A Knowledge-based System for Intelligent Support in Pharmacogenomics Evidence Assessment: Ontology-driven Evidence Representation and Retrieval

Chia-Ju Lee, PhD[1], Beth Devine, PhD, MBA, PharmD[1,2], Peter Tarczy-Hornoch, MD, FACMI[1,3,4]
[1]Department of Biomedical Informatics and Medical Education, [2]Department of Pharmacy, [3]Department of Pediatrics, [4]Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA

Abstract
Pharmacogenomics holds promise as a critical component of precision medicine. Yet, the use of pharmacogenomics in routine clinical care is minimal, partly due to the lack of efficient and effective use of existing evidence. This paper describes the design, development, implementation and evaluation of a knowledge-based system that fulfills three critical features: a) providing clinically relevant evidence, b) applying an evidence-based approach, and c) using semantically computable formalism, to facilitate efficient evidence assessment to support timely decisions on adoption of pharmacogenomics in clinical care. To illustrate functionality, the system was piloted in the context of clopidogrel and warfarin pharmacogenomics. In contrast to existing pharmacogenomics knowledge bases, the developed system is the first to exploit the expressivity and reasoning power of logic-based representation formalism to enable unambiguous expression and automatic retrieval of pharmacogenomics evidence to support systematic review with meta-analysis.

Introduction

Pharmacogenomics is the study of how genetic variants affect a person's response to a drug. The rapid advances in pharmacogenomics research have made pharmacogenomics one of the genomics-based innovations that has great potential to contribute to improving people's health and reducing health care costs by increasing drug efficacy and safety[1]. Yet, the adoption of pharmacogenomics in routine clinical care is relatively low[2], partly due to the perception that there is insufficient evidence to determine the value of pharmacogenomics and the lack of effective and efficient use of already existing evidence[3,4].

Systematic review with meta-analysis is a well-established methodology used in evidence-based medicine that assesses the findings of a collection of studies that address a similar research question of interest in order to provide a more precise estimate of the effect of interventions or risk factors on patients' outcomes[5]. Generally, the review process is time-consuming and labor-intensive and involves the following generally manual steps[6,7]: a) conducting a comprehensive literature search, b) screening articles to identify relevant studies, c) extracting quantitative data and other essential elements from included studies, d) synthesizing the extracted data when they are acquired from sufficiently similar clinical context, e) rating the quality and strength of evidence, and f) interpreting the synthesized results. Informatics approaches such as natural language processing, machine learning and text mining have been applied to improve the efficiency of conducting a systematic review by reducing the burden of manual efforts in tasks of literature screening and data extraction[8-10]. However, there remains considerable room for further improvement, particularly in the area of representing the extracted primary evidence in a semantically computable formalism to enable intelligent support in initial and ongoing updating of evidence retrieval, synthesis and interpretation. In particular a system that leverages semantically computable formalisms would greatly facilitate the addition of new evidence and the reassessment of the conclusions factoring in the new evidence.

Knowledge representation and reasoning is a sub-domain of artificial intelligence that is concerned with encoding knowledge into semantically computable formalisms that can be efficiently manipulated by reasoning programs so that computers can demonstrate human-like abilities. During the past decade, Web Ontology Language (OWL) has been developed by combining the Semantic Web technologies and logic-based representation formalisms to advance computer interpretability of Web content[11]. OWL-encoded ontologies provide shared conceptualizations and controlled vocabularies of a domain of interest which allow for formal representation and automatic reasoning. Because of its expressivity and reasoning ability, research efforts are encouraged to exploit the advanced features of OWL in developing more intelligent systems that assist human decision making.

Considering the time-consuming and knowledge-intensive nature of pharmacogenomics evidence assessment,

the idea of developing a knowledge-based system for intelligent initial and ongoing support in evidence assessment emerges intuitively from the perspective of biomedical informatics. We hypothesized that a knowledge-based system with the following three critical features can assist effective and efficient evidence assessment, and therefore facilitate timely decisions on adoption of pharmacogenomics in clinical practice. First, the information provided by the knowledge-based system should be clinically relevant evidence, which means that evidence related to clinical validity and clinical utility of pharmacogenomics should be accumulated in the system. Second, the information provided by the knowledge-based system should be acquired through an evidence-based approach, which means that primary evidence acquired from empirical research should be collected and synthesized through methodologies established in comprehensive systematic reviews. Third, the information provided by the knowledge-based system should be semantically computable, which means that a knowledge-based system should take full advantage of the expressivity and reasoning power of logic-based knowledge representation formalisms such as OWL 2 DL[11] so that pharmacogenomics knowledge is unambiguously represented and accumulated in a knowledge base which allows for automatic reasoning.

Upon reviewing existing pharmacogenomics knowledge bases including the Pharmacogenomics Knowledgebase (PharmGKB)[12], the PharmacoGenomics Mutation Database (PGMD)[13] and the DrugBank database[14], we discerned that none of them fully meets the critical features of our envisioned pharmacogenomics knowledge-based system (**Table 1**). This gap motivated us to design and develop the knowledge-based system described in this paper *de novo*, aiming to provide intelligent assistance for pharmacogenomics evidence assessment.

