

Big Data in Biomedical Informatics

Lei Liu

Shanghai Center for Bioinformation Technology

Shanghai Academy of Science and Technology

Abstract

The emergence of genome-wise high-throughput technologies have lead to the creation of the interdisciplinary field between biotechnology and information technology, where bioinformatics comes out. With the application of these techniques in medical research, the concept of "translational medicine" has obtained more and more attentions and concerns. How to introduce the results of basic biomedical researches into clinical practices? How to promote basic researches with the needs of clinical practices? These questions are the core issues in translational medicine. To solve these problems, information technology is the key. This makes the intercross and fusion among the clinical practice, biology, and information science happen again and leads to the formation of the Biomedical Informatics.

The biggest challenge for current medical information systems is to reduce the redundancy of the data and functions, to seek the Inter-connection and Inter-communication technology for different data structures in the complex business environment and different system architectures, and to build an efficient and flexible content management system. To meet these challenges would facilitate the building of a high-quality information system integrating clinical data and basic biomedical research data. This is the core of translational medicine and the key to obtain the transformation of "from the laboratory bench to hospital bed". This information system should include three parts: de-identified clinical data repository (de-id CDR), omics databases including analysis platform, and tissue bank (disease and normal control) adhering to the requirements of research design.

The clinical data acquisition, storage, management, and applications related health information technologies are essential for the 21st century medical and health services. During the healthcare informization projects with the aim to improve the quality of care and control health care costs (or conduct clinical researches), there is an "irreconcilable conflict:" between personalized clinical data collection and enhanced medical workflows control. The automatic or semi-automatically form filling through the docking PACS, HIS, EMR and LIS systems with information system for research can significantly improve the accuracy of the data and work efficiency. However, in existing medical information systems, there are still a large number of very important clinical data exists only in the form of free text documents and reports. To solve this problem, medical natural language processing technology has become a hot and big challenge field in biomedical informatics research. The automatic processing of medical text in English has been undertaken for many years, and some of the achievements have been transformed into commercial applications. In China, medical record documentation of automatic processing technology-oriented research is still very rarely reported. Considering the huge number of population size in China, Chinese medical language processing technology must be one of promising study fields in translational medicine research.