Translational Knowledge Discovery Between Drug Interactions and Pharmacogenetics

Heng-Yi Wu[1], Aditi Shendre[2,*], Shijun Zhang[2], Pengyue Zhang[2], Lei Wang[2], Desta Zeruesenay[3], Luis M. Rocha[4,5], Hagit Shatkay[6], Sara K. Quinney[3,7], Xia Ning[2,8,*] and Lang Li[2,*]

Clinical translation of drug-drug interaction (DDI) studies is limited, and knowledge gaps across different types of DDI evidence make it difficult to consolidate and link them to clinical consequences. Consequently, we developed information retrieval (IR) models to retrieve DDI and drug-gene interaction (DGI) evidence from 25 million PubMed abstracts and distinguish DDI evidence into *in vitro* pharmacokinetic (PK), clinical PK, and clinical pharmacodynamic (PD) studies for US Food and Drug Administration (FDA) approved and withdrawn drugs. Additionally, information extraction models were developed to extract DDI-pairs and DGI-pairs from the IR-retrieved abstracts. An overlapping analysis identified 986 unique DDI-pairs between all 3 types of evidence. Another 2,157 and 13,012 DDI-pairs and 3,173 DGI-pairs were identified from known clinical PK/PD DDI, clinical PD DDI, and DGI evidence, respectively. By integrating DDI and DGI evidence, we discovered 119 and 18 new pharmacogenetic hypotheses associated with *CYP3A* and *CYP2D6*, respectively. Some of these DGI evidence can also aid us in understanding DDI mechanisms.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Several studies have explored different informatics approaches to mine drug interactions data from the biomedical literature. However, none of them have distinguished the drug-drug interaction (DDI) evidence into *in vitro* pharmacokinetic (PK), clinical PK, and clinical pharmacodynamic (PD) studies, which can impede the translational scope of drug interactions research.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ The goal of this study was to retrieve and extract DDI and drug-gene interaction (DGI) evidence from the biomedical literature and distinguish the DDI study types into *in vitro* PK, clinical PK, and clinical PD studies. Additionally, the integrated DDI and DGI evidence were used to determine knowledge gaps that could enable the generation of novel DDI or DGI and adverse drug event (ADE)-related hypotheses.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ This study adds to the existing knowledge by providing (i) a novel algorithm that extracts drug interaction evidence from diverse DDI and DGI studies, (ii) a method to distinguish the different types of DDI studies, and (iii) an integrated drug interactions data that enables knowledge discovery through generation of novel genetic hypotheses or molecular DDI mechanisms.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ The integrated knowledge generated by our study is valuable for translational research in drug interaction studies. It can facilitate future studies that help in improving our understanding of DDI-related ADEs through the detection of novel genetic or molecular mechanisms. The validated hypotheses can then be evaluated for potential clinical applications in the future.

Drug-drug interactions (DDIs) are one of the major causes of adverse drug events (ADEs) and have been demonstrated as a public health burden.[1,2] With increasing rates of poly-pharmacy, the incidence of DDIs is most likely to increase, and, thus, drug interaction research remains essential.[3] Current DDI studies investigate different but complimentary scopes of drug interactions: *in vitro* pharmacokinetics (PK), clinical PK, and clinical pharmacodynamics (PD).[4–6] *In vitro* PK studies investigate DDI-related molecular mechanisms, such as metabolic enzymes or drug transporter proteins using recombinