Original Article

Access this article online
Quick Response Code:

Website:
www.carcinogenesis.com
DOI:
10.4103/jcar.jcar_21_02_10

Cancer Genomics and IT Infrastructure for data storage and analysis

Mishal Sohail[1*], Pakiza Ikram[1]

Abstract

In the pursuit of precision oncology, the science of cancer genomics, which is distinguished by its investigation of the genetic landscape of tumors, has become indispensable. This study explores the mutually beneficial interaction between cutting-edge IT infrastructure and cancer genomics, revealing the revolutionary implications for cancer research, diagnosis, and therapy. Our knowledge of the molecular details underlying oncogenesis has been completely transformed by cancer genomics, which is based on the thorough examination of genetic material from cancer cells. The potential of precision oncology, in which therapeutic interventions are customized based on the distinct genetic composition of every patient's tumor, is contingent upon the smooth incorporation of extensive genomic datasets into the clinical decision-making procedures. Given the volume of genetic data produced by cancer research, a strong IT infrastructure that can handle issues with data management, storage, and computation is required. The research study also covers the wider ramifications of IT integration and cancer genomics in population health, epidemiology, and the moral issues related to the proper use of sensitive genetic data. It highlights the collaborative nature of this project and the need for data sharing, standardized formats, and interoperability in order to advance the subject. The research concludes by highlighting the many implications of cancer genomics and IT infrastructure, both as a scientific endeavour and as a commitment to revolutionizing cancer treatment. With science and technology coming together to drive us towards customized, targeted therapies, every genetic sequence represents a ray of hope in the never-ending fight against cancer.

Keywords:

Cancer genomics (CG), IT Infrastructure (ITI), Data Storage (DS) Analysis, Technology (T)

Introduction

The word "cancer Genomics" can be explained in these words "It is the study of structure of DNA, its sequence and its expression of genes to point out the difference between cancerous cells and normal cells." The word IT infrastructure can be explained in these words as "a complex and interconnected system of hardware, components, software and facilities which help in better delivery of different outcomes related to business or Healthcare aspects". This study covers the knowledge related to cancer Genomics and the use of IT infrastructure that can be used for data storage and Analysis. The word cancer means uncontrolled cell division. Usually, there are different types of internal and external interaction present in the human body which affects the normal functioning of genetic material by causing changes in it[1]. Cancer is still a difficult and ever-changing threat to mankind, having tormented us for generations. But as more sophisticated tools become available, the field of cancer research and therapy is changing dramatically. Cancer genomics, a science that studies the genetic nuances of tumours to solve the riddles of their genesis, development, and response to therapy, is at the vanguard of this transformation. Cancer genomics is essentially the study of a cancer cell's genome, which is its whole collection of genes. Genomic research explores the molecular causes of cancer, in contrast to previous cancer research, which frequently concentrated on the location and appearance of tumours. Scientists are able to pinpoint the precise genetic alterations that propel the growth of cancer by deciphering the genetic code of tumours. The Human Genome

[1] Department of Computer Science, Superior University Lahore, Lahore, Pakistan

Address for correspondence:
Mishal Sohail, Department of Computer Science, Superior University Lahore, Lahore, Pakistan,

Submitted: 26-Sep-2022
Revised: 18-Oct-2022
Accepted: 03-Nov-2022
Published: 21-Dec-2022

How to cite this article: Sohail M, Ikram P. Cancer Genomics and IT Infrastructure for data storage and analysis. J Carcinog 2022; 21(2):69-76

a massive endeavour to map the whole human genome, marked the beginning of the field of cancer genomics. With the development of technology, attention turned to comprehending the genetic changes underlying cancer. The understanding that cancer is a group of different illnesses, each with unique genetic abnormalities, rather than a single, homogenous illness, created new opportunities for personalized treatment. The change in genetic material is called mutation. When these different types of carcinogens, such as occupational carcinogens and environmental carcinogens, enter in body by different means, they directly affect the functioning and structure of DNA, thus causing mutation in them. These mutations result in the production of an abnormal number of cells in the body because of uncontrolled cell division in the body.

