NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 3, ISSUE 5, May.-2016

REVIEW OF ASSESSMENT OF REAL TIME GENOMIC DATA GENERATED BY NANOPORE SEQUENCING TECHNOLOGY AND ITS CONSEQUENCES FOR THE FREEDOM OF THE INDIVIDUAL

Noorah Kazi
Department of Information Technology, Pravara Rural Engineering College, Loni, India

ABSTRACT :

The next generation sequencing technologies have revolutionised the generation of raw sequence data of the human genome. Sequencing of human genome can now be completed in a short span of time using the latest next-generation sequencing technology. Data of the size of terabytes and more is generated during the sequencing operation in real time. The complexity of data manipulation and management is a huge challenge facing the researchers today. The review provides an insight into the available next generation sequencing technologies. A review of assessment of real time genomic data generated by nanopore sequencing technology and its consequences for the freedom of the individual brings out the need for real time data analysis algorithms of genomic data. The research concludes by the impact of nanopore sequencing technology and its consequence on individual freedom.

INTRODUCTION:

Sequencing technologies have provided a major breakthrough in the field of genomics. The advancement of sequencing technologies has made it possible to sequence genomes at a remarkable rate and at modest cost. It has brought about a significant transformation in the field of personalized medicine, cancer genome analysis, drug discovery and has made a favorable impact on well-being of society by making genetic study possible. The completion of Human Genome project (Genome.gov, 2015) in the year 2003 marked an important milestone in the era of genomics. This has been accomplished with the help of Sanger sequencing, a first generation sequencing method widely used until recently. Sanger Sequencing has been developed by Sanger and colleagues in the year 1977 (Sanger, Nicklen and Coulson, 1977).

Several sequencing platforms have emerged since then to perform high throughput and low cost sequencing. Schadt and his colleagues have classified sequencing technologies into three generations (Schadt, Turner, & Kasarskis, 2010).The process of sequencing DNA in first generation sequencing technology includes three phases sample preparation, physical sequencing and reassembly. The second generation sequencing technologies require amplification of target DNA and involve massive parallel sequencing. These have advanced to provide high throughput at modest cost (Metzker, 2009). The third generation sequencing technologies include single molecule sequencing methods. Unlike the first and second generation technologies these do not require the steps of amplification and expensive sample

preparation. Nanopore sequencing technologies have emerged promising in this field. It has advanced amongst the third generation technologies as an exceptional technology for low cost and high throughput sequencing.

EXISTING SYSTEMS:

NANOPORE SEQUENCING TECHNOLOGY:

The concept of nanopore sequencing technology was independently proposed by Deamer at University of California and George Church at Harvard University (Branton et al., 2008).The concept is to electrophoretically drive a strand of DNA or RNA through a nanopore of suitable diameter. Each nucleobase modulates a distinct ionic current while passing through the nanopore. The change in modulating current helps in identifying the sequences. This concept has been implemented successfuly by the recently released Oxford Nanopore Technologies (ONT) MinION sequencing device (Nanoporetech.com, 2015).It establishes the potential of the nanopore sequencing platform. Recent studies using the device have shown that the device has been able to generate long reads and it has been possible in resolving the organization of cancer-testis gene family. As it is the first hand held, nanopore-based single molecule sequencer available to researchers, it is important to consider the MinION in the assessment of nanopore sequencing technology. The device is expected to be available at a low cost and produces data in real-time. The data generated from Oxford minION is stored in cloud and analysed using services from Metrichor(Metrichor.com, 2015).This cloud-based software uses hidden Markov models(HMMs) to infer sequences.

MinION:

The MinION sequences individual DNA molecules, providing very long read lengths to help overcome some of the drawbacks of short-read sequencing (Jain et al., 2015). POISSON algorithm (Szalay & Golovchenko, 2015) has been recently used to accurately sequence DNA bacteriophage upto 99% accuracy. The cloud platform provides a secure and moderately expensive infrastructure for users to run complete analysis in a short time. A collection of tools and techniques for analysis of nanopore sequencing data are have been recently developed (Loman and Quinlan, 2014) (Watson et al., 2014)..

The research today is facing the challenge of processing and analyzing massive volume of nanopore sequencing data generated in real time and available in the cloud. The availability of genetic information about individuals also have an impact on the freedom of an individual. Issues such as confidentiality, discrimination based on genetic characteristics need to be addressed.

SCOPE:

Although there have been remarkable advancement and commercialization of the SGS technologies they require sample preparation and complex algorithms for analysis. Nanopore sequencing technology requires inexpensive sample preparation. The generation of genomic data in real time by nanopore sequencers makes it a challenging task to analyse this data in real time and infer potential information from it. Analysis of genomic data is essential to

extract information about gene functions, role of them in rare diseases and in the development of the personalized medicine field. Personalized diagnosis and therapy has become commercially available due to the drop in sequencing costs, and with data analysis and sequencing speed going up (Qin et al., 2015).

