IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Big Data Might Just Cure Cancer - The research and the reality

Sudhir Allam, *Sr. Data Scientist, Department of Information Technology, USA*

***Abstract***—*This paper demonstrates how big data might be instrumental in curing cancer. Cancer is a disease that continues to cause many deaths all over the world due to its difficulty in detecting, diagnosis, and treatment. Figures from the World Health Organization show that the number of cases diagnosed in the U. S in 2018 was 1,735,350 while the number of cancer deaths was 609,640 [1]. Looking at these statistics demands more action in terms of research on various aspects of the disease especially early detection and finding its cure. To find the cure for cancer, this paper proposes the application of big data in predicting the cancer sequences and identifying its targeted therapy and cure. The application of big data transforms the way treatment decisions is made through an understanding of cancer's inner working [1]. The proposed big data analytical technique to treat cancer applies the information retrieved from multiple cancer types which will be vital in identifying the potential cure for cancer. The findings from this research show that the use of high-powered computer systems to collect and process large amounts of biological data has the potential to significantly impact progress in finding a cure for cancer. The main big data technique for discovering novel drugs could be molecular profiling. It's a method that scientists use to examine specific patients' cancerous cells to distinguish genetic variants and biomarkers that are unique to them. Biological markers are molecules that can be used to confirm the existence of a disorder or illness, and they are also being used to monitor how well an individual respond to treatments [1,2]. The amount of druggable tumor-specific genetic anomalies has increased significantly over the last decade, with biological markers matching interventions providing a major protective effect in many cancer types. As a result, molecular biology has been critical not just for tumor detection and pathogenesis, but also for therapeutic decision-making in clinical practice. The science of precision oncology has been transformed by the advent of next-generation sequencing technology and the growing number of large-scale cancer molecular profiling projects in many medical and research institutions across the world.*

**Keywords:** *Big data, cancer, molecular profiling, genome editing,*

## CONTEXT OF THE RESEARCH

According to the American Cancer Society's "2014 Cancer Facts & Statistics" study, it was projected that about 585,720 people could die from cancer in the United States.

That meant that one out of every 54 persons could die. In general, cancer numbers are alarming regardless of whether they offer positive or bad news, but in the context of the disease, science is bearing down on treatment for several of the different forms of cancer. For the last two and a half centuries, we have seen an incredible number of momentous advances in the fight against cancer, a disease known to the human race for thousands of years. The evidence found on this timeline provides an understanding of several important findings in the study of cancer. According to one report, naked mole-rats may be used in laboratories to learn why they are cancer immune. Studies conducted on the naked mole-rat family mole rats also found that their cells function differently than other kinds of rodent cells, making them extraordinarily unique and resistant to the development of cancer. Cancer cells in naked mole-rats have a very limited ability to replicate and long-life span is connected to their anatomy. in 2018, the FDA gave the approval to a tumor-targeting larotrectinib (a medication specifically developed to target NTRK gene fusions). The use is limited to patients who have deteriorated anywhere on the body following prior therapy with NTRK gene fusions who have unknown mutations. Secondly, it has been shown that molecularly targeted cancer treatments may be useful regardless of where the tumor is found. The medicine, named Kymriah (Tisagenlecleucel) has been approved to treat children and young adults under the age of 25 who have not improved with other therapies with acute B-cell lymphoblast leukemia. It is believed that scientists are currently focusing on a generalizable form of CAR T-cell therapy for adults and many other kinds of cancers.

## I. INTRODUCTION

The impact of cancer has been felt by every nation around the world. This is a disease that continues to wreak havoc in people's lives at every chance it gets. Statistics reveal that the disease is not getting any better among the people but instead is getting worse as the number of cases and deaths keeps increasing even with the interventions put in place. From 2013-1017, the rate of new cases in the United States was reported to be at 442.3 per 100,000 people while the mortality rate was 158.3 for every 100,000 people [2]. More men (189.5per 100,000men) were affected by the mortality rate than women (135.7 for every 100,000). The country spent over $150.8billion in national expenditure in 2018 for cancer care and these costs will likely increase in the future as the population increases [2]. As the treatments become more sophisticated, it will also increase the cost of care. Across the world, cancer greatly affects countries with a higher life expectancy, standard of living, and educational level. The number of new cancer cases across the world was 18.1million while the cancer-related deaths were 9. 5million [3]. These trends show that solutions need especially treatment for the disease need to be found. This paper aims to explore how big data can be a useful tool in fighting the threats of cancer through gene profiling of molecular factors of the disease.

