Transcriptomics: Open Access

Big Next Generation Sequencing Data

Jagajjit Sahu[1] and Anupam Das Talukdar[2]*

[1]Distributed Information Centre, Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, India
[2]Department of Life Science and Bioinformatics, Assam University, Assam, India

Commentary

Science has kept changing its form from observational to experimental to data-driven in the field of life science. With the advancement of Next Generation Sequencing (NGS) technology, new findings are coming up with a great amount of responsibilities whereas storing and analysing these data is concern Li [1], Stephens et al. [2]. During last decade, the cost of sequencing has reduced heavily by allowing access to more scientists. A simple search in PUBMED can provide the scenario of exponential growth of the number of reports published using NGS technology. However, the deposition of raw data in the public domain is increasing dramatically outstripping the proper annotation of these data which is still half-cooked or ambiguous. The data scientists have already started facing difficulty in envisaging the scientific standpoint of handling the data deluge. The only solution to sail across this flood of data is to develop efficient and flexible algorithms which can analyse the raw data and extract meaningful information. Already approaches like compressive genomics, cloud computing, NoSQL, etc. have been coined to deal with the big data issue. Compressive algorithms help in reducing the task of computing on redundancy data by allowing direct computation on the compressed data Loh [3]. This approach can also be implemented with tools such as Basic Local Alignment Search Tool (BLAST) to achieve sublinear analysis. Cloud computing is basically an alternative to the economic and efficiency problems of the common user who always has to think of upgrading his available computational facilities to handle the high-throughput data Zhou [4]. Researchers have also started using NoSQL to store the data in a more classified way. Unlike the available relational databases (MySQL), NoSQL stores data using graphs, objects and many more which provides an user-friendly as well as more informative view to the large-scale data Have [5]. Especially graph databases such as AllegroGraph, Neo4J, etc. are being preferred by bioinformaticians. While it comes to the analysis of massive data, Neural network approaches (Nns) owe their dynamic efficiency towards all types of biological data Chen [6]. The underlying principle of Nns is the machine learning approaches which enhance the algorithms to recognize patterns, classify the data and so many other features. The traditional way of bioinformatics analysis has become obsolete. Systems biology combines the computational tools, statistical and mathematical models along with high-throughput techniques to analyze the core components in a biological systems and bring out the most significant information such as various regulatory networks along with functions of specific regulators like miRNAs in the network Li et al., [7]. The available computational facilities are not enough to handle the big NGS data; however, there should be more focus on development of powerful algorithms so that the researchers would be able to know where they are heading with their own data.

References

1. Li Y, Chen L (2014) Big Biological Data: Challenges and Opportunities. Genomics Proteomics Bioinformatics 12: 187-189.

2. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015) Big Data: Astronomical or Genomical? PLOS Biology 13: e1002195.

3. Loh PR, Baym M, Berger B (2012) Compressive genomics. Nat Biotechnol 30: 627-630.

4. Zhou S, Liao R, Guan J (2013) When cloud computing meets bioinformatics: a review. J Bioinform Comput Biol 11: 1330002.

5. Have CT, Jensen LJ (2013) Are graph databases ready for bioinformatics? Bioinformatics 29: 3107-3108.

6. Chen K, Kurgan LA (2012) Neural Networks in Bioinformatics: Handbook of Natural Computing. Spinger-verlag Berlin Heidelberg, Germany.

7. Li Z, Qin T, Wang K, Hackenberg M, Yan J, et al. (2015) Integrated microRNA, mRNA, and protein expression profiling reveals microRNA regulatory networks in rat kidney treated with a carcinogenic dose of aristolochic acid. BMC Genomics 16: 365.

***Corresponding author:** Anupam Das Talukdar, Department of Life Science and Bioinformatics, Assam University, Silchar-11, Assam, India, Tel: +91-9401416452; E-mail: anupam@bioinfoaus.ac.in