International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 1, January 2026)

Genomics and Big Data: Opportunities and Challenges in the Era of Digital Transformation

Monisha T[1], Nuthana Jain V[2], Shruti Awasthi3, Preethi Rajesh[4], Chethana V Chalapathy[5]

[1,2]*M.Sc, Student Biotechnology, Department of Life Sciences, Garden City University, Bengaluru*
[3,5]*Professor, Department of Life Sciences, Garden City University, Bengaluru*
[4]*Professor and Head, Department of Life Sciences, Garden City University, Bengaluru*

***Abstract*-- The combination of genomics and big data analysis has revolutionized modern medical research, promoting digitalized data in life sciences. Technological advances in high-throughput sequencing, artificial intelligence (AI), and machine learning (ML) have enhanced the research in genomics, big data, and interpretation. These enhancements have increased the importance of big data in genomics for precision medicine to predict and diagnose disease. This kind of information has an application in various fields such as disease detection, targeted therapies, and pharmacogenomics. However, this acceleration of the genomic data has its own challenges in terms of storage, interoperability, standardization, and ethical governance. Cloud computing and scalable digital infrastructure are arising as solutions to store and interpret large datasets effectively. To provide safety to the patient's data and build trust, the legal, privacy, and ethical issues remain crucial.**

***Keywords*-- Genomics, Big Data, Precision Medicine, Pharmacogenomics, Cloud Computing, Data Storage, Interoperability, Ethical Governance, Data Privacy.**

I. INTRODUCTION

Genomics is the study of a complete set of DNA in the organism. Recent advancement in technologies has reshaped genomics, enabling for data driven research.

With the help of computational tools and data-driven research methods, genomics is leading in the field of modern biological sciences. Before the field of technology advancement genomic study was time consuming and had to be done manually, limiting both speed and scale. With advancement in technology and digital transformation the scope of genomics has expanded i.e. characterized by automation, high performance analytics , A journey from traditional laboratories to data centric ecosystem (Mardis, 2017).

The application of genomics has advanced from descriptive to predictive and personalized intervention. This change has led researchers not only to understand genetic variations but also to implement in health sectors i.e targeted gene therapies predict the potential disease and monitor disease development. This transformation in the field has supported the growth of genomics data which is now measured in petabytes, creating demands for both computational infrastructure and cross disciplinary cooperation (Stephens et al., 2015).

Figure 1: Approaches to genomics research

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 1, January 2026)

The figure contrasts two approaches to genomics research. Traditional methods are manual, time-consuming, and limited in scalability, whereas data-driven approaches rely on automation and high-performance computation to enable predictive, large-scale, and personalized genomic analyses.

II. Role Of Big Data In Modern Genomic Research

Genomics serves as the backbone for modern genomic discovery. Each clinical trial, sequencing project and population level study generates huge datasets that contain valuable biological insights. All these datasets contain transcriptomicsprofiles,proteomic data, genomic sequence and phenotypic records. Big data analytics helps researchers and scientists to understand the complex relationship between traits and genes whose small scale analysis was impossible (Hasin et al., 2017).

Big data is beyond storage- it is helpful in integration and interpretation. Using techniques like datamining , pattern recognition, and statistical modeling this helps researchers and scientists to discover biomarkers and investigate the genes interacting. Moreover, big genomic datasets like the cancer genome atlas (TCGA) and UK biobank have become global resources for open science and collaborative innovation.

Big data brings challenges in terms of reproducibility, ethical governance, heterogeneity and it emphasizes on the need for standardizing the workflow and transparent data sharing practices (Schwarz et al., 2020).

III. Data Generation And High-Throughput Sequencing Technologies

Development in the field of next generation sequencing (NGS) and Third generation sequencing (TGS) has shown growth in genomic data production. Illumina ,PacBio, and Oxford Nanopore technology allows sequencing of billions of DNA fragments. Simultaneously reducing cost per genome (Goodwin et al., 2016). High -throughput systems like metagenomics , real time pathogen surveillance and population scale studies have brought opportunities that were unimaginable a decade ago.

