

Efficiency of CYP2C9 Genetic Test Representation for Automated Pharmacogenetic Decision Support

V. G. Deshmukh¹; M. A. Hoffman²; C. Arnoldi²; B. E. Bray¹; J. A. Mitchell¹

¹University of Utah, School of Medicine, Department of Biomedical Informatics, Salt Lake City, UT, USA;

²Cerner Corporation, Kansas City, MO, USA

Keywords

Pharmacogenetics, clinical decision support systems, SNP, allele

Summary

Objectives: We investigated the suitability of representing discrete genetic test results in the electronic health record (EHR) as individual single nucleotide polymorphisms (SNPs) and as alleles, using the CYP2C9 gene and its polymorphic states, as part of a pilot study. The purpose of our investigation was to determine the appropriate level of data abstraction when reporting genetic test results in the EHR that would allow meaningful interpretation and clinical decision support based on current knowledge, while retaining sufficient information in order to enable reinterpretation of the results in the context of future discoveries.

Methods: Based on the SNP & allele models, we designed two separate lab panels within the laboratory information system, one con-

taining SNPs and the other containing alleles, built separate rules in the clinical decision support system based on each model, and evaluated the performance of these rules in an EHR simulation environment using real-world scenarios.

Results: Although decision-support rules based on allele model required significantly less computational time than rules based on SNP model, no difference was observed on the total time taken to chart medication orders between rules based on these two models.

Conclusions: Both, SNP- and allele-based models, can be used effectively for representing genetic test results in the EHR without impacting clinical decision support systems. While storing and reporting genetic test results as alleles allow for the construction of simpler decision-support rules, and make it easier to present these results to clinicians, SNP-based model can retain a greater amount of information that could be useful for future reinterpretation.

The interpretation of existing genetic data within the context of emerging scientific knowledge would require data abstraction at various levels such as single nucleotide polymorphisms (SNPs), alleles, haplotypes, etc., in accordance with the corresponding discoveries. While genetic data can be represented using Bioinformatics Sequence Markup Language (BSML) [9], clinical findings are typically stored using SNOMED CT [10] and LOINC [11], and although the latter two vocabularies could also be used for reporting results of genetic tests, they lack the granularity required to describe genetic findings in sufficient detail that would allow meaningful clinical inference [12]. The Clinical Bioinformatics Ontology (CBO) is a semantically structured controlled medical vocabulary that enables the standardized reporting of clinical molecular diagnostics in a consistent, machine-readable format. The CBO contains manually curated, pre-coordinated concepts that have been annotated with orthogonal mapping to bioinformatics databases in the form of facets [13], and is therefore suitable for the generation of executable knowledge needed for performing advanced queries, inference logic and knowledge discovery in the area of clinical molecular diagnostics [12, 13].

An allele is one of many alternate forms of a gene that occupies a given locus on a chromosome, and can consist of one or more polymorphisms in a gene relative to a reference sequence. An alternate form of a gene consisting of more than one SNP can be efficiently described using an allele name (or symbol) in lieu of a list of the constituent polymorphisms. The Cytochrome P-450 (CYP) enzyme system is part of a common metabolic pathway for several important classes of drugs; and differences in drug metabolism have been attributed to various SNPs [14–16] (e.g.: CYP2C9 1076 C > T, 1188delA, etc.), which can also be described

Correspondence to:

Vikrant G. Deshmukh, M.Sc., M.S.
University of Utah, School of Medicine
Department of Biomedical Informatics
26 South 2000 East Room 5775
Salt Lake City, UT 84112
USA
E-mail: Vikrant.Deshmukh@hsc.utah.edu

Methods Inf Med 2009; 48: 282–290

doi: 10.3414/ME0570

received: May 4, 2008

accepted: November 11, 2008

prepublished: March 31, 2009

1. Introduction

With the completion of the human genome project [1], and continued discoveries in the genetics of human diseases, a number of molecular-genetic tests have now become available [2], and the integration of these tests in clinical care will be a major step toward delivering personalized medicine [3, 4]. The storage, retrieval and reporting of genetic test results, as well as

their integration within the clinical environment pose unique challenges [5], and the HL7 [6] clinical genomics standard (CGS) has been proposed to address such issues [7]. One of the strengths of the CGS model is the ‘encapsulate and bubble-up’ approach, in which raw genomic data are reported along with the genetic test results, while additional interpretations can bubble-up as new knowledge and data become available [8].

Methods Inf Med 3/2009

by their representative alleles [17, 18] (e.g.: CYP2C9*2, *3, etc.). Warfarin, a commonly prescribed oral anticoagulant, is a vitamin-K antagonist [19] metabolized by CYP2C9 [20], with a narrow therapeutic index, and significant inter-patient variability in dose response, due to which, it has been under-utilized [21]. The variability in dose response to Warfarin has been partially explained by SNPs in CYP2C9 and VKORC1 [22] genes, and the United States Food and Drug Administration's (FDA) new labeling on all Warfarin products [23] underscores the importance of pharmacogenetics in general, and of this use-case in particular.

Since the availability of genetic tests for CYP2C9 and VKORC1 genes, dosing algorithms that incorporate results of these two tests are now available [24], and their implications in Warfarin dosing may be conceptually represented by applying the hierarchical knowledge model [25] (► Fig. 1). In Figure 1, the results of molecular assays used to detect known SNPs [14] constitute raw data, which may be represented by the corresponding alleles [18] that subsume the SNPs (information), which may then be interpreted according to the expected phenotype [15] as slow-metabolizers of Warfarin (knowledge), which may be understood by clinicians as having elevated risk of bleeding complications [26] in Warfarin therapy (understanding), who may then adjust the Warfarin dosage [24] by using a combination of dosing nomograms, pharmacogenetics, the physiological condition of the patient, and their experience in treating patients with similar conditions (wisdom). Alternatively, genetic findings may be reported with recommendations for dose adjustment, which could provide clinical context for the results.

