EDITORIAL

Bioinformatics in Biological Big Data Era

The major feature of our current life sciences, is the rapid increase of biological data, which are presented in many forms, and reflect the characteristics of biological systems at various levels, including genome, transcriptome, epigenome, proteome and metabolome *etc*. This is the so-called biological big data we are facing. Biological big data bring both challenges and opportunities to bioinformatics. Tools and techniques for analyzing big biological data enable us to translate massive amount of information into a better understanding of the basic biomedical mechanisms, which can further be applied to translational or personalized medicine.

This thematic issue with a theme of "Bioinformatics in Biological Big Data Era" aims at extensively showing the latest development and achievements in Bioinformatics in this biological big data era. The ten papers in this thematic issue were selected from the 1st CCF Bioinformatics Conference (CBC 2016), which was sponsored by China Computer Federation (CCF). The selected papers cover methods and algorithms for processing biological big data in addressing various bioinformatics issues. Before submitted to the special issue, these papers have gone through strict reviewing organized by CBC 2016. Further reviewing was organized by the guest editors.

In what follows, we give a brief review of the 10 papers included in this thematic issue.

Liu *et al.* in their paper "Sparse linear modeling kinase inhibition network for predicting combinatorial drug sensitivity in cancer cells" used a sparse linear model called uncertain group sparse representation (UGSR) to infer essential kinases governing the cellular responses to drug treatments, based on the massively collected drug-kinase interactions and drug sensitivity datasets over hundreds of cancer cell lines [1].

In the paper "TagNovo: A dictionary based approach to predict peptide theroy spectra" Wang *et al.* presented a new theoretical spectrum prediction model called TagNovo, which builds a "tag dictionary" from exiting spectrum library and is used for theory spectrum prediction [2].

In the paper "Large-scale Investigation of Long Noncoding RNA Secondary Structures in Human and Mouse" by Guo *et al.* the authors conducted a large-scale investigation of lncRNA secondary structures especially for hairpin structural motif in human and mouse based on computational prediction using the RNAfold software, and found that the secondary structures of lncRNAs have many characteristics, most of which are similar with those in mRNAs [3].

Nie *et al.* in their paper "Prediction of protein S-Sulfenylation sites using a deep belief network" developed a computational method DBN-Sulf to effectively predict S-sulfenylation sites by using optimally extracted properties based on Deep Belief Network (DBN) with Restricted Boltzmann Machines (RBMs). DBN-Suf shows significantly better performance than the existing methods [4].

In the paper "Feature identification for phenotypic classification based on genes and gene pairs", Su, Zhang and Pan proposed a new algorithm called FSGGP to select both feature genes and feature gene pairs on the binary-value gene expression data [5].

Chan *et al.* in their paper "MyPhi: Efficient Levenshtein Distance Computation on Xeon Phi based Architectures" introduced MyPhi, an ultra-fast implementation of the Myers algorithm on Intel Xeon Phi based architectures for efficiently computing Levenshtein Distance between genome sequences [6].

The paper "A Metric on the Space of Rooted Phylogenetic Trees" by Wang and Guo proposed a new metric on the space of rooted phylogenetic trees, which can be calculated in polynomial time with the size of the compared trees [7].

Liao *et al.* developed a method to classify Small GTPases and non-small GTPases in their paper "Classification of small GTPases with hybrid protein features and advanced machine learning techniques" [8]. In the paper "Identification of Attention Deficit/Hyperactivity Disorder in Children Using Multiple ERP Features", Li *et al.* used non-invasive event-related potential (ERP) features for Attention deficit hyperactivity disorder (ADHD) prediction [9]. The paper "Low Rank Representation and its application in bioinformatics" by You, Cai and Huang reviews the theoretical and numerical models based on low rank representation and their applications in bioinformatics area [10].

2212-392X/18 $58.00+.00

This thematic issue is the result of many people's contribution, support and cooperation. We appreciate the authors for submitting their works to this thematic issue, and we thank the reviewers for their hard work in reviewing the papers. We also thank the Editor in Chief and the staff of Current Bioinformatics for their valuable help to make this issue possible.

REFERENCES

[1] Liu H, Luo L, Cheng Z, Sun J, Guan J, Zheng J, Zhou S. Group-sparse modeling drug-kinase networks for predicting combinatorial drug sensitivity in cancer cells. Curr Bioinform 2018; 13(5): 437-43.

[2] Wang Y, Zhang J, Bu D, Sun S. TagDict: prediction of theoretical spectra of peptides based on a tag dictionary. Curr Bioinform 2018; 13(5): 444-9.

[3] Guo X, Gao L, Wang Y, Chiu DKY, Wang B, Deng Y, Wen X. Large-scale investigation of long noncoding RNA secondary structures in human and mouse. Curr Bioinform 2018; 13(5): 450-60.

[4] Nie L, Deng L, Fan C, Zhan W, Tang Y. Prediction of protein S-sulfenylation sites using a deep belief network. Curr Bioinform 2018; 13(5): 461-7.

[5] Su Y, Li Y, Zhang Z, Pan L. Features identification for phenotypic classification based on genes and gene pairs. Curr Bioinform 2018; 13(5): 468-78.

[6] Chan Y, Xu K, Lan H, Schmidt B, Peng S, Liu W. MyPhi: efficient levenshtein distance computation on xeon phi based architectures. Curr Bioinform 2018; 13(5): 479-86.

[7] Wang J, Guo M. A metric on the space of rooted phylogenetic trees. Curr Bioinform 2018; 13(5): 487-91.

[8] Liao Z, Wan S, He Y, Zou Q. Classification of small GTPases with hybrid protein features and advanced machine learning techniques. Curr Bioinform 2018; 13(5): 492-500.

[9] Li W, Zhou T, Zou L, Lu J, Liu H, Wang S. Identification of attention deficit/hyperactivity disorder in children using multiple ERP features. Curr Bioinform 2018; 13(5): 501-7.

[10] You Y, Cai H, Chen J. Low rank representation and its application in bioinformatics. Curr Bioinform 2018; 13(5): 508-17.

Shuigeng Zhou
(Guest Editor)
Fudan University
Shangai
China

Fa Zhang
(Guest Editor)
Institute of Computing Technologies
Chinese Academy of Sciences
Beijing
China

Le Zhang
(Guest Editor)
Southwest University
Chongqing
China