The drop in the sequencing cost has made genomic data accessible for democratic access. This rapid expansion of data introduces a new set of subsequent challenges, including the essential tasks of data preprocessing, correcting inherent errors, annotating complex metadata, and ensuring the long-term archival of these massive datasets. Management of raw sequencing reads, quality control files and genomes assembled do not need only computational capacity but also robust bioinformatics pipelines and metadata standards (Logsdon et al., 2020)

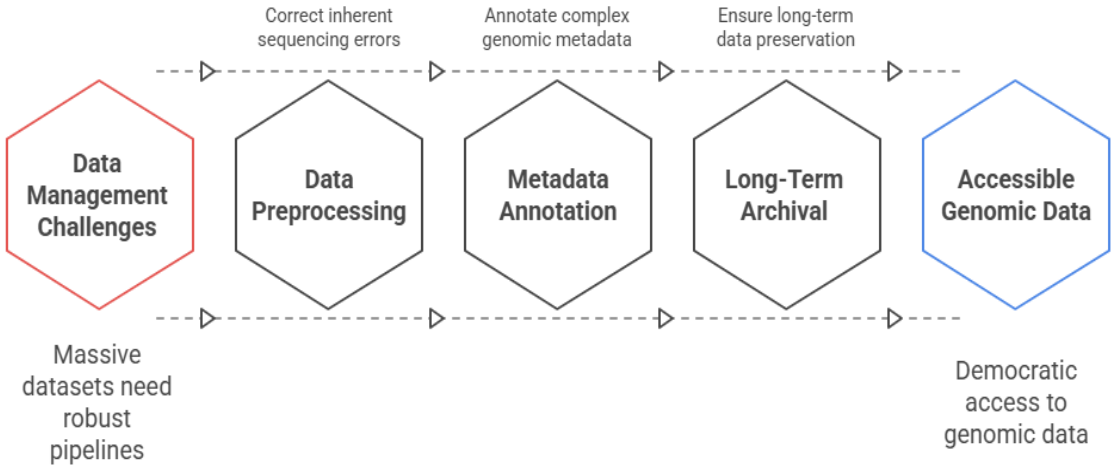

Figure 2: Democratize genomic data through next-generation and third-generation sequencing (NGS/TGS)

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 1, January 2026)

The image outlines the workflow required to democratize genomic data through next-generation and third-generation sequencing (NGS/TGS). It begins with the major challenge of managing massive sequencing datasets, followed by essential steps such as data preprocessing, metadata annotation, and long-term archival. Together, these processes ensure error correction, accurate annotation, and secure preservation, ultimately enabling broad and equitable access to high-quality genomic data.

IV. Data Storage, Management, And Integration Challenges

The exponential growth in genomic information has created a critical bottleneck, positioning data storage and management as one of the most formidable challenges facing the field.

A raw sequence from a single human genome (200) gigabytes can copy more data while large scale collaborative projects now can manage data in petabytes. The traditional storage system is insufficient for handling large data, distributed databases, and data compression algorithms.

Integration adds another layer of complexity. Genomic research increasingly requires combining heterogeneous data sources—genomics, epigenomics, transcriptomics, and clinical metadata—into unified frameworks. Without standardized formats and metadata protocols, interoperability becomes limited. Furthermore, privacy and ethical considerations—especially in clinical genomics—demand secure data governance models that comply with frameworks such as the General Data Protection Regulation (GDPR) and HIPAA (Beard et al., 2022).

Figure 3: Data Storage, Management, And Integration Challenges

V. AI And Machine Learning In Genomic Data Analysis

Artificial Intelligence (AI) and Machine Learning (ML) are transforming the way researchers examine and interpret genomic information. These technologies make it possible to uncover subtle patterns hidden within complex and multi-dimensional datasets.

By doing so, AI-based systems support a range of analytical tasks such as identifying genetic variants, annotating gene functions, and predicting phenotypic traits (Libbrecht& Noble, 2015). In recent years, advanced deep learning models—particularly convolutional and recurrent neural networks—have been successfully applied to problems including genome annotation, gene expression prediction, and the discovery of new drug targets.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 1, January 2026)

The growing use of ML in genomics has also accelerated the movement toward precision medicine. Predictive algorithms can help clinicians estimate disease risks, select the most effective treatment options, and develop personalized care plans tailored to individual genetic profiles. Despite these advances, the effectiveness of AI tools in genomics still depends largely on the quality and diversity of the data they learn from. Equally important are transparency, interpretability, and fairness; without them, algorithms risk amplifying existing biases or inequities in healthcare (Rajkomar et al., 2019).

