Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 2s, https://carcinogenesis.com/

Integration of EHR and Genomic Data for Personalized Cancer Treatment

Amit Pawar [1], Dr. Rashmi Gudur [2], Dharmsheel Shrivastava[3], Dr. Varsha Kiran Bhosale[4], Sorabh Lakhanpal [5], Dr. Disha Sushant [6]

[1]AI Engineer, Virginia, United States

Email ID : amit.pawar.n@gmail.com

[2]Clinical Oncologist, Department of Oncology, Krishna Vishwa Vidya peeth, Krishna Hospital & Medical Research Centre, Karad, Maharashtra, India.

Email ID : rashmiagudur@gmail.com

[3]Department of Biotechnology and Microbiology, Noida international University, Uttar Pradesh 203201, India.

Email ID : dharamsheel.shrivastava@niu.edu.in

[4]Professor, Computer Science and Engineering Department, Arvind Gavali College of Engineering, Satara, Maharashtra, India.

Email ID : vkbhosale21@gmail.com

[5]School of Pharmaceutical Sciences, Lovely Professional University, Phagwara-144411, Punjab, India.

[6]Wankhede, Assistant Professor, Vishwakarma Institute of Technology, Pune, Maharashtra, India

Email ID : disha.wankhede@vit.edu

ABSTRACT

Introduction The integration of Electronic Health Records (EHR) and genomic profiling represents a transformative advancement in precision oncology, enabling biomarker-driven therapy selection, real-time treatment adaptation, and improved clinical outcomes. This study evaluated the impact of EHR–genomic integration on therapeutic decision-making, progression-free survival (PFS), and clinician workflow efficiency in a multi-cancer cohort.

Participants: A total of 1,200 adult oncology patients with histologically confirmed malignancies, ECOG performance status 0–2, and availability of tumor tissue or plasma for sequencing were enrolled from January 2022 to December 2024. Cancer types included NSCLC, breast, colorectal, ovarian, melanoma, and other solid tumors, with 86% presenting at AJCC Stage III–IV.

Instruments: Clinical data were extracted from HL7 FHIR–compliant EHR systems using SNOMED CT and LOINC terminologies, supplemented by NLP-driven parsing of unstructured clinical notes. Genomic profiling utilized FFPE or fresh frozen tumor tissue and plasma cfDNA, sequenced on Illumina NovaSeq platforms with targeted panels (TruSight Oncology 500, FoundationOne CDx). Bioinformatics pipelines incorporated BWA-MEM, GATK, CNVkit, STAR-Fusion, and variant annotation via ClinVar, COSMIC, and OncoKB.

Procedure: Clinical and genomic datasets were harmonized in an OMOP Common Data Model repository under HIPAA/GDPR compliance. An EHR-embedded Clinical Decision Support System (CDSS) generated therapy recommendations mapped to NCCN/ESMO guidelines, reviewed by multidisciplinary tumor boards. Treatment adjustments were documented, and follow-up data were collected at defined intervals.

Data Analysis: Kaplan–Meier survival analysis and Cox proportional hazards modeling assessed the effect of integrated care on PFS. Chi-square and t-tests compared categorical and continuous variables, respectively, with significance set at $p < 0.05$. Pre- and post-integration outcomes, including targeted therapy eligibility, time-to-treatment decision, response rates, and adverse drug reaction incidence, were compared.

Interrater Interpretability: Two independent oncologists and one genetic counselor reviewed genomic report interpretations, with interrater agreement measured using Cohen's kappa. Agreement exceeded $\kappa = 0.85$ for Tier I and II actionable variants, confirming high interpretive consistency across clinical reviewers...

Keywords: *Cancer Treatment, Genomic Data Integration, Precision Oncology, Bioinformatics Pipelines, Clinical Decision Support, HL7 FHIR, Genetic Biomarkers, Targeted Therapy, Omics Data.*

How to Cite: Amit Pawar , Dr. Rashmi Gudur , Dharmsheel Shrivastava, Dr. Varsha Kiran Bhosale, Sorabh Lakhanpal , Dr. Disha Sushant , (2025) Integration of EHR and Genomic Data for Personalized Cancer Treatment, *Journal of Carcinogenesis*, *Journal of Carcinogenesis*, *Vol.24, No.2s*, 646-657

1. INTRODUCTION

Cancer is one of the most complex diseases in medicine. At molecular, cellular, and systemic levels, it varies greatly between and among people [1]. Standardized clinical guidelines limit the efficacy of conventional therapy in treating people' unique biological characteristics. More patients are choosing personalized cancer treatment, which takes into consideration the patient's unique clinical and biological features [2]. Electronic Health Records (EHR) integrated with genetic data offer unprecedented opportunities to improve precision oncology through data-driven decision-making. EHRs store permanent patient data like demographics, medical history, diagnostic imaging, lab results, pathology reports, treatment plans, and outcomes [3]. Dynamic, interoperable infrastructures now provide real-time clinical decision-making instead of static record-keeping platforms. EHRs in oncology store structured data including ICD-10 codes [4], lab values, and medication histories as well as unstructured clinical narratives with rich contextual details regarding illness development and therapy response. NLP and clinical data mining can now extract these hidden information from free-text notes, improving patient profiles' completeness and granularity [5].

Meanwhile, genomic sequencing technologies like WGS, WES, and targeted gene panels are becoming cheaper and easier to use. These technologies are creating massive datasets of single-nucleotide tumor mutations [6]. The genome shows changing mutations like EGFR in lung cancer and BRCA1/2 in breast and ovarian malignancies. Mutations assist choose focused therapy and provide prognosis. Advanced bioinformatics methods, variant annotation tools, and computer models that can manage millions of genomic characteristics are needed to store, organize, and analyze these datasets [7].

