HIDDEN IN PLAIN SIGHT: IDENTIFICATION OF GENOMIC INFORMATION WITHIN EHR ALLERGY DOCUMENTATION

By: Joey Milbrandt; PharmD Candidate;
SIUE School of Pharmacy

ABSTRACT

Pharmacogenomics (PGx) can improve medication safety and efficacy, but its impact is limited by how genomic data are stored within electronic health records (EHRs). This study evaluated the presence of genomic information within EHR allergy documentation. A retrospective analysis of 223 EHR allergy records was conducted using keyword-based identification of genomic indicators within free text reaction fields. Identified data were categorized into clinically relevant groups, including G6PD deficiency and pseudocholinesterase deficiency. Genomic indicators were present in 78.9% of records (n = 176), most commonly G6PD deficiency (36.9%) and pseudocholinesterase deficiency (23.3%). These were frequently associated with high-risk medications such as succinylcholine, sulfa antibiotics, and opioids. Demographic variation was also observed, particularly among Black or African American patients. Despite their prevalence, genomic data were inconsistently documented and embedded in unstructured text, limiting clinical usability. These findings support the need for structured genomic data integration within EHR systems to enhance clinical decision-making and precision medicine.

Introduction

The Growing Role of Pharmacogenomics in Clinical Care

Pharmacogenomics (PGx) is a rapidly growing aspect of medicine specifically related to its importance within precision medicine. Precision medicine involves the use of the PGx data to provide safer and more effective medication treatment by considering the genetic differences that can lead to variations in drug metabolism, patient response, and drug toxicity between patients. This allows clinicians to improve outcomes and reduce adverse drug events. Clinical guidelines developed by the Clinical Pharmacogenetics Implementation Consortium (CPIC) helps convert genetic test results into prescribing recommendations. This creates a usable pathway for clinicians to incorporate genomic data into routine care (Caudle et al., 2014).

Although the idea seems nice, actual implementation of PGx data remains limited. A major barrier is how genomic data is captured and stored within the electronic health record (EHR). This limits the ability of clinicians to access and use this information effectively. For PGx to be properly used, genomic information must be available in a structured, standardized format that helps support clinical decision-making. Without this structure, established guidelines cannot be consistently applied in practice (Blagec et al., 2018).

Recent literature highlights that PGx implementation is expanding globally, with increasing diversity in how health systems attempt to integrate genomic data into clinical workflows. However, there is still a consistent gap that remains between the availability of genomic information and its ability to be used in the healthcare settings (Volpi et al., 2018).

Evidence from Real-World EHR Data: Hidden Genomic Information

Although PGx is often considered novel and unique, EHR data suggests that genomic and pharmacogenomic information already exists within clinical records. The issue arises when looking at the inconsistent and unstructured storage of this information. In many healthcare systems, genomic indicators are not stored in dedicated fields but instead appear indirectly within allergy lists, adverse drug reaction entries, or clinical notes (Overby et al., 2010).

For example, patients with conditions such as glucose-6-phosphate dehydrogenase (G6PD) deficiency, pseudocholinesterase deficiency, or cytochrome P450 enzyme variations have these findings documented within free text reaction descriptions. Similarly, pharmacogenomic test results, such as CYP2D6 metabolizer status or HLA-associated drug risks, may be recorded sporadically across notes, allergy modules, or scanned documents rather than in structured genomic sections. This inconsistent documentation limits the ability of clinicians and clinical decision support (CDS) systems to reliably identify and act on genomic risk factors when necessary.

Ongoing Challenges: Structure, Accessibility, and Infrastructure

In health systems pursuing integration of PGx information, significant barriers. These challenges include inconsistent data, a lack of standardized storage, and limited interoperability within the different systems. Efforts to incorporate PGx into EHRs, such as implementations leveraging HL7 standards and FHIR Genomics frameworks, demonstrate that structured genomic data can support effective CDS tools and improve medication safety (Alterovitz et al., 2020).

However, these solutions are not yet widely adopted. A major limitation is that genomic information is frequently stored within unstructured text fields, making it difficult to query, analyze, or integrate into automated workflows. As a result, healthcare systems often contain genomic information that remains inaccessible and underutilized. This gap highlights the need for an improved informatics infrastructure to support the standardization of storage and use of genomic data within the EHR (Volpi et al., 2018).

