

PREFACE

PICB: facing the new challenge and opportunity of big data

Zefeng Wang*

CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

* Correspondence: wangzefeng@picb.ac.cn

The CAS-MPG Partner Institute for Computational Biology (PICB) is a research institute established by the Max Planck Society of Germany and Chinese Academy of Sciences in 2005. At that time, it is the only institute in the world that is dedicated primarily to the interdisciplinary field of computational biology. The overarching research goal of PICB is to understand and interpret the principal of life using mathematic language. This institute is the fruit of the rich history of scientific collaboration between China and Germany, and reflects the visions of the leading scientists in both countries.

RESEARCH CONCEPT

PICB was established with initial emphasis on developing theoretical tools and mathematic algorithm to model life. With the evolution of the computational biology field, this institute has transformed from an exclusively theoretical research institute into rather a merger of theoretical and experimental biology though recruiting new research groups that conduct both “dry” and “wet” research. The recent rapid expansion of biological big data further transformed the field of computational biology from a hypothesis-driven science toward a data-driven science. Such shift of research paradigm presents both a challenge and a unique opportunity, and the innovative use of big data has been involved in the research of all groups in PICB.

Currently PICB has three focal areas of research: (i) integrative study of gene regulation using multi-omics data, (ii) system model and simulation for the regulation of complex trait, and (iii) human evolution and adaptation model using population genetics. These three directions represent integrative analysis of multi-omics data, computational systems biology, and population genetics/evolution, all of which will flourish on the foundation of bio-medical big data. The groups of PICB study diverse types of key biological questions involving innovative analyses of bio-medical big data, including global regulation of gene expression via epigenetics control and RNA, systematic modeling of human development and ageing, kinetic modeling of plant metabolism, adaptation mechanism underlying human divergence and history. In addition, PICB researchers also seek to develop new algorithms and network modeling approaches to analyze multi-omics data that includes diverse data-types like transcriptome data, gene regulation data, epigenetic data, proteomics, and metabolomics data.

ESTABLISHING THE BIG DATA CENTER

The rapid accumulation of big data in life science and medicine has transformed every filed of biology in recent years. Extensive and innovative analysis of bio-medical big data is a key to the success of almost every large research initiative in life science, and has also provided a unique opportunity to the field of computational biology. To embrace the challenge of big data, PICB has taken this unique opportunity by starting a scientific

mission to translate big data into novel knowledge of human biology. In the past few years, PICB has been advocating for the construction of a centralized national-level bio-medical big data infrastructure, and we are very encouraged by the responses of Chinese government that recognizes the crucial and urgent needs for this national infrastructure. In July 2016, the construction of a national bio-medical big data center was formally approved as one of national science and technology infrastructure in the 13th five-year strategic development plan of China. This infrastructure should achieve superb capacity of safe data storage, standardized data processing, systematic data integration across multiple data types and in-depth data mining.

To prepare for this national initiative, PICB has established a bio-med big data center. A new faculty, Dr. Yixue Li, was recruited to PICB as the director of this big data center. Several faculties were recruited by PICB to work exclusively for the big data center. In addition, Dr. Guoping Zhao was recruited as an adjunct member of PICB to work as the leading investigator of the big data center in PICB. PICB has also obtained additional space from SIBS to accommodate the bio-med big data center. Our bio-med big data center seeks to serve as a hub for the storage, standardization and sharing of bio-medical big data (such as multi-omics data), and also provide various data analysis service to the entire research community. This center also serves as a prototype for the national infrastructure of bio-medical big data, which will be formally launched after an open call of competitive proposals.

To facilitate the construction of national infrastructure of bio-medical big data, SIBS/PICB signed a contract with Gui'an government in June 2017 to build a data storage/backup node in the national big data industrial park in Gui'an, Guizhou Province. The operation of the facility in Gui'an can save energy consumption by 80% per year. In addition, the long-term plan will include building additional nodes of the infrastructure to form a network over the entire country to provide regional service. PICB is running the big data center in both sites of Gui'an and Shanghai, which is supported by each local government.

FUTURE PROSPECTS

With the new challenge and opportunity of expanding biological big data, PICB has taken a unique position to play a leading role in this exciting area and serves as a national and international hub for new technologies in data achieving, integration and analysis. Significant efforts must continue to maintain PICB as a highly respected and influential international research institute in the field of computational biology at big data era. The institute will initiate excellent international research platforms on data-driven biological science and help attract outstanding scientists in computational biology.