The word "big data" is becoming more common in our daily lives. However, every discussion of it has a particular connotation, depending on what we are using it. Big Data is It's the unprocessed data that researchers will use to come up with new innovative ideas [3]. To have valuable information, Big Data must be processed in the same way as raw crude must be refined to drive a vehicle.

Big data has now penetrated certain components of the health sector, owing to various specific needs, especially in research institutions and the diagnosis of rare diseases. Approximately 5% to 7% of Americans suffer from one of the 7,000 disorders deemed rare by medical practitioners [4]. Since a single health facility will be unable to provide the requisite number of patients to research the disorder, big data has become one of the most effective ways in terms of collecting data from various hospitals to build a profile about various cancer types. Big data is now delivering evidence that is useful to patients. This paper will therefore look at how big data can be useful in finding the cure for cancer. The main sectors covered will include the significance of big data in genetic research, how medical records are important in building a profile on the disease, genome and geniting editing, big data initiatives examples of research done to fight the disease.

## IV. RESEARCH PROBLEM

What this research looks at solving are the possible ways in which big data can be utilized in finding a cure for cancer. This topic is significant because the impacts of cancer around the world are dire. Big data is one of the proposed ways in which cancer can be treated given that a lot of data is already available to study and understand the dynamics of the disease. Research about the disease has been going on for decades but it will be a breakthrough if the data is looked at closely to create a profile in which therapeutic interventions can be found. Cancer can come in different ways for instance from environmental effects, or hereditary factors which makes it more of a threat [5].

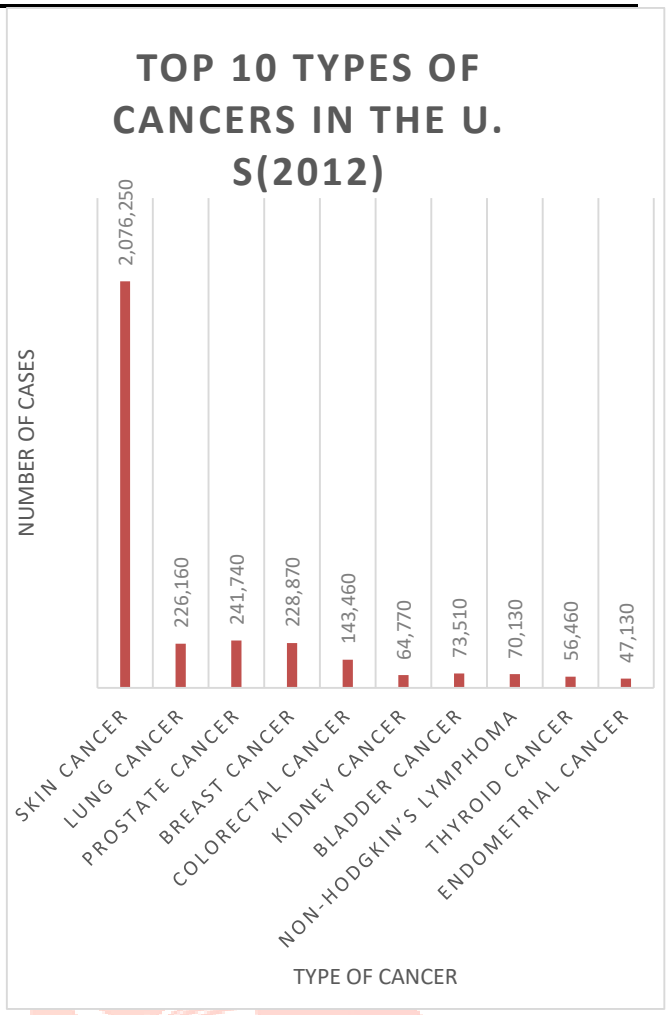


Fig i: Top 10 types of cancers in the U. S (2012)

### *A. Building the research portfolio*

To contextualize how Big Data can assist in the treatment of cancer, this research carried out a systematic review of literature. The first step was to establish what the research would try to answer which made it the main focus of the exploratory searches in the database. Some of the Keywords that were used are Big Data, cancer, molecular profiling, genome editing. The main databases that were looked at for preliminary search included PubMed, Cochrane Library and EMBASE. The preliminary searches confirmed the combination of the key words and the final searches were done. Duplicate journals were filtered based on themes outside scope, publication year and number of citations.