Why a Dedicated Genomic Informatics Section in the EHR Matters

The limitations observed in current EHR documentation highlight the need for a dedicated informatics storage structure for genomic data. A structured genomic section within the EHR could:

- Improve safety and efficacy: Enable CDS tools to identify gene-drug interactions and prevent adverse drug reactions
- Improve usability: Present genomic data in a clear, usable format within clinician workflows
- Support standardization: Promotes consistent representation of genomic data across patients and systems
- Enable scalability: Facilitate data sharing, research, and population health analysis

Methods / Solution

Study Design

This project is a retrospective descriptive analysis of how genomic and pharmacogenomic information is represented within structured EHR allergy data. This project will evaluate how genomic information is shared and stored within routine clinical documentation. In this case, it will specifically focus on entries within allergy and adverse reaction records. The goal is to assess the

consistency, structure, and usability of this information to determine whether a dedicated genomic informatics section within the EHR is warranted.

Genomic information was identified within the electronic health record (EHR) allergy dataset through a structured keyword approach. Because the EHR does not contain a dedicated field for genomic or pharmacogenomic data within the allergy module, relevant information was found within free text reaction descriptions. To capture these data, records were queried for presence of keywords suggestive of genetic and enzymatic variation.

Keywords included in the search include “CYP,” “metabolize,” “enzyme deficiency,” “genotype,” and “pharmacogenetic”. These are commonly associated with known genomic or metabolic differences which can impact a patient’s response to a specific drug. Any record containing one or more of these words was pulled and inserted into an excel document. The data points were then classified as containing genomic information or not containing genomic information. For example, multiple instances of “metabolic acidosis” appeared due to the inclusion of “metabolism” into the query. These entries were then further categorized into broader genomic groups (e.g., G6PD deficiency, different Factor deficiencies, pseudocholinesterase deficiency) to support structured analysis.

This approach allowed for the clinically relevant genomic information to be identified within unstructured EHR data. This highlights both the presence of pharmacogenomic information and the limitations of its current documentation format.

Data Source and Population

The study utilizes a dataset extracted from the electronic health record (EHR) consisting of allergy and adverse reaction entries. The dataset includes both structured and unstructured data elements commonly documented within the EHR allergy module, including:

- Medication
- Reaction details documented in free text fields
- Recorded severity levels
- Patient demographic information (race)

All records within the dataset were included in the analysis. Identification of genomic information relied on keyword searches of free text reaction descriptions.

Outcomes

- **Primary Outcomes:**
    - Proportion of EHR allergy records with "key" words containing genomic or pharmacogenomic indicators identified through keyword searches of free text reaction fields
    - Distribution of genomic information types (e.g., G6PD deficiency, CYP-related metabolism, pseudocholinesterase deficiency) identified within EHR documentation
- **Secondary Outcomes:**
    - Association of keyword genomic indicators with specific medications and most common genomic indicators based on patient race.

Descriptive statistics

Performed using Excel on EHR allergy data, with genomic indicators identified through keyword searches of free text reaction fields. Analyses will include:

- Counts and percentages of records containing keyword genomic indicators vs. non-genomic records
- Frequency and distribution of specific genomic conditions (e.g., G6PD deficiency, CYP-related metabolism, pseudocholinesterase deficiency)
- Association of genomic indicators with specific medications, including identification of drugs most frequently linked to genomic documentation
- Distribution of genomic indicators across patient race categories to identify the most common genomic patterns within different demographic groups

Deliverables

- A summary of genomic information stored within EHR allergy data
- Evidence supporting the need for a dedicated genomic informatics section
- Recommendations for improving structure, accessibility, and clinical integration of genomic data

Results

Primary Outcome: A total of 223 EHR allergy records were analyzed to identify the presence of genomic or pharmacogenomic indicators using a keyword approach applied to free text reaction fields. Of these, 176 records (78.9%) contained at least one keyword suggestive of genomic or metabolic variation, while 47 records (21.1%) did not contain identifiable genomic indicators. (Figure 1)

