Integration of Genomic Data into Electronic Health Records for Enhanced Clinical Decision Support: Past, Present, & Future

Sanhita Chundury, Shreya Jain, Sahil Haresh Ratwani

Submitted to: Journal of Medical Internet Research
on: February 19, 2025

Table of Contents

Integration of Genomic Data into Electronic Health Records for Enhanced Clinical Decision Support: Past, Present, & Future

Sanhita Chundury[1*] BSA; Shreya Jain[1*] BS; Sahil Haresh Ratwani[1*] PharmD

[1] Weill Cornell Medicine New York US
[*]these authors contributed equally

Corresponding Author:
Sanhita Chundury BSA

Weill Cornell Medicine
402 East 67th Street
New York
US

Abstract

Background: Integrating genomic data into electronic health records (EHRs) and clinical decision support (CDS) systems marks the beginning of a new era of precision medicine. Leveraging tailored insights from patient genomic data opens the door for patient care to have enhanced diagnosis, personalized treatment plans, and optimized medication management.

Objective: This systematic review aims to examine the evolution of genetic data integration into Electronic Health Records by analyzing past developments, current implementation strategies, and future opportunities. It seeks to review historical milestones, identifying key advancements in genomic data collection, storage, and early integration efforts. It assesses current clinical practice and evaluates existing technologies and policies that shape the standards of genomic data interoperability. It also aims to highlight the technical, ethical, and regulatory challenges that hinder effective genomic integration and propose potential strategies and recommendations to address them.

Methods: This systematic review follows PRISMA 2020 guidelines to ensure transparency and reproducibility. A comprehensive search was completed on the PubMed database with a predefined query and relevant inclusion and exclusion criteria.

Results: A total of 48 articles were reviewed, with 27 included in the final review. Identified themes included technical integration, clinical utility, precision medicine, and human genome sequencing. These findings emphasize the role of AI, multi-omics data, APIs, and current work in maximizing the role of genetic data in clinical practice through integration in Electronic Health Records.

Conclusions: The integration journey is riddled with challenges such as data standardization, interoperability, data privacy, and resource constraints, which hinder the effective utilization of genomic information. This systematic review explores historical milestones, current practices, and future implementations in genomic data integration. It emphasizes the potential of emerging technologies like artificial intelligence, multi-omics data, next-generation sequencing, APIs, and automation to bridge existing gaps. Genomic integration can revolutionize healthcare delivery by harnessing these technologies, ensuring equitable and improved patient outcomes.

(JMIR Preprints 19/02/2025:72864)
DOI: https://doi.org/10.2196/preprints.72864

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Integration of Genomic Data into Electronic Health Records for Enhanced Clinical Decision Support: Past, Present, & Future

I. Abstract

Background:
Integrating genomic data into electronic health records (EHRs) and clinical decision support (CDS) systems marks the beginning of a new era of precision medicine. Leveraging tailored insights from patient genomic data opens the door for patient care to have enhanced diagnosis, personalized treatment plans, and optimized medication management.
Objectives:
This systematic review aims to examine the evolution of genetic data integration into Electronic Health Records by analyzing past developments, current implementation strategies, and future opportunities. It seeks to review historical milestones, identifying key advancements in genomic data collection, storage, and early integration efforts. It assesses current clinical practice and evaluates existing technologies and policies that shape the standards of genomic data interoperability. It also aims to highlight the technical, ethical, and regulatory challenges that hinder effective genomic integration and propose potential strategies and recommendations to address them.
Methods:
This systematic review follows PRISMA 2020 guidelines to ensure transparency and reproducibility. A comprehensive search was completed on the PubMed database with a predefined query and relevant inclusion and exclusion criteria.
Results:
A total of 48 articles were reviewed, with 27 included in the final review. Identified themes included technical integration, clinical utility, precision medicine, and human genome sequencing. These findings emphasize the role of AI, multi-omics data, APIs, and current work in maximizing the role of genetic data in clinical practice through integration in Electronic Health Records.
Conclusions:
However, the integration journey is riddled with challenges such as data standardization, interoperability, data privacy, and resource constraints, which hinder the effective utilization of genomic information. This systematic review explores historical milestones, current practices, and future implementations in genomic data integration. It emphasizes the potential of emerging technologies like artificial intelligence, multi-omics data, next-generation sequencing, APIs, and automation to bridge existing gaps. Genomic integration can revolutionize healthcare delivery by harnessing these technologies, ensuring equitable and improved patient outcomes.

Keywords: Genomic Data, Interoperability, Clinical Decision Support, EHRs
Abbreviations:

- EHR: Electronic Health Record
- CDS: Clinical Decision Support
- NGS: Next-Generation Sequencing
- AI: Artificial Intelligence
- PRISMA: Preferred Reported Items for Systematic Reviews and Meta-Analyses
- API: Application Programming Interface

Data Availability: The data used in this study were obtained from publicly available sources. Further details can be provided upon request.