Although the above knowledge model may work well in the short-term, with an increasing use, and evolving knowledge of the implications of these test results in clinical practice, the complexity of clinical information is likely to increase [27]. In addition, the issues with navigating information sources that enable genotype-to-phenotype type translation of such knowledge into clinical practice [28] necessitate the use of clinical decision support systems (CDSS) at the point of care [29, 30]. CDSS require that the genetic test results, as well as the interpretations reported in the Laboratory Information System (LIS) and the

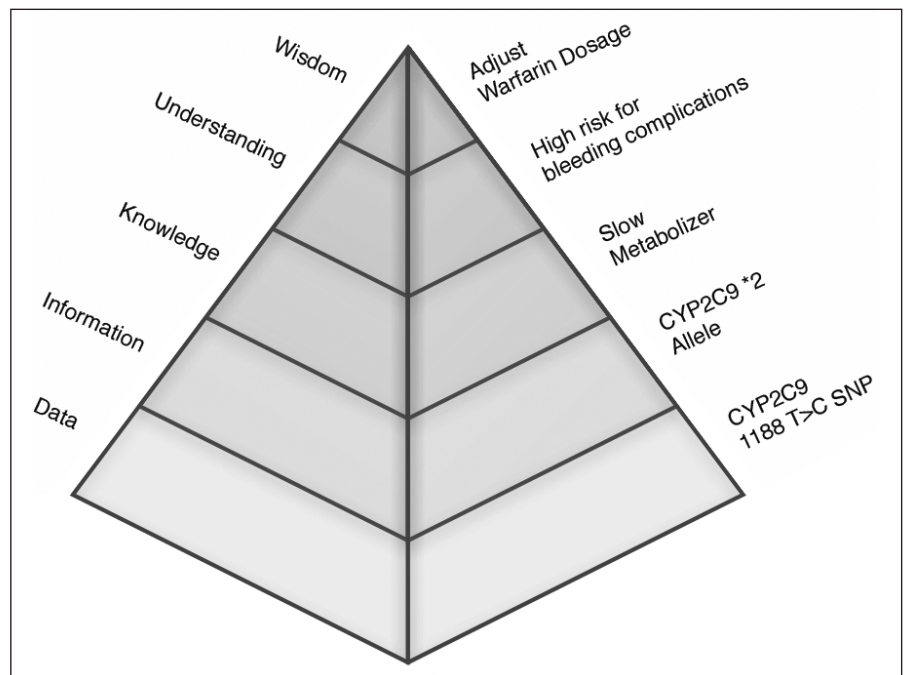


Fig. 1 Conceptual levels of data abstraction: A set of SNPs can be considered data, the allele containing the SNPs as information, the resulting slow-metabolizer phenotype as knowledge, the implications on bleeding complications as understanding, while the overall need to adjust Warfarin dosage based on all of the above as wisdom.

EHR be in discrete, concise and machine-readable format. Further, with the increasing availability of genetic testing in external reference labs [2] (or potentially direct-to-consumer genetic testing facilities [31, 32]) results of these genetic tests may not have been entered in the same LIS as the rest of their results, in which case, these genetic test results would then have to be collected as part of their history & physical (H&P) examination. However, having H&P as plain text would negate the benefit of having a point-of-care CDSS, particularly for genetic test results, which often tend to be more complex than other clinical findings [27], and it would become necessary to capture any genetic information provided by the patient in a discrete, coded format, rather than as a textual narrative.

In the absence of an appropriate level of data abstraction, the sheer amount and complexity of information generated by genetic testing have the potential to overwhelm existing EHR systems as well as the clinical end-users. In the present work, we investigate the suitability of reporting genetic data in the EHR at the level of SNPs and alleles by comparing these two data models from the perspective of clinical decision support, report-

ing within the EHR, and suitability for integrating future discoveries.

2. Methods

Our pilot project involved the CYP2C9 gene and the corresponding alleles and SNPs known to have clinical significance in Warfarin therapy. The initial prototyping was performed in a simulation environment at Cerner Corporation headquarters in Kansas City, MO, and then the decision-support component was reconstructed in a live clinical system environment at the University of Utah Hospital, Salt Lake City, UT. Although all software testing was performed using Cerner software, our methods are generalizable, and can be evaluated using any EHR system that integrates a point-of-care clinical decision support system (CDSS), and is independent of the underlying computational environments, databases, etc. (the environments that we prototyped and tested in have different underlying hardware and software architectures). All the decision support rules used in our study are available for download as standard Arden syntax files at <http://>