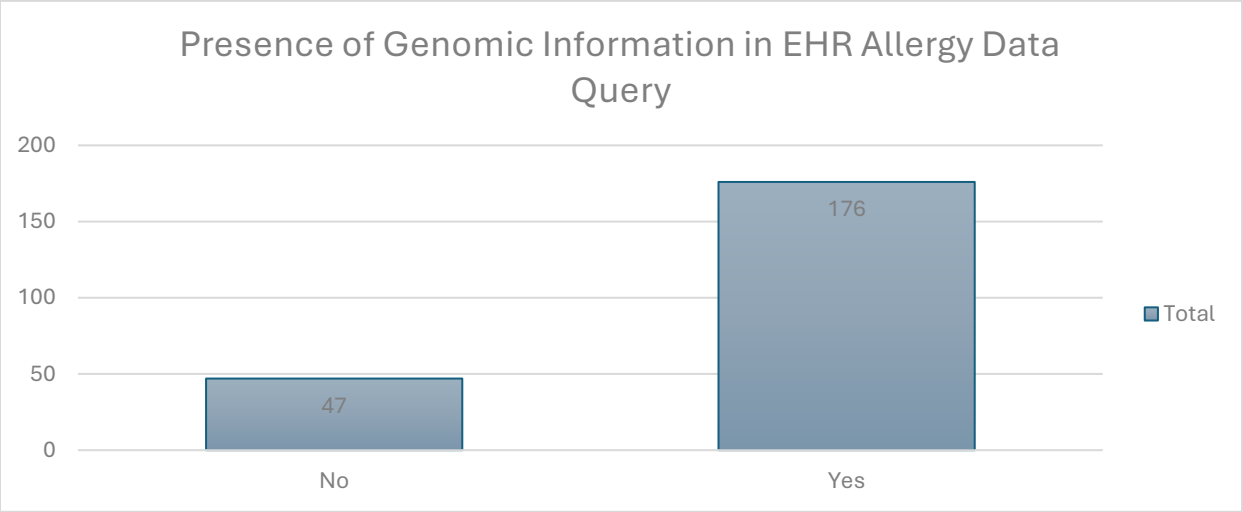

FIGURE 1

Secondary Outcomes

Among the 176 records identified as containing genomic indicators, multiple types of genomic and pharmacogenomic conditions were observed within the EHR allergy data. The most frequently identified condition was G6PD deficiency (n = 65, 36.9%), followed by pseudocholinesterase deficiency (n = 41, 23.3%). A substantial portion of entries were categorized as “other” genomic indicators (n = 55, 31.3%), reflecting a wide range of inconsistently documented conditions. (Figure 2)

Less frequently observed genomic indicators that were specifically mentioned included coagulation disorders such as von Willebrand factor deficiency (n = 5, 2.8%), as well as rarer entries such as Factor VII, VIII, and IX deficiencies, and pharmacogenomic findings including **HLA-A 31:01, UGT2B15, and TMPT variants*, each appearing in a small number of records (<1%). (Figure 2)

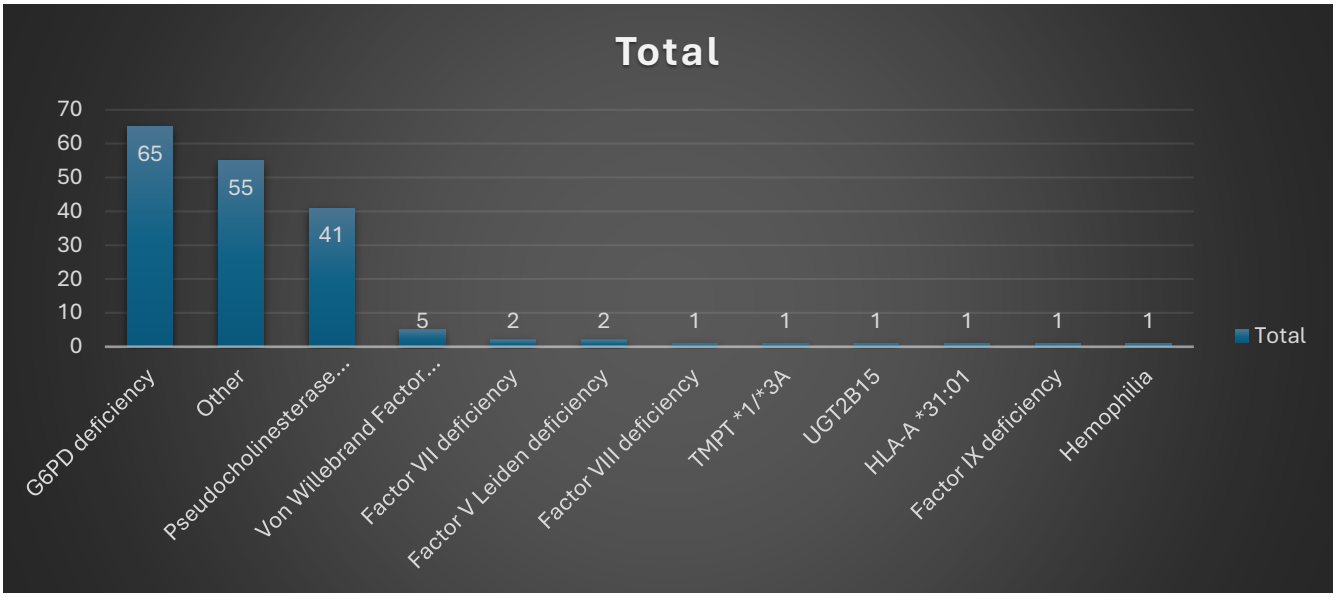

FIGURE 2

Genomic information identified within EHR allergy records were associated with a wide range of medications, with several drugs demonstrating strong and clinically relevant patterns. The most frequently associated medication was succinylcholine (n = 37), which was almost independently linked to pseudocholinesterase deficiency. This reflected an established pharmacogenomic interaction. (Table 1)