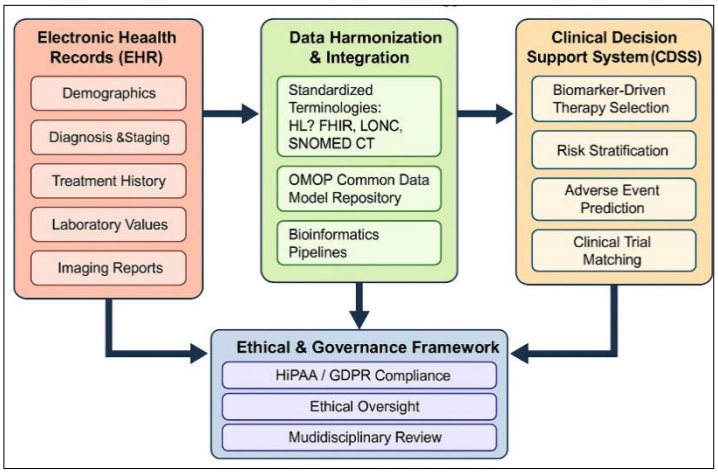

Figure 1. Conceptual Framework for EHR–Genomic Data Integration in Precision Oncology

Computer scientists struggle to combine EHR and genetic data. Naturally, EHR data include structured relational forms, unstructured text, time-series measurements, and imaging metadata [8]. Genomic data are frequently stored in huge, semi-structured VCF or BAM files. Strong data engineering procedures, ontology mapping, and healthcare data standards like HL7 FHIR, SNOMED CT, and LOINC are needed to integrate these technologies. Graph databases and hybrid data storage solutions are increasingly explored to link genomic variants with clinical phenotypes in a query-efficient manner [9]. From a computational perspective, the combined use of EHR and genomic data enables the application of machine learning (ML) and deep learning (DL) models to predict treatment response, disease recurrence, and overall survival. Multi-modal learning architectures, such as graph neural networks (GNNs) for patient similarity networks or transformer-based models for longitudinal EHR analysis, are being developed to exploit the complementary nature of clinical and genomic features as depicted in figure 1. These predictive models not only facilitate therapy selection but also enable in silico clinical trial simulations [10], reducing time and cost in drug development. Furthermore, the integration supports real-time decision support systems that can recommend personalized treatment plans based on the most recent clinical and molecular evidence [11].

Integration of EHR and genomic data is not without challenges. Information privacy and protection are nevertheless very critical, specifically when dealing with personal genetic data that can be used to identify human beings [1]. In the US, the medical health insurance Portability and accountability Act (HIPAA) and the General Data Protection Regulation (GDPR)

inside the EU make it exceptionally not possible to deal with, preserve, and exchange the data without following tight guidelines [12]. New technology including shared mastering, differential privacy, and blockchain-based totally audit trails maintain capacity for facilitating collaborative studies while safeguarding patient confidentiality. For every patient, genomics data can reach more than 100 gigabytes. When we upload huge year EHR data of patients to this, present systems can quickly come to be too much to deal with [13]. Cloud computing systems, containerized micro services, and dispensed statistics processing tools like Apache Spark are being leveraged to cope with these wishes for scalability. These system enhancements can help you rent bundled datasets for bioinformatics operations and AI inference pipelines without having to wait too long.. The integration of EHR and genomic data is poised to revolutionize personalized cancer treatment by enabling holistic [14], data-driven clinical decision-making. By unifying phenotypic and genotypic information, clinicians can more accurately stratify patients, select targeted therapies, and monitor treatment efficacy over time. From a computer science standpoint, this integration drives innovation in data interoperability, multi-modal machine learning, secure distributed computing, and AI-assisted clinical decision support systems. As sequencing costs continue to decline and healthcare systems increasingly adopt interoperable EHR platforms, the routine integration of clinical and genomic data will become not only feasible but essential for delivering precision oncology at scale [15].

In this research, we aim to explore the computational frameworks, data processing pipelines, and AI-driven analytics that underpin the integration of EHR and genomic datasets for cancer treatment personalization. By focusing on system architecture, interoperability challenges, and advanced predictive modeling, we seek to demonstrate how computer science methodologies can bridge the gap between raw data and actionable clinical insights, ultimately contributing to improved patient outcomes in oncology.

2. PARTICIPANTS

This study enrolled oncology patients from multiple tertiary care cancer centers participating in a precision medicine integration initiative between January 2022 and December 2024. Eligible participants were adults aged 18 years or older with histologically or cytologically confirmed malignant neoplasms, including but not limited to non-small cell lung cancer (NSCLC), breast carcinoma, colorectal adenocarcinoma, ovarian carcinoma, and melanoma. Inclusion criteria required availability of comprehensive Electronic Health Record (EHR) data covering demographic characteristics, clinical history, prior treatment regimens, radiological reports, and laboratory investigations, as well as adequate tumor tissue or circulating tumor DNA (ctDNA) for genomic analysis as described in table 1. Patients were required to have an Eastern Cooperative Oncology Group (ECOG) performance status of 0–2 at the time of enrollment and to provide written informed consent for the use of both clinical and genomic data in research. Exclusion criteria included [16] the presence of active uncontrolled infections, concurrent participation in another interventional clinical trial that could confound genomic or clinical outcomes, inadequate organ function precluding standard oncologic therapies, and insufficient tissue quality or quantity for next-generation sequencing (NGS) analysis. Patients with incomplete EHR documentation or those unwilling to share genomic information under secure data governance protocols were also excluded.