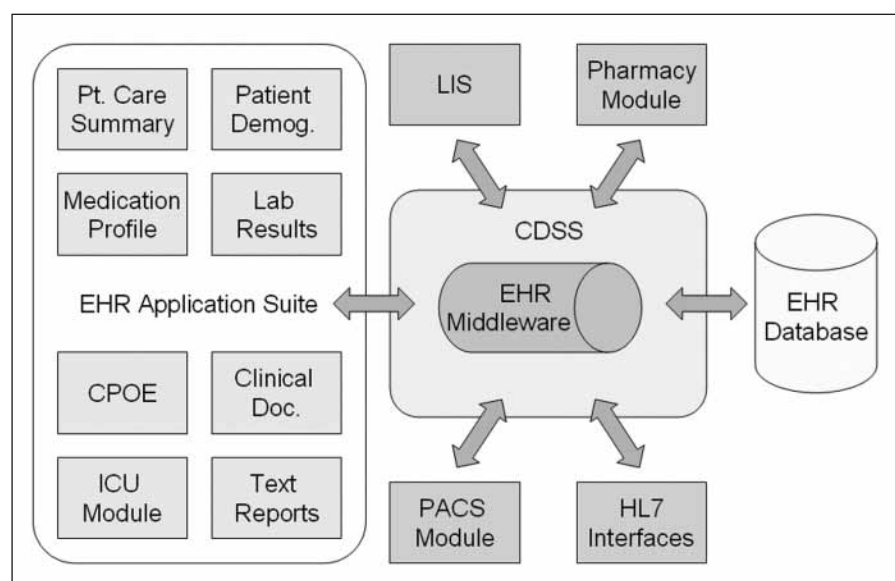


Fig. 2 Block diagram of information architecture: The above schematic shows a simplified, scaled-down version of the various components in our EHR system. The EHR system contains several integrated modules, and these communicate with other components and with the database through a middleware layer. The CDSS component resides in the middleware layer, and is available within most components of the EHR.

gram in ► Figure 2 shows a simplified, scaled-down schematic of various components of the EHR that are relevant to the present work. Laboratory orders can be placed in the CPOE module of the EHR, or in the LIS modules, and the results of lab tests can be charted in the LIS. One of the differences in the EHR environments that were used during prototyping and testing was that our testing environment at the University of Utah Hospital receives lab results over an HL7 interface, whereas the prototyping environment at Cerner Corporation shared a common database with the LIS module through the EHR middleware. Regardless of the setup, however, once charted, the results are then automatically sent to the EHR application, where they appear under the lab results section in the patient's chart. The same genetic results can also be charted as part of discrete patient-care documentation within the EHR itself as part of the clinical documentation module, which also posts these results to the exact same place within the database through the middleware, which is important when considering scenarios where genetic test results may have been provided directly by the patients themselves as results from other independent labs [2] during their history and physical examination. The integrated, point-of-care CDSS module in the EHR which operates within the middleware layer is able to consume the results, and communicate the alerts to several different applications, regardless of the application/method in which these were posted to the EHR, making our methods even more generalizable. Medication orders can be placed in the EHR, or in a separate pharmacy module, and the clinical decision support system runs in the background in both of these modules.

During prototyping at Cerner Corporation, we developed two lab panels in the LIS (► Table 1), a panel for reporting test results as alleles (allele panel) and another panel for reporting test results as SNPs (SNP panel). Discrete genetic test results within each of these lab panels were mapped to their corresponding pre-coordinated concepts within the CBO, using the CYP2C9*2A allele as an example (► Table 2). E.g. the definition of the CBO allele concept CYP2C9.004 contains several individual SNPs represented by the respective CBO concepts, and also the synonym CYP2C9*2A, which represents the allele.

Allele	Allele Subtype	SNP: cDNA	SNP: Genomic DNA
*1	*1A	None	None
	*1B		2665_2664delTG; 1188T>C
	*1C		1188T>C
	*1D		2665_2664delTG
*2	*2A	430C>T	1188T>C; 1096A>G; 620G>T; 485T>A; 484C>A; 3608C>T
	*2B		2665_2664delTG; 1188T>C; 1096A>G; 620G>T; 485T>A; 484C>A; 3608C>T
	*2C		1096A>G; 620G>T; 485T>A; 484C>A; 3608C>T
*3	*3A	1075A>C	1911T>C; 1885C>G; 1537G>A; 981G>A; 42614A>C
	*3B		1911T>C; 1885C>G; 1537G>A; 1188T>C; 981G>A; 42614A>C
*4		1076 T>C	42615T>C
*5		1080 C>G	42619C>G
*6		818delA	10601delA

Table 1 CYP2C9 alleles, SNPs and lab panels (adapted from the Human Cytochrome P450 (CYP) Allele Nomenclature Committee's website) [18]

informatics.bmi.utah.edu/cyp2c9/suppl/. The latest version of the CBO is available for download at <http://www.clinbioinformatics.org>.

The Cerner® Millennium® platform consists of several modules that serve different functions, leveraging a common application and database infrastructure. The block dia-

Within the allele panel, the concept representing CYP2C9*2A was used as-is for storing one discrete data element, whereas in the SNP panel, the individual concepts that represent SNPs contained in this concept each had their own separate, discrete data elements. The LIS also allowed automated interpretation of genetic results as ‘homozygous normal’, ‘heterozygous affected’ and ‘homozygous affected’, based on the results entered for each copy of the allele or SNP in the corresponding lab panels. Upon electronically signing the results charted in a given lab panel for a test patient, these results and their automated interpretations were posted to the EHR. During testing at the University of Utah, the same genetic test panels were recreated within the discrete clinical documentation module of the EHR itself, and results were charted using that module.

Decision-support rules based on each data model were built in the CDSS module, using the logic illustrated in Figure 3A. In addition to the individual rules based on allele- & SNP-models, two other generic rules were created, one of which triggered upon adding a medication to the list of medication orders (rule ‘A’), and another which triggered upon actually signing the medication order (rule ‘D’). Individual rules based on allele- & SNP-models (rules ‘B’ and ‘C’ respectively) were set to trigger in response to the placing of Warfarin orders, and the rule-evaluation criteria were slightly different for each rule (Figure 3B), with the allele-rule checking for lab values indicating CYP2C9 *2, *3, or *6 alleles (Table 1, column 1), and the SNP-rule checking for all the corresponding SNPs for each of these alleles (Table 1, columns 3/4). The execution order of these rules were set so that during the process of adding Warfarin to a patient’s list of medication orders, the first rule that fired was the generic rule ‘A’, followed by either ‘B’ or ‘C’ depending on whether we were testing the allele-model or the SNP-model, and then finally rule ‘D’. The common element within each of these rules was an action that created a database time-stamp, with precision in milliseconds, so that the difference between three time-stamps in each test case would give us the actual amount of time needed to evaluate the rule, the time taken to respond to the medication alert (Figure 4), and the total time needed to complete individual orders.