These abnormal cells accumulate to form tumors in the body. There are two types of these tumors depending upon the location of these tumors. If these tumors do not leave their position and thus have no movement to other body parts, these tumors are termed benign tumors, but if these tumors move to other body parts and the formation of tumors also starts in other body parts, then these tumors are called malignant tumors, and these tumors are the main reason for causing cancer in the body[2]. With the advancement in science and technology, the diagnosis and treatment of cancer have become effective by use of this technology. The main problem which is associated with cancer is that the diagnosis of cancer takes place, which decreases the chances for effective treatment. But with the help of different technologies, now it is possible to get an idea of cancer for early diagnosis[3]. Artificial Intelligence-based technologies are used which detect and diagnose cancerous cells with less chance of error. One of the most important technologies is deep learning, which is abbreviated as DL, which is associated with the diagnosis and treatment of cancer. The other technology is based on medical images of tumors, which help to diagnose the stage and type of cancer that will help to formulate long-term treatment of cancer. Cancer in the body takes place in three steps: initiation, promotion, and progression. The process of initiation is the start of cancer in the body in the form of mutation in genetic material[4].

The second step, promotion is the production of abnormal cells in the body, and the third step, progression is the spread of cancer in the whole body. This study covers the use of cancer Genomics for studying the genetic material of cancerous cells. As we know, the problem of cancer is associated with basic genetic level. Thus, study of the sequence of DNA is very necessary to understand the causes, symptoms, time duration, and treatment of different diseases such as cancer. There are some startling differences between cancerous cells and normal cells[5]. Precision oncology, a paradigm that aims to tailor cancer treatment based on the distinct genetic profile of each patient's tumors, is one of the pillars of cancer genomics. Precision oncology seeks to alter the one-size-fits-all approach of traditional cancer medicines by identifying the precise genetic abnormalities causing a patient's cancer and treating them with customized therapies. There is much potential in this move away from a general therapeutic method and towards a more individualized one. It raises the standard of patient care overall by reducing probable adverse effects and increasing therapeutic effectiveness. However, the smooth incorporation of enormous volumes of genetic data into the therapeutic decision-making process is essential to the success of precision oncology. On the one hand, the sheer amount of genetic data necessitates the use of sophisticated computing tools and storage systems that can handle the volume and complexity of analysis. However, this abundance of data also creates opportunities for the discovery of new biomarkers and treatment targets, as well as for the improvement of our comprehension of the complex interactions between genetics and cancer.

The cancerous cells have a high growth rate, abnormal shape, and defective DNA as compared to normal cells in the body. The technology of cancer Genomics is quite related to studying these sequences of DNA that help differentiate between cancerous cells and normal cells. This technology is also termed precision medicine because it uses the study of genetic changes in patients. This analysis is based on a minor molecular level that leads to a better understanding of the stage and type of cancer in the human body[6]. There are many benefits of cancer Genomics, such as it helps to prescribe medicines to cancer patients according to their genetics. It also leads to the study of the composition of that carcinogen along with the mechanism of action of that carcinogen in normal tissues to form tumors or cancerous tissues in the body. The other part of the study is IT infrastructure, which also helps in data storage and Analysis for better cancer treatment. There are three primary areas for cancer treatment at medical healthcare centers[7]. These three areas are research, clinical care, and education, which all help to meet the goal of effective cancer treatment.

In this regard, IT infrastructure can be better technology for research area because the research area is the basic pillar for cancer treatment. Research areas help to understand the type of carcinogens before time for better diagnosis that leads to effective treatment of cancer. Nowadays, physicians are moving towards medical healthcare records in contrast to physical paperwork because electronic medical records have much more importance as compared to paperwork[8]. Robust and sophisticated IT infrastructure that makes data storage, processing, and interpretation easy is the foundation of cancer genomics research. Several important factors need to be carefully considered in order to develop an efficient IT system. A scalable and secure system is

needed to store and manage the enormous amounts of genetic data generated in cancer research. To suit the unique requirements of a research institute or healthcare facility, on-premises servers, cloud storage, or a hybrid method can be used. AWS, Microsoft Azure, and Google Cloud are a few examples of cloud systems that have the ability to increase storage capacity in response to demand. They offer redundancy, guaranteeing accessibility even in the event of hardware failures and data integrity. But in order to prevent unwanted access and data breaches, the sensitivity of genetic data necessitates strict security measures like encryption and access controls. Because genomic research involves a lot of computing, it takes a lot of processing power to identify the intricate patterns hidden within large datasets. For the execution of complex algorithms and simulations, cloud-based computing resources or HPC clusters are essential. These electronic medical health records can be easily accessed and analyzed. These medical health records can be stored on a computer for a time.

These electronic medical health records can be of multiple use at the same time. But there are also some problems related to cancer Genomics, such as it cannot test and study all types of cancer that make it of limited use. The second challenge related to cancer Genomics is that it requires skilled clinical experts to analyze DNA sequences. The third challenge related to cancer Genomics is that it is based on computer systems that are not easily accessed by each cancer patient. Moreover, this technology for cancer diagnosis and treatment is costly now, which makes it of limited use[9]. If these challenges are coped with effectively, cancer Genomics and IT infrastructure can be proved effective in alleviating the increasing level of cancer cases across the world. It will also lead to better future aspects for the treatment of other diseases along with cancer. These two technologies are true evidence of increasing reliance on science and technology for the treatment of cancer and other diseases[10].

Research objective

The main objective of this study is to understand the use of cancer Genomics and IT infrastructure for Data storage and Analysis for the treatment of cancer along other diseases. This study has also overviewed the different aspects related to cancer Genomics and IT infrastructure for Data Storage and Analysis. This study ensures that cancer Genomics and IT infrastructure can be effectively used for data storage and Analysis for treatment of cancer.

The research study describes that Cancer Genomics and IT Infrastructure for data storage and analysis. The research is divided into five chapters: the first portion describe the introduction part related to the IT infrastructure, etc. this portion represents the objective of the research, the second section present a literature review the third portion describes the research methodology. The fourth portion represents the results and their descriptions. This portion present applications related to the IT infrastructure. the last section summarized overall research study and present some recommendations.

Literature Review

Cancer Genomics and IT Infrastructure for Data Storage and Analysis

Cancer Genomics is the study of the DNA sequence and expression of genes between the host cells and tumor cells. Oncogenomic is the study of that part of genomics that deals with the cancer-associated with the genes. The profiling of the molecular genomics has been used to study the cancer associated genes. It helps in learning that how normal cells have been changed into the cancer cells[11]. The cancer genomes have provided the extraordinary results that studies the complexity of genomes that trigger the growth of cancer cells. This genomic revolution has enabled us to know the development and progression of the cancer cells in humans. It opened new arena for the diagnosis and treatment of the cancer and changed the ways as people and clinicians look at the cancer[12]. The analytical pipeline is accelerated by the use of parallel processing, a method that splits huge datasets into smaller jobs that are handled simultaneously. This not only speeds up research but also makes it possible to investigate a variety of genomic data-related topics, such as finding mutations and understanding the complex regulatory systems that control cancer cells. Processing and storage of genetic data are not the only aspects of its effective management. Ensuring the quality, traceability, and repeatability of research findings requires an extensive data management system. Researchers can collaborate and share data more easily by tracking changes over time due to version control methods. Validating research findings and upholding openness need data provenance, which records the source and modification of data. It is impossible to negotiate adherence to privacy laws, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US.