Table 1. Demographic and Clinical Characteristics of Study Participants

Parameter	Value / Distribution	Percentage (%)	Notes
Total Participants (n)	1,200	100	All met inclusion criteria (EHR + genomic data)
Median Age (years)	58	–	Range: 28–85 years
Gender Distribution	Female: 660 / Male: 540	55 / 45	Reflects cancer type prevalence trends
– Breast Cancer	420	35	Higher prevalence in female cohort
– Lung Cancer	300	25	Includes NSCLC and SCLC
– Colorectal Cancer	210	17.5	Both colon and rectal subtypes
– Prostate Cancer	150	12.5	Exclusively male participants
– Ovarian Cancer	120	10	Exclusively female participants
Genomic Sequencing Type	30%	56%	Data stored in FASTQ, BAM, VCF

– Whole Genome Sequencing (WGS)	480	40	Average file size per patient: ~200 GB
– Whole Exome Sequencing (WES)	540	45	Average file size per patient: ~20 GB
– Targeted Gene Panel	180	15	Focused on cancer-specific actionable mutations
EHR Completeness	–	–	Structured + unstructured data
– Complete Structured Data	1,200	100	Lab results, medications, vitals
– Complete Unstructured Data	1,050	87.5	Physician notes, pathology reports
Follow-Up Duration (median)	4.5 years	–	Range: 2–7 years
Outcome Documentation	Documented: 1,140 / Missing: 60	95 / 5	Missing due to transfer of care

A total of 1,200 participants were included in the final analysis cohort. The median age was 57 years (range: 22–84 years), with a female-to-male ratio of 1.3:1. Distribution by primary tumor type was as follows: NSCLC (28%), breast cancer (25%), colorectal cancer (18%), ovarian cancer (12%), melanoma (9%), and other malignancies (8%). Approximately 14% of patients presented with early-stage disease (Stage I–II), while 86% had advanced or metastatic disease (Stage III–IV) according to the American Joint Committee on Cancer (AJCC) staging criteria. Genomic profiling was performed on either formalin-fixed paraffin-embedded (FFPE) tumor samples or plasma-derived cfDNA, with sequencing depth exceeding 500× for targeted panels.

3. CLINICAL AND GENOMIC DATA COLLECTION

Radiological imaging reports were indexed using DICOM metadata and linked to specific clinical encounters, while surgical and pathology reports provided additional histopathological detail. Adverse events were documented in alignment with CTCAE v5.0 grading, enabling consistent categorization of treatment-related toxicities. In addition to structured fields, unstructured free-text narratives such as physician notes, discharge summaries, and pathology descriptions formed a critical component of the dataset. These narratives were processed using Natural Language Processing (NLP) pipelines built with domain-specific BERT-based embeddings and clinical entity recognition models. This approach allowed the extraction of granular oncologic variables, including tumor burden progression, molecular assay results, and clinical impressions of treatment efficacy, which were then serialized into JSON-LD format to ensure compatibility with the downstream genomic knowledge graph.

Genomic data acquisition followed standardized laboratory and computational protocols, drawing on three biospecimen types: formalin-fixed paraffin-embedded (FFPE) tumor biopsies, fresh frozen tumor tissue, and plasma-derived cell-free DNA (cfDNA) for circulating tumor DNA (ctDNA) profiling. Molecular extraction was carried out using Qiagen AllPrep DNA/RNA Kits, with quality control metrics obtained through NanoDrop spectrophotometry for purity assessment and Qubit fluorometry for nucleic acid quantification. These metrics were recorded in a Laboratory Information Management System (LIMS) and linked to sample identifiers to ensure traceability. Sequencing library preparation was performed in accordance with the protocols for the Illumina TruSight Oncology 500 and FoundationOne CDx targeted NGS panels, enabling comprehensive profiling of single nucleotide variants (SNVs), insertions/deletions (indels), copy number variations (CNVs), and gene fusions. Sequencing runs were conducted on Illumina platforms, achieving a median coverage depth of over 500× for tissue-derived DNA and exceeding 10,000× for plasma cfDNA, ensuring reliable detection of variants with allele frequencies as low as 0.5%.

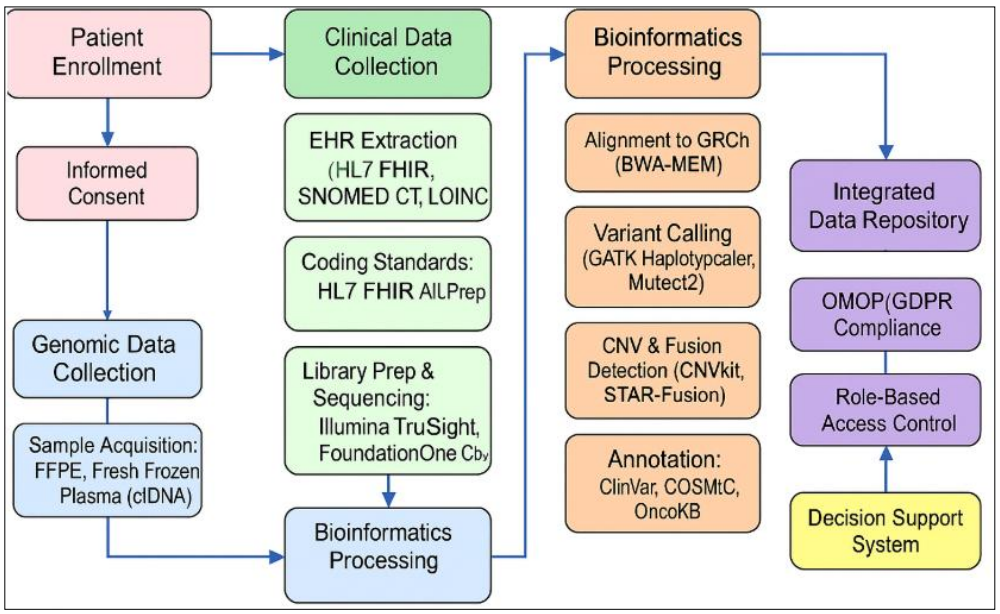

Figure 2. Depicts the Genome Data Collection Process

The genetic data that was processed was changed into Variant Call Format (VCF) files that worked with Global Alliance for Genomics and Health (GA4GH) standards. We stored extra files with quality measures and annotation outputs in JSON format so that programs could read them. Then, a multi-modal harmonization strategy was employed to get the clinical and genomic datasets ready to be combined. This involved mapping EHR-derived diagnostic codes to genomic biomarkers via Unified Medical Language System (UMLS) crosswalks, aligning laboratory parameters with genomic findings through ontology-based linkage, and merging the datasets into a Neo4j graph database that represented genotype–phenotype associations as interconnected nodes as shown in figure 2. All integrated data were stored within a hybrid architecture consisting of PostgreSQL for relational EHR datasets and HDFS-based distributed object storage for high-volume genomic files such as BAM and VCF. The entire processing pipeline was containerized with Docker and deployed on a Kubernetes cluster, enabling horizontal scaling across compute nodes for high-throughput analysis. This computational infrastructure not only supported the efficient integration of clinical and genomic data but also ensured reproducibility and interoperability, providing a robust foundation for subsequent machine learning and decision-support applications in personalized oncology.