Table 1: Overview of identified gaps in current pharmacogenomics knowledge bases

Features of the envisioned pharmacogenomics knowledge-based system		PharmGKB	PGMD	DrugBank
Clinically relevant evidence	Clinical validity	Y	Y	Y
	Clinical utility	Y	N	N
Evidence-based approach	Primary evidence	Y	Y	Y
	Sufficient information for meta-analysis	N	Y	N
	Risk of bias assessment	N	N	N
	Synthesized evidence	Y	N	N
	Explicit inclusion criteria	N	N	N
Semantically computable formalism	Logic-based formalized ontology	N	N	N
	Ontology-committed knowledge base	N	N	N
	Question answering by automatic reasoning	N	N	N

PharmGKB: the Pharmacogenomics Knowledgebase; PGMD: the PharmacoGenomics Mutation Database; DrugBank: the DrugBank database. Y: abbreviation of "yes", indicating that the knowledge base meets the specified features, N: abbreviation of "no", indicating that the knowledge base does not meet the specified features.

Methods

Conceptual modeling of the domain of pharmacogenomics evidence assessment

To address the aforementioned features of clinically relevant evidence and an evidence-based approach, we proposed a basic information structure for developing the conceptual model of the domain of pharmacogenomics evidence assessment using a faceted analysis approach[15]. This information structure is composed of five building blocks, namely, information entities, information components, concepts, relations and terms. **Figure 1** illustrates that in the domain of pharmacogenomics evidence assessment, an information entity is composed of information components, an information component is expressed by relation-concept pairs, and relations and concepts are substantiated by terms to express the intended meaning. Based on the information needs in conducting systematic reviews with meta-analyses[6,7], the information entities in the conceptual model include publication, study, and evidence, and the minimal set of information components to describe these intended information entities include study population, drug therapy, comparison, outcome, genetic variation, study design, effect estimation, risk of bias assessment and bibliographical information of publication.
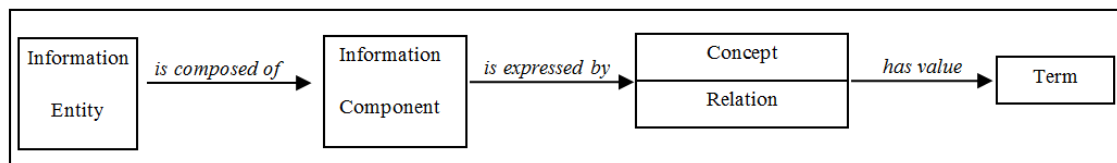

Figure 1: Basic structure of the conceptual model and its building blocks for conceptualization of the domain of pharmacogenomics evidence assessment

We created operational definitions of evidence of clinical validity and utility and deployed a fine-grained characterization of these two types of pharmacogenomics evidence acquired from empirical pharmacogenomics studies in clopidogrel and warfarin therapies to identify concepts, relations and terms that are essential for modeling the domain of pharmacogenomics evidence assessment. References cited by the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for clopidogrel and warfarin therapies[16,17] were used as the major sources from which we selected original research articles for manual extraction of concepts, relations and terms. Review articles cited in the two CPIC guidelines were used for backward citation tracking to identify relevant articles that did not directly cited in the guidelines' reference list. Articles recently published after the release of the CPIC guidelines were also sought.

Implementation of a pharmacogenomics knowledge-based system

Our knowledge-based system for pharmacogenomics evidence assessment consists of three core components, i.e., an ontology, a knowledge base and a reasoner (**Figure 2**). The primary aim of the knowledge-based system is to enable formal representation and automatic retrieval of pharmacogenomics evidence to assist in meta-analysis, which lays the foundation for further applications in systematic review such as classification of homogeneous evidence and interpretation of clinical significance of evidence. We adopted OWL 2 DL[11] as our formal representation language, used Protégé[18] as an ontology editor, and leveraged HermiT[19] as a reasoner to implement the knowledge-based system.
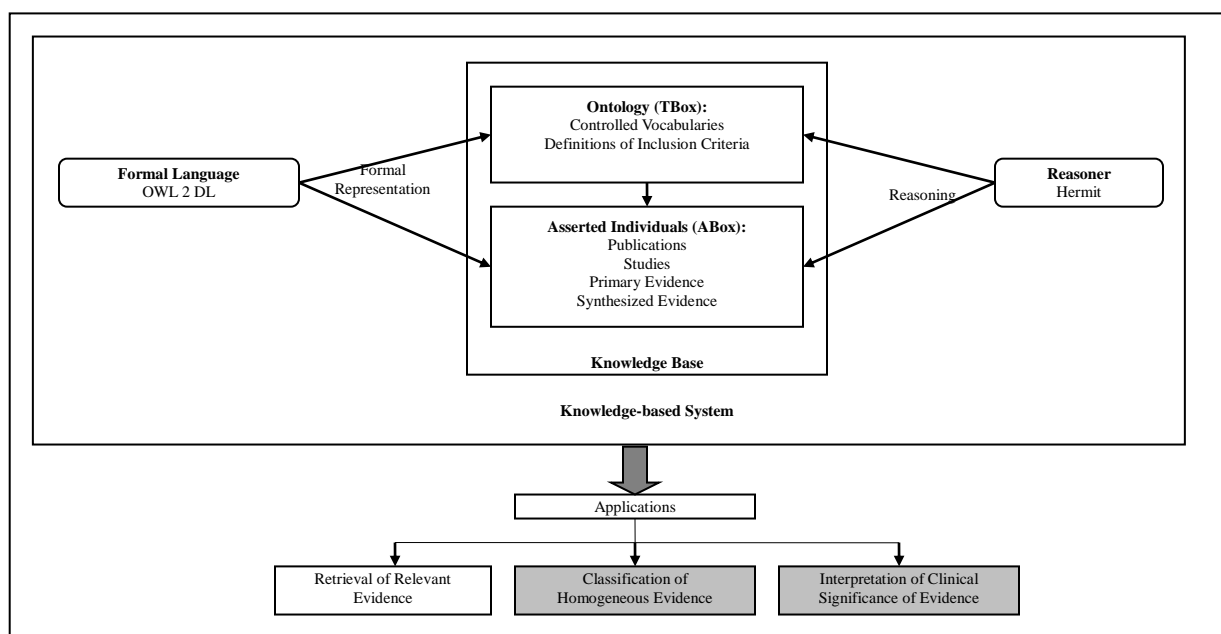