Strong data governance procedures guarantee the moral and legal treatment of patient data, giving patients and researchers peace of mind. Cybercriminals find the wealth of genetic data to be an alluring target. IT infrastructure is the foundation of cancer genomics research; hence it has to be protected with strong cybersecurity protocols. Data encryption offers an extra degree of security for both data in transit and data at rest. The presence of genome related research has enabled the targeted changes in humans and tests that recognize the changes in the patients. This approach is referred to as precision medicine, which is specified for individual patients at one time[13]. Among different kinds of the

cancer tumors in patients, the genomes have found the unexpected similarities across the tumor cases. For this purpose, the treatment is provided to the located areas of the tumors. But the growing study of the cancer has enabled the production of the drugs that is given to the patients without specified area in which the cancer is grown[14]. The genes effect the signaling of the cells, RNA Splicing, Metabolism, and regulation of the epigenomic. Historically, it was known that cancer was caused by chromosome disarrangement in humans, which causes the cells to grow rapidly and cause tumors in the human body. But in the present day, it is considered as the genome disease that shows the existence of the cancer-causing disease in humans[15].Portal such as CBio cancer Genome Portal provides the access to the people to gather information about the large data set of cancer genomes. It provides the 5000 samples of tumors. This IT infrastructure allows the world to gather information in no time from software generated programs[16]. The data set of the Cancer genomes can be gathered from the small-scale internet tools. The visualization of the interactive image's tools is based on the web-based portals which provides a new vision of genomes[17].

The presence and mutation of the proteins revealed through the genome sequencing which change the nature of the cancer tissues[18]. Stress is the basic factor that exerts more pressure on human behavior. With time, the organisms have developed features to cope with the changing environment. The results are the remodeling of the genome and physiological or regulation of genes[19]. There is a more pathological structure that has single stranded DNA Gemi virus, which has effects on the structure of genes in organisms. The DNA replicate itself with the qualities and features it has in it. Sometimes the malfunctioning of the historical genes has impacts on the future genes[20]. For resolving the matter of data privacy and security, the cloud genome services are adopted. Related to the protection of the sensitive data sets in genome categories, the cloud computing is adopted for maximum gaining the fruits[21]. Unauthorized access can be avoided in part by implementing access controls that specify who can access datasets and carry out tasks. It is essential to conduct regular security audits and upgrades to find vulnerabilities and strengthen the system against new and emerging cyber threats. Establishing a strong defense against such attacks requires close cooperation between cybersecurity and IT specialists.

Collaboration between several academic institutions, medical providers, and pharmaceutical businesses is common in the field of cancer genomics. Ensuring interoperability across disparate IT systems is crucial for smooth data transfer and teamwork. Genome data standard formats, including the FASTQ and VCF formats, make interoperability easier by giving disparate systems a similar language. APIs, or application programming interfaces, improve connection even further by facilitating efficient communication across various platforms and software. Interoperability pushes us closer to defeating the complexity of cancer by accelerating the speed of discovery through the dismantling of silos and encouraging data exchange.International Cancer Genome Consortium (ICGC) is the collective efforts to feature the genome activities in 50 different types of abnormalities of cancer cells. With the help of the Bio Mart software, the portals are available to its members for developing their own data bases. It also provides open access to the people for microRNAs, gene expressions, and somatic mutation[3]. In 2016, the NCI Genomic Data Commons (GDC) was launched which provide access to the number of cancer data set genome to help the researchers. Till today, the Genome Data Commons (GDC) contains almost data set of 14500 patients. The data submitted can be downloaded at any time by the researchers[2]. With the advancement in the technology, the researcher of single cell of cancer has been transformed into the data set of the genomes.