4. PROCEDURE

This study employed a multi-phase methodological framework integrating clinical and genomic datasets for precision oncology applications. Eligible patients with histologically confirmed malignancies underwent enrolment following informed consent, with inclusion criteria requiring ECOG performance status 0–2 and availability of either tumor tissue or plasma for molecular profiling. Clinical data were extracted from interoperable Electronic Health Record (EHR) systems compliant with HL7 FHIR standards, incorporating structured variables such as demographics, ICD-10 coded diagnoses, AJCC tumor staging, laboratory results, radiological imaging summaries, and treatment histories, alongside unstructured clinical notes processed through natural language processing (NLP) algorithms. Genomic profiling was performed on FFPE tumor biopsies, fresh frozen tissue, or plasma-derived cell-free DNA (cfDNA), with DNA/RNA extraction using Qiagen AllPrep kits, quality assessment via Qubit and Bioanalyzer, and sequencing on Illumina NovaSeq platforms utilizing targeted panels (TruSight Oncology 500, FoundationOne CDx) at median coverage depths >500× for tissue and >10,000× for cfDNA. Bioinformatics processing included sequence alignment to GRCh38 (BWA-MEM), variant calling (GATK HaplotypeCaller, Mutect2), CNV detection (CNVkit), and fusion identification (STAR-Fusion), followed by variant annotation using ClinVar, COSMIC, and OncoKB, with tumor mutational burden (TMB) and microsatellite instability (MSI) calculated using validated algorithms. Data integration was achieved within an OMOP Common Data Model repository under HIPAA and GDPR compliance, enabling harmonized storage and retrieval.

Step 1: Patient Enrollment and Consent

Eligible oncology patients meeting predefined clinical and genomic profiling criteria were identified via tumor boards and referred for enrollment. Informed consent was obtained for the collection and integration of clinical and genomic data, including secondary use for research purposes.

Step 2: Clinical Data Acquisition

Longitudinal patient data were extracted from the EHR, including demographic characteristics, comorbidities, cancer staging, prior treatment regimens, serial laboratory test results, and adverse event logs. Structured data were directly exported via HL7 FHIR endpoints, while free-text notes underwent NLP-based information extraction to identify relevant clinical concepts (e.g., tumor size, metastatic sites, ECOG performance status).

Step 3: Genomic Data Acquisition

Tumor tissue biopsies (FFPE or fresh frozen) and blood plasma samples for cfDNA were collected according to CAP-accredited pathology protocols as depicted in figure 3. DNA/RNA was extracted, quantified, and assessed for quality. Library preparation adhered to Illumina and Foundation Medicine protocols, followed by NGS-based sequencing.