II. Introduction

Electronic health records (EHRs) and clinical decision support (CDS) systems represent the forefront

of modern healthcare advancement, facilitating the organization and application of patient data for evidence-based decision-making. The rapid adoption of such technologies signifies a substantial step toward precision medicine, where treatments are tailored to specific patient characteristics. Despite the demonstrated benefits of EHRs and CDS, their potential remains underutilized without integrating genomic data.

Genomic data encompasses information derived from an individual's DNA, such as mutations and gene variations that influence health risks, traits, and treatment responses. This data is essential to patient outcomes as it helps clinicians understand an individual's predisposition to diseases, their reactions to specific treatments, and potential drug interactions or adverse effects. Clinicians' ability to effectively use this information in clinical practice relies on the seamless availability of genomic data in EHRs and CDS systems.

This systematic review examines the past, present, and future of integrating genomic data into EHRs for CDS systems, providing a comprehensive perspective of how genomic data is transformative in enhancing personalized medicine. By analyzing different stages of integration, it identifies previous successes and failures, highlighting gaps, limitations, and opportunities for innovation.

III. Methods

The methodology our systematic review adheres to is the PRISMA 2020 guidelines to ensure transparency and reproducibility. A comprehensive search was conducted in the PubMed database using the following query:

> ("Electronic Health Records"[mesh] OR "EHR"[tw]) AND (“Clinical decision support”[tw] OR "clinical decision support systems"[tw] OR "CDSS"[tw]) AND ("genomic data"[tw] OR "genomic data integration"[tw] OR “genomic information”[tw] OR “genetic information”[tw] OR “genetic data”[tw])

We targeted articles published from 2016 onward to focus on advancements following the adoption of next-generation sequencing (NGS) and the emergence of artificial intelligence in genomics, ensuring the inclusion of studies incorporating the latest technologies. Articles were included if they described genomic data integration into EHRs, its impact on pharmacogenomic treatment or CDS, and outcomes such as treatment efficacy or adverse drug events. Eligible study designs included randomized controlled trials, cohort studies, case-control studies, systematic reviews, and meta-analyses, while exclusion criteria removed studies unrelated to EHR-based genomic integration or clinical applications.

A total of 48 articles were retrieved from the query. Two were excluded after title and abstract screening for apparent irrelevance, and seven were excluded after full-text review due to limitations such as insufficient data, inappropriate study design, irrelevance to the research question, or excessive bias. Ultimately, 27 articles were included in the review, categorized into themes focusing on technical integration, clinical utility, and their implications for the study's overarching research question. Data extraction followed a standardized approach, prioritizing systematic reviews and high-ranking articles in the search results to maximize relevance and objectivity. These 27 articles were reviewed in-depth to identify key features and relevance metrics, discussed later in the paper to address the study's research question comprehensively.

IV. Results

A. Historical Milestones

Integrating genomic data into healthcare is driven by the more significant trend and focus on precision medicine. Precision medicine aims to provide personalized treatments based on patient characteristics, underscoring that treatments have a differential outcome based on a patient's genetic profile. Precision medicine's rise in popularity can be manifested in federal policies like the US Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, where the importance of genomic integration into EHRs in precision medicine is a key area of focus. Similarly, Turkey's Health Information Program, set in place in 2003, aimed to improve the integration of genetic data and the interoperability of health data exchange, demonstrating a global trend towards genetic data integration in EHRs and application in medical practice [12].

Current approaches to genomic recommendations are primarily manual and case-specific, making them time-inefficient and unsustainable as genetic data increases with the widespread implementation of whole-genome sequencing. The American College of Medical Genetic and Genomics (ACMG) identified 73 actionable genes linked to 71 conditions over various major drug interaction classes, including metabolic, neoplastic, and structural interactions. The high variability and complexity underscores the need for an automated, scalable way to integrate genomic data and efficiently map the data to automated, scalable clinical decision support systems for physicians [7].

However, as with any nascent implementation, early challenges exist in implementing genetic data into EHRs. From a policy perspective, challenges have risen due to uncoordinated privacy regulations and a need for more standardization. In the UK, privacy regulations are "upstream," focusing on maintaining privacy during data collection. On the other hand, in the US, "downstream" privacy regulations focus on disclosure rights after data is collected and stored. This difference creates systematic barriers to the smooth implementation of genomic data into EHRs. Other policies that underscore the complexity of genomic data integration include GINA, the Genetic Information Nondiscrimination Act of 2008. GINA is designed to prevent genetic discrimination in health insurance and employment by prohibiting health insurers from using genetic data to determine coverage or premiums and restricting employers from making employment decisions based on genetic information [15].

Other efforts to address the need for a standardization of genetic information exchange include the introduction of HL7 v3 and SNOMED in 2018, aimed at standardizing medical and administrative data exchange globally. Overall, the primary driver for automation is the need for a standardized knowledge base and ontology, similar to the existing drug-drug interaction system standard. This will allow a consistent, equitable, bias-free, and scalable application for genomic recommendations.