Table 2 Concepts and relationships in the Clinical Bioinformatics Ontology using CYP2C9*2A allele as an example [12]

CBO Concept 1	Relationship	CBO Concept 2
Human Allele	Subsumes	CYP2C9.0004
CYP2C9.0004	Synonym	CYP2C9*2A
CYP2C9.0004	Has constituent variant	CYP2C9.c.-1188T>C
CYP2C9.0004	Has constituent variant	CYP2C9.c.-430C>T
CYP2C9.0004	Has constituent variant	CYP2C9.c.-1096A>G
CYP2C9.0004	Has constituent variant	CYP2C9.c.-620G>T
CYP2C9.0004	Has constituent variant	CYP2C9.c.-485T>A
CYP2C9.0004	Has constituent variant	CYP2C9.c.-484C>A

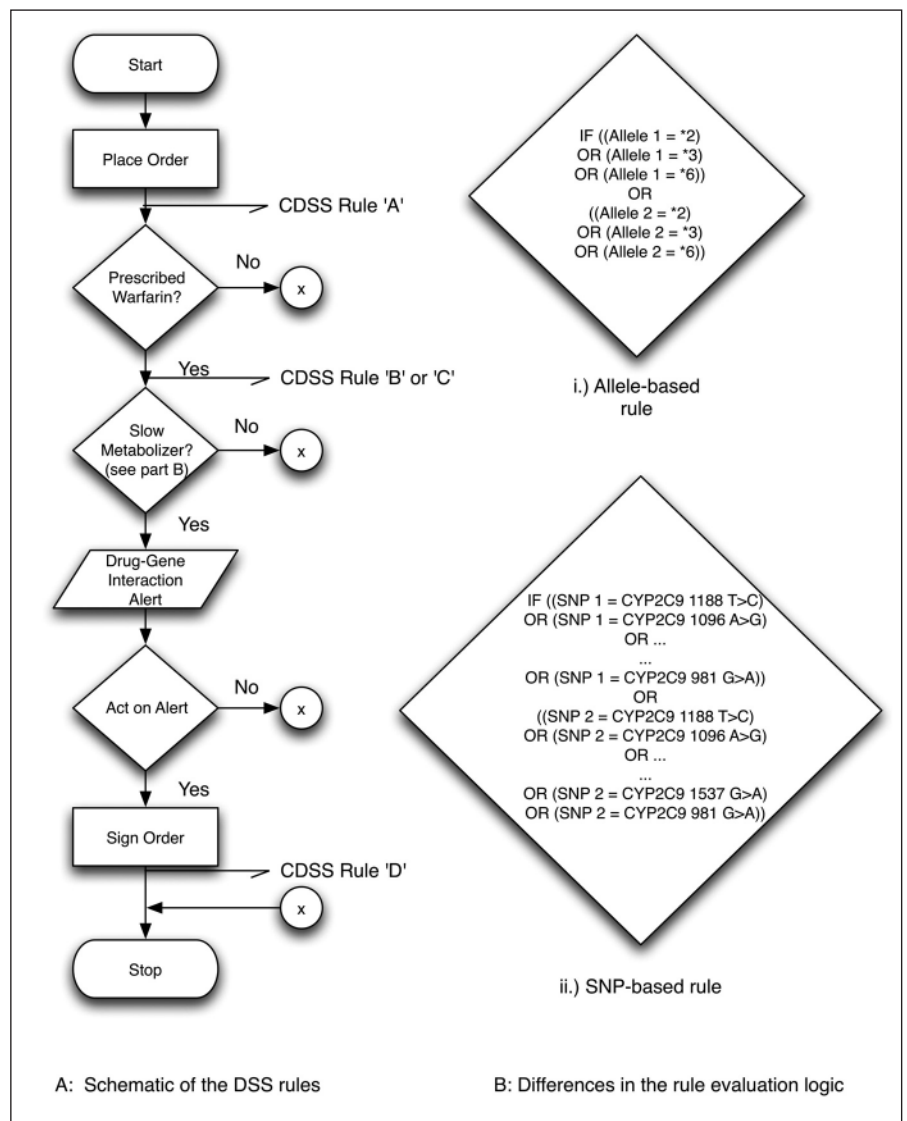


Fig. 3 Testing methodology. A: The overall testing method showing the rules being triggered upon placing an order for Warfarin on a test patient on whom CYP2C9 Allele/SNP results were available; B: Overall differences in the rule evaluation logic for the Allele & SNP rule.