Figure 2: Fundamental architecture and intended application scenarios of the developed pharmacogenomics knowledge-based system. The two applications highlighted by grey blocks are proposed for future research.

The aforementioned conceptual model of pharmacogenomics evidence assessment served as the blueprint for constructing the ontology. Based on a commonly cited guide for constructing an OWL 2 ontology[11], we derived mapping principles to convert the varieties of building blocks of the conceptual model into the constructs of an OWL ontology, i.e., classes, properties and individuals. The individual information entities extracted for deriving the conceptual model served as the test materials to construct the knowledge base. Using the constructs encoded in the OWL ontology, and constructors (i.e., restrictions and operators) supported by OWL 2 DL, we derived representation patterns for asserting individual information entities with heterogeneous information content. Our major considerations while deriving the representation patterns were to avoid computational inefficiency caused by over-representation and irrelevant retrieval of individuals caused by under-representation.

Evaluation of the implemented knowledge-based system on ontology-driven pharmacogenomics evidence retrieval

In order to provide a proof-of-concept that the developed knowledge-based system is capable of providing intelligent support in retrieving relevant pieces of pharmacogenomics evidence for systematic review with

meta-analysis, a convenience sample of 9 systematic reviews[20-28] that investigated the association between genetic variations and responses to clopidogrel was obtained from the reference list of the CPIC clopidogrel guideline[16]. A collection of 33 meta-analyses were selected from these reviews and used as test cases to evaluate the precision and efficiency of the ontology-driven approach to evidence retrieval.

The reported criteria for including individual pieces of evidence into each test case of meta-analyses were extracted from the respective review articles. Ontology-driven evidence retrieval was implemented first by formally representing these inclusion criteria as the necessary and sufficient conditions of defined classes using the constructed OWL ontology. Then the HermiT reasoner embedded in Protégé was manually triggered to perform instance checking over the implemented knowledge base to retrieve those individual pieces of evidence that match the definition of each defined class. The results of ontology-driven evidence retrieval were evaluated in terms of precision and efficiency. Precision was calculated as the percentage of retrieved individual pieces of evidence that are relevant to the inclusion criteria specified for each respective meta-analysis. The relevance was judged by BD, one of the authors of this paper. Efficiency was measured by the computing time taken by HermiT reasoner to perform the reasoning tasks, which was captured from Protégé Command Prompt.

Results

Table 2 provides an overview of basic statistics on evidence source, conceptual model, ontology metrics and asserted individual information entities of the developed knowledge-based system. A total of 73 empirical research articles were selected as evidence source, from which three types of intended information entities were identified, including 73 pieces of publications, 82 pieces of studies and 445 pieces of evidence.

Table 2: Overview of statistics of data source, conceptual model, ontology and asserted individual information entities of the developed knowledge-based system

Evidence Source		Building Blocks of Conceptual Model		Metrics of Ontology and Asserted Individual IE	Ontology	Asserted Individual IEs
Publication	73	Information entity (IE)	3	DL expressivity	$\mathcal{ALCRF(D)}$	$\mathcal{ALCRQ(D)}$
-clopidogrel	51	Information component	9	Class	306	-
-warfarin	22	Concept	30	Object property	69	-
Study	82	Relation	49	Datatype property	12	-
- clopidogrel	57	Term	282	Individual	9	667
- warfarin	25			SubClassOf axioms	289	-
Evidence	445			EquivalentClasses axioms	9	-
- clopidogrel	285			SubObjectPropertyOf axioms	27	-
- warfarin	160			SubPropertyChainOf axioms	11	-
				SubDatatypePropertyOf axioms	5	-
				FunctionalDatatypeProperty axioms	7	-
				DatatypePropertyRange axioms	7	-
				ClassAssertion axioms	9	2670
				ObjectPropertyAssertion axioms	-	1187
				DatatypePropertyAssertion axioms	-	1522

Conceptual model

Fine-grained characterization of this collection of individual information entities yielded 30 concepts, 49 relations, and 282 terms to describe the 9 intended information components. By organizing these extracted building blocks, we derived a conceptual model for representing the domain of pharmacogenomics evidence assessment (**Figure 3**). Three types of information entities are independent yet inter-related. Specifically, `Evidence` is related to `Study` via the relation of "*is acquired from*", and `Study` is in turn related to `Publication` via "*is reported in*". Each type of information entity is described by specific information component modules, with `Publication` described by publication module, `Study` described by modules of study population, study design, drug therapy, and risk of bias assessment, and `Evidence` described by modules of comparison, genetic variation, outcome, and effect estimation. Each information component module is expressed in a layered structure that is composed of multiple relation-concept pairs. When the conceptual model is used to express concrete information entities, concepts and relations are directly substantiated by terms commonly used in a variety of clinical, pharmacological or genomic domains. Thus the meaning of each individual real-world information entity could be explicitly and precisely expressed. For example, to describe a study population of "*patients who were treated with clopidogrel for acute coronary syndrome*", the concept of `Person` is substantiated by the term of `Patient`, the concept of `Drug` by the term of `Clopidogrel`, and the concept of `Disease` by the term of `Acute Coronary Syndrome`. The developed conceptual model was validated by fitting two original articles[29,30] and two systematic reviews[20,31]. The preliminary