B. Current Practices

Today, current integration practices for genomic data occur through a variety of methods. While some practices are still in developmental stages, others have undergone significant refinement and adaptation to align with the ever-changing technological and healthcare standards. The storage and retrieval of genomic data aims to ensure that information is accessible for clinical decision-making and the ability to use the information for research and population health management. In many healthcare settings, genomic test results, such as multiplex panels, are reported in PDF format or physically mailed and faced to providers. Such approaches limit the usability of genomic data for CDS or secondary research purposes. As described in *Integrating cancer genomic data into electronic health records*, written by Warner, Jain, and Levy, in cancer care, many molecular signatures and molecular profiling reports are scanned as PDFs within EHR systems or are kept as copies in patient files. This prevents a seamless electronic integration with EHRs or into clinical

workflows [3].

Despite these limitations, an initiative known as eMERGE III has made significant strides in incorporating actionable clinical genetic findings. Funded by the National Human Genome Research Institute (NHGRI), eMERGE III focuses on genetic variants that influence medication dosing, disease risk management, and preventative health strategies. Linking genetic data with phenotypic information from EHRs paves the way for advancing personalized medicine. To date, eMERGE III has identified 35 genes strongly associated with diseases such as hereditary cancer syndromes, cardiovascular conditions, and pharmacogenomics-related traits. eMERGE III prioritizes genetic variants in genes such as BRCA1/2 linked to hereditary breast and ovarian cancer. Identifying these mutations allows for targeted interventions in individuals who may also present similar genetic variants. Similarly, loss-of-function variants in the CYP2C19 gene can mitigate medication efficacy and inform alternative treatment and prevention strategies, ensuring clinicians receive actionable alerts to improve patient outcomes.

Beyond the efforts of eMERGE III, other initiatives highlight how genomic data is integrated and utilized in healthcare today. One exemplary case is the implementation of the PennChart system at Penn Medicine for patients with Lynch Syndrome (LS), an inherited cancer syndrome. LS increases the risk of colorectal, endometrial, and other cancers, which makes it necessary for regular surveillance and follow-up. The system integrates genetic data, enabling the automated application of genomic indicators for LS, such as variants in MLH1, MSH2, or MSH6. The tool includes clinician-facing alerts embedded into EHRs, informing providers with automated notifications of when patients are due for specific surveillance activities such as colonoscopies or genetic program visits. Initial results showed high sensitivity (96.4%) and specificity (91.0%) in identifying patients due for colonoscopic surveillance. The adherence to surveillance was shown to have a positive association with the use of reminders on the clinician and patient portals. The system’s automated updates ensure up-to-date CDS recommendations, improving adherence to cancer risk management guidelines and life-saving surveillance activities [5].

C. Architecture

To integrate genomic data into electronic health records (EHRs), sophisticated architectural frameworks are required to manage the amount and complexity of genomic data, guarantee interoperability, and enable reliable clinical decision support (CDS) systems. Efforts over the last few decades have focused on several key architectural pipelines.

Scalable storage solutions for genomic test data are critical databases tailored for genomic data to guarantee adequate data storage and retrieval, following guidelines such as those set forth by the Human Genome Variation Society (HGVS) for variant annotation. For instance, the eMERGE III network harmonized clinical sequencing data across multiple sites, improving coordination and delivery for clinical use [1].

Standardization is integral for interoperability. The Fast Healthcare Interoperability Resources (FHIR) provides a standardized framework for efficient data exchange across EHR and CDS systems [21]. Lau-Min et al. developed a CDS tool that leverages EHR platforms and links them with genomic indicators for Lynch Syndrome, as introduced earlier, which improves surveillance and medication adherence [10]. The FHIR Genomics Operations captures the complexity and heterogeneity of genomic data, simplifying the development of sophisticated architectural frameworks. Robert et al. showcased a successful attempt to convert vcf files generated from sequencing results to a standardized FHIR format by developing a tool called ‘vcf2fhir’[22]. The tool converts vcf variants into HL7 FHIR format reports. It works on simple variants like SNVs, MNVs,

and Indels, along with zygosity and phase relationships for autosomes, sex chromosomes, and mitochondrial DNA. Despite the tool's limitations, it shows great promise in translating genomic data into a standardized format. This tool can be incorporated into CDS pipelines.These elements of an architectural framework need to be effectively incorporated into a clinical workflow in a way that not only aids clinical decision-making but also enhances the efficacy, quality, and safety of patient care.

Additionally, due to the size and complexity of genomic data, ancillary genomic systems have been developed that can manage and process genetic information separately from the central EHR systems using specialized optimized solutions. The PennChart Genomics Initiative (PGI) illustrates how multi-disciplinary collaboration enables the integration of both germline and somatic data, providing a framework for delivering precision medicine. The PGI linked over 420 unique computerized genetic testing orders, reducing ordering times by 6 minutes [16].