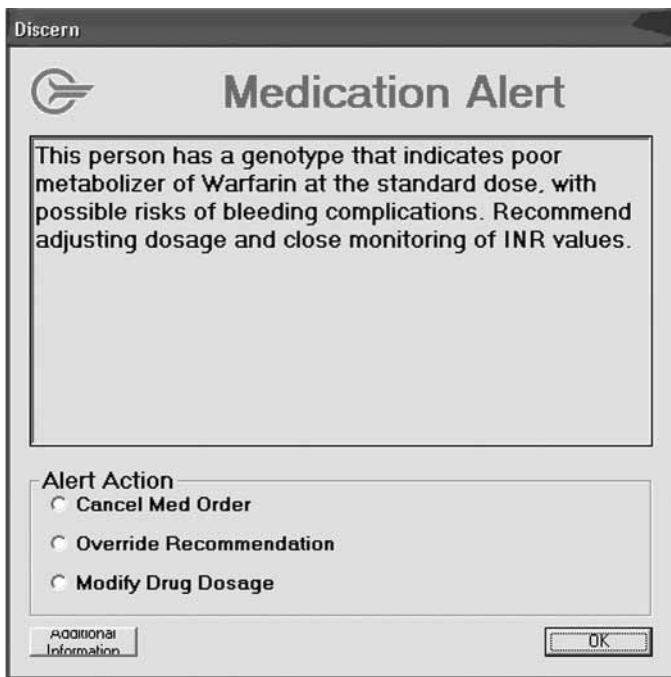


Fig. 4 Medication alert: Medication alert triggered by adding Warfarin to the list of medication orders on a test patient with results for CYP2C9 Allele/SNP results

The CDSS rules ‘B’ and ‘C’ were tested in isolation from one another by enabling one, while having disabled the other, and using test patients for whom the corresponding SNPs

or alleles were reported as positive through the LIS or the clinical documentation modules. This was done in order to prevent test conditions where either of these rules could

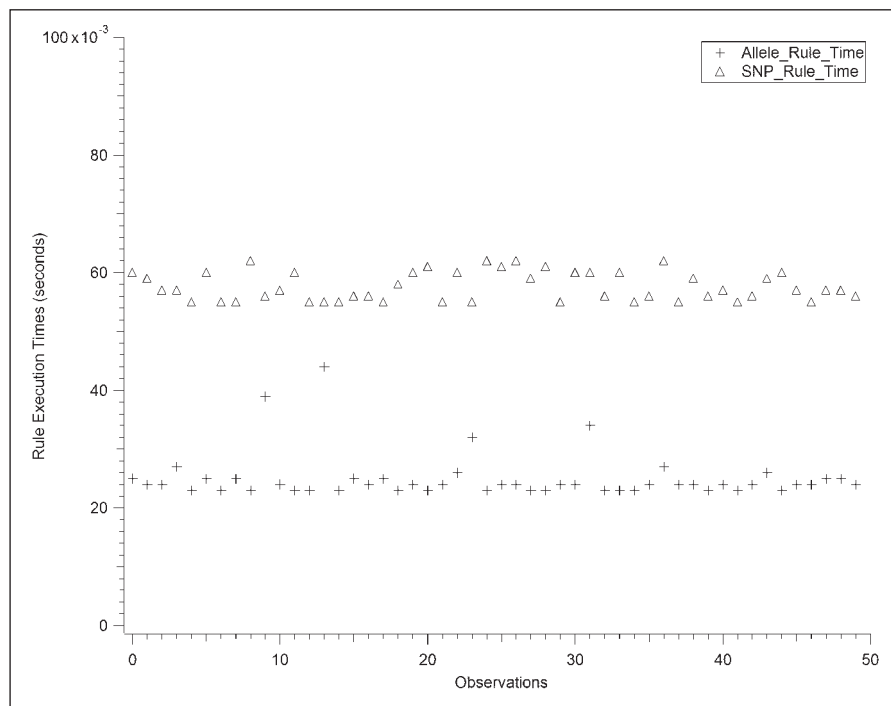


Fig. 5 Rule execution times: Rule execution times were measured as the difference between the database time-stamps on adding Warfarin to the list of the test patient’s medication orders, and the time needed to complete evaluation of the rule logic.

be triggered simultaneously, in order to avoid any potential impact on each other, so that the triggering of the rules followed an order A-B-D for the allele-model and A-C-D for the SNP-model. It is possible to place the same medication order through either the inpatient pharmacy system module or the Computerized Provider Order Entry (CPOE) module, and so as to allow the execution of the rule. The CPOE module is integrated tightly with the main EHR application, and the drug-gene interaction alerts were primarily intended to be seen by physicians at the point where they would place the orders. Thus, for our testing, we chose to place these orders using the CPOE module of the EHR system, rather than the pharmacy module. Each order for Warfarin placed in this manner was later discontinued, and this process was then repeated so that there were no active orders for Warfarin on the given test patient at the time of placing another order. This was done in order to avoid triggering other existing error-checking mechanisms such as therapeutic duplication checking and dose-range checking, which could have potentially introduced confounders by interfering with rule execution.

The execution of the rule in each case generated a popup alert (Fig. 4) indicating a medication warning due to an underlying genetic condition, with options to accept or cancel the order, or ignore the recommendation, and a link for additional information describing the importance of these genotypes in Warfarin therapy [26]. Although it was also possible to make numeric dosing recommendation based on pharmacogenetic data [24], this was not included in our tests to minimize confounding. Over 50 orders were placed to evaluate each rule in this manner for Warfarin on each test patient, with the rule triggering on every event, and the rule execution times recorded as time-stamps in the EHR database using database triggers. In order to further minimize confounding due to differences in system utilization during testing, the simulations were performed during the same time periods of a day to account for system load.

Other aspects of reporting genetic test results in the EHR, such as the formats for reporting results and interpretations to the clinicians were also considered, although a formal evaluation was not performed as part of the present study.

3. Results

3.1 Rule Execution Times

The rule execution times were determined by examining the differences in database time-stamps from the point of rule triggering to the completion of rule-logic evaluation and generating an alert in the EHR. The results for rules based on both allele- and SNP-models are plotted in ►Figure 5. These measurements were logged in milliseconds, and the average rule execution times were 25.06 ms ($n = 50$) for the allele rule and 57.64 ms ($n = 50$) for the SNP rule (►Table 3). Using a two-tailed Student t -test of two samples assuming equal variance, the p -value was $<6.14679E-71$ ($\alpha = 0.05$, $df = 98$), and thus, there was a significant difference between the means for rule-evaluation times for allele-model and SNP-model.