verification results showed that our model is adequate for annotating primary pharmacogenomics evidence and inclusion criteria for meta-analysis.
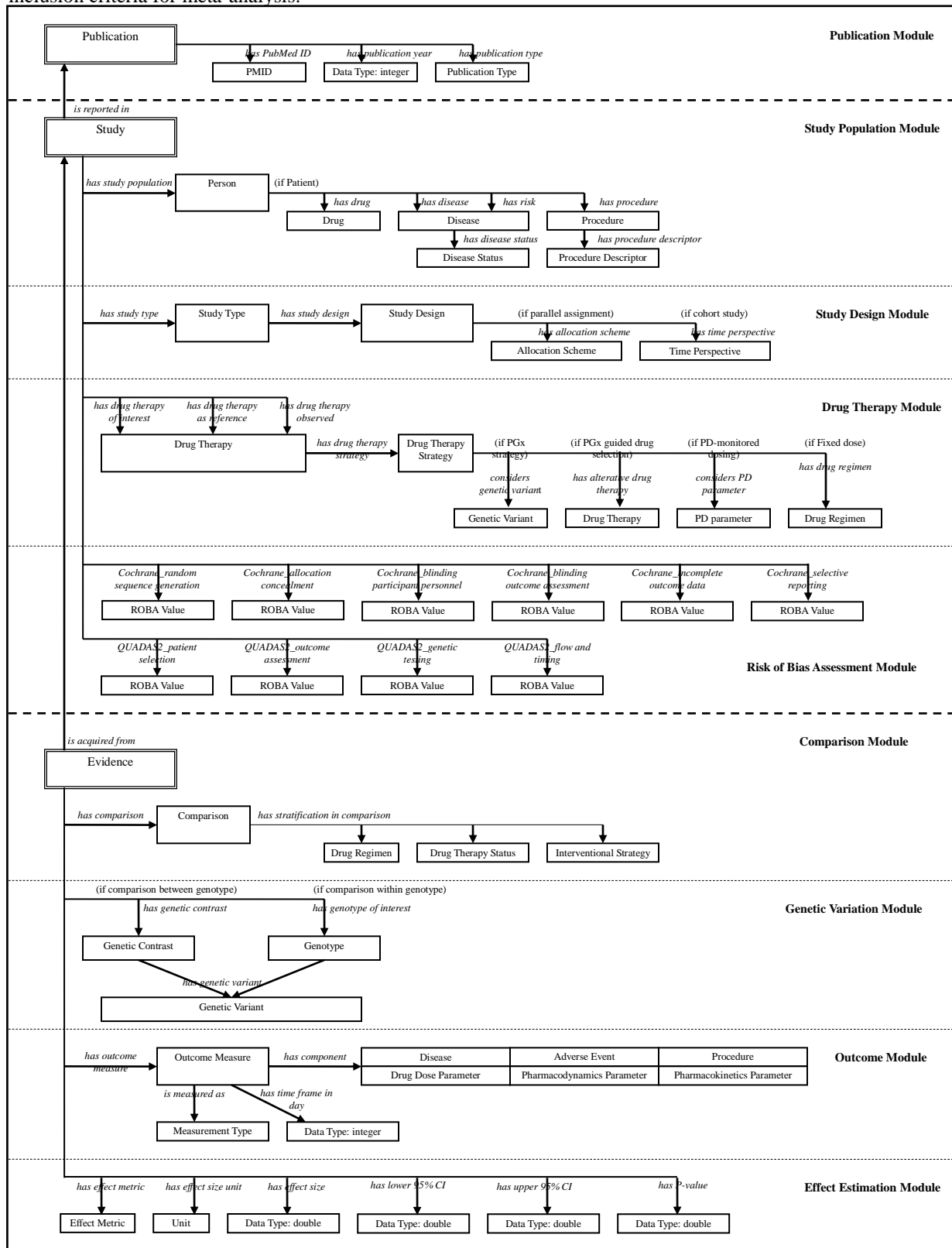

Figure 3: Conceptual Model of Pharmacogenomics Evidence Assessment. Double-lined squares: information entities, single-lined squares: concepts, arrows: relations. Dotted lines divide the entire model into 9 modules, each corresponding to one information component.

Ontology

The constructed ontology contains 396 constructs, including 306 classes, 69 object properties, 12 data properties, and 9 individuals (**Table 2**). By following the information structure illustrated in **Figure 3**, these constructs could be used to formally represent publications, studies and evidence that were involved in assessing the evidence of clinical validity and utility in the domain of clopidogrel and warfarin pharmacogenomics.