The development and growth of technology related to the single cell sequencing helps the researchers to study the more complex data sets. Since 1665, biologist such as the Leeuwenhoek has studied cancer cell under the microscope. Earliest the observations are made on the basis of the differences between the cancer cells[22]. Cancer is basically related as the genome disease. The cancer researchers conducted to find new avenues of the treatments. It arises with the accumulation of mutation in the gene's series with time. Till time no single research has been found for describing the genes mutation extensively[23]. The field of cancer genomics demonstrates the beneficial relationship between scientific advancement and technological innovation. With more knowledge of the genetic foundation of cancer, there will be a greater need for cutting edge IT infrastructure. From the creation of genetic data to the production of actionable insights, a well-balanced combination of storage options, computational capabilities, data management techniques, and cybersecurity measures is needed. Though there are many opportunities, the difficulties are as daunting. A strong IT infrastructure can enable cancer genomics, which might transform cancer therapy and usher in a new era of personalized medicine. It's more than simply science; it's a dedication to bettering the lives of patients, deciphering the secrets of cancer, and eventually defeating a powerful antagonist. The combination of cancer genomics and IT infrastructure predicts a hopeful future where every genetic sequence puts us one step closer to beating cancer, even as we traverse this complex environment.The personalized medicine of the days world for the cancer patients are becoming widely used due to the mutation in the somatic and genetic aberration. The availability of the personalized medicine

is based on the genome data set in the form of genetic aberrations. WGS is the basic technology that basic for the sequencing of genomes the cancer. For the sequencing of the genome, steps are followed as preparation of the sample, physical sequencing, and reconstruction. By the preparation of the sample by breaking it down into fragments[24].

The cancer Genome Atlas (TCGA) is the cancer genomic program that features 20000 cancer samples of various types. It is based on the DNA sequencing and expression of genes between tumor cells and normal host cells[25]. For the analysis of the single cells in the mutation of the cancer cells helps in analyzing the single cell variations of the cancer cells. Genetic changes and the environmental mutation results in the heterogeneity of cancer cells within the tumor cells. It opens new avenues of studying and classifying the single cells technologies[26].

The recent surge in the genomes of the data in expanding the classification of cancer in patients. The cancer cells are described in the form of the genome data set. The certain cancers and molecular subset of the Cancerian cell. The eligibility of the molecular data helps in generating the sequencing the cancer cells[27].

Table (1) Correlations

		Cancer Genomics 1	Cancer Genomics 2	Cancer Genomics 3	IT Infrastructure 1	IT Infrastructure 2	data storage
Cancer Genomics 1	Pearson Correlation	1	.167	.227	-.190	-.199	.245
	Sig. (2-tailed)		.246	.112	.186	.165	.087
	N	50	50	50	50	50	50
Cancer Genomics 2	Pearson Correlation	.167	1	.488**	.300*	-.225	.240
	Sig. (2-tailed)	.246		.000	.034	.117	.093
	N	50	50	50	50	50	50
Cancer Genomics 3	Pearson Correlation	.227	.488**	1	-.089	.116	.411**
	Sig. (2-tailed)	.112	.000		.538	.421	.003
	N	50	50	50	50	50	50
IT Infrastructure 1	Pearson Correlation	-.190	.300*	-.089	1	-.009	-.379**
	Sig. (2-tailed)	.186	.034	.538		.953	.007
	N	50	50	50	50	50	50
IT Infrastructure 2	Pearson Correlation	-.199	-.225	.116	-.009	1	.177
	Sig. (2-tailed)	.165	.117	.421	.953		.220
	N	50	50	50	50	50	50
data storage	Pearson Correlation	.245	.240	.411**	-.379**	.177	1
	Sig. (2-tailed)	.087	.093	.003	.007	.220	
	N	50	50	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The above result present correlation coefficient analysis results describing the Pearson correlation values, the significant values, and the number of observation rates of each indicator. The data storage and IT infrastructure present negative and some positive correlation between them its rates are -0.199, -0.225, 0.116, -0.009 respectively. The cancer genomics shows that negative effects of health. The foundation of this infrastructure is made up of cloud platforms, high-performance computer clusters, and strict cybersecurity protocols that guarantee the safe, scalable, and effective analysis of genetic data. Key applications of the marriage of cancer genomics and IT infrastructure are examined in this research, such as personalized treatment plans, biomarker discovery-based early detection and diagnosis, innovative contributions to clinical trials and therapeutic innovation, and the role of oncology informatics in optimizing healthcare workflows

Table (2) Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.951	32.510	32.510	1.951	32.510	32.510
2	1.444	24.074	56.583	1.444	24.074	56.583
3	1.175	19.583	76.167	1.175	19.583	76.167
4	.667	11.117	87.284			
5	.496	8.275	95.559			
6	.266	4.441	100.000			

Extraction Method: Principal Component Analysis.