Predicting how cancer affects the cells of a patient and finding its cure involves the application of Big Data mining. The concept has proved to be effective in tracking patterns of a disease and therefore will be vital in understanding the patterns of a cancer sequence.

## V. LITERATURE REVIEW

### *A. Cancer Database*

About 14 million individuals in the United States now have or have had cancer at any point in their lives. The figure is projected to grow to 18 million over the next 4 years or so [6]. Their medical history, blood tests, and tumor tissue samples have been collected or will be collected in the future to determine their treatment options. Each patient has a unique cancer background including symptoms, family history, diagnoses, medications, tumor characteristics, genes, tests, and other variables. The data obtained from many patients contributes to a significant amount of information. Big data studies investigate how these vast catalogs of data may be accumulated and analyzed to find relatively undiscovered linkages that could

enhance the clinical outcomes [6]. For instance, a person who is diagnosed with a rare melanoma in a remote health facility —meets an oncologist with no experience in handling this form of cancer, looks for information by utilizing an online resource finding a procedure that benefited related patients who have experienced with a similar rare form of cancer. Another consideration is a patient with breast cancer where the tumor bears a genetic defect most often seen in lung cancer, prompting an oncologist to prescribe a treatment that supported a group of lung cancer patients who experienced a similar mutation. These kinds of interactions are exceedingly unusual [7,8]. Additionally, as the number of patients diagnosed with cancer increases, it will be common to find those with similar types of cancer or common types of tumor. Accumulated genetic mutations have been associated with many types of cancers. One project that is vital in helping with the treatment of cancers is the AACR Project GENIE which was designed to look at patients with similar molecular tumor characteristics like mutations. Clinicians utilize this project when deciding the treatment options for the patient. A patient's cancer profile will reveal how they will respond to treatment. The AACR Project GENIE registry had over 17,000 medical records when it was released in November 2015 and by 2020, the AACR was expected to include 100,000 records [9]. Seven major cancer organizations have signed on to the first step of the AACR Project GENIE, which involves utilizing the data to address a study issue, including whether a treatment for a certain genetic variation in one form of cancer is indeed beneficial for the same variant in another type of cancer. Specific cancer centers and clinics have traditionally not exchanged or been allowed to exchange data with other institutions.

### *B. A Patient-Centered Approach*

Data takes a variety of forms in large-scale programs, but information often comes from patients. They may choose to discuss their clinical history, which may provide details regarding their general health and symptoms, as well as genotypes regarding normal and tumor cells. Tissues such as blood or tissue specimens may be donated to a biorepository or tumor repository to be examined and may be shared with other research organizations. The information collected for these studies could include a patient's medication choices as well as the explanations of whether these therapies were chosen or not. In certain instances, the details might also contain patient queries and feedback [9].

To collect and utilize data, hospitals must take all reasonable steps like obtaining consent from the patients—often granted as general consent for using genomic and tissue samples in research and experiments. Informed consent arrangements can inform the patient which information would be exchanged and rendered confidential. Consent clauses can therefore warn patients that they can determine when to disclose their information as well as how to keep it secure. For doctors, a portal like CancerLinQ may include medical treatment guidelines [10]. For instance, if cancer with a certain gene mutation develops resistance to a certain medication, the oncologist can want to stop using that medication.

Big-data programs offer patients the ability to engage in studies that might enhance therapies for them and future generations. "The good part regarding CancerLinQ is the fact that it captures all the data that might otherwise be destroyed, deposited in the local registry [10]. Any initiatives leveraging big data's crowd-sourcing capacity explore other facets of a therapeutic relationship.

### *C. How big data changes genetic research*

Big data and its associated applications shape the way we conduct operations around the globe, from start-ups to Fortune 500 companies. Whatever area you work in, or the scale of the enterprise, data collection, data management, and interpreting the data is vital and has a wide-ranging impact. The range of data-producing tools today provides market perspective in the infrastructure environment we live in; whether from a portal, social networking, or online shopping, which can be utilized to optimize business operations and engagement. Big data's effect isn't limited to industrial sectors; it also improves how genetic testing is conducted. Big data utility ensures it is used in nearly every field around the globe, including hospitals, pharmaceutical engineering, and genetic research [11]. Technological developments also allowed scientists to rapidly develop, preserve and interpret data that might take years to collect.

New biomedical methods, like next-generation genome sequencing, generate vast amounts of data and contribute to medical discoveries, but scientists are unable to cope up with the current streams. For instance, the National Institutes of Health initiated the 'Precision Medicine Initiative' and Big Data to Knowledge Initiative (BD2K) to establish genetically driven care of customized, precise medicine for better preventive, early diagnosis, and treatment options for common chronic diseases [11]. It aims to do this to categorize and collect the whole genome sequence, proteins, cell population, metabolites, RNA, DNA, and related behavioral details while collecting and connecting electronic medical records and information of a population of 1 million Americans. It's plenty of numbers [11]. Realistic use of data mining in genetic science is huge, but the biggest obstacle is turning large data into valuable knowledge which can be used for education and research [12]. To be a genetic scientist involves dealing with algorithms that process Big Data in the genetic engineering and data analysis software in today's quick transition into a large data environment.

### *D. Proposed molecular profiling to treat cancer*

This paper proposes the application of genomic profiling as a potential big data technique in treating cancer. Cancer is caused by genetic mutation and therefore the treatment options should involve looking at the genetic makeup of the cancer cells. To arrive at finding the cure for cancer, some procedures must be followed to determine the genetic sequence of the tumor cells.

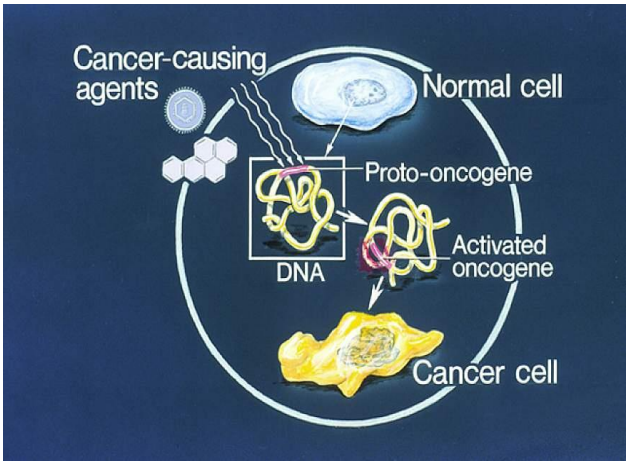


Fig ii: An illustration of how cancer-causing agents lead to a cancer cell

The tumour is removed from the body tissues through a biopsy. Cancer cells are separated and removed from the tissues and sequenced in the laboratory to extract their DNA sequence. With the aid of the QIAamp DNA mini kit, the genomic tumor DNA will be extracted from

the portions of formalin-fixed paraffin-embedded (FFPE) tumor tissues [13]. The sequenced genetic profile is examined by highly advanced tools for alterations to determine how the tumor behaves. The DNeasy Blood Kit can also be used to remove the DNA from leukocytes. The ctDNA sequencing libraries that utilize KAPA DNA Library Preparation Kits are developed [13]. Anomalies may be studied in a laboratory to decide if they fit identified mutations that might have led to treatments or where data indicates a possible therapeutic choice not normally considered. When a match exists, it might be possible for the oncologist to use the findings to prescribe therapies employed in the past to treat the same alterations. The oncologist will discuss the findings as well as any recommendations for potential therapeutic strategies and will develop an individual treatment plan. A minimum means effective depth of 300 times for tissues and 1000 times for plasma samples is needed for the highly targeted capture sequencing [14]. The average successful coverage depth of 1,295 times in tissues and 2,014 times in plasma samples, as well as 299 times for germline DNA samples, is needed in the analysis [1].

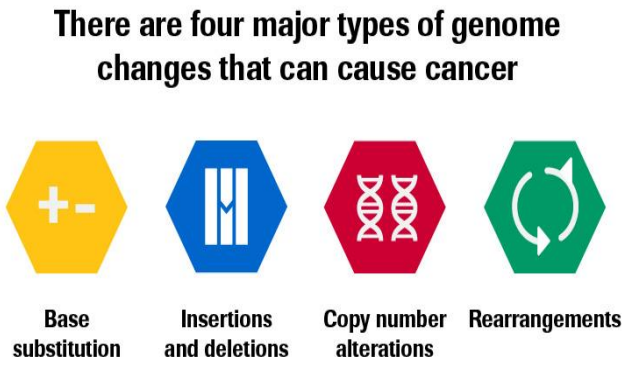


Fig iii: Genome changes that cause cancer

The cancer database is consulted by using genes identified in 175 non-redundant trials, like the TCGA and non-TCGA trials, without any samples overlapping, to assess the prevalence of target mutants in unusual tumors. The downloading and filtering of mutations in certain genes would be performed. TGAs observed concurrently include simple replacements, brief insertions and deletions, focal gene amplifications, homozygotic removals (altering copy numbers), and genome reorganization. Synonymic varieties recognized germline varieties in dbSNP, and varieties occurring in the Exome Sequencing Project at the population frequency of >1 percent are removed [14]. Germlines have been described as pathogenic, potentially pathogenic, unclear significance, possibly benign and benign, and have been identified in the ACMG guidance. For more study, only level 1-4 mutations are preserved. The overall number of targeted mutations for each mutation would be split into total sample quantities for each cancer to determine the frequency. Present physician procedure approaches cancer based on the tissue from which cancer arose. The doctor will help adapt cancer treatment to the cancer genetic makeup by researching the genetics of cancer. In May 2017 for instance, the FDA authorized the usage of an immunotherapy medication for cancer dependent purely on the genomic profile of cancer regardless of the nature of cancer [15]. 49% of solid tumors have been found to have at least one genetically mutated mutation, as a targeted treatment for cancer with these particular genetic disorders already exists [15]. In certain instances, cancer genetic screening may assist a practitioner in preventing the usage of costly and harmful chemotherapy. For instance, early breast cancer will contribute to a 46 percent decrease in the number of patients who require further postoperative chemotherapy in genetic tests in medically high-risk patients. The genetic profile of cancer helps certain patients to enter different clinical trials such as the NCI-MATCH and ASCO-TAPUR studies and several more [16]. One trial has shown that 11 percent of cancer patients are matched and participated in a clinical study. These research studies may contribute to better recovery and research revelations on medications. Many survivors and their relatives are often given some relief from the agony of their cancer by supporting prospective patients in some possible clinical trials.

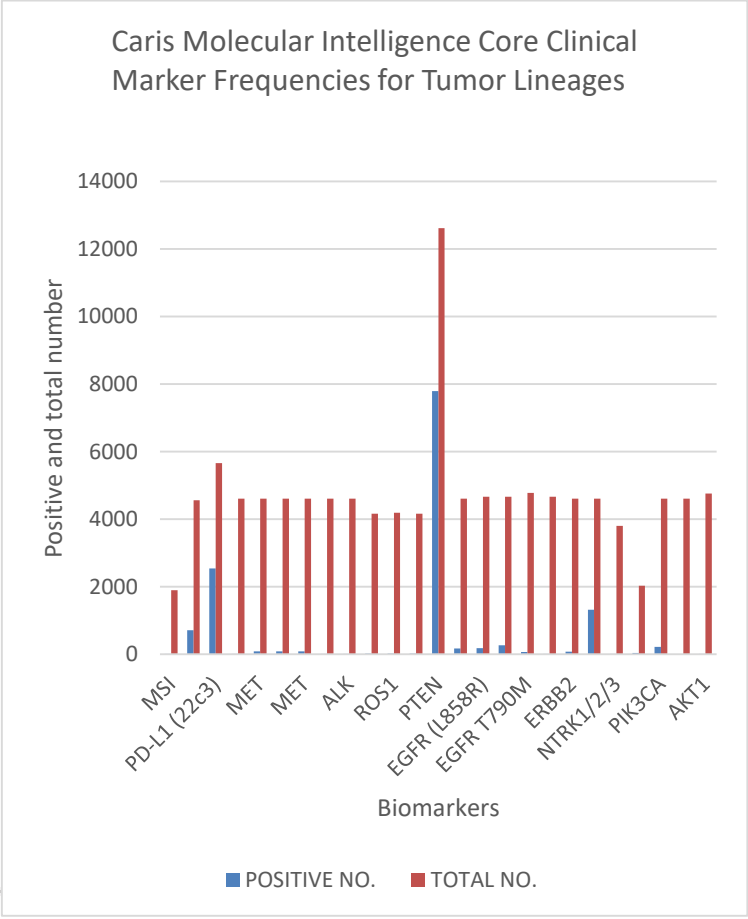


Fig iv: The graph above shows the frequencies of Non-small-cell lung carcinoma biomarkers. These biomarkers can be applied in to trace cancer lineages and therefore firmly establish the mutational status of this type of cancer.