As shown in Table 2, the ontology features the use of several axioms, including `SubClassOf`, `EquivalentClass`, `SubObjectPropertyOf`, and `SubPropertyChainOf` axioms, to facilitate reasoning for evidence retrieval. For example, we used `SubClassOf` axioms to construct the class of `Disease` with 6-level depth of class hierarchy, where the bottom-level classes are more specific than the top-level classes. Thus a piece of evidence annotated with specialized disease terms could be retrieved by inclusion criteria defined with broad disease terms. We used `EquivalentClasses` axioms to define acute coronary syndrome (ACS) as equivalent to the union of ST-segment elevation myocardial infarction (STE_MI), non-ST-segment elevation myocardial infarction (NSTE_MI) or unstable angina (UA). Thus inclusion criteria that are specified with ACS as disease characteristics of patients will retrieve not only evidence that is exactly annotated with ACS but also those annotated with STE_MI, NSTE_MI or UA. We used `SubObjectPropertyOf` axioms to represent more specific relations. For example, the object property *hasDrugTherapy* represents a general relation between a study and a drug therapy under investigation. Subproperties such as *hasDrugTherapyObserved*, *hasDrugTherapyOI* and *hasDrugTherapyRef* were created to specify a drug therapy that was investigated under an observational study, or given to the experimental arm, or given to the control arm respectively. Thus inclusion criteria that are specified with *hasDrugTherapyObserved* will retrieve exactly those evidence acquired from observational studies. We used `SubPropertyChainOf` axioms to connect individuals by a chain of properties. For example, an individual of evidence `Ie` is linked to an individual of study `Is` via object property *isAcquiredFrom*, and `Is` is linked to a risk-of-bias-assessment (ROBA) value `low` on random sequence generation via object property *hasROBA_Cochrane_RandomSequenceGeneration*, by linking these two properties to form a property chain, `Ie` will be automatically inferred the ROBA value of `low`.

Knowledge base

The constructed knowledge base contains 73, 82 and 445 individual pieces of asserted publications, studies and evidence respectively. These information entities were formally represented via class assertion axioms, object property assertion axioms, and datatype property assertion axioms (**Table 2**). **Figures 4, 5 and 6** illustrate respectively the formal representation of individual pieces of publication, study and evidence that were extracted from the article [Kimmel et al., 2013][30].

Figure 4 illustrates the formal representation of an individual publication labeled as `pub_24251361`, expressing that it is a full-text refereed journal article that was published in 2013 and its PubMed identifier is 24251361.

Figure 4: Example of assertion of an individual piece of publication. Screenshot extracted from Protégé.
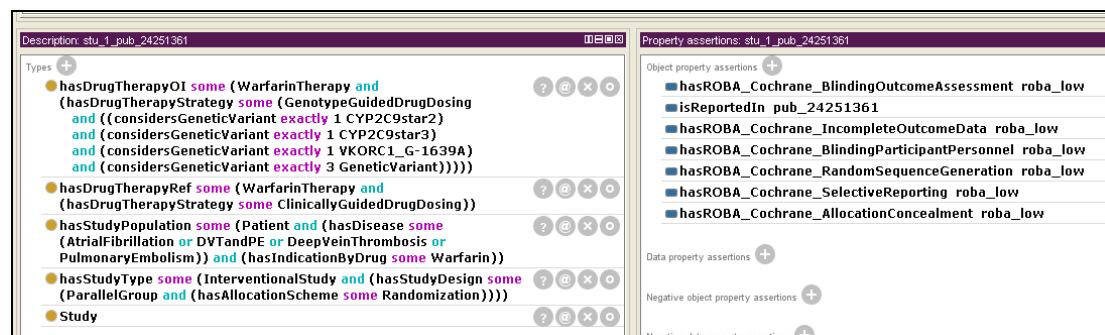

Figure 5: Example of assertion of an individual piece of study. Screenshot extracted from Protégé.

Figure 5 illustrates the formal representation of an individual study labeled as `stu_1_pub_24251361`, which was reported in the publication `pub_24251361` that has been asserted in **Figure 4**. This study is expressed as a randomized and paralleled controlled clinical trial that aimed to investigate a genotype-guided warfarin therapy considering three genetic variants (*CYP2C19*2 and *CYP2C19*3 and *VKORC1-1639G/A*) versus clinically guided warfarin dosing in patients with atrial fibrillation or deep vein thrombosis or pulmonary embolism or deep vein thrombosis & pulmonary embolism. In addition, the risk of bias in this particular study was assessed using Cochrane assessment tool, with low risk of bias in each of the six criteria.

Figure 6 illustrates the formal representation of an individual piece of evidence labeled as `evi_01_pub_24251361_stu_1`, which was acquired from the study `stu_1_pub_24251361` that has been asserted in **Figure 5**. This evidence is expressed as comparison between two drug therapies (which could be known by its link to `stu_1_pub_24251361`). The outcome measure was the percentage of time of international normalized ratio in the therapeutic range up to the follow-up of 28 days. The effect was measured as absolute difference between two group means and was estimated as -0.2% with 95% confidence interval -3.4% to 3.1% and P-value of 0.91. In addition, some information was inferred for this evidence (as shown in the highlighted blocks) through its linkage with `stu_1_pub_24251361`, e.g., the publication `pub_24251361` from which it was extracted, and the risk of bias assessment values of the study `stu_1_pub_24251361` from which the evidence was acquired.
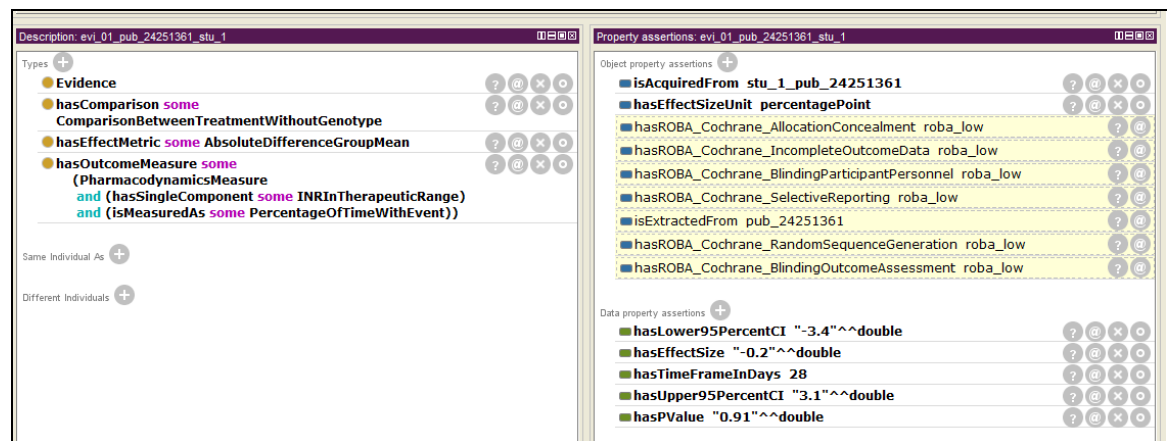