Other commonly associated medications included aspirin (n=23) and sulfa antibiotics (n=17), both of which were frequently linked to G6PD deficiency. Nonsteroidal anti-inflammatory drugs (NSAIDs) (n=12) showed associations with multiple genomic conditions, including G6PD deficiency and different coagulation disorders. This highlights the complexity of documentation within this class. (Table 1)

Additional medications such as metformin (n = 8), morphine (n = 7), and codeine (n = 6) were also commonly associated with genomic information frequently described as altered drug metabolism. Limited information in these sections made it difficult to determine type of metabolism present (rapid, slow, etc.). This led to them being included in the category labeled “other”. (Table 1)

TABLE 1

Medication	Total Count	Primary Genomic Association
Succinylcholine	37	Pseudocholinesterase deficiency
Aspirin	23	G6PD deficiency
Sulfa antibiotics	17	G6PD deficiency
NSAIDs	12	G6PD / coagulation disorders
Metformin	8	Metabolic-related variation
Morphine	7	Altered metabolism (CYP-related)
Codeine	6	Altered metabolism (CYP2D6)
Lidocaine	6	Metabolic variability
Nitrofurantoin	5	G6PD deficiency
Primaquine	4	G6PD deficiency

Analysis of genomic indicators by patient race revealed distinct patterns in the distribution of documented conditions. Among Black or African American patients (n = 37), the majority of genomic indicators were associated with G6PD deficiency (n = 35, 94.6%), making it the most dominant condition in this group. (Table 2)

Compared to the Black or African American demographic, White or Caucasian patients (n=114) demonstrated a more diverse distribution of genomic indicators. The most common findings in this group included pseudocholinesterase deficiency (n = 34) and entries categorized as “other” genomic indicators (n = 52), in addition to G6PD deficiency (n = 14). (Table 2)

Among Asian patients (n = 8), genomic indicators were primarily associated with pseudocholinesterase deficiency (n = 6) and G6PD deficiency (n = 2). Smaller groups, including patients categorized as “Other” or “Patient Declined,” also showed the presence of G6PD deficiency and less frequent genomic conditions. (Table 2)

TABLE 2

Patient Race	Total Records	Most Common Genomic Indicator	Count
White or Caucasian	114	Other / Pseudocholinesterase	52 / 34
Black or African American	37	G6PD deficiency	35
Asian	8	Pseudocholinesterase deficiency	6
Other	8	G6PD deficiency	7
Patient Declined	7	G6PD deficiency	5

Overall, the analysis demonstrates that genomic and pharmacogenomic information is inconsistently present within EHR allergy and adverse reaction records. This information does span a wide range of conditions, with G6PD deficiency and pseudocholinesterase deficiency representing the most common findings and are often directly associated with clinically relevant medications such as succinylcholine, sulfa antibiotics, and opioid analgesics.

Additionally, demographic analysis revealed distinct patterns in the distribution of genomic indicators, particularly the strong association between G6PD deficiency and Black or African American patients, as well as greater variability in documentation among White or Caucasian patients.

Despite the presence of clinically meaningful genomic information, the findings consistently highlight that this data is stored within unstructured free text fields, lacks standardization, and is represented inconsistently across records.

Discussion

This analysis shows that genomic and pharmacogenomic information is already stored within EHR allergy and adverse reaction documentation but is not stored in a way that allows for consistent clinical use. Instead of being stored in structured, computable fields, genomic indicators are almost exclusively documented within free text reaction descriptions. This limits their visibility and usefulness within routine clinical workflows.

From a pharmacy perspective, the findings highlight the presence of very useful gene-drug relationships which are extremely important when it comes to medication safety and efficacy. Common patterns, such as associations between pseudocholinesterase deficiency and succinylcholine, or G6PD deficiency and sulfa-containing medications, reflect established pharmacogenomic interactions. However, because these relationships are documented inconsistently and located within narrative text, they are not available in a way that the information can be optimized and used by clinical decision support systems. As a result, opportunities to prevent adverse drug reactions or optimize therapy based on genetic factors may be missed.