3.2 Times Required for Completing Medication Orders

When testing the decision support rules based on each model, these were executed in a predetermined sequence (A-B-D or A-C-D), with the very first rule being triggered on adding any medication to the list of medication orders, followed by either the allele-/SNP-based rule, and finally the rule triggered upon signing the medication order. The difference in time-stamps from the time Warfarin was added to the list of medication orders (rule A) and the time the order was signed (rule D) was the total medication order time measured in seconds, and the mean times for the two models were 6.603 s for the allele-model, and 6.578 s for the SNP-model (Table 3). Using a two-tailed Student t -test of two

samples assuming equal variance, the p -value was 0.904 ($\alpha = 0.05$, $df = 98$), and thus, there was no significant difference between means for total time required for signing medication orders in these two experiments.

Similarly, the difference in time-stamps from the time that the allele-/SNP-based rule was triggered (rule B or C) to the time that the medication order was signed (rule D) was the reaction time to the alert generated by the EHR, which was measured in seconds. The mean times for the two models were 6.578 s for the allele-model, and 6.520 s for the SNP-model (Table 3). Using a two-tailed Student t -test of two samples assuming equal variance, the p -value was 0.784 ($\alpha = 0.05$, $df = 98$), and thus, there was no significant difference between means for reaction times in responding to the alerts in the EHR in these two experiments.

3.3 Reporting Genetic Results in the EHR

Genetic test results for both panels were reported under the category of 'labs' within the EHR. For the allele panel (Table 1), 12 discrete results were generated per panel ordered: one per each copy of the allele (e.g.: Copy 1: CYP2C9*2 'Present'; Copy 2: CYP2C9*2 'Absent') per allele. The report for the SNP panel (Table 1) was much more extensive due to the number of SNPs, with one discrete result per copy of the SNP (e.g.: Copy 1: CYP2C9 1096A>G 'Absent'; Copy 2: CYP2C9 1096A >G 'Present'), thus generating about 32 discrete genetic test results in the EHR per test order.

4. Discussion

The integration of genetic data in clinical care is an important step toward delivering personalized medicine, and point-of-care CDSS that enable recommendations based on these data will serve as important means of realizing this goal. Given the inherent complexity of genetic data and the need to have concise, human-interpretable guidelines at the point-of-care, it will be necessary to present these data in a form that can be consumed by front-line clinicians, and it will therefore be necessary to abstract these data. However, each level of data-abstraction going from the complete DNA sequence onward to SNPs, alleles, haplotypes, etc. comes with tradeoffs in terms of current vs. future usability of the findings, performance of the CDSS, loss of information that may be important in secondary use, etc. In the present study, we have considered one such scenario involving data abstraction by comparing two data models for reporting genetic data in the EHR based on SNPs & alleles, and have considered some of the potential implications of choosing either of these data models in CDSS involving genetic data at the point-of-care by tackling a real-world clinical problem involving CYP2C9 polymorphisms & Warfarin dosing in a real-world EHR environment.

4.1 Rule Execution Times

From the CDSS database time-stamps, it was estimated that the average rule-execution times for the allele model were 25.06 ms, while that for the SNP model were 57.64 ms (Table 3), which were both within acceptable limits for interactive software applications, and would not have a negative impact on end-

Table 3
Differences in rule execution times, reaction times and total times

Time	Sample size	Mean	Standard deviation	Degrees of freedom	Two-tailed t-statistic ($t_c = 1.96$)	p-value ($\alpha = 0.05$)
Allele rule execution	50	25.06 ms	2.431 ms			
SNP rule execution	50	57.64 ms	3.997 ms	98	49.245	6.15E-71
Allele alert reaction time	50	6.578 s	0.908 s			
SNP alert reaction time	50	6.520 s	1.182 s	98	0.275	0.784
Allele total time	50	6.604 s	0.909 s			
SNP total time	50	6.578 s	1.182 s	98	0.120	0.904

user applications if these rules were triggered in isolation. In spite of the significant differences in rule execution times, differences between the total rule execution times as well as the reaction times to the EHR alerts for the Allele-/SNP-models were insignificant, since these two times, measured in seconds, differed from the rule execution times by at least two orders of magnitude. In other words, within our experiments, there was plenty of room to accommodate more complex CDSS rules before they could have had a noticeable impact on the end-user applications. However, in a clinical system with multiple CDSS rules being evaluated on multiple patients, the overall rule execution times could still vary, and possibly impact the performance and responsiveness of the front-end applications. Some of these potential issues with performance could be addressed by consolidating two or more CDSS rules into a single rule which runs faster, but such an approach could potentially create new problems by adding to the rule complexity, and maintaining such rules over time.

