Shodhsamhita: Journal of Fundamental & Comparative Research
Vol. II, Spl.Issue-01, Dec-2021
ISSN: 2277-7067

"HEALTH BIG DATA AND PRECISION MEDICINE: IMPORTANCE AND CHALLENGES"

AMULIA P M

Assistant Professor, Naipunnya School of Management – Cherthala-Kerala

Email: amuliamarkose@gmail.com

ABSTRACT

Precision Medicine is a novel approach in healthcare which takes into account a person's health data to make effective prediction regarding his health and to improve treatment of diseases. According to Blagoj (2018) this type of personalized, predictive, participatory and preventive medicine is based on using electronic health records (EHRs) and huge amounts of complex biomedical data and high quality –omic data. This data if efficiently integrated, cleaned and analysed could revolutionise health care system with precision medicine. Further to improving health care system, electronic health data can be used in precision medicine for grouping and sub-grouping of patients based on similar health data. According to Pedro and Tatiana (2021) with the coming of big data, the fields of precision medicine and public health are converging into precision public health. The paper focuses on the importance of precision medicine specifying the different data types and data mining tools used in precision medicine. This paper also reviews the different challenges that could arise when using health data in precision medicine.

Keywords: Precision medicine, EHR (Electronic health record), -omic data, biomarker, precision public health

INTRODUCTION

The treatment method followed by the healthcare system today is 'one size fits all' [2] approach. A new model in healthcare system, 'precision medicine model' is focused on personalising the healthcare system. In addition to genomic data, precision medicine analyses the health records and environmental details of a person. This emerging approach could improve healthcare by providing personalised, predictive, participatory and preventive medicine. With precision medicine earlier intervention and advanced diagnostics of diseases and economical treatments are possible. Treatment received today is based on trial and error method which does not consider person to person differences. This is why medicine prescribed for one patient may not be effective for another even for the same disease.

Precision medicine looks at the root cause of the illness rather than treating the symptoms. The gene determines the treatment received.

Furthermore, precision public health could analyse large amounts of health data to predict and warn large populations about a particular disease. This would require a large amount of health data analysis at the population level.

LITERATURE REVIEW

Humans are different so medicine prescribed for one person may not be effective for another, even for the same disease. Many diseases are linked to genes, life style, and environment. Certain gene changes can cause disease. Genetic mutations (permanent change in one or more specific genes) can cause diseases. If a person inherits a genetic mutation that causes a certain disease, then he or she will usually get the disease. Other changes or differences in genes, called genetic variants, may increase or decrease a person's risk of developing a particular disease. One person's heart disease, diabetes, or cancer act differently from another's. Knowing how genes and diseases interact can help predict and fine-tune treatments to make them work better. Some people have inherited conditions that make them more susceptible to certain diseases. If these conditions are identified early enough then the chances of them contracting the disease can be greatly reduced.

According to paper 'Review of Medication Therapy for the Prevention of Sickle Cell Crisis' 2018 Currently, universal treatment approaches for SCD(Sickle cell disease) revolve

around infection prevention through the use of antibiotics, pain killers, vaccines, and education; blood transfusions for prevention of stroke and hydroxyurea. The medicines used in treatments can have many side effects. 6 billion people world-wide are affected by sickle cell disease.

Sickle cell anemia is an inherited disease caused due to defective hemoglobin where there aren't enough healthy red blood cells to carry oxygen throughout your body. The hemoglobin becomes sickle shaped which can cause severe pain, organ failure, stroke and death.

About 5% to 10% of breast cancer is due to gene mutations that are inherited from parent. Many life threatening diseases such as cancer, diabetes, alzeimer's etc. are related to gene alterations.

TYPES OF HEALTH BIG DATA AND ITS IMPORTANCE

Terabytes of data is available in health care today which contains abundant information. Biological health data can be categorised as -omic data and electronic health record data. -Omic data [7] contains a comprehensive catalog of molecular profiles (e.g. genomic, transcriptomic, epigenomic, proteomic, and metabolomic). The goal of analysing -omic data is to extract molecular profiles, identify statistically significant molecules and molecular interactions and find significant biomarkers.