The above result demonstrates that the total variance analysis result describes the initial eigenvalue and extraction sums of squared values of each component. The total values are 1.951, 1.444, 1.175, 0.667, 0.496 also that 0.266 shows that 26%, 49%, 66% variance rates between them. The cumulative percentage values are 32.510, 56.58, 95.55 all of them are present positive cumulative values of each indicator. The result also presents that extraction sum of squared rates are 32%, 24%, 19% the cumulative rates are 56%, 76% and 32%.

Table (3) Component Matrixa

	Component 1	2	3
Cancer Genomics 1	.552	-.009	-.557
Cancer Genomics 2	.606	.682	.141
Cancer Genomics 3	.793	.116	.309
IT Infrastructure 1	-.293	.773	.364
IT Infrastructure 2	-.020	-.475	.774
data storage	.751	-.377	.132

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

The component matrix reveals significant positive rates for cancer genomics components 1, 2, and 3, with values of 0.552, 0.606, and 0.793, indicating positive rates of 55%, 60%, and 79% respectively. Conversely, the IT infrastructure exhibits negative rates for each component, with values of -0.293, -0.020, -0.475, and -0.377. These negative rates underscore potential challenges or drawbacks associated with the IT infrastructure in relation to the analyzed components.

Table (4) Test Statistics

	Cancer Genomics 1	Cancer Genomics 2	Cancer Genomics 3	IT Infrastructure 1	IT Infrastructure 2	data storage
Chi-Square	16.080[a]	8.920[b]	8.680[b]	21.280[b]	2.880[c]	16.360[b]
df	3	2	2	2	1	2
Asymp. Sig.	.001	.012	.013	.000	.090	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 12.5.
b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 16.7.
c. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.0.

The above result describes that test statistical analysis result describes that chi square analysis the significant analysis presents that 100% significant level between them. The chi square values of cancer genomics 1 are 16.080, 8.920, and 8.680. All values show a positive chi square. The IT infrastructure 1,2 presents that 21.280 and 2.880. the data storage presents that 16.360 respectively the overall probability rate is 0.000 respectively.

Applications

The broad objective of enhancing cancer diagnosis, therapy, and outcomes is supported by the widespread use of cancer genomics and cutting-edge IT infrastructure. Now let's explore a few important uses:

Tailored Intervention Techniques

Genomic profiling: Physicians can detect certain mutations causing cancer by examining the genetic composition of tumors. By using this information to inform the selection of targeted medicines, treatment efficacy is maximized.

Drug Development: New medications that precisely target the genetic changes found in some tumors are developed due to genomic discoveries, ushering in a new age of precision medicine.

Early Identification and Assessment:

Biomarker Discovery: The identification of biomarkers, or indications of the presence or development of disease, is aided by genomic data. Because of these indicators, treatments may be made when the illness is most curable, enabling early discovery. Genetic alterations that cause cancer are found by means of cancer genomics. This information supports fundamental research and advances our understanding of the biochemical processes behind various cancer forms.

Investigating Tumor Heterogeneity: Genetic differences between distinct cells within tumors are revealed by genomic data. Comprehending this variability is essential to developing all-encompassing treatment plans.

Patient Stratification: By classifying patients according to the genetic makeup of their tumors, genomic data enables more focused and efficient patient enrolment in clinical trials.

Adaptive Clinical Trials: Real-time modifications based on new genetic discoveries are made possible by the IT infrastructure that enables adaptive clinical trial designs. This quickens the process of developing new drugs.

Health Information Systems: By streamlining the administration of patient data, integrated IT systems guarantee that medical professionals have access to thorough clinical and genetic data for well-informed decision-making.