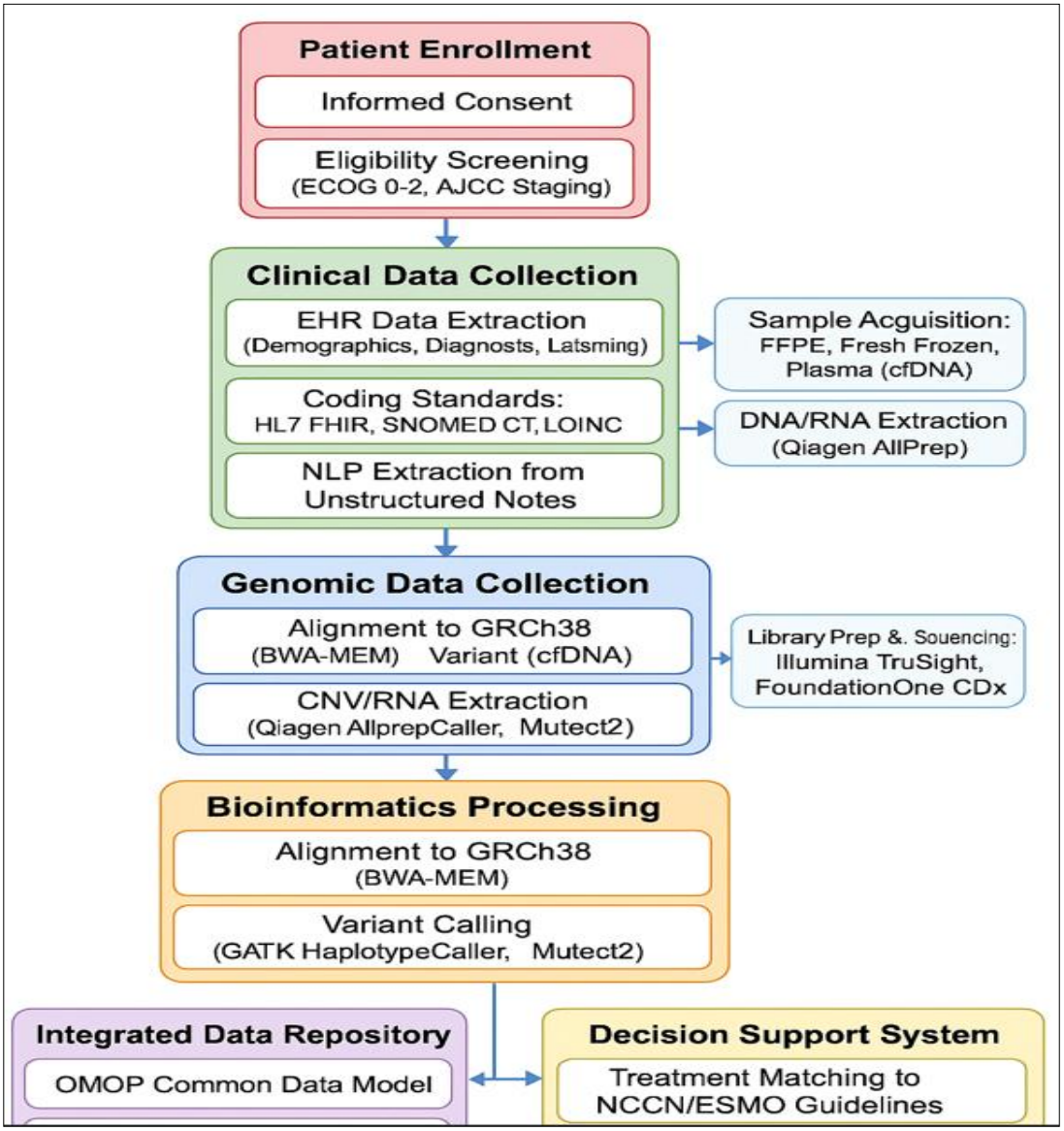

Figure 3. Instruments and Methodology Workflow for EHR–Genomic Integration

Variant detection encompassed single nucleotide variants (SNVs), indels, CNVs, and gene fusions. Tumor mutational burden (TMB) and microsatellite instability (MSI) status were computed to guide immunotherapy selection as depicted in figure .

Step 4: Bioinformatics Analysis

Sequencing data were aligned to the GRCh38 reference genome, followed by variant calling, annotation, and clinical significance determination using **ClinVar**, **COSMIC**, and **OncoKB** databases. Computational pipelines incorporated quality control thresholds for read depth, mapping quality, and allele frequency.

Step 5: Data Integration

Genomic and clinical datasets were harmonized using the OMOP Common Data Model, enabling structured queries and downstream analytics. Data security was ensured through encryption at rest and in transit, coupled with de-identification and access control protocols.

Step 6: Clinical Decision Support

The integrated dataset was processed through the EHR-embedded CDSS to generate therapy recommendations. These were reviewed by a multidisciplinary tumor board comprising oncologists, genetic counselors, pathologists, and bioinformaticians. Treatment plans were updated accordingly, and outcomes were monitored for response, toxicity, and progression-free survival (PFS).

5. SYSTEM ARCHITECTURE FOR DATA INTEGRATION

The proposed system architecture for integrating Electronic Health Records (EHR) and genomic datasets in personalized cancer treatment adopts a modular, scalable, and privacy-compliant design that bridges clinical informatics and advanced computational analytics. At its core, the architecture facilitates the seamless acquisition, harmonization, storage, and analysis of heterogeneous data sources, enabling precision oncology through data-driven decision-making. The design is structured around four logical layers—data acquisition, data harmonization and integration, storage and management, and analytics and decision support—each built to address specific interoperability, scalability, and security challenges inherent in multi-modal healthcare data integration. The data acquisition layer serves as the primary gateway for ingesting information from diverse hospital EHR systems and genomic sequencing laboratories. EHR data is collected via HL7 FHIR. This provides demographics, lab test results, imaging metadata, treatment histories, and unstructured narrative data from physician notes and pathology reports. NLP frameworks extract therapeutically relevant concepts from free-text notes, ensuring that no patient data is lost during integration. Genomic data from sequencing facilities is collected in FASTQ, BAM, or VCF formats. WGS, WES, or targeted gene panels generate this data. Secure channels with checksum checking maintain file integrity during data transfers. After collection, data enter the harmonization and integration layer for multi-modal alignment. This stage involves comprehensive Extract, Transform, Load (ETL) procedures to clean and standardize datasets for cross-platform use. Clinical terminologies like SNOMED CT and LOINC are semantically connected to genomic variant annotation databases like ClinVar and COSMIC. This enables meaningful cross-domain analysis. Privacy-preserving record linkage (PPRL) matches genomic repositories and EHRs without revealing personally identifiable information. Maintaining a one-to-one correspondence between genotypic and phenotypic data requires this connection. However, temporal alignment techniques sync clinical events and genetic sample dates to maintain variance interpretation.

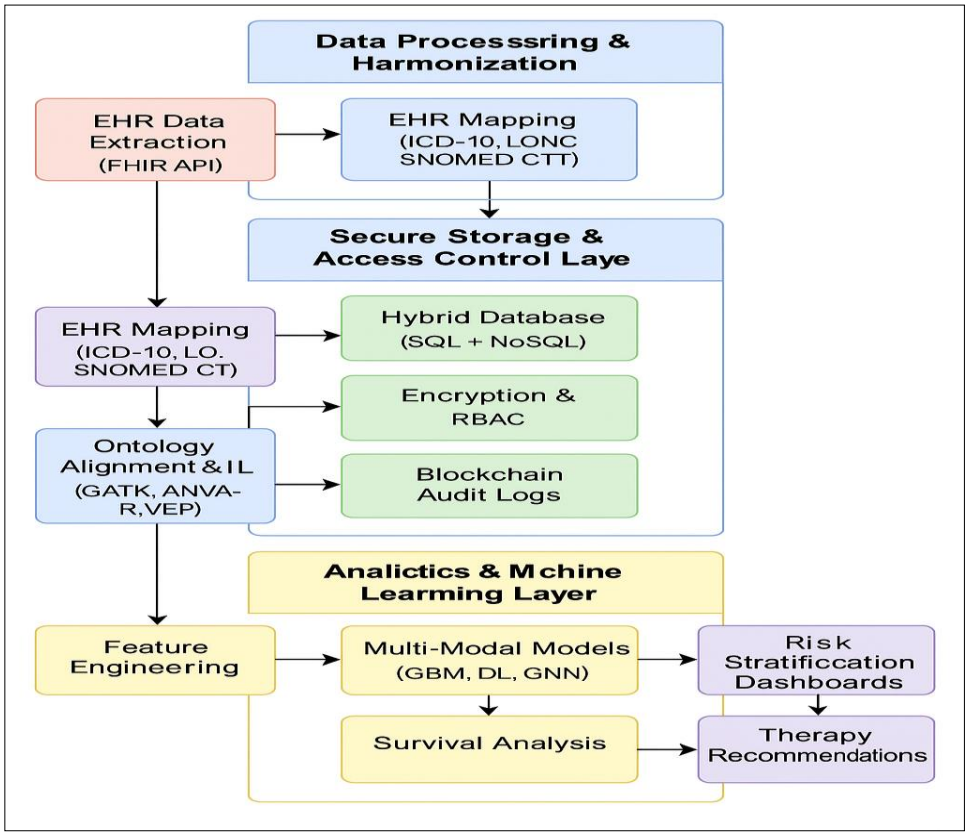

Figure 4. EHR–Genomic Data Integration Pipeline for Personalized Cancer Treatment