Electronic health record (EHR) [6] data can be unstructured, semi structured or structured. EHR data includes clinical data such as observation data, evaluation data or action data which includes surgery report or lab test results. This form of heath data could be collected and recorded by humans (such as doctor, nurse or patient) or using machine (such as thermometer, monitoring system, scanning system or other IoHT (Internet of Health Things) devices. Other patient data like demographical details, life style and food factors are also considered in preventive and predictive analysis.

DATA COLLECTION AND CASE STUDIES

SICKLE CELL DISEASE

Each molecule of haemoglobin contains four globin proteins (2 beta globin and 2alpha globin). HBB gene provides instructions for making a protein called beta-globin. Alpha-

globin is produced from another gene called HBA. Sickle cell disease is caused by replacement of glutamic acid by valine at 6th position of the beta globin protein chain of the haemoglobin molecule of chromosome 11.

BREAST CANCER

About 5% to 10% of breast cancer is due to gene mutations that are inherited from parent. The most common cause of hereditary breast cancer is due to inherited mutation in BRCA1 or BRCA2 gene. The association between breast cancer and mutation in BRCA1 and BRCA2 tumour suppressor genes has allowed investigators to quantify lifetime risk of developing cancer. BRCA1 gene normally acts to restrain the growth of cells in the breast but if mutated is the cause of 50% to 80% of breast cancer by age 70 [3] and more than 80% of inherited breast and ovarian cancers.

Humans have approximately 30,000 genes in a cell. The human genome contains approximately 3 billion of the base pairs, which reside in the 23 pairs of chromosomes within the nucleus of all our cells

CHALLENGES IN HEALTH BIG DATA

Big Data analytics on health data has huge benefits when it comes to improving the quality of health care but it has to overcome many challenges.

This voluminous amount of health data are gathered from different sources. Clinical data of a person, such as lab test results, may be available at multiple laboratories while his surgery details, medication and diagnostic details might be available at a particular hospital. Present health care system is a disconnected system where data resides in isolated islands or 'silos'. Healthcare workers have no access to data needed. This complex data that is heterogeneous [3], structured, semi-structured or unstructured, discrete or continuous, if efficiently integrated and analysed can improve the quality of health care.

Maintaining data quality and reliability is one of the challenges in precision medicine since health data is available in different 'data silos'[1] in different format.

Doctor diagnosis can be written or dictated clinical notes describing the patient condition which contribute to unstructured health data. Data recorded by health devices can be discrete or continuous. Data collected after lab tests is discrete whereas data collected by a blood pressure detecting device is continuous. According to Pierluigi (2019) these data silos

form a scattered "island" view requiring retrieval of information from different systems in order to reconstruct the full view on the patient's health pathway.

Another problem is that data format can change from time to time. A form available to a doctor for entering patient details may change from time to time. It might contain new variables or certain existing variables may have become irrelevant with time and may be deleted from the available form.

Electronic health data are gathered from medical devices, IoHT (Internet of Health Things) devices and entered manually by doctors, nurses, health workers or patient. Although large amount of health data is gathered from devices, in clinical databases there is still a vast amount of data that is entered manually. So the accuracy of the EHR data greatly relies on the person who enters the data and on the health care devices that record the data. Reading and interpretation errors and lack of standardisation errors can occur while entering observation or evaluation data.

Quality of -omic data is controlled by biological, instrumental and environmental factors such as sample contamination, noise and batch effect.