4.2 Rule Complexity

With the growing complexity of knowledge in the genetics of human diseases, CDSS rules will also tend to become more complex, and this complexity was readily apparent in the differences in rule logic between the SNP & allele models (► Fig. 3B). The complexity of CDSS rules can be addressed by constructing an executable knowledge base of SNPs, alleles, haplotypes, etc., and the relevant clinical effects, interpretations and recommendations, so that pharmacogenetic decision support could be driven by stored, updateable knowledge instead of hard-coded logic such as that used in our rules. The PharmGKB, a publicly available, searchable online resource, is one such pharmacogenetic knowledge base [33] that contains current information on the relationships between drugs, diseases and genes, but in order to be used as a knowledge base for CDSS rules, the rules based on these relationships themselves will have to be formalized and stored in an executable format. Molecular-genetic vocabularies such as the CBO, which contain pre-coordinated concepts for genetic findings (e.g.: the CBO concept CYP2C9.c.430C>T implies a change

from Cytosine to Thymine at position 430 in the cDNA of CYP2C9 gene) will have to be combined with other clinical vocabularies such as SNOMED CT [10] to adequately describe the effect as well as the recommendations that will be required for enabling pharmacogenetic decision support at the level necessary for formalizing these rules. However, at present, most clinical vocabularies lack the granularity and coverage needed to describe the effects and interpretations of molecular-genetic tests as well as recommendations based on these results, and may require considerable improvement before they can be used for effectively representing these rules in an executable knowledge base.

4.3 Precision of Allele Assignment

With DNA sequencing becoming the gold standard for genetic information, any other form of capturing SNP findings is subject to imprecision. For instance, an allele-specific PCR panel could generate SNP results that are interpreted as a specific allele. Subsequent DNA sequencing could identify a SNP that was not specifically targeted in the initial panel and require the correction of the allele assignment. Therefore, it is important to consistently document the method used to generate an allele assignment.

A vocabulary concept descriptive of a finding that represents an allele can be semantically related to other concepts that represent the corresponding SNPs for that allele. Using the allele concepts for reporting the results of genetic tests does not exclude the possibility of other SNPs being detected by a more comprehensive assay, but the burden falls on the LIS to be configured to accurately describe the methodology. Some allele terminologies imply phylogenetic relationships between allele names, which present a problem of precision, and therefore accurate descriptions of these methodologies becomes important. E.g.: subtypes of the CYP2C9*2 allele, *2A & *2B, are identical from a functional perspective, causing the same overall change in the cDNA (Table 1). However, the *2B allele includes a few more SNPs in addition to those found in *2A, and reporting the results as only the relevant allele *2 would lead to a loss of this information, which may

be important to retain in the context of future discoveries. Further, allele nomenclature itself is not consistent across different genes, and it may be more suitable to report all individual SNPs within the LIS and the EHR, in order to retain the maximum amount of information. The CBO addresses these concerns by creating unique concepts for each allele regardless of functional significance and through the use of a phylogenetically neutral naming convention that is consistent for all genes represented (modeled after the allele convention utilized by OMIM).

In the light of recent discoveries [34] in the genetics of human diseases, it is important to retain as much information as possible about both the findings and the methods used to generate those findings. Although the allele model allows the construction of simpler CDSS rules (Fig. 3B), this model requires clear presentation of the method in order to prevent loss of information that may be useful in ‘bubbling up’ interpretations of existing data, in the future. The HL7 CGS approach has provisions for capturing all possible genetic information for the explicit purpose of allowing future reinterpretation; however, since the EHR system in question is built around a database-driven architecture, it does not implement HL7 information models directly, like a majority of present-day EHRs. Using an ontology like the CBO, which is structured around biological observations in conjunction with another codification system (e.g. LOINC) to describe the actual method used to collect such observations could possibly reduce such information loss.

4.4 Reporting Genetic Results

Reporting discrete genetic test results in the EHR could pose some unique problems for clinicians. Unlike other lab results such as ‘International Normalized Ratio’ (INR), where the results can be interpreted directly within the context provided by the normal ranges, and have traditionally been a part of clinicians’ training, genetic test results such as those obtained during CYP2C9 testing may require additional clinical recommendations in addition to the results themselves. This becomes particularly evident when considering the SNP panel in our simulations, where 32 discrete results could be reported as part of a

single assay, compared to 12 that can be reported as part of the allele panel. In each of these cases, the only clinically relevant piece of information is that having certain SNPs would predispose the patient to slower metabolism of the drug Warfarin, thereby increasing their risk of bleeding complications during anticoagulation therapy (Fig. 1), thereby necessitating dose adjustments. In the light of the constantly improving molecular-genetic diagnostic methods, it is important to retain as much information as possible, so that future re-interpretations of existing results could be performed in the proper context of the sensitivity/specificity of these assays. However, presenting all these discrete results may not be of much direct value to clinicians at the point-of-care, and may be counter-productive, whereas presenting the same results as a decision-support alert along (Fig. 4) with a suitable recommendation at the point of ordering Warfarin may be far more desirable.

4.5 Limitations

Although we considered the CYP2C9 gene and its implications in anticoagulation therapy, the CYP allele nomenclature itself is unique in some ways, and frequently, genetic variations may be described by haplotypes rather than alleles, as is the case with VKORC1 gene, which would then subsume the constituent SNPs in a manner similar to the alleles described in this work. However, with regard to differences in complexity of rule design for CDSS, and scenarios involving loss of information depending on the level of data abstraction, it is still possible to generalize these findings. Scenarios involving multiple genes, alleles and haplotypes were not considered in the present study, and these could further add to the complexity of CDSS rules that were considered in the evaluation of the two models.

