Science and Technology of China (UESTC) in 1991, and the PhD degree of computer science from Fudan University in 2000. He is currently a professor in the School of Computer Science, Fudan University, Shanghai, China. He served at the Shanghai Academy of Spaceflight Technology from 1991 to 1997, as an engineer and a senior engineer (since

August 1995), respectively. He was a post-doctoral researcher in the State Key Lab of Software Engineering, Wuhan University, from 2000 to 2002. His research interests include data management, data mining, machine learning, and bioinformatics. He has published more than 200 papers in domestic and international journals (including *IEEE TKDE*, *IEEE TPDS*, *IEEE TCBB*, *NAR*, *Bioinformatics*, *Plos One*, *BMC Bioinformatics*, *DKE*, *PRE*, *EPL*, *EPJB*, etc.) and conferences (including SIGMOD, VLDB, ICDE, ICDCS, SIGKDD, SIGIR, SODA, AAAI, IJCAI, RECOMB etc.). Currently, he is a member of the IEEE, ACM, and IEICE. More information about him can be found at <http://admis.fudan.edu.cn/~sgzhou>.



**Yi-Ping Phoebe Chen** (SM'2007) received the BInfTech degree with first class honors and the PhD degree in computer science (bioinformatics) from the University of Queensland. She is a professor and chair in the Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia. She is the chief investigator of the ARC Centre of Excellence in Bioinformatics. She is currently working on knowledge discovery technologies and is especially interested in their application to genomics

and biomedical science. Her research focus is to find the best solutions for mining, integrating, and analyzing complex data structure and functions for scientific and biomedical applications. She has been working in the area of bioinformatics, health informatics, multimedia databases, query system, and systems biology, and has coauthored more than 200 research papers, with many published in top journals and conferences. She is steering committee chair of the Asia-Pacific Bioinformatics Conference (founder) and the International Conference on Multimedia Modelling. She has been on the program committees of more than 100 international conferences, including top ranking conferences such as ICDE, ICPR, ISMB, CIKM, etc. More information about her can be found at <http://homepage.cs.latrobe.edu.au/ypchen/index.htm>. She is a senior member of the IEEE.