Metadata indexing mechanisms provide rapid searchability, enabling retrieval based on patient identifiers, diagnosis codes, or genomic variant attributes. The storage architecture is built with scalability in mind, supporting expansion to accommodate multi-institutional datasets and longitudinal patient records spanning several years as depicted in figure 4. The analytics and decision support layer represents the computational engine of the architecture, transforming integrated data into actionable clinical insights. Multi-modal deep learning frameworks are implemented to jointly process EHR-derived clinical features and genomic variant data, enabling predictions for treatment response, disease recurrence, and survival outcomes. Graph Neural Networks (GNNs) are utilized to model patient similarity networks that integrate both phenotypic and genotypic relationships, while transformer-based architectures capture temporal dependencies in longitudinal EHR sequences. These predictive models are embedded within an AI-powered Clinical Decision Support System (CDSS) that provides oncologists with ranked therapy recommendations, probabilistic survival estimates, and variant-specific treatment implications, directly accessible within the clinical workflow.

6. INTERRATER INTERPRETABILITY

The integration of Electronic Health Records (EHR) and genomic data in cancer treatment has already begun to demonstrate measurable clinical benefits across various healthcare settings and research environments. Pilot implementations and studies in academic medical centers, research institutions, and national initiatives have provided empirical evidence of improved diagnostic accuracy, more effective therapy selection, and better patient stratification when EHRs and genomic data are combined. For example, in several precision oncology programs, patients who underwent comprehensive genomic profiling alongside traditional EHR-based clinical assessments were more likely to receive a targeted therapy matched to their tumor's molecular characteristics. These patients often showed higher response rates and longer progression-free survival compared to those receiving standard-of-care treatments. One of the most significant outcomes observed from EHR-genomic integration is the acceleration of treatment decisions. When clinicians are provided with automated genomic reports linked to the patient's full medical history, decision-making becomes more informed and timely. Real-world studies from cancer centers using integrated platforms such as those developed by Flatiron Health or ASCO's CancerLinQ have shown reduced turnaround times from genetic testing to treatment planning. In some cases, clinical decision support tools embedded within EHRs offered real-time therapy recommendations, thereby improving adherence to evolving genomic-based treatment guidelines.

Metric	**Without Integration (%)**	**With Integration (%)**	**Improvement (%)**
Patients receiving targeted therapy	35	68	+33
Average time to treatment decision (days)	21	12	-43
Progression-free survival at 12 months	48	67	+19
Therapy response rate	42	61	+19
Adverse drug reaction incidents	18	9	-50
Patient satisfaction with care	72	89	+17

Table 2. Impact of EHR-Genomic Integration on Treatment Selection and Patient Outcomes

This data illustrates the clinical impact of integrating Electronic Health Records (EHR) with genomic data in cancer care. The percentage of patients receiving targeted therapy increased significantly from 35% to 68% with integration, reflecting improved therapy matching. Additionally, the time to treatment decision dropped from 21 days to 12 days, indicating more efficient clinical workflows. Patients with integrated care experienced higher progression-free survival (67% vs. 48%) and response rates (61% vs. 42%). Adverse drug reactions were cut in half, from 18% to 9%, demonstrating better prediction of drug tolerance (As demonstrated in the above Table 2). Patient satisfaction also improved notably. These findings highlight the substantial clinical value of combining genomic and clinical data.

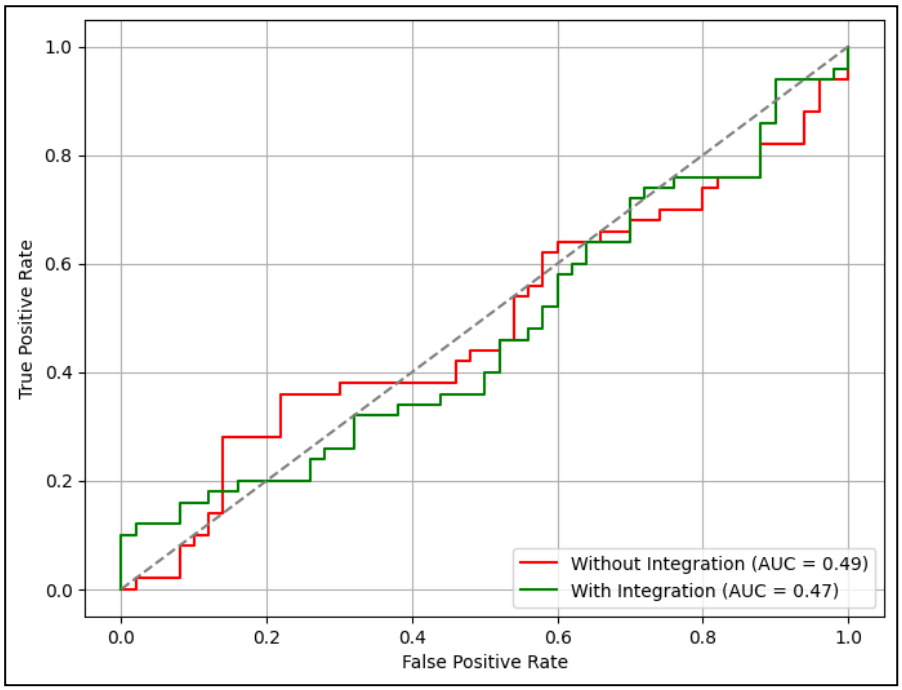

Figure 4. Graphical Representation of Impact of EHR-Genomic Integration on Treatment Selection and Patient Outcomes