D. CDS Enhancements

Integrating genomic data into CDS systems marks a transformative step in precision medicine, leveraging genomic insights within EHRs for accurate diagnoses, personalized treatment plans, and optimized prescription behavior while minimizing the risk of drug-drug interactions. A notable example of this integration is outlined by Dolin, Boxwala, and Shalaby, who describe pharmacogenomics (PGx)-specific CDS service architecture, which effectively incorporates genetic information to improve clinical prescription practices.

At the core of the PGx CDS service lies FHIR (Fast Healthcare Interoperability Resources), a standard that facilitates seamless data exchange between EHRs and CDS platforms. FHIR enables the integration of heterogeneous data from diverse health systems, ensuring that the genetic information is incorporated into the clinical workflow. The PGx CDS service is triggered through CDS Hooks. This standard connects clinical workflow tasks– medication, patient view, and order review, to external CDS services. This event-driven approach ensures that genetic data is considered at the point of care during clinically relevant tasks.

When the CDS service is triggered, the EHR sends an FHIR medication resource containing the RxNorm code for the prescribed drug. The CDS system queries relevant knowledge bases to retrieve standardized drug information, including any potential genetic interactions with the medication. The service analyzes the patient's genetic information for drugs in the PGx medication value sets. The CDS interfaces with patient data access APIs to retrieve pertinent patient details such as genetic variants and alleles from the Genomic Archiving and Communication System (GACS). GACS provides information on gene variants and the associated alleles, which can be queried for genotype data. Allele-variant conversion rules are applied to ensure accurate mapping of genetic data to actionable star alleles.

Recommendations provided by the system are based on the CPIC Level A guidelines, which govern the evaluation of drug-gene interactions. Based on these guidelines, the system assesses whether the patient's genetic data can be used to predict the optimal medication response. This ensures that the prescribed medication suits the patient's genetic profile. The PGx CDS service returns its CDS Hook Information findings to the EHR. It appears as a dialog box before the clinician completes the medication order, presenting actionable recommendations. For example, it suggests adjusting the medication dosage based on the patient's genetic profile or suggesting an alternative medication if necessary [8].

This system offers several advantages further bolstered by the ubiquity of PGx data. However,

several challenges must still be addressed to implement genetic data into CDS systems successfully. A key issue is the lack of standardization in the semantics and interoperability of FHIR and CDS Hooks, which complicates the seamless integration of PGx data into healthcare systems. Additionally, the human genome variation society nomenclature (HGVS) presents significant challenges in its application within CDS systems. There are multiple valid representations for a single gene variant, leading to inconsistencies in the interpretations of a specific gene. The technical complexity of HGVS nomenclature also requires specialized knowledge, creating a barrier to widespread adoption and usability amongst clinicians. Finally, there is an additional challenge along the theme of interoperability, specifically the lack thereof in fully integrating the GACS architecture into EHRs [9].

In a related study, Nielsen et al. (2020) examined a similar approach to integrating genetic data into CDS systems, similarly working to identify preventable adverse events (PAEs) caused by drug-gene interactions. This system also leveraged the CPIC Level A guidelines to flag potential PAEs, allowing for the identification of individuals at higher risk for adverse drug reactions (ADRs). Integrating genetic data into CDS systems significantly improves the predictive accuracy of machine learning models, underscoring the potential of pharmacogenomics in clinical decision support for personalized treatment [17].

E. Future Implementations

The future of genomic data integration into electronic health records (EHRs) and clinical decision support (CDS) systems is driven by advancements in artificial intelligence (AI), machine learning (ML), and next-generation technologies. These key technologies drive the enhancement of the quality of care by making genomic data more accessible, actionable, and accurate for clinical decision-making.

AI and Machine Learning in CDS for Genomic Data

AI and machine learning are revolutionizing industries, and their application to CDS systems improves the ability to analyze complex genomic datasets. One promising application is transfer learning in natural language processing (NLP), where deep learning models can be fine-tuned for clinical tasks. A prominent example is Bidirectional Encoder Representations from Transformers (BERT), a model developed by Google AI. BERT is trained on clinical data from the MIMIC-III database using clinical notes to process unstructured clinical text, such as discharge summaries, clinical notes, labs, and patient histories. By integrating BERT into EHR systems, artificial intelligence healthcare providers can automatically transform unstructured, free-text data into structured information so that it can trigger real-time alerts or recommendations for patient care. For instance, notes documenting genetic test results can be processed to highlight any genetic risks or suggest alternative therapies based on the patient's profile, improving the speed and accuracy of decision-making [4].