The ultimate aim of trials is to convert genetic discoveries into theoretically successful treatment outcomes for patients today. Besides that, a slew of proof-of-principle experiments are either underway or in the works. One approach to maximizing the use of this expertise is real-time national and foreign collaborations with cancer researchers and pharmaceutical firms to conduct broad-based profiling and contextualise tailored patient treatments. Simultaneously, data pooling is necessary utilizing universal data-sharing technologies to optimize the usefulness of these observations and produce broad pools of data. Successes and mistakes will offer a more full view, and the outcome will get us closer to successful clinical trials to find cancer cures.

## VI. FUTURE IN THE UNITED STATES

This research will change how the research will be conducted in many research centers in the United States. More researchers need to focus more on big data analysis to find a match between one cancer sequence to that in the cancer database to establish how a targeted treatment of the mutated cells will be conducted. Cancer is a major health issue that experts in the U.S. and abroad strive relentlessly to overcome. Many researchers are optimistic that they would be able to draw on decades of knowledge and recent developments to reduce cancer's impact even faster. The development and efficiency of gene sequencing knowledge have significantly increased – ensuring that researchers and physicians will now find out whether an individual is particularly at risk for some cancers, and what medication will ideally fit cancer patients. The acceptance of vaccination to help deter complications from the human papillomavirus (HPV), which causes cervical cancers, was another significant step ahead [17]. Targeted treatment, immunotherapy, and cancer detection technologies have also seen significant advancements. In addition to medications including ipilimumab and nivolumab, researchers are working on a new immunotherapy technique that involves manipulating a patient's T cells to selectively target cancer cells more effectively. The technique is showing positive outcomes for the treatment of relapsed B cell acute lymphoblastic leukemia and other blood cancers, and it may also be effective in the treatment of solid tumors [18].

## VII. ECONOMIC BENEFITS

Finding the cure for cancer will be a tremendous milestone in saving lives. To begin with, if cancer could be treated, the demand for cancer drugs will increase. In 2016, the total cost of cancer treatment and medications was 113 billion dollars, with the United States responsible for about 46% of the amount [19]. The economy of the United States will benefit by selling the drugs through its pharmaceutical companies. Furthermore, a cancer cure will increase the life span of millions of Americans, which is beneficial by making sure that more people work. Every month, nearly 63 million individuals earn Social Security payments, and when more younger generations become entitled to benefits, the demand for Social Security payments increases faster than the number of people who will pay into the scheme. The discovery of cancer treatment will also overburden the Social Security system, resulting in potential claimants earning fewer funds than they donated, putting yet another pressure on the economy.

## VIII. CONCLUSION

This paper explored how big data can play a vital role in finding a cure for cancer. The conclusion that can be made from this research is the fact that cancer sequencing is the way to go in finding a cure for cancer. The availability of a massive amount of data stored in the U.S database will be vital in identifying a match between different types of cancers. The application of precision targeting will help identify the mutations in genes and therefore make it easy in coming up with targeted treatments for cancerous cells. The more we understand about metastasis, the higher the possibility of potential therapeutic strategies. In the war towards cancer, modern methods to taming the immune system are moving us closer to a world in which cancer is a curable illness. More developments are expected to revolutionize treatment for cancer: targeted vaccination, cell therapy, genome editing, and microbiome therapies. It is becoming increasingly clear that there will be no single "remedy." Instead, each patient would be handled separately, depending on their particular criteria. However, for precision treatment to become a possibility, a wide variety of medications must be accessible to treat cancer around the board.

## References


[1] World Health Organization, "Cancer: Key Facts", *Who.int*, 2018. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] R. Siegel, K. Miller and A. Jemal, "Cancer statistics, 2018", *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7-30, 2018.

[3] Cancer Atlas, "The Burden of Cancer", *The Cancer Atlas*, 2018. [Online]. Available: https://canceratlas.cancer.org/the-burden/the-burden-of-cancer/.

[4] National Organization For Rare Disorders, "About the Day - Rare Disease Day - NORD (National Organization for Rare Disorders)", *NORD (National Organization for Rare Disorders)*, 2018. [Online]. Available: https://rarediseases.org/rare-disease-day/about-the-day/.

