Editorial

Latest Computational Techniques for Big Data Era Bioinformatics Problems

In recent years, computational methods have been employed extensively in bioinformatics researches, including protein classification, motif detection, phylogenetics tree reconstruction, traditional Chinese medicine analysis, *etc.* In term of the explosion of various biology omics data, traditional bioinformatics algorithms fall behind the big data era.

Big data computational techniques have advanced quickly over the past few years. Several novel methods were reported in the scholar and industry fields. For example, affinity propagation was published in Science as a novel clustering algorithm, and deep learning has become a hot topic in the predictions and classifications which is capable of processing big data. Parallel mechanisms, such as Spark and Hadoop, are also developed to speed up the algorithms. Computer scientists devote themselves to the advanced large scale data analysis. However, the applications on bioinformatics and genomics are limited and fall behind the techniques.

This special issue will target the recent large-scale computational techniques reviews in genomics applications. The editors expect to collect a set of recent advances in the related topics, to provide a platform for researchers to exchange their innovative ideas and genomics data.

In this thematic issue, four outstanding reviews were presented from different countries and regions, including P.R.China, USA, and Germany.

Feng *et al.* gave a review of ancestral genome reconstruction on the whole genome level. Comparative genomics, evolutionary biology, and cancer researches all require tools to infer the genome content and gene ordering of ancestral genomes by using genomic structural variants, such as rearrangements, duplications, gene gains and gene losses. There are mainly two kinds of computational approaches in the ancestral genome reconstruction study. These approaches are either based on the evolutionary distance/ events among genomes, or the similarities of homologies /adjacencies. In this paper, Feng *et al.* reviewed the principles, algorithms and further improvements of these approaches; they also provided reviews on the advantages and limitations of these approaches in dealing with different datasets, evolutionary event and reconstruction problems.

Du summarized computational resources for predicting protein submitochondrial locations. Over the last two decades, a large number of software resources have been released to predict protein submitochondrial locations. These resources include online predictors, databases, visualization utilities and fundamental tools for developing new predictive models. Du reviewed the motivations, biological backgrounds, basic concepts, theoretical framework and computational implementations of widely used resources. They also provided a comprehensive comparison of the available resource to facilitate the effective application of these resources in practical studies.

Fu *et al.* gave a survey on computational methods on identification of transcription factor binding site. Difficulty on identification of transcription factor binding site dues to the short sizes of motifs, compared with those hundreds or thousands *bp* background noise sequences. Moreover, the motif instance of a transcription factor is likely to mutate partially. In this paper, Fu *et al.* summarized the computational methods of prediction the transcription factor binding site in recent years.

Each paper in this special issue was extensively peer-reviewed by more than two external reviewers. I would like to thank all the authors for contributing their work to our hot thematic issue and all the reviewers for their time and efforts. At last, I would like to thank the Editor-in-Chief of Current Genomics, Prof. Christian Néri, and Editorial Office of Bentham Science Publishers, for their support during the whole processes.

Quan Zou

Guest Editor

School of Computer Science and Technology
Tianjin University
Tianjin 300354
P.R. China