Multi-omics Data Integration for Patient Profiles

Another advancement is integrating multi-omics data into CDS systems, enabling a more holistic understanding of patient health. "Omics" refers to the large-scale study of biological molecules that make up an organism. It can include various disciplines such as genomics, proteomics, epigenomics, or microbiomics. The integration of these diverse data types enhances personalized medicine. We can help identify disease subtypes by examining the interaction between genetic variants (genotype) and observable traits (phenotype) by analyzing data from various biological layers. For instance, in cancer patients, a multi-omics approach can help identify subtypes of cancers that may not be visible only through genetic data alone. Genomics, proteomic, and metabolic data can work together to help identify patterns in patient health that can predict responses to specific therapies.

For example, a BRCA1 mutation associated with breast cancer can impair the BRCA1 protein, which is responsible for DNA repair, leading to increased oxidative stress. This metabolic change represents a distinct molecular subtype of breast cancer. CDS systems can use genomic data to identify the specific BRCA1 mutation, confirm its impact on protein function using proteomics, and recognize the associated metabolic signatures using metabolomics–providing actionable multi-omics-driven recommendations tailored to the disease at a molecular level with superior precision [4].

Advancements in Next-Generation Sequencing (NGS)

Next-generation technologies have already transformed genomic medicine, and further advancements are projected to enhance the precision and utility of genomic data in clinical settings. DeepVariant, an analysis pipeline tool developed by Google AI, uses deep neural networks to improve the accuracy of calling genetic variants from sequencing data. Integrating such a high-fidelity variant into EHRs can enhance the quality of genomic data available to healthcare providers. DeepVariant can accurately identify variations, such as single nucleotide polymorphisms (SNPs) and insertions/deletions in DNA sequences. It takes input from a high-throughput DNA sequencing platform, typically as a binary alignment map (BAM) file. It converts the sequencing data into images where each image represents a genomic locus, reference sequence, aligned reads, and base qualities. These images are analyzed by a convolutional neural network (CNN), trained to classify each genomic position by matching the reference genome or if it contains a variant. This model then outputs a variant cell format (vcf) file identifying variants with performance that is often comparable to or better than that of human experts. By ensuring more accurate information on genetic variants, DeepVariant and other NGS tools help clinicians make better-informed decisions in precision medicine, where slight differences in genetic variants can significantly improve patient outcomes and health [8]. As NGS continues to evolve, integration into EHRs will reduce costs and turnaround times, making them more accessible across healthcare systems. This will allow clinicians to regularly use genomic data to diagnose diseases and improve treatment recommendations for cancers, rare genetic disorders, and cardiovascular diseases [4].

APIs and Middleware for Data Sharing

APIs and middleware solutions are becoming increasingly common in addressing the challenges of fragmented genomic data across healthcare systems. APIs facilitate the exchange and usage of genomic data across various systems, applications, and platforms. It allows seamless communications between systems, allowing healthcare professionals to incorporate genomic information into their healthcare, research, and clinical decision-making workflows. APIs such as NCBI E-utilities and Ensembl REST API provide access to genomic datasets, enabling retrieval of sequences, variants, and annotations. SMART on FHIR Genomics API and HL7 FHIR Genomics Module are frameworks embedded into EHRs for genomic insights into FHIR-compliant EHR systems. APIs support interoperability and facilitate integration while automating data analysis and retrieval to reduce manual effort. By bridging the gap between raw data and clinical application, APIs are a stepping stone in incorporating genomic data into patient profiles [21].

Middleware acts as an intermediary layer between systems, translating, standardizing, and orchestrating workflows. Using standardized schemas like HL7 FHIR, OMOP, or vcf, middleware ensures communication between systems with different protocols or formats, such as REST APIs, HL7 messages, or database formats. Solutions, such as standalone Foundation Medicine and Caris Life Science portals offer middleware that integrates genomic profiling data into clinical workflows, supporting precision oncology through actionable insights [3].

Automating Genomic Recommendations in CDS

As genomic data is integrated into EHR systems, precision medicine is moving toward automating genomic recommendations directly embedded into clinical decision support systems. Such recommendations are implemented manually, relying on clinicians' memory, guidelines, or consultations with other specialists. However, with the expansive volume and complexity of genomic data, it is becoming increasingly difficult for clinicians to make decisions informed by all relevant data, underscoring the need for an automated solution. Genomic data recommendations are grouped into five categories: cardiovascular, metabolic, neoplastic, medication interactions, and structural conditions. The grouping ensures that recommendations are consistent and care is equitable amongst all patient populations. Creating an automated system for clinical recommendations is a way to implement a centralized knowledge base to apply genetic insights that can assist with the prevention of disease manifestations, surveillance, and treatment management [7].

Another innovative approach is the Personalized Medicine-Therapy Optimization Model, more commonly known as PM-TOM, which is a clinical decision support system that uses data from the Electronic Medical Record Repository of the Personal Genome Project, DrugBank, and the Comprehensive Toxicogenomics Database. It employs a heuristic algorithm to analyze complex interactions between drugs, genes, and medical conditions. The algorithm outputs a Comprehensive Adverse Drug Interaction (CADI) indicator that helps minimize potential adverse interactions and create more personalized treatment plans. By incorporating PM-TOM into CDS systems, healthcare providers can automate identifying adverse interactions and create more precise treatment plans [23].