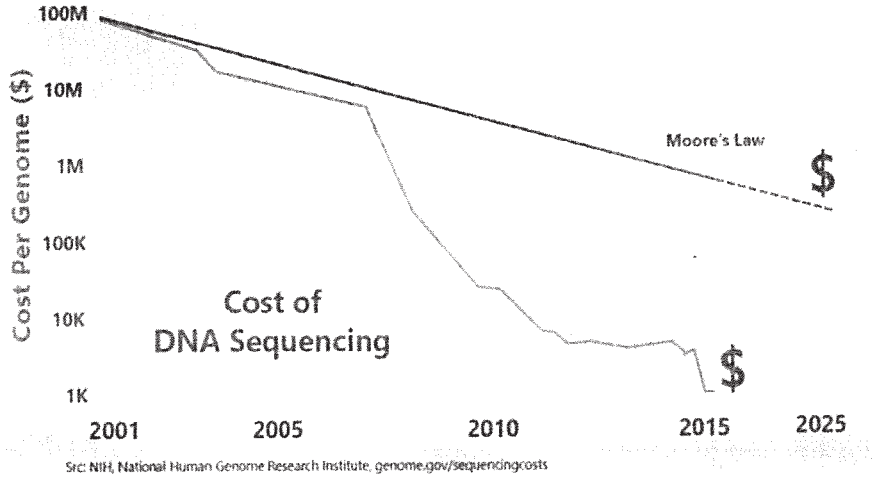
Another challenge is security and privacy [9] issues. The owner of health data is the person itself. So using a person's health data by an organisation for data analytics can raise privacy and security challenges.

FINDINGS

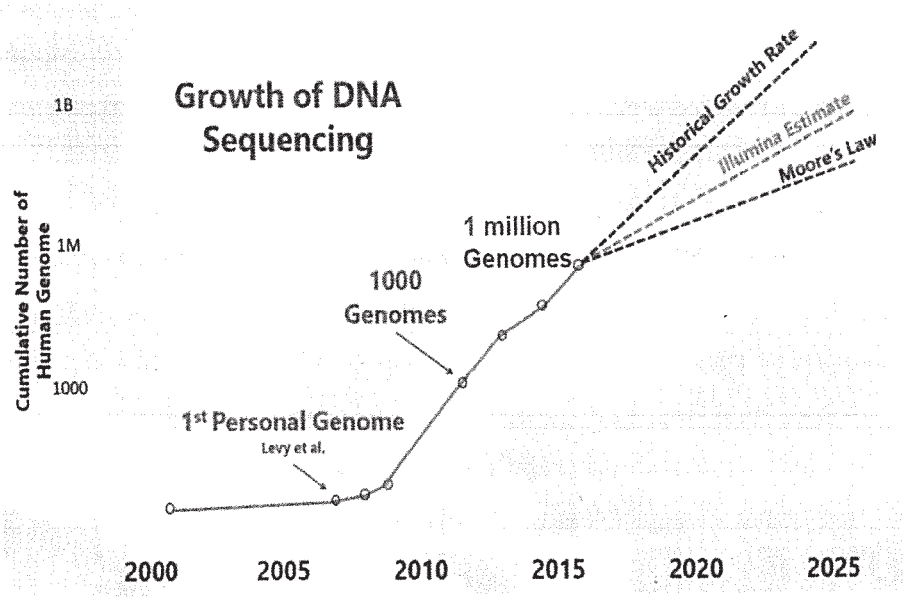
Precision medicine technique combines data science and medicine. It uses genome manipulation and patient data and has the potential to prevent and cure many diseases including breast cancer, sickle cell anemia. By identifying the order and sequence of the four bases within the DNA it is possible to extract molecular profiles, identify statistically significant molecules and molecular interactions and find significant biomarkers. -omic data can be accumulated using Next generation sequencing (NGS), mass spectrometry (MS) or microarray and nuclear magnetic resonance (NMR) techniques. DNA sequencing can be used to locate single nucleotide polymorphism (SNP or snips)[8], double nucleotide polymorphism (DNP) or triple nucleotide polymorphism (TNP). SNP's are used in genome wide association studies (GWAS) [8].

With advancement in technology the cost of DNA sequencing is dropping every day and more and more people are doing DNA sequencing.

COST OF DNA SEQUENCING



GROWTH OF DNA SEQUENCING



Big data, AI, the genome, and everything (sponsored by Microsoft) Vijay Narayanan (Microsoft)

Precision medicine can identify and reduce the risk of developing inherited cancer. Biomarker testing used in precision medicine supports targeted treatment options that can reduce the damage caused by traditional treatment techniques. Biomarker testing looks for

genes, proteins or other substances that support the growth of cancer cells. PARP (Poly ADP ribose polymerase) inhibitors are new chemotherapeutic agents developed at targeting cancer cells in BRCA1 mutation cancer cells over normal cells. PARP inhibitors interfere with certain enzymes that help cancer cells repair. Blocking these enzymes allows the cancer cells to die. These inhibitors are targeted therapies. They target cancer cells and have less effect on healthy cells than traditional chemotherapy.