VI. Personalized Medicine And Clinical Applications

Genomics and big data have exceptionally developed the concept of personalised medicine, which laid a foundation for the modern healthcare sector. Personalised medicine reframes clinical treatment, resolution, and prevention techniques based on the genetic and molecular features of an individual as an alternative to uniform medications.

The clinical applications of this prototype are more prominent in the fields of pharmacogenomics, oncology and disease management signifying the possibilities of interpreting genomic data into a substantial welfare of the patient (Ginsburg & Willard, 2009).

The approach has been most prominent in the field of oncology where the tumor gene sequencing led to targeted therapies upon mutations in genes such as BRAF, EGFR, or ALK, directly targeting the molecular reasons of cancer with reduced side effects compared to traditional chemotherapy (Wang et al., 2019). Pharmacogenomics is another significant field that applies personalized medicine. Mutations in genes encoding specific enzymes, receptors, and transporters can contribute to the patient's response to a particular drug. For example, a person with particular CYP2C19 mutants can poorly metabolize drugs like clopidogrel, leading to reduced efficacy (Sadee& Dai, 2005). Personalised medicine has revolutionized the diagnosis and management of rare diseases. Families with a history of undiagnosed diseases often termed as "diagnostic odyssey", are diagnosed using whole-genome and whole-exome sequencing (Taylor et al., 2011).

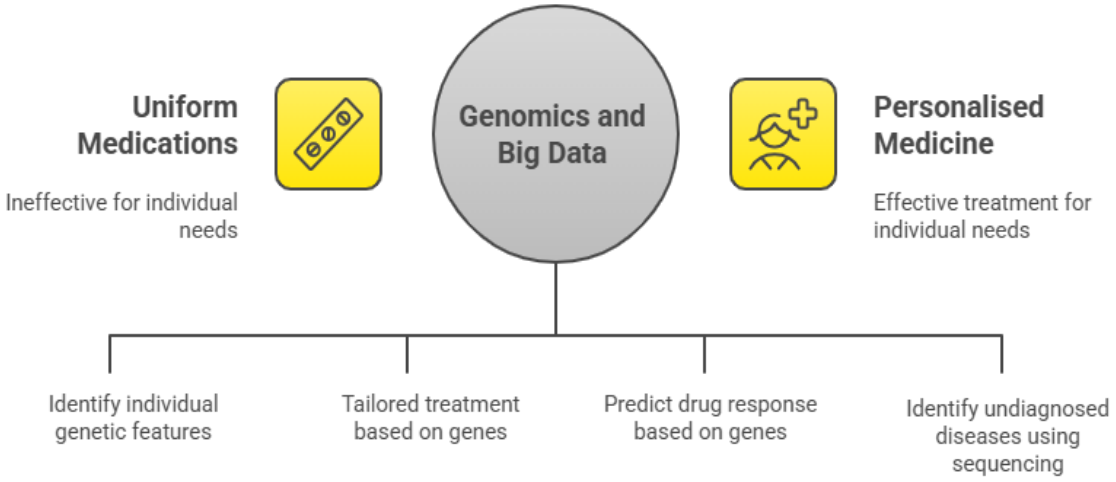

Figure 4:Genomics and big data transform healthcare from uniform medications to personalised medicine

The image illustrates how genomics and big data transform healthcare from uniform medications to personalised medicine. By analysing individual genetic features, genomic sequencing enables tailored treatments, prediction of drug responses, and identification of undiagnosed diseases, leading to more effective and patient-specific medical care.

VII. Ethical, Legal, And Privacy Concerns In Genomic Data Sharing

Advancements in genomics and big data provide a powerful platform for biomedical research, enhance precision medicine, and facilitate early detection of rare diseases that contribute to public health.