5. Conclusions

The present work represents one of the first efforts at exploring the real-world application of genetic data in the EHRs using decision support, and the issues we have considered represent a few among the myriad of ques-

tions that will arise from the increased use of genetic data in clinical care in the future. We evaluated two data models for CDSS rules in the EHR on the basis of their performance, complexity, loss of information and reporting within the EHR. Although there was a significant difference between the computational times needed for evaluating rules based on the allele model and the SNP model, this difference, being in the order of milliseconds, did not translate into a significant difference in the time taken to place a Warfarin order. CDSS rules based on the SNP data model are inherently complex, and will be difficult to maintain with the continuous addition of new knowledge in this domain. Although the allele model allowed for simpler clinical decision support and clinical reporting, maintaining an allele nomenclature locally can be a challenge over time. The issue is further complicated by incorrect assignment of allele concepts during system implementations. At the present time, due to the lack of a pharmacogenetic knowledge base containing rules and recommendations in an executable, machine-readable format, as well as a consortium of experts to maintain such a resource, it may be necessary to hard-code many of the decision-support rules involving pharmacogenetic data, thus necessitating the abstraction of genetic test results for use in EHRs; and the appropriate level of data abstraction will ultimately have to be decided on a per-gene basis.

Acknowledgments

This research was supported by an education and travel award from the Cerner Corporation to the University of Utah Department of Biomedical Informatics, and by the University of Utah Health Sciences Information Technology Services. The authors are also grateful to Lisa Prins from the University of Utah Hospital, and to Nick Smith, Scott Haven and Ginger Kuhns from Cerner Corporation.

References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431 (7011): 931–945.
2. GeneTests.org at the University of Washington. <http://www.genetests.org/>. 1–15–2007.

3. Personalised medicines: hopes and realities. <http://www.royalsoc.ac.uk/displaypagedoc.asp?id=15874>. 9–1–2005. The Royal Society (UK).
4. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003; 422 (6934): 835–847.
5. Mitchell DR, Mitchell JA. Status of clinical gene sequencing data reporting and associated risks for information loss. *J Biomed Inform* 2007; 40 (1): 47–54.
6. Health Level 7. <http://www.hl7.org>. 3–2–2007.
7. Shabo A. Introduction to the Clinical Genomics Specifications and Documentation of the Genotype Topic. HL7 Clinical Genomics SIG DSTU Update 2 (Genotype topic update 1)[V0.9]. 11–5–2006. HL7.org.
8. Shabo A, Dotan D. The seventh layer of the clinical-genomics information infrastructure. *IBM Systems Journal* 2007; 46 (1): 57–67. International Business Machines Corporation.
9. Bioinformatic Sequence Markup Language. <http://www.bsml.org>. 3–2–2007.
10. SNOMED CT. <http://www.snomed.org/snomedct/>. 3–2–2007.
11. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update. *Clin Chem* 2003; 49 (4): 624–633.
12. Hoffman M, Arnoldi C, Chuang I. The clinical bioinformatics ontology: a curated semantic network utilizing RefSeq information. *Pac Symp Biocomput* 2005; 139–150.
13. Clinical Bioinformatics Ontology Whitepaper. <https://www.clinbioinformatics.org/cbopublic/>. 2004. Cerner Corporation.
14. Furuya H, Fernandez-Salguero P, Gregory W, et al. Genetic polymorphism of CYP2C9 and its effect on warfarin maintenance dose requirement in patients undergoing anticoagulation therapy. *Pharmacogenetics* 1995; 5 (6): 389–392.
15. Rettie AE, Wienkers LC, Gonzalez FJ, Trager WF, Korzekwa KR. Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9. *Pharmacogenetics* 1994; 4 (1): 39–42.
16. Takahashi H, Echizen H. Pharmacogenetics of CYP2C9 and interindividual variability in anticoagulant response to warfarin. *Pharmacogenomics J* 2003; 3 (4): 202–214.
17. Stubbins MJ, Harries LW, Smith G, Tarbit MH, Wolf CR. Genetic analysis of the human cytochrome P450 CYP2C9 locus. *Pharmacogenetics* 1996; 6 (5): 429–439.
18. Oscarson M, Ingelman-Sundberg M. CYP alleles: a web page for nomenclature of human cytochrome P450 alleles. *Drug Metab Pharmacokinet* 2002; 17 (6): 491–495.
19. Bell RG, Sadowski JA, Matschiner JT. Mechanism of action of warfarin. Warfarin and metabolism of vitamin K 1. *Biochemistry* 1972; 11 (10): 1959–1961.
20. Kaminsky LS, Zhang ZY. Human P450 metabolism of warfarin. *Pharmacol Ther* 1997; 73 (1): 67–74.
21. Horton JD, Bushwick BM. Warfarin therapy: evolving strategies in anticoagulation. *Am Fam Physician* 1999; 59 (3): 635–646.
22. Rost S, Fregin A, Ivaskevicius V, et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 2004; 427 (6974): 537–541.
23. United States Food and Drug Administration. FDA Approves Updated Warfarin (Coumadin) Prescribing Information. 8–16–2007.

24. Sconce EA, Khan TI, Wynne HA, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 2005.
25. Ackoff RL. From Data to Wisdom. *Journal of Applied Systems Analysis* 1989; 16: 3–9.
26. Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 1999; 353 (9154): 717–719.
27. Gutmacher AE, Collins FS. Welcome to the genomic era. *N Engl J Med* 2003; 349 (10): 996–998.
28. Mitchell JA, McCray AT, Bodenreider O. From phenotype to genotype: issues in navigating the available information resources. *Methods Inf Med* 2003; 42 (5): 557–563.
29. Martin-Sanchez F, Maojo V, Lopez-Campos G. Integrating genomics into health information systems. *Methods Inf Med* 2002; 41 (1): 25–30.
30. Mitchell JA. The impact of genomics on E-health. *Stud Health Technol Inform* 2004; 106: 63–74.
31. Navigenics. <http://www.navigenics.com/>. 9–3–2008.
32. 23andMe. <https://www.23andme.com/>. 9–3–2008.
33. Klein TE, Altman RB. PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base. *Pharmacogenomics J* 2004; 4: 1.
34. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007; 446 (7132): 153–158.