Like Blood-banks, Bio-banks in a safe and well controlled system could store and manage genomic data. These biobanks should function under proper protocols and should be organised, systematised and searchable. Guidelines and tools for integrating health data should be maintained.

This could help doctors to build a genetic description of their patient. Other patient details such as life style, environment exposure, food habits, data from IoHT device, observations and evaluation reports, lab and clinical reports, social life data could also be included.

Precision medicine analyses health data and predicts the outcome for an individual or a group of people. Precision medicine takes into account gene-gene interaction, gene-environment interaction to provide personalised, predictive heath care. Multiple social and environmental factors such as air pollution, food habits, excercising habits, exposure to contaminated products, exposure to asbestos etc. could also be included into the patient medical history. Guidelines and tools for integrating health data should be maintained. This could help doctors to build a genetic description of their patient.

Pharmacogenomics [4] is the combination of genomics and pharmacology. It focuses on how a person's genes affect his response towards a particular medicine. This study helps doctors prescribe drug that is most effective and safe for a person by analysing his DNA.

PRECISION MEDICINE IN PUBLIC HEALTH

The details of similar patients could be linked together using a common factor to create a disease network. Clustering based on genomic, clinical, demographic, ethnicity/race, age and gender can further advance precision medicine towards public health. Precision medicine approach identifies high risk individuals in general population, offering some promise of improving chances for disease prevention in specific groups. Heriditary form of breast cancer and ovarian cancer has been noted in certain ethnic groups such as Ashkenazi jewish population.

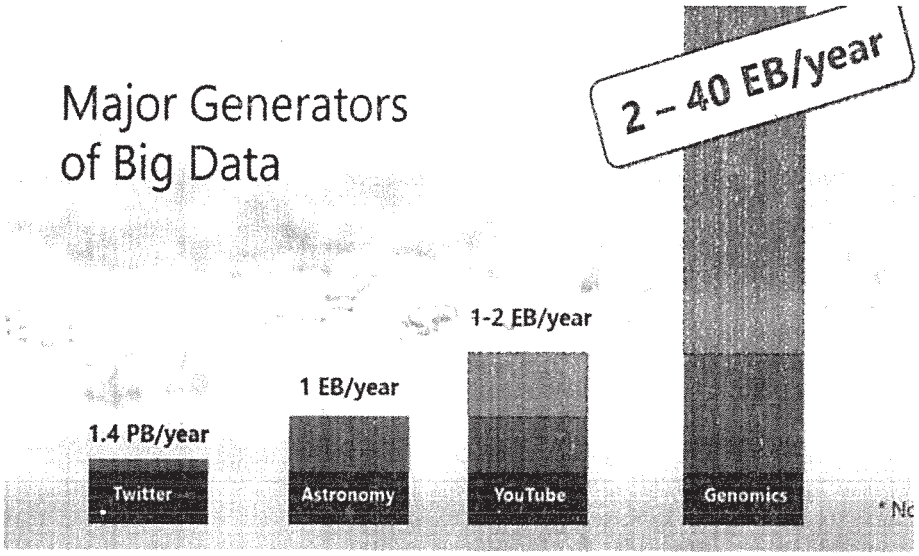
Patient disease network could generate collective intelligence to be used in identifying and studying specific group of patients. The details of similar patients could be linked together using a common factor to create a disease network.

Through genome sequencing, infectious diseases can be tracked and controlled. The DNA fingerprint [3] of the germ can be identified and doctors can prescribe the exact drug for the germ that is causing the sickness. Public health workers can also warn public through social media or other methods against the germ and the source of the germ thereby preventing additional cases.