An increase in the storage and sharing of genomic data leads to various ethical, legal, and privacy disputes, which must be addressed to ensure safety of the data provided by the individual and their families.

A chronic problem is the risk of privacy and re-identification. Genomic data carries the individual's information along with the inherent data, increasing the risks for privacy. Even when the data is de-identified, advancements in recent technologies, along with the amalgamation of other datasets, such as hereditary, demographic, or biometric, allow the data to be re-identified (Kaye, 2012; Takashima et al., 2018). There are also psychological concerns for the individual or their family members, who are uninterested in knowing the information regarding the genetic risks (Bonomi et al., 2020). There are laws such as the European Union's General Data Protection Regulation (GDPR) that force necessity for data security and usage. But it provides an exemption for scientific research, allowing the use of pseudonymized genomic data under controlled access conditions (Kaye, 2012; Wang et al., 2017).

VIII. Interoperability And Standardization Of Genomic Databases

Interoperability and standardization are now identified as important principles for using genomic data in clinical research and medicine. In the generation of digital modernization, the proportions of genomic data have increased drastically, but they lack standardization of the formats and the shared frameworks of this information. Interoperability ensures that the raw sequence data can be shared and analyzed in laboratories and research institutions.

Preliminary research showed how scattered genomic data and lack of interoperability hindered teamwork and research. The fundamental cross-database queries in microbial genomics were repressed by contrasting nomenclature systems and discordant frameworks of the metadata (Romano et al., 2005). This highlighted that cultivating common data models and correlating guidelines was a required step for data integration. As the complexity of genomic data increased, the requirement of standardization extended. The potential interoperability not only depends on the files such as FASTQ, BAM, and VCF but is also administered by vocabularies, ontologies, and metadata frameworks that connect the genomic data with the biological and clinical information (Masseroli et al., 2016)

IX. Cloud Computing And Scalable Infrastructure For Genomics

Cloud computation and flexible infrastructure have transformed genomic research providing scalability, accessibility, and digitalization for large- scale data analysis. Gene sequencing produces large datasets which consume massive storage capacity and processing requirements. Cloud computation provides a flexible source that can automatically increase based on the computational requirements, which results in the effective management of large genomic data without the constraints of fixed-capacity servers (Schatz et al., 2010).

The flexibility of cloud-based systems allows the researchers to work with advanced genomic procedures such as read alignment and comparative genomics, which reduces the evaluation period (Langmead et al., 2018).The integration of big data strategies into bioinformatics workflows has optimised operations by using parallel processing and error-proof data processing. These strategies have authorised the implementation of digitally comprehensive bioinformatics algorithms through cloud computation with enhanced efficiency and replicability (Langmead et al., 2018). Further research on hybrid and multi-cloud strategies are attaining recognition, integrating the local information for confidential data with public cloud sources for adaptable computing (Stephens et al., 2015).

X. Future Directions: Digital Transformation In Precision Healthcare

Digital transformation is evolving rapidly in recent times in response to precision medicine but this is not visible in the case of genomics and big data. Today, genomic sequencing is more quick and affordable, the prediction of disease, diagnosis and treatment is possible with the help of precision medicine available in regular clinics (Stoumpos et al., 2023). The integration of multi-omics data and electronic health records (EBRs), help in understanding the interrelation between genomic data and phenotypic characters. The EHRs can be linked with the genomic information which enables researchers to study the hereditary risks and diagnostic methods (Robertson et al., 2024). Under the considerations of ethics, digitalization strategies can be used in the generation of patient friendly devices such as habiliment and kits. The combination of these devices with genomic data, the medications can be personalized accordingly (Stoumpos et al., 2023).

Eventually, the final goal of this digitalization in precision medicine is to design innovative devices with ethical considerations, must be affordable, and patient friendly. In the near future, we would be able to predict and diagnose diseases using this technology with precision and accuracy (The Role of Electronic Health Records in Advancing Genomic Medicine, 2021).

REFERENCES

[1] Mardis, E. R. (2017). Next-generation sequencing platforms. Annual Review of Analytical Chemistry, 10, 387–404.

