

The Challenges from Biological Big Data

Yixue Li

Shanghai Center for Bioinformation Technology

Shanghai Academy of Science and Technology

In scientific view, now computational science is already become a third paradigm of scientific studies, together with experimental science, and then theoretical science. Then, for many problems, the theoretical models and experimental results grew too complicated to solve analytically, and people had to start simulating and data mining. Thousand years ago: science was empirical describing natural phenomena, Last few centuries: theoretical branch using models, generalizations. Last few decades: a computational branch simulating complex phenomena. Today: scientific studies are generating a whole lot of data, the “big data” and data exploration unify theory, experiment, and simulation. Especially in biological studies, data captured by various high-throughput instruments and platforms, such as “Omics” technology. The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration. Now we are already in big data era, all of us must to face the challenges.