Volunteers could be used in collecting medical records, genomic data and environment data for the biobanks. To improve accuracy and to reduce noise in the recorded health data, different data cleaning techniques such as binning, regression analysis, outlier analysis, clustering and classification methods could be used. Binning on mean, median or boundary values can be done to reduce noise in health data.

Specific clinical facilities and precautionary measures can be provided based on the grouping as sample size increases.

Advancement in big data technologies has led to the development of tools and platforms such as Apache Hadoop, Apache Spark, and Apache Storm that can analyse rapidly generated, diverse amount of data. Apache Parquet is a columnar storage file format that is well suited for genomic data. Amazon EMR (Amazon Elastic MapReduce) is a managed cluster platform that simplifies running big data frameworks, such as Apache Hadoop and Apache Spark to process and analyze vast amounts of data.



Big data, AI, the genome, and everything (sponsored by Microsoft) Vijay Narayanan (Microsoft)

CONCLUSION

Precision medicine delivers right treatment at the right time to the right person. Precision medicine by applying genomics and other relevant technologies analyses and identifies the biomarkers of specific diseases and can provide personalised solutions and targeted therapy. Precision medicine does not negate traditional medicine since traditional medicine is the basis of precision medicine. With more advancement in precision medicine diseases like cancer, diabetes, alzheimer's etc. can have early interventions and targeted treatment. Other factors such as patient's medical records, life style, environment conditions, race, gender, age etc. can be used in categorising patients. With advancement in big data technologies and tools zeta bytes of data could be analysed accurately and quickly. This can be further expanded at population level to provide better public precision health care system. Precision medicine will soon become an important part of health care system.

REFERENCES

1. From biobank and data silos into a data commons: convergence to support translational medicine, Rebecca Asiimwe, Stephanie Lam, Samuel Leung, Shanzhao Wang, Rachel Wan, Anna Tinker, Jessica N. McAlpine, Michelle M. M. Woo, David G. Huntsman and Aline Talhouk Published online 2021 Dec 4. doi: 10.1186/s12967-021- 03147-z https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8645144

2. The future of precision medicine: towards a more predictive personalized medicine Olivier Elemento DOI:10.1042/ETLS20190197 2020 https://www.pubmed.ncbi.nlm.nih.gov/32856697
3. From Big Data to Precision Medicine Tim Hulsen, Saumya S. Jamuar, Alan R. Moody, Jason H. Karnes, Orsolya Varga, Stine Hedensted, Roberto Spreafico, David A. Hafler and Eoin F. McKinney, Front. Med., 01 March 2019. https://www.frontiersin.org/articles/10.3389/fmed.2019.00034
4. Personalized medicine in breast cancer: pharmacogenomics approaches, Shabnam Jeibouei, Mohammad Esmael Akbari, Alireza Kalbasi, Amir Reza Aref, Mohammad Ajoudanian, Alireza Rezvani, and Hakimeh Zali Published online 2019 May 27. doi: 10.2147/PGPM.S167886. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6549747/
5. Precision Medicine from a Public Health Perspective Volume 39, 2018 Ramaswami, pp 153-168, Ramya Ramaswami, Ronald Bayer, and Sandro Galea
6. Big Data Analytics in Medicine and Healthcare ,Blagoj Ristevski and Ming Chen Published online 2018 May 10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6340124/#j_jib-20030_ref_006
7. Omic and Electronic Health Records Big Data Analytics for Precision Medicine Po-Yen Wu, Chih-Wen Cheng, Chanchala D. Kaddi, Janani Venugopalan, Ryan Hoffman, Members, IEEE, and May D. Wang, Senior Member IEEE, IEEE Trans Biomed Eng. Author manuscript; available in PMC 2018 Mar 20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5859562
8. Genomics for Disease Treatment and Prevention Cinnamon S. Bloss, Ph.D, Dilip V. Jeste, M.D, and Nicholas J. Schork, Ph.D, Author manuscript; available in PMC 2012 Mar 1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3073546
9. Legal, Ethical, and Financial Dilemmas in Electronic Health Record Adoption and Use Dean F. Sittig, PhD and Hardeep Singh, MD, MPH, Pediatrics,2011Apr https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065078