[2] Stephens, Z. D., et al. (2015). Big Data: Astronomical or Genomical? PLoS Biology, 13(7), e1002195.

[3] Hasin, Y., Seldin, M., &Lusis, A. (2017). Multi-omics approaches to disease. Genome Biology, 18(1), 83.

[4] Schwarz, R. F., et al. (2020). Big data and machine learning in genomic medicine. Bioinformatics Advances, 1(1), vbab016.

[5] Ginsburg, G. S., & Willard, H. F. (2009). Genomic and personalized medicine: Foundations and applications. Translational Research, 154(6), 277–287.

[6] Goodwin, S., McPherson, J. D., &McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351.

[7] Logsdon, G. A., Vollger, M. R., &Eichler, E. E. (2020). Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21(10), 597–614.

[8] Beard, N., Naugler, C., & Smith, B. (2022). Data governance in genomic research: Ethical and legal considerations. Journal of Medical Ethics, 48(3), 210–217.

[9] Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16(6), 321–332.

[10] Rajkomar, A., Dean, J., &Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

[11] Wang, Q., Armeni, P., Gori, D., &Graziadio, S. (2019). The challenges of clinical implementation of personalized medicine. Human Genetics, 138, 27–37.

[12] Sadee, W., & Dai, Z. (2005). Pharmacogenetics/genomics and personalized medicine. Human Molecular Genetics, 14(2), R207–R214.

[13] Taylor, M. R. G., Carnes, J., &Basson, C. T. (2011). Genetic testing in cardiovascular medicine. Clinical and Translational Medicine, 1(1), 28.

[14] Takashima, K., Maru, Y., Mori, S. et al. Ethical concerns on sharing genomic data including patients' family members. BMC Med Ethics 19, 61 (2018).

[15] Bonomi, L., Huang, Y. &Ohno-Machado, L. Privacy challenges and research opportunities for genomic data sharing. Nat Genet 52, 646–654 (2020).

[16] Kaye, J. (2012). The tension between data sharing and the protection of privacy in genomics research. Annual Review of Genomics and Human Genetics, 13, 415-431.

[17] Wang, S., Jiang, X., Singh, S., Marmor, R., Bonomi, L., Fox, D., ...&Ohno- Machado, L. (2017). Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. Annals of the New York Academy of Sciences, 1387(1), 73-83.

[18] Romano, P., Dawyndt, P., Piersigilli, F., & Swings, J. (2005). Improving interoperability between microbial information and sequence databases. BMC bioinformatics, 6(Suppl 4), S23.

[19] Timme, R. E., Wolfgang, W. J., Balkey, M., Venkata, S. L. G., Randolph, R., Allard, M., & Strain, E. (2020). Optimizing open data to support one health: best practices to ensure interoperability of genomic data from bacterial pathogens. One Health Outlook, 2(1), 20.

[20] Masseroli, M., Kaitoua, A., Pinoli, P., &Ceri, S. (2016). Modeling and interoperability of heterogeneous genomic big data for integrative processing and querying. Methods, 111, 3-11.

[21] Schatz, M. C., Langmead, B., &Salzberg, S. L. (2010). Cloud computing and the DNA data race. Nature biotechnology, 28(7), 691-693.

[22] Langmead, B., & Nellore, A. (2018). Cloud computing for genomic data analysis and collaboration. Nature Reviews Genetics, 19(4), 208-219.

[23] Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ...& Robinson, G. E. (2015). Big data: astronomical or genomical?.PLoS biology, 13(7), e1002195.

[24] Stoumpos, A. I., Kitsios, F., &Talias, M. A. (2023). Digital transformation in healthcare: Technology acceptance and its applications. International Journal of Environmental Research and Public Health, 20(4), 3407.

[25] Robertson, A. J., et al. (2024). It is in our DNA: Bringing electronic health records and genomics closer together. Annual Review of Genomics and Human Genetics, 25, 1–24.

[26] The Role of Electronic Health Records in Advancing Genomic Medicine. (2021). Annual Review of Genomics and Human Genetics, 22, 89–113.

Corresponding Author: Chethana V Chalapathy

chethana.v@gcu.edu.in