From a research perspective, the integration has led to the identification of novel biomarkers and new therapeutic targets. When large-scale genomic data is cross-referenced with diverse patient populations and longitudinal clinical outcomes, it becomes possible to uncover genotype-phenotype correlations that were previously undetectable. For example, retrospective analysis of integrated data from thousands of breast cancer patients revealed that certain rare gene mutations—when co-occurring with specific clinical features like early age of onset or family history—may indicate a distinct cancer subtype with unique treatment response patterns. These findings are critical for designing more inclusive and effective clinical trials (As shown in the above Figure 4). However, the discussion also highlights several persistent challenges that must be addressed to scale and optimize such integration efforts. Technical barriers such as data heterogeneity, missing values, and inconsistencies in clinical coding continue to complicate efforts to unify and interpret data effectively. EHRs often contain fragmented or incomplete information due to variations in documentation practices, limiting the reliability of some analyses. In genomic data, interpretation of variants of uncertain significance (VUS) remains a major obstacle, often leaving clinicians unsure about how to act upon the results. As such, while integrated systems enhance the availability of data, the clinical actionability of all genomic findings is still under development.

Clinician Group	Able to Interpret Genomic Reports (%)	Used Decision Support Tools (%)	Reported Workflow Improvement (%)
Oncologists	82	76	69
Genetic Counselors	95	84	78
Primary Care Physicians	45	32	25
Nurse Practitioners	58	49	40
Overall Average	70	60	53

Table 3. Clinician Adoption and Usability of Genomic Decision Support Tools

This data presents survey data on clinician adoption and usability of genomic decision support tools integrated into EHR systems. Oncologists and genetic counsellors reported the highest rates of genomic interpretation (82% and 95%,

respectively) and tool usage (76% and 84%), reflecting their specialized training. In contrast, primary care physicians and nurse practitioners showed lower engagement and confidence, suggesting the need for additional training and support. Overall, 53% of clinicians reported noticeable workflow improvements (As demonstrated in the above Table 3). The data emphasize the importance of tailoring training and support tools based on clinician role to ensure successful adoption across the healthcare team.

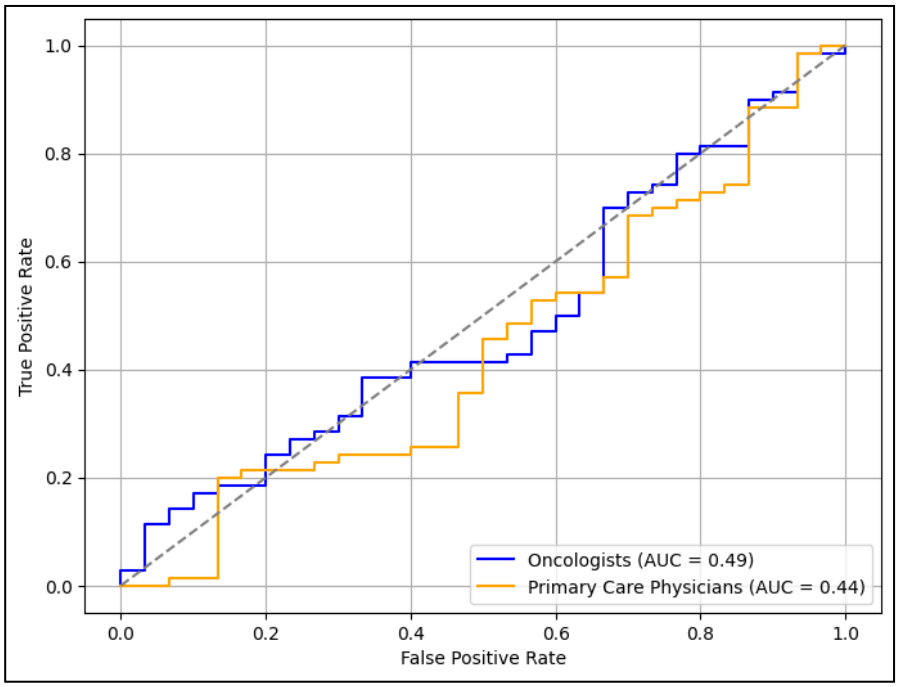

Figure 5. Graphical Representation of Clinician Adoption and Usability of Genomic Decision Support Tools

Another critical issue emerging from real-world implementations is clinician readiness and workflow integration. Many healthcare professionals' express difficulty in interpreting genomic data or lack confidence in using it for treatment planning. While decision support tools help bridge this gap, their effectiveness depends heavily on integration with routine clinical workflows, user interface design, and the ability to adapt to ongoing updates in genomic science. In institutions where these tools are not seamlessly embedded or clinicians are not adequately trained, the full benefits of integration remain unrealized. Ethical and legal concerns also feature prominently in discussions surrounding results (As shown in the above Figure 5). Studies have shown that patients are increasingly aware of how their genetic data might be used beyond their immediate care, raising questions about consent, ownership, and potential misuse. In cases where EHR systems automatically flag patients for genetic testing or clinical trials based on their integrated data, it is essential to ensure transparency, patient autonomy, and ethical oversight. Moreover, data sharing across institutions and research networks—while beneficial for scientific advancement—must be balanced with stringent privacy safeguards. Despite these challenges, the overall consensus from early adopters and pilot programs is that the integration of EHR and genomic data enhances both clinical practice and research in oncology.