Figure 6: Example of assertion of an individual piece of evidence. Screenshot extracted from Protégé.

The construction of our knowledge base features the design of representation patterns to enable representation of heterogeneous information content of complicated information components. For example, the representation patterns for describing 6 types of information content in the drug therapy module are summarized in **Table 3**. Each type of information content is represented by an anonymous class expression, which is described by object properties, property restrictions, classes used as property values, and operators that link multiple property values, as appropriate. The exemplary drug therapies illustrated in **Figure 5** were asserted based on these representation patterns. It is worth mentioning that these representation patterns are capable of describing the highly heterogeneous drug therapies investigated in clinical pharmacogenomics studies. In the pilot implementation of 82 individual studies, the representation patterns were successfully used to represent a total of 35 different types of drug therapies.

Table 3: Representation patterns for describing information content of drug therapy

Information content	Object property	Property restriction	Class used as property value (possible number of values)	Operator used to link multiple values
Drug therapy	*hasDrugTherapy* with subproperties	Existential restriction	DrugTherapy (single or multiple)	or
Drug therapy strategy	*hasDrugTherapyStrategy*	Existential restriction	DrugTherapyStrategy (single)	Not applicable
Genetic variant considered in genotype-guided strategy	*considersGeneticVariant*	Qualified cardinality restriction	GeneticVariant (single or multiple)	and
Alternative drug therapy in genotype-guided drug selection	*hasAlternativeDrugTherapy*	Existential restriction	DrugTherapy (single or multiple)	or
Pharmacodynamic parameter monitored	*monitorsPharmacodynamicsParameter*	Existential restriction	PharmacodynamicsParameter (single)	Not applicable
Drug regimen	*hasDrugRegimen*	Existential restriction	DrugRegimen (single or multiple)	and/or

Performance of ontology-driven evidence retrieval

Table 4 illustrates the implementation and result of ontology-driven evidence retrieval, using one test case meta-analysis selected from the review article of [Singh et al., 2012][27] as an example. The inclusion criteria extracted from the review article are summarized in the upper left part. Ontology-based formal representation of the extracted inclusion criteria is presented in the lower left part. Implementation of 9 defined classes to represent the inclusion criteria of 9 meta-analyses is presented in the middle part (marked by brackets). After triggering the HermiT reasoner, the retrieved relevant evidence could be viewed by clicking each specific defined class. As shown in the right part, 22 pieces of relevant evidence were retrieved for the defined class named as `MACE_CYP2C19star2_CADandPCI_Singh`.

Table 4: Example of ontology-driven evidence retrieval

Inclusion criteria extracted from review article [Singh et al., 2012]	
Publication year	< 2011
Publication type	Refereed journal article or conference abstract
Study population	Patient with coronary artery disease and percutaneous coronary intervention
Study design	Randomized parallel-controlled trial or prospective cohort study
Drug therapy	Clopidogrel therapy with standard dose regimen
Genetic contrast	Carrier of at least one CYP2C19*2 allele versus noncarrier
Outcome	Incidence of major adverse cardiovascular events

Formal representation of inclusion criteria as a defined class named as MACE_CYP2C19star2_CADandPCI_Singh

Evidence and (((hasComparison some ComparisonBetweenGenotypeWithinDrugTherapyOI) and (isAcquiredFrom some (Study and (hasDrugTherapyOI some ClopidogrelTherapy)))) or ((hasComparison some ComparisonBetweenGenotypeWithinDrugTherapyObserved) and (isAcquiredFrom some (Study and (hasDrugTherapyObserved some ClopidogrelTherapy)))) or ((hasComparison some ComparisonBetweenGenotypeWithinDrugTherapyRef) and (isAcquiredFrom some (Study and (hasDrugTherapyRef some ClopidogrelTherapy))))) and (hasGeneticContrast some (CarrierOfAtLeast1VsNoncarrier and (hasGeneticVariant some CYP2C19star2))) and (hasOutcomeMeasure some (ClinicalEfficacyMeasure and (hasMultipleComponent some (AdverseEvent or Disease or Procedure)) and (isMeasuredAs some IncidenceOfEvent))) and (isAcquiredFrom some (Study and (hasStudyPopulation some (Patient and (hasDisease some CoronaryArteryDisease) and (hasProcedure some PCI))) and (hasStudyType some ((InterventionalStudy and (hasStudyDesign some (ParallelGroup and (hasAllocationScheme some Randomization)))) or (ObservationalStudy and (hasStudyDesign some (Cohort and (hasTimePerspective some Prospective)))))) and (isExtractedFrom some (Publication and (hasPublicationType some (ConferenceAbstract or RefereedJournalArticle)) and (hasPubYear some integer[< 2011])))

Implementation and result of ontology-driven evidence retrieval*

* Screenshots extracted from Protégé. The 9 defined classes marked by the brackets represent the inclusion criteria of 9 meta-analyses.