By incorporating automation into CDS systems, clinicians can quickly generate appropriate recommendations, even before a patient has undergone genetic counseling. For instance, if a patient's genetic profile reveals a variant associated with Marfan Syndrome in the FBN1 gene, the system can automatically detect the variant and recommend that the clinician inform the patient to undergo regular echocardiograms and consider pharmacological interventions, such as beta-blockers. This information would seamlessly integrate into the patient profile, allowing the clinician to review and act promptly. The entire process would operate in the background as the patient's genomic data is incorporated into their record. Such automation reduces the burden on the clinician's memory and supports the efficient and accurate identification of potential conditions and treatment options. These advancements represent a pivotal step toward a healthcare system capable of leveraging real-world data to enhance and personalize patient care [7].

F. Challenges

This analysis of genomic data integration demonstrates the potential of translating genetic insights into routine clinical care. Programs like eMERGE III and systems like PennChart highlight how genetic insights can be utilized to improve precision medicine. These initiatives ensure clinicians and patients are better equipped to manage severe conditions and illnesses, such as cancers or pharmacogenomic traits. Despite these advances with EHRs and clinical decision support systems, genomic data in healthcare records is still an evolving concept. While the standards set in stone have facilitated significant progress, many institutes still need to rely on unstructured formats of genomic data such as lab results as scanned results or PDFs. This, ultimately, hinders the broader utility of genomic data. These advancements are not without their challenges. Limitations related to data standardization, interoperability, privacy, and resource constraints continue to challenge the seemingly seamless integration of genomic data into medical records. We explain each of the significant limitation realms below:

Data Standardization

One of the primary obstacles is the need for standardized formats to represent genomic data. As noted, many genomic test results are delivered as unstructured PDFs or scanned documents, which limits their ability to contribute to EHRs or CDS systems. This lack of uniformity makes it challenging to leverage available genomic data. Although HL7 FHIR, LOINC, and SNOMED-CT have been developed to standardize healthcare data, their adoption across healthcare systems must be more consistent. Without standardization, genomic data cannot easily trigger actionable alerts or recommendations within CDS systems [13].

Interoperability

The lack of standardization is closely tied to interoperability challenges surrounding genomic data in EHRs. Many systems store genomic data in unstructured formats or separate systems, creating data silos that hinder healthcare providers from accessing a complete patient profile. This fragmentation limits their ability to leverage genomic data in real-time decision-making. Even with the implementation of interoperability frameworks like FHIR, genetic data encoding remains inconsistent across institutions. This emphasizes the need for standardized, consistent approaches to ensure true interoperability [19].

Privacy

Genomic data presents unique privacy concerns due to its sensitive nature. While all medical records require protection, genetic information carries implications for family members and future generations. Breaches in data can result in significant ethical and legal implications and ensuring compliance with regulations like HIPAA is an ongoing challenge. While encryption and consent mechanisms are in place to protect genomic data, more is needed, given the scale of integration into healthcare systems [5].

Resource Constraints

Resource constraints are a prominent barrier that affects various aspects of the implementation, adoption, and utilization of genomic data in EHRs. Genomic testing, infrastructure development, and the workforce needed to support genomic testing are impacted by resource availability. Advanced genomic tests, such as next-generation sequencing (NGS), can be expensive. For example, tumor-normal comparisons to identify mutations can double the cost of testing. These expenses, along with the lack of insurance reimbursement, leave healthcare providers and patients to bear the burden of the finances. This financial strain can exacerbate disparities, particularly in underserved populations and smaller healthcare facilities that cannot afford such testing. Additionally, workforce training for genomic testing has proven challenging, as seen with Penn Medicine's LS PennChart system. Even well-funded institutions like Penn Medicine faced difficulties implementing the system, highlighting the barriers smaller institutions may face in adopting such technologies [3].

V. Discussion

Integrating genomic data into EHR systems and enabling the development of CDS tools marks a transformative leap in healthcare, paving the way for personalized medicine through precision diagnostics and treatment. This review explores the significant milestones in integrating genomic data into EHRs with initiatives like eMERGE III and PennChart that exemplify practical applications. The development of these systems has advanced real-time clinical decision support, enabling functionalities like recommending alternative medications based on genetic profiles or improving medication adherence in specific disease groups. Seamlessly integrating genomic insights with clinical workflows will demonstrate the potential it contains and enhance patient outcomes.

Standardized frameworks like HL7 FHIR have been crucial in advancing integration, allowing for more efficient and interoperable data exchange across various healthcare systems. However, the reliance on unstructured data formats limits the full potential of genomic data in clinical workflows. Addressing these limitations involves integrating both structured and unstructured data into EHRs. Once achieved, this integration has the potential to enhance the utility of CDS systems, as evidenced by the notable improvements seen in medication and surveillance activity adherence when genomic data is effectively utilized.

Despite the advancements, several challenges continue to hinder the seamless integration of genomic data into EHRs. A primary obstacle remains to be data standardization, with many genomic test results still being reported in PDF format or as scanned documents. This lack of standardization leads to interoperability issues, creating data silos that prevent healthcare providers from accessing a complete patient profile in real time and make it challenging to automate clinical workflows. Privacy concerns are also paramount, given the sensitive nature of genetic information, which can have implications for those related to the genetic data. Robust encryption protocols and consent mechanisms are of utmost importance to protect this data, especially as the scale of genomic data integration grows. Additionally, resource constraints– such as the high cost of genomic testing and the need for specialized workforce training–pose significant barriers, particularly for smaller healthcare facilities or underserved populations.

Integrating artificial intelligence (AI) and machine learning (ML) into CDS systems holds immense promise. Technologies like BERT for natural language processing can convert unstructured clinical text into actionable insights, enhancing the speed and accuracy of decision-making. The potential for multi-omics data integration, combining genomic data with proteomics, epigenomics, and other biological data, offers a holistic view and understanding of patient health, enabling highly personalized treatment strategies. Next-generation sequencing (NGS) technologies, such as DeepVariant, are expected to refine the accuracy of genetic variant calling, making genomic data more reliable and actionable. The development of APIs and middleware solutions will continue to bridge the gap between disparate systems, facilitating real-time data exchange and enhancing the utility of genomic information in clinical settings. By automating clinical workflows with the help of genomic recommendations, we can alleviate the burden on clinicians, fostering a shift towards more proactive and precise patient care.

As genomic data integration into clinical decision support systems continues to evolve and become more prevalent, it becomes necessary to evaluate healthcare policies addressing privacy and security concerns thoroughly. Policies like the GINA in the U.S. and similar global regulations need to evolve alongside the advancements in genomic technology. The evolution of policies and rules must balance privacy protections and responsible usage of such sensitive data. Standardization efforts, such as adopting HL7 FHIR frameworks, should be encouraged and implemented universally to provide a seamless flow of genomic information. Concurrently, new educational initiatives are essential to equip healthcare providers with the knowledge and skills to interpret genomic data effectively and adopt these new frameworks. Financial barriers to genomic testing also require policy interventions to ensure that technologies are accessible to all. Strategies such as insurance coverage or subsidies can help to prevent deepening health disparities and mitigate costs. Looking ahead, the future of healthcare will see a shift towards a learning health system where genomic data continuously informs and refines clinical care. To successfully achieve this, it is necessary to adopt a collaborative approach between technology developers, healthcare providers, and policymakers to put down the foundation of precision medicine and its transformative potential.

VI. References

1. eMERGE Consortium. Electronic address: agibbs@bcm.edu. and eMERGE Consortium. “Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network.” American journal of human genetics vol. 105,3 (2019): 588-605. doi:10.1016/j.ajhg.2019.07.018
2. Freimuth, R R et al. “Implementing Genomic Clinical Decision Support for Drug-Based Precision Medicine.” CPT: pharmacometrics & systems pharmacology vol. 6,3 (2017): 153-155. doi:10.1002/psp4.12173
3. Warner, Jeremy L et al. “Integrating cancer genomic data into electronic health records.” Genome medicine vol. 8,1 113. 26 Oct. 2016, doi:10.1186/s13073-016-0371-3
4. Krumm, Niklas, and Brian H Shirts. “Technical, Biological, and Systems Barriers for Molecular Clinical Decision Support.” Clinics in laboratory medicine vol. 39,2 (2019): 281-294. doi:10.1016/j.cll.2019.01.007
5. Sitapati, Amy et al. “Integrated precision medicine: the role of electronic health records in delivering personalized treatment.” Wiley interdisciplinary reviews. Systems biology and medicine vol. 9,3 (2017): 10.1002/wsbm.1378. doi:10.1002/wsbm.1378
6. Lin, Jing, and Kee Yuan Ngiam. “How data science and AI-based technologies impact genomics.” Singapore medical journal vol. 64,1 (2023): 59-66. doi:10.4103/singaporemedj.SMJ-2021-438
7. Murugan, Mullai et al. “Genomic considerations for FHIR®; eMERGE implementation lessons.” Journal of biomedical informatics vol. 118 (2021): 103795. doi:10.1016/j.jbi.2021.103795
8. Dolin, R H et al. “A Pharmacogenomics Clinical Decision Support Service Based on FHIR and CDS Hooks.” Methods of information in medicine vol. 57,S 02 (2018): e115-e123. doi:10.1055/s-0038-1676466
9. Cimino, James J. “Planning for Actionable Precision Medicine.” Studies in health technology and informatics vol. 310 (2024): 244-248. doi:10.3233/SHTI230964
10. Lau-Min, Kelsey S et al. “Development of an Electronic Health Record-Based Clinical Decision Support Tool for Patients With Lynch Syndrome.” JCO clinical cancer informatics vol. 7 (2023): e2300024. doi:10.1200/CCI.23.00024
11. Rasmussen, Luke V et al. “Practical considerations for implementing genomic information resources. Experiences from eMERGE and CSER.” Applied clinical informatics vol. 7,3 870-82. 21 Sep. 2016, doi:10.4338/ACI-2016-04-RA-0060
12. Şık, Ayhan Serkan et al. “Assessing the readiness of Turkish health information systems for integrating genetic/genomic patient data: System architecture and available terminologies, legislative, and protection of personal data.” Health policy (Amsterdam, Netherlands) vol. 125,2 (2021): 203-212. doi:10.1016/j.healthpol.2020.12.004
13. Johnson, Darren et al. “Genetically guided precision medicine clinical decision support tools: a systematic review.” Journal of the American Medical Informatics Association : JAMIA vol. 31,5 (2024): 1183-1194. doi:10.1093/jamia/ocae033
14. Brunner, Mauricio et al. “Bioinformatics Architecture for Integrating Genomics Data into Electronic Health Records.” Studies in health technology and informatics vol. 310 (2024): 996-1000. doi:10.3233/SHTI231114
15. Mascia, Cecilia et al. “OpenEHR modeling for genomics in clinical practice.” International journal of medical informatics vol. 120 (2018): 147-156. doi:10.1016/j.ijmedinf.2018.10.007
16. Wiley, Ken et al. “A research agenda to support the development and implementation of genomics-based clinical informatics tools and resources.” Journal of the American Medical Informatics Association : JAMIA vol. 29,8 (2022): 1342-1349. doi:10.1093/jamia/ocac057
17. Kulchak Rahm, Alanna et al. “User testing of a diagnostic decision support system with machine-assisted chart review to facilitate clinical genomic diagnosis.” BMJ health & care informatics vol. 28,1 (2021): e100331. doi:10.1136/bmjhci-2021-100331

18. Hoffman, James M et al. “Developing knowledge resources to support precision medicine: principles from the Clinical Pharmacogenetics Implementation Consortium (CPIC).” Journal of the American Medical Informatics Association : JAMIA vol. 23,4 (2016): 796-801. doi:10.1093/jamia/ocw027
19. Zebrowski, Alexis M et al. “Qualitative study of system-level factors related to genomic implementation.” Genetics in medicine : official journal of the American College of Medical Genetics vol. 21,7 (2019): 1534-1540. doi:10.1038/s41436-018-0378-9
20. Kidwai-Khan, Farah et al. “Pharmacogenomics driven decision support prototype with machine learning: A framework for improving patient care.” Frontiers in big data vol. 5 1059088. 15 Nov. 2022, doi:10.3389/fdata.2022.1059088
21. Dolin, Robert H et al. “Introducing HL7 FHIR Genomics Operations: a developer-friendly approach to genomics-EHR integration.” Journal of the American Medical Informatics Association : JAMIA vol. 30,3 (2023): 485-493. doi:10.1093/jamia/ocac246
22. Dolin, Robert H et al. “vcf2fhir: a utility to convert VCF files into HL7 FHIR format for genomics-EHR integration.” BMC bioinformatics vol. 22,1 104. 2 Mar. 2021, doi:10.1186/s12859-021-04039-1
23. Kulenovic, Adnan, and Azra Lagumdzija-Kulenovic. “Using PM-TOM for the Minimization of Adverse Drug and Gene Interactions in Therapies for Common Multi-Diseases.” Studies in health technology and informatics vol. 272 (2020): 205-208. doi:10.3233/SHTI200530
24. Williams, Marc, et al. Evaluating the Impact of a User-Friendly Report of Genetic Testing Results for Parents of Children with Genetic Disorders. Patient-Centered Outcomes Research Institute (PCORI), September 2018. doi:10.25302/9.2018.CD.13046987
25. Danahey, Keith et al. “Simplifying the use of pharmacogenomics in clinical practice: Building the genomic prescribing system.” Journal of biomedical informatics vol. 75 (2017): 110-121. doi:10.1016/j.jbi.2017.09.012
26. Floyd, James S, and Bruce M Psaty. “The Application of Genomics in Diabetes: Barriers to Discovery and Implementation.” Diabetes care vol. 39,11 (2016): 1858-1869. doi:10.2337/dc16-0738
27. Stellmach, Caroline et al. “Creation of a structured molecular genomics report for Germany as a local adaption of HL7's Genomic Reporting Implementation Guide.” Journal of the American Medical Informatics Association : JAMIA vol. 30,6 (2023): 1179-1189. doi:10.1093/jamia/ocad061