Data Visualization Tools: By making complicated genomic data easier to grasp, user-friendly interfaces and data visualization tools help researchers and doctors make sense of the data.

Both epidemiology and population health

Genomic Epidemiology: By incorporating genomic data into population health studies, researchers can learn more about the genetic variables impacting the incidence and course of cancer in certain populations. Targeted public health initiatives, such as screening programs and customized preventative efforts based on individual genetic risk factors, are informed by genomic information.

Instruction and Practice

Bioinformatics Training: A competent workforce is needed, given the convergence of cancer genomics and IT. Initiatives for training and education make sure that scientists and medical personnel are prepared to use genetic data. Essentially, the use of IT infrastructure and cancer genomics is not limited to one area but rather encompasses a variety of clinical, scientific, and public

health endeavors. This multifaceted strategy demonstrates a dedication to fully use genetic knowledge possible for the good of people, communities, and the worldwide cancer research community.

Conclusion

In summary, the combination of state-of-the-art IT infrastructure with cancer genetics represents a turning point in our efforts to comprehend, treat, and eventually eradicate cancer. The path from deciphering the complexities of the human genome to applying that understanding to customized cancer treatment is proof of the unwavering quest of scientific advancement and innovation. Precision oncology, which is driven by the knowledge obtained from genetic data, has the potential to completely change the way we treat cancer. Customizing treatments to the distinct genetic composition of every patient's tumor is a paradigm change that offers better overall patient experience by reducing side effects while simultaneously increasing efficacy. But with this revolutionary potential also come great problems, especially around IT infrastructure.

Large-scale genomic data generation calls for interoperable, secure, and scalable technologies. The core components of this infrastructure are strong data storage, high-performance computing, careful data management, and steadfast cybersecurity measures. Working together is essential as we negotiate this complex terrain. Accelerated discovery is facilitated by the sharing of genetic data across academic institutions, healthcare providers, and industrial partners. Standardized formats, easy data integration, and interoperability provide a collaborative environment across institutional borders that advances our understanding of the complexity of cancer. It is imperative that sensitive genetic data be handled securely and ethically considering these difficulties. Achieving a balance between security and accessibility guarantees that the hunt for scientific knowledge stays in line with moral principles, fostering confidence in both researchers and the people whose data is being contributed to these enormous projects.

This mutually beneficial journey between science and technology comes to an end, but it also marks a first step towards a day when cancer will no longer be an unbeatable enemy but rather a problem that can be addressed with specialised, focused, and potent therapies. A future where every genetic sequence is a ray of hope, leading us into a new chapter in the fight against cancer, is shown by the merging of cancer genomics with IT infrastructure. The unrelenting quest to comprehend the genetic complexities of cancer is not only a scientific imperative, but also a commitment to revolutionising the field of oncology and, consequently, the lives of those impacted by this unforgiving disease, as we stand at the nexus of discovery and implementation.

References

1. R. L. Grossman, "Data lakes, clouds, and commons: A review of platforms for analyzing and sharing genomic data," *Trends in Genetics,* vol. 35, no. 3, pp. 223-234, 2019.
2. S. Wilson *et al.*, "Developing cancer informatics applications and tools using the NCI genomic data commons API," *Cancer research,* vol. 77, no. 21, pp. e15-e18, 2017.
3. J. Zhang *et al.*, "International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data," *Database,* vol. 2011, p. bar026, 2011.
4. J. W. Lau *et al.*, "The Cancer Genomics Cloud: collaborative, reproducible, and democratized—a new paradigm in large-scale computational research," *Cancer research,* vol. 77, no. 21, pp. e3-e6, 2017.
5. B. Langmead and A. Nellore, "Cloud computing for genomic data analysis and collaboration," *Nature Reviews Genetics,* vol. 19, no. 4, pp. 208-219, 2018.
6. L. H. Saal *et al.*, "The Sweden Cancerome Analysis Network-Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine," *Genome medicine,* vol. 7, no. 1, pp. 1-12, 2015.
7. S. M. Reynolds *et al.*, "The ISB Cancer Genomics Cloud: a flexible cloud-based platform for cancer genomics research," *Cancer research,* vol. 77, no. 21, pp. e7-e10, 2017.
8. A. M. Noor, L. Holmberg, C. Gillett, and A. Grigoriadis, "Big Data: the challenge for small research groups in the era of cancer genomics," *British journal of cancer,* vol. 113, no. 10, pp. 1405-1412, 2015.
9. C. Wilks *et al.*, "The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data," *Database,* vol. 2014, p. bau093, 2014.
10. I. V. Hinkson, T. M. Davidsen, J. D. Klemm, I. Chandramouliswaran, A. R. Kerlavage, and W. A. Kibbe, "A comprehensive infrastructure for big data in cancer research: accelerating cancer research and precision medicine," *Frontiers in cell and developmental biology,* vol. 5, p. 83, 2017.
11. B. L. Weber, "Cancer genomics," *Cancer cell,* vol. 1, no. 1, pp. 37-47, 2002.
12. N. E. Navin, "Cancer genomics: one cell at a time," *Genome biology,* vol. 15, no. 8, pp. 1-13, 2014.
13. B. Tran *et al.*, "Cancer genomics: technology, discovery, and translation," *J Clin Oncol,* vol. 30, no. 6, pp. 647-660, 2012.
14. L. Chin, J. N. Andersen, and P. A. Futreal, "Cancer genomics: from discovery science to personalized medicine," *Nature medicine,* vol. 17, no. 3, pp. 297-303, 2011.
15. L. A. Garraway and E. S. Lander, "Lessons from the cancer genome," *Cell,* vol. 153, no. 1, pp. 17-37, 2013.
16. E. Cerami *et al.*, "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data," *Cancer discovery,* vol. 2, no. 5, pp. 401-404, 2012.
17. M. J. Goldman *et al.*, "Visualizing and interpreting cancer genomics data via the Xena platform," *Nature biotechnology,* vol. 38, no. 6, pp. 675-678, 2020.
18. M. R. Stratton, P. J. Campbell, and P. A. Futreal, "The cancer genome," *Nature,* vol. 458, no. 7239, pp. 719-724, 2009.
19. A. Madlung and L. Comai, "The effect of stress on genome regulation and structure," *Annals of Botany,* vol. 94, no. 4, pp. 481-495, 2004.
20. S. G. Lazarowitz and R. Shepherd, "Geminiviruses: genome structure and gene function," *Critical Reviews in Plant Sciences,* vol. 11, no. 4, pp. 327-349, 1992.

21. S. Thiebes, G. Kleiber, and A. Sunyaev, "Cancer genomics research in the cloud: a taxonomy of genome data sets," in *Proceedings of the 4th International Workshop on Genome Privacy and Security*, 2017.
22. L. R. Yates and P. J. Campbell, "Evolution of the cancer genome," *Nature Reviews Genetics,* vol. 13, no. 11, pp. 795-806, 2012.
23. M. P. Schroeder, A. Gonzalez-Perez, and N. Lopez-Bigas, "Visualizing multidimensional cancer genomics data," *Genome medicine,* vol. 5, no. 1, pp. 1-13, 2013.
24. S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer genomics & proteomics,* vol. 15, no. 1, pp. 41-51, 2018.
25. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz Jr, and K. W. Kinzler, "Cancer genome landscapes," *science,* vol. 339, no. 6127, pp. 1546-1558, 2013.
26. A. Saadatpour, S. Lai, G. Guo, and G.-C. Yuan, "Single-cell analysis in cancer genomics," *Trends in Genetics,* vol. 31, no. 10, pp. 576-586, 2015.
27. R. Simon and S. Roychowdhury, "Implementing personalized cancer genomics in clinical trials," *Nature reviews Drug discovery,* vol. 12, no. 5, pp. 358-369, 2013.