The evaluation of ontology-driven evidence retrieval in terms of precision and efficiency is encouraging. Overall, 33 test cases of ontology-based evidence retrieval achieved a precision rate of 100%. The computing time taken to retrieve relevant evidence for each systematic review approximately ranged from 9 to 23 seconds (**Table 5**).

Table 5: Evaluation of efficiency of ontology-driven evidence retrieval

Systematic review	Singh 2012[27]	Jang 2012[26]	Bauer 2011[23]	Zabalza 2012[25]	Jin 2011[22]	Hulot 2010[20]	Sofi 2011[21]	Holmes 2011[24]	Yamaguchi 2013[28]
Number of meta-analysis included	9	6	4	4	3	3	2	1	1
Total number of evidence retrieved for all included meta-analyses	57	58	44	31	19	22	23	31	16
Approximated computing time (seconds)	23	21	21	16	18	17	18	11	9

Note: The retrievals were tested on a personal laptop (Intel Corei7-4700MQ 2.4GHz Processor, 16 GB DDR3 Ram and a 64-bit version of Windows 8.1).

Discussion

This paper presents a knowledge-based system that adopts OWL 2 DL as the representation language to enable ontology-based representation of primary evidence and ontology-driven retrieval of relevant pieces of evidence for conducting systematic review with meta-analysis. The unique features of this system are elaborated as follows.

First, the system was developed based on a conceptual model of pharmacogenomics evidence assessment that considers different dimensions of information needs and thus accommodates different types of information in a unified model. Considering that both clinical validity and clinical utility evidence are essential to integrate pharmacogenomics into clinical practice, the conceptual model was designed to enable annotation of evidence related to association between genetic variant and drug response as well as evidence related to effectiveness of genotype-guided drug therapies. Moreover, to address the information needs for conducting systematic reviews with meta-analyses, the conceptual model was designed to enable annotation of primary evidence along with its study

context and provenance as well as to allow annotation of inclusion criteria for retrieving relevant evidence. To our best knowledge, none of the existing pharmacogenomics knowledge bases is capable of annotating all of these types of information in a single information model.

Second, the system exploited the expressivity and reasoning ability of OWL 2 DL to deliver an ontology and a number of representation patterns, which collectively allow complex and heterogeneous pharmacogenomics evidence to be unambiguously represented. Thereby, the formally represented primary evidence could be classified at different levels of specificity as defined by different research questions. Nevertheless, it was challenging to derive representation patterns that avoid computational inefficiency caused by over-representation and irrelevant retrieval caused by under-representation. We also identified some cases in that the meaning of our intended retrieval criteria were not expressible in the representation patterns we designed.

Third, the system could represent inclusion criteria as defined classes and embed them into the ontology, which allows the users of the system to acquire the most updated profile of evidence of their interests each time newly extracted pieces of evidence are added in the knowledge base. This is achieved via the OWL 2 DL reasoner's capability of automatic reasoning. This feature is most beneficial in view of the evolving nature of the development of pharmacogenomics and the recurrent needs to assess the change of evidence over time.

Our preliminary work has several limitations. The scope of knowledge base was limited to clinical validity and utility of clopidogrel and warfarin pharmacogenomics. Some useful information was missing in the conceptual model, such as age and ethnicity of study population. The evidence asserted in the knowledge base is not exhaustive, but to serve as representative examples to provide a proof-of-concept of the design, development, implementation, and evaluation of the envisioned knowledge-based system. No informatics tool has been developed to automatically export retrieval results from Protégé to existing statistical software that supports meta-analysis.

Through evaluation of its performance using real-world test cases, the preliminary pharmacogenomics knowledge-based system has proven to be an effective and efficient approach to retrieve relevant primary evidence for conducting systematic review with meta-analysis. Future research to enhance its applicability is proposed as follows. The scope that limited to clopidogrel and warfarin pharmacogenomics should be expanded to include other domains, particularly cancer pharmacogenomics. The information component modules should be refined to express more useful information. Moreover, the application scenarios should be extended to address the subsequent steps in the process of a comprehensive evidence assessment, such as formal representation of synthesized evidence to enable semantic computation of the clinical significance of genetic variants in predicting drug response and improving patient outcome. With the enhanced applicability, the knowledge-based system might greatly improve the efficiency of pharmacogenomics evidence assessment, and ultimately increase the adoption of pharmacogenomics in routine clinical care.