[5] E. Susman, "NCCN Incorporates Aromatase Inhibitors into Breast Cancer Guidelines", *Oncology Times*, vol. 27, no. 9, pp. 52-53, 2005.

[6] . de Moor, A. Mariotto, C. Parry, C. Alfano, L. Padgett, E. Kent, L. Forsythe, S. Scoppa, M. Hachey and J. Rowland, "Cancer Survivors in the United States: Prevalence across the Survivorship Trajectory and Implications for Care", 2021. .

[7] Y. Xu, H. Liu, J. Chen and Q. Zhou, "Comparisons between the National Comprehensive Cancer Network (NCCN) non-small-cell lung cancer (NSCLC) Clinical Practice Guidelines (Chinese version), the NCCN original edition, and the European Society for Medical Oncology NSCLC Guidelines in 2009", *Thoracic Cancer*, vol. 1, no. 2, pp. 83-86, 2010.

[8] J. Kent, "Big Data Analytics May Lead to More Precise Cancer Treatments", *HealthITAnalytics*, 2018. [Online]. Available: https://healthitanalytics.com/news/big-data-analytics-may-lead-to-more-precise-cancer-treatments. [Accessed: 29- Mar- 2021]

[9] AACR Project GENIE Consortium, "AACR Project GENIE: Powering Precision Medicine through an International Consortium", *Cancer Discovery*, vol. 7, no. 8, pp. 818-831, 2017.

[10] G. Sledge, R. Miller and R. Hauser, "CancerLinQ and the Future of Cancer Care", *American Society of Clinical Oncology Educational Book*, no. 33, pp. 430-434, 2013.

[11] L. Etheredge, "Rapid Learning For Precision Medicine: A Big Data To Knowledge (BD2K) Initiative | Health Affairs Blog", *Healthaffairs.org*, 2013. [Online]. Available: https://www.healthaffairs.org/do/10.1377/hblog20130221.028472/full/. [Accessed: 29- Mar- 2021]

[12] M. Padek, N. Mir, R. Jacob, D. Chambers, M. Dobbins, K. Emmons, J. Kerner, S. Kumanyika, C. Pfund, E. Proctor, K. Stange and R. Brownson, "Training scholars in dissemination and implementation research for cancer prevention and control: a mentored approach", *Implementation Science*, vol. 13, no. 1, 2018.

[13] R. Margolis, L. Derr, M. Dunn, M. Huerta, J. Larkin, J. Sheehan, M. Guyer and E. Green, "The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data", *Journal of the American Medical Informatics Association*, vol. 21, no. 6, pp. 957-958, 2014.

[14] E. Bonnet, M. Moutet, C. Baulard, D. Bacq-Daian, F. Sandron, L. Mesrob, B. Fin, M. Delépine, M. Palomares, C. Jubin, H. Blanché, V. Meyer, A. Boland, R. Olaso and J. Deleuze, "Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples", *PLOS ONE*, vol. 13, no. 4, p. e0195471, 2018.

[15] K. Morrissey, T. Yuraszeck, C. Li, Y. Zhang and S. Kasichayanula, "Immunotherapy and Novel Combinations in Oncology: Current Landscape, Challenges, and Opportunities", *Clinical and Translational Science*, vol. 9, no. 2, pp. 89-104, 2016.

[16] C. Hudis, "Clinical trials, health information technology platform to test power of precision medicine", *Healio.com*, 2015. [Online]. Available: https://www.healio.com/news/hematology-oncology/20150702/clinical-trials-health-information-technology-platform-to-test-power-of-precision-medicine. [Accessed: 29- Mar- 2021]

[17] J. Kaiser, "New NCI director expects big data to revolutionize cancer research, care", *Science*, 2017.

[18] H. Pegram, E. Smith, S. Rafiq and R. Brentjens, "CAR therapy for hematological cancers: can success seen in the treatment of B-cell

acute lymphoblastic leukemia be applied to other hematological malignancies?", *Immunotherapy*, vol. 7, no. 5, pp. 545-561, 2015.

[19] M. Kirkwood, "The State of Cancer Care in America, 2016: A Report by the American Society of Clinical Oncology | JCO Oncology Practice", *Ascopubs.org*, 2021. [Online]. Available: https://ascopubs.org/doi/10.1200/JOP.2015.010462.