A set of existing algorithms used for genomic data analysis needs to be studied. Since research is still in its early stages for real time data analysis of genomic data generated through nanopore sequencing the research will focus on studying existing algorithms as well as developing novel algorithms for real time data analysis in cloud. While large datasets containing genomic data of individuals is becoming available on cloud it becomes essential to consider privacy risks associated with it. With access to algorithms and techniques it is principally possible to identify the individual based on genetic information available freely.

The article 1 of "Universal Declaration on the Human Genome and Human Rights" states that the human genome is the heritage of humanity. It is the recognition of their inherent dignity and diversity [13].It is important to consider the issues such as discrimination based on genetic characteristics, confidentiality of genetic data held for research and protection of an individual's identity. Having entered the era of big data and genetic information being available in the cloud with less or no security it becomes important to assess the impact of these technologies on the freedom of an individual.

PROPOSED SYSTEM:

The proposed system involves assessment of the nanopore as well as considers its impact on the freedom of individual. The impact of using privacy protecting algorithms and genomic data sharing during analysis is being studied.

The research studies the existing data analysis algorithms used for genomic data and develop novel genomic real time data analysis algorithms. The study is a combination of theoretical and practical research work. The research initially assesses the nanopore sequencing technology. It will involve working with big genomic data on the cloud platform. Ethical considerations while collection of genomic datasets and working on the datasets with respect to consent, confidentiality and sharing of results should be considered. The genomic datasets generated by nanopore sequencing technology are required so that data analysis can carried out on genomic nanopore data. The design of existing algorithms being used for data analysis in nanopore sequencing will be studied and factors responsible for accuracy, speed and cost will be considered. The available computational tools used for data will need to be installed and prepared for study. It will include analysis of the existing algorithms used in genomic data analysis. An initial step to protect the individual and ways in which one can protect the privacy and freedom of an individual will be studied.

The rights, benefits and risks of individuals contributing to research needs to be declared and explained clearly to participants. This is considered important for protection of privacy of individuals. New algorithms designed as part of the study will be implemented and tested on genomic data generated using nanopore sequencing technology.

CONCLUSION:

The proposed system will provide researchers with new algorithms and tools for analysis of genomic data. It will be helpful in considering issues in personalized genomics. The algorithm implemented as a part of this project will help in analysis of rare diseases and eventually help towards drug discovery. This research will contribute towards future work and development in the field of nanopore sequencing technologies.

REFERENCES:

1] Branton, D., Deamer, D., Marziali, A., Bayley, H., Benner, S., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S., Krstic, P., Lindsay, S., Ling, X., Mastrangelo, C., Meller, A., Oliver, J., Pershin, Y., Ramsey, J., Riehn, R., Soni, G., Tabard-Cossa, V., Wanunu, M., Wiggin, M. and Schloss, J. (2008). *"The potential and challenges of nanopore sequencing"*. Nat Biotechnol, 26(10), pp.1146-1153.

2] Genome.gov, (2015). *"All About The Human Genome Project (HGP)"*. [online] Available at: http://www.genome.gov/10001772 [Accessed 14 Jun. 2015].

3] Jain, M., Fiddes, I., Miga, K., Olsen, H., Paten, B. and Akeson, M. (2015). *"Improved data analysis for the MinION nanopore sequencer"*. Nature Methods, 12(4), pp.351-356.

4] Loman, N. and Quinlan, A. (2014). *"Poretools: a toolkit for analyzing nanopore sequence data"*.Bioinformatics, 30(23), pp.3399-3401.

5] Metrichor.com, (2015). Metrichor. [online] Available at: http://metrichor.com [Accessed 8 Jul. 2015].

6] Metzker, M. (2009). *"Sequencing technologies — the next generation"*. Nat Rev Genet, 11(1), 31-46.'

7] Ohchr.org, (2015). *"Declaration on the Human Genome and Human Rights"*. [online] Availableat:http://www.ohchr.org/EN/ProfessionalInterest/Pages/HumanGenomeAnd HumanR ghts.aspx [Accessed 10 Jul. 2015].

8] Sanger, F., Nicklen, S. and Coulson, A. (1977). *"DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences*", 74(12), pp.5463-5467.

9] Schadt, E. E., Turner, S., & Kasarskis, A. (2010). *"A window into third-generation sequencing. Human Molecular Genetics"*, 19(R2), 227-240.

10] Szalay, T., & Golovchenko, J. A. (2015). *"A de novo DNA Sequencing and Variant Calling Algorithm for Nanopores"*. 1-16.

11] Watson, M., Thomson, M., Risse, J., Talbot, R., Santoyo-Lopez, J., Gharbi, K. and Blaxter, M. (2014). "*poRe: an R package for the visualization and analysis of nanopore sequencing data"*. Bioinformatics, 31(1), pp.114-115.

12] Yiming, Q., Krishna, Y. H., Jing, Q., Bin, Y., & Junwen, W. (2015). *"The Current Status and Challenges in Computational Analysis of Genomic Big Data"*. Big Data Research, 2(1), 12-18.