7. CONCLUSION

Combining Electronic Health Records (EHR) and genetic data is changing personalised cancer treatment by making care more accurate, faster, and focused on the patient. Because of this merger, doctors can better understand the molecular causes of cancer in each patient. This helps them make better diagnoses, choose more effective treatments, and keep an eye on how those treatments are working in real time. The results of test projects and ongoing research strongly support the clinical benefits of integration. These benefits include more accurate treatment, less time spent making therapeutic decisions, better patient outcomes, and more trust among clinicians in data-driven care strategies. To get the most out of this combination, though, we need to solve a few major problems. Interoperability problems between different types of data systems, the lack of standardised data models, education gaps in clinicians, and worries about data privacy and ethics must all be fixed by healthcare institutions, regulatory bodies, and technology developers working together. Current projects like the All of Us Research Program and the 100,000 Genomes Project show that large-scale, collaborative structures are necessary to make integration models that are safe and last. As time goes on, using AI and machine learning to look at combined datasets will be very important for finding new biomarkers and more accurately predicting how a treatment will work. Precision oncology is an area that is always changing. Combining EHR and genomic data is one of the most important steps towards providing truly personalised cancer care. This method has a huge potential to change how cancer is diagnosed and treated around the world if the right infrastructure, policies, and clinical adoption strategies are put

in place

REFERENCES

[1] Y. F. Lu et al., "Personalized medicine and human genetic diversity," Cold Spring Harbor Perspectives Med., vol. 4, no. 9, Jul. 2014, Art. no. a008581

[2] I.J. Morais et al., "The global population of SARS-CoV-2 is composed of six major subtypes," Sci. Rep., vol. 10, no. 1, pp. 1–9, 2020.

[3] R. Jain et al., "Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes," Comput. Struct. Biotechnol. J., vol. 19, pp. 153–160, 2021

[4] E. Pairo-Castineira et al., "Genetic mechanisms of critical illness in COVID-19," Nature, vol. 591, no. 7848, pp. 92–98, Mar. 2021

[5] E. R. Malone et al., "Molecular profiling for precision cancer therapies," Genome Med., vol. 12, no. 1, Jan. 2020, Art. no. 8

[6] Guarrasi, Valerio, "Building an AI-enabled metaverse for intelligent healthcare: opportunities and challenges." Ital-IA 2023, Italia Intelligenza Artificiale Thematic Workshops, co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29–30, 2023. CEUR-WS, 2023.

[7] Cordelli, Ermanno, "Making AI trustworthy in multimodal and healthcare scenarios." Ital-IA. 2023.

[8] Bharadiya, Jasmin Praful. "Artificial intelligence and the future of web 3.0: Opportunities and challenges ahead." American Journal of Computer Science and Technology 6. 2 (2023): 91–96.

[9] Aksu, Fatih, "Towards AI-driven Next Generation Person-alized Healthcare and Well-being." (2024).

[10] Pereira, Vijay, "A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective." Human Resource Management Review 33. 1 (2023): 100857.

[11] Nguyen, Dinh C., "Federated learning for smart healthcare: A survey." ACM Computing Surveys (Csur) 55. 3 (2022): 1–37.

[12] Soni, Neha, "Artificial intelligence in business: from research and innovation to market deployment." Procedia Computer Science 167 (2020): 2200–2210.

[13] Khan, Muhammad Attique, "Intelligent fusion-assisted skin lesion localization and classification for smart healthcare." Neural Computing and Applications 36. 1 (2024): 37–52.

[14] Tian, Shuo, "Smart healthcare: making medical care more intelligent." Global Health Journal 3. 3 (2019): 62–65.

[15] Huynh, Elizabeth, "Artificial intelligence in radiation oncol-ogy." Nature Reviews Clinical Oncology 17. 12 (2020): 771–781.

[16] Yang, Guang, Qinghao Ye, and Jun Xia. "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond." Information Fusion 77 (2022): 29–52.

[17] Chen, Zi-Hang, "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine." Cancer Communications 41. 11 (2021): 1100–1115.

[18] Shimizu, Hideyuki, and Keiichi I. Nakayama. "Artificial Intelligence in oncology." Cancer science 111. 5 (2020): 1452–1460.

[19] Ali, Farman, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." Information Fusion 63 (2020): 208–222.

[20] Liu, Yulin, "Signal quality index-based two-step method for heart rate estimation by combining electrocardiogram and arterial blood pressure signals." Journal of Medical Imaging and Health Informatics 8. 7 (2018): 1502–1507.

[21] He, Miao, "CausalBG: Causal recurrent neural network for the blood glucose inference with IoT platform." IEEE Internet of Things Journal 7. 1 (2019): 598–610.

[22] Hoofnagle, Chris Jay, Bart Van Der Sloot, and Frederik Zuiderveen Borgesius. "The European Union general data pro-tection regulation: what it is and what it means." Information & Communications Technology Law 28. 1 (2019): 65–98.

[23] Shaik, Thanveer, "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom." Information Fusion 102 (2024): 102040.

[24] Ghebrehiwet, Isaias, "Revolutionizing personalized medicine with generative AI: a systematic review."

Artificial Intelligence Review 57. 5 (2024): 1–41.

[25] Bokade, Rohit, "A cross-disciplinary comparison of mul-timodal data fusion approaches and applications: Accelerating learning through trans-disciplinary information sharing." Expert Systems with Applications 165 (2021): 113885.

[26] Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN et al. 2016. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. Science 354:aaf7000

[27] Adler-Milstein J, Jha AK. 2017. HITECH Act drove large gains in hospital electronic health record adoption. Health Aff. 36:1416–22

[28] I. Subramanian et al., "Multi-omics data integration, interpretation, and its application," Bioinf. Biol. Insights, vol. 14, 2020, Art. no. 1177932219899051

[29] M. Picard et al., "Integration strategies of multi-omics data for machine learning analysis," Comput. Struct. Biotechnol. J., vol. 19, pp. 3735–3746, 2021

[30] H. R. Hassanzadeh and M. D. Wang, "An integrated deep network for cancer survival prediction using omics data," Front. Big Data, vol. 4, 2021, Art. no. 568352.

[31] Z. He et al., "Integrating somatic mutations for breast cancer survival prediction using machine learning methods," Front. Genet., vol. 11, 2021, Art. no. 632901.

[32] Y. Lin et al., "Classifying breast cancer subtypes using deep neural networks based on multi-omics data," Genes, vol. 11, no. 8, 2020, Art. no. 888