One informatics challenge identified in this study is the lack of standardization not only in where this information is being stored but also how genomic information is represented within these different sections. The same condition may appear in multiple formats (e.g., “G6PD deficiency,” “G-6-PD deficiency,” or are meant to be implied through medication restrictions), and pharmacogenomic test results are often recorded alongside general clinical observations without clear distinction. This variability introduces uncertainty, reduces data quality, and makes it difficult to collect genomic information across patient populations.

The variability observed across patient demographics further highlights some of the limitations in current documentation practices. While certain genomic conditions showed expected distribution patterns, the inconsistency in how information is recorded suggests that differences in documentation may reflect provider practices or system limitations rather than true biological variation. This raises concerns about reliability. Inconsistent documentation may lead to unequal access to the benefits of pharmacogenomic related care.

All in all, these findings point to a gap between the presence of genomic information and its effective use in clinical care. Addressing this gap will require the development of a dedicated genomic informatics infrastructure within the EHR. A structured genomic section that is specifically designed to capture standardized genetic results, metabolizer status, and gene-drug interactions would allow for integration with clinical decision support tools, improve data accessibility, and improve the consistency of documentation. This would not only support safer prescribing practices but also provide easier access to information leading to future developments in research, population health, and precision medicine.

Limitations

Several limitations should be considered when interpreting the findings of this analysis. First, the identification of genomic and pharmacogenomic information was based on a keyword approach applied to free text reaction fields, which may have resulted in misclassification. Some records containing genomic relevance may not have included the selected keywords and were therefore not captured, while others may have been incorrectly categorized as genomic related based on ambiguous language.

Second, the dataset represents a single EHR sample, which may limit its generalizability to other institutions with different documentation practices and/or patient populations. Variability in how clinicians document allergy and adverse reaction information may also influence the presence, or absence, and format of genomic indicators observed in this analysis.

Finally, this analysis does not confirm whether the documented genomic indicators were pulled from pharmacogenomic testing or just inferred clinically based on patient reactions to medications. As a result, the accuracy and validity of some results cannot be fully verified. Despite these limitations, the findings remain valuable in illustrating current documentation patterns and highlighting informatics gaps within the EHR.

Conclusion

This project helps show that genomic and pharmacogenomic information is frequently present within the EHR but is largely stored in unstructured and inconsistently documented formats. Clinically meaningful gene-drug relationships, particularly those involving drug metabolism and enzyme deficiencies, are being captured in

routine care, yet their current storage location limits their accessibility and ability to be integrated into clinical decision-making.

The findings highlight a critical disconnect between the availability of genomic information and its usability within the EHR. Without standardized structure and dedicated storage, this information remains difficult to properly identify, interpret, and use in a way that allows for continuous improvements in medication safety and therapeutic outcomes.

To fully realize the potential of pharmacogenomics in clinical practice, there is a clear need for a dedicated genomic informatics section within the EHR. Ultimately, by turning this fragmented and inconsistent genomic documentation into structured, actionable data is an important step towards advancing precision medicine, improving pharmacy practice, and optimizing patient care.

References:

Alterovitz, G. *SMART on FHIR Genomics: facilitating standardized clinico-genomic apps.* J Am Med Inform Assoc. 2015 Nov;22(6):1173-8. doi: 10.1093/jamia/ocv045. Epub 2015 Jul 21. PMID: 26198304; PMCID: PMC11737838.

Blagec, K. *Implementing pharmacogenomics decision support across seven European countries: The Ubiquitous Pharmacogenomics (U-PGx) project.* J Am Med Inform Assoc. 2018 Jul 1;25(7):893-898. doi:10.1093/jamia/ocy005. PMID:29444243; PMCID:PMC6016647.

Caudle, KE. *Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process.* Curr Drug Metab. 2014 Feb;15(2):209-17. doi:10.2174/1389200215666140130124910. PMID: 24479687; PMCID: PMC3977533.

Overby, CL. *Opportunities for genomic clinical decision support interventions.* Genet Med. 2013 Oct;15(10):817-23. doi: 10.1038/gim.2013.128. Epub 2013 Sep 19. PMID: 24051479; PMCID: PMC3858176.

Volpi, S. *Research Directions in the Clinical Implementation of Pharmacogenomics: An Overview of US Programs and Projects.* Clin Pharmacol Ther. 2018 May;103(5):778-786. doi: 10.1002/cpt.1048. Epub 2018 Mar 30. PMID: 29460415; PMCID: PMC5902434.