

Big Data Needs Big Ideas: Towards Formation of Genome Information Society through Biomedical Genomics and Environmental Metagenomics

Takashi Gojobori
Center for Information Biology,
National Institute of Genetics, Mishima
e-mail: tgojobor@nig.ac.jp

As we know, Next-generation sequencing (NGS) Technologies provides us with unique opportunities of examining health conditions by biomedical genomics and of monitoring the surrounding environments by micro-organismal metagenomics. However, we are confronted by difficulties in appropriate handling of the so-called big data, because NGS technologies produce an enormously large amount of sequence data in conducting both researches of genomics and metagenomics.

To resolve these essential difficulties, we have to challenge at least a few number of issues; (1) construction of database with contents-rich annotation, (2) developments of time-efficient algorithms and pipelines for processing a huge amount of sequence data, and (3) setting up of scientifically significant problems. The last issue is the most critically important because production of high-throughput data does not make sense without scientific problems to be solved.

I would state it as “Big Data needs Big Ideas,” because I believe that data-driven scientific discoveries are impossible without any working hypotheses. Of course, appropriate working hypotheses may be obtained once a large amount of data is examined. Thus, Big Data and Big Ideas should be conceived to be inter-relationship with a positive-spiral structure. In the present talk, I would propose, as the vision of our future, new formation of the “genome information-oriented society” by taking biomedical genomics approaches to construction of one-of-care systems and marine metagenomics approaches to understanding of a diversity of sea water microorganisms as concrete examples.