References

1. Green ED, Guyer MS. Charting a course for genomics medicine from base pairs to bedside. *Nature*. 2011;470(7333):204-13.
2. Shuldiner AR, Relling MV, Peterson JF, Hicks JK, Freimuth RR, Sadee W, et al. The Pharmacogenomics Research Network Translational Pharmacogenetics Program: overcoming challenges of real-world implementation. *Clin Pharmacol Ther*. 2013;94(2):207-10.
3. Pirmohamed M. Pharmacogenetics: past, present and future. *Drug Discov Today*. 2011;16(19-20):852-61.
4. Sadee W. Pharmacogenomic biomarkers: validation needed for both the molecular genetic mechanism and clinical effect. *Pharmacogenomics*. 2011;12(5):675-80.
5. Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD. Chapter 1: Introduction. In: Higgins JPT, Green S (eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (*Version 5.1.0)*. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.
6. Higgins JPT, Green S (eds.). *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)*. The Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org.
7. Agency for Healthcare Research and Quality. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(14)-EHC063-EF. 2014. Available from: www.effectivehealthcare.ahrq.gov.
8. Kiritchenko S, de Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. *BMC Med Inform Decis Mak*. 2010;10:56. doi: 10.1186/1472-6947-10-56.

9. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. *Syst Rev*. 2015;4:78. doi: 10.1186/s13643-015-0066-7.

10. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. *Syst Rev*. 2014;3:74. doi: 10.1186/2046-4053-3-74.

11 Hitzler P, Krötzsch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 web ontology language primer, second edition. W3C recommendation. Available from: http://www.w3.org/TR/owl2-primer/.

12. Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. *Methods Mol Biol*. 2013;1015:311-20.

13. Kaplun A, Hogan JD, Schacherer F, Peter AP, Krishna S, Braun BR, et al. PGMD: a comprehensive manually curated pharmacogenomic database. *Pharmacogenomics J*. 2016;16(2):124-8.

14. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2011;39(Database issue):D1035-41.

15. Hjørland B. Facet analysis: the logical approach to knowledge organization. *Information Processing and Management*. 2013;49(2):545-57.

16. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 Update. *Clin Pharmacol Ther*. 2013;94(3):317-23.

17. Johnson J, Gong L, Whirl-Carrillo M, Gage B, Scott S, Stein C, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin Pharmacol Ther*. 2011;90(4):625-629.

18. Protégé. Available from: http://protege.stanford.edu. [Accessed August 23, 2016].

19. Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z. HermiT: An OWL 2 reasoner. *J Automated Reasoning*. 2014;53:245-69.

20. Hulot JS, Collet JP, Silvain J, Pena A, Bellemain-Appaix A, Barthelemy O, et al. Cardiovascular risk in clopidogrel-treated patients according to cytochrome P450 2C19*2 loss-of-function allele or proton pump inhibitor coadministration: a systematic meta-analysis. *J Am Coll Cardiol*. 2010;56:134-43.

21. Sofi F, Giusti B, Marcucci R, Gori AM, Abbate R, Gensini GF. Cytochrome P450 2C19*2 polymorphism and cardiovascular recurrences in patients taking clopidogrel: a meta-analysis. *Pharmacogenomics J*. 2011;11:199-206.

22. Jin B, Ni HC, Shen W, Li J, Shi HM, Li Y. Cytochrome P450 2C19 polymorphism is associated with poor clinical outcomes in coronary artery disease patients treated with clopidogrel. *Mol Biol Rep*. 2011;38:1697-702.

23. Bauer T, Bouman HJ, van Werkum JW, Ford NF, ten Berg JM, Taubert D. Impact of CYP2C19 variant genotypes on clinical efficacy of antiplatelet treatment with clopidogrel: systematic review and meta-analysis. *BMJ*. 2011;343:d4588.

24. Holmes MV, Perel P, Shah T, Hingorani AD, Casas JP. CYP2C19 genotype, clopidogrel metabolism, platelet function, and cardiovascular events: a systematic review and meta-analysis. *JAMA*. 2011;306:2704-14.

25. Zabalza M, Subirana I, Sala J, Lluis-Ganella C, Lucas G, Tomas M, et al. Meta-analyses of the association between cytochrome CYP2C19 loss- and gain-of-function polymorphisms and cardiovascular outcomes in patients with coronary artery disease treated with clopidogrel. *Heart*. 2012;98:100-8.

26. Jang JS, Cho KI, Jin HY, Seo JS, Yang TH, Kim DK, et al. Meta-analysis of cytochrome P450 2C19 polymorphism and risk of adverse clinical outcomes among coronary artery disease patients of different ethnic groups treated with clopidogrel. *Am J Cardiol*. 2012;110:502-8.

27. Singh M, Shah T, Adigopula S, Molnar J, Ahmed A, Khosla S, et al. CYP2C19*2/ABCB1-C3435T polymorphism and risk of cardiovascular events in coronary artery disease patients on clopidogrel: is clinical testing helpful? *Indian Heart J*. 2012;64:341-52.

28. Yamaguchi Y, Abe T, Sato Y, Matsubara Y, Moriki T, Murata M. Effects of VerifyNow P2Y12 test and CYP2C19*2 testing on clinical outcomes of patients with cardiovascular disease: A systematic review and meta-analysis. *Platelets*. 2013;24:352-61.

29. Wallentin L, James S, Storey RF, Armstrong M, Barratt BJ, Horrow J, et al. PLATO investigators. Effect of CYP2C19 and ABCB1 single nucleotide polymorphisms on outcomes of treatment with ticagrelor versus clopidogrel for acute coronary syndromes: a genetic substudy of the PLATO trial. *Lancet*. 2010;376(9749):1320-8.

30. Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, et al. COAG Investigators. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med*. 2013;369(24):2283-93.

31. Stergiopoulos K, Brown DL. Genotype-guided vs clinical dosing of warfarin and its analogues: meta-analysis of randomized clinical trials. *JAMA Intern Med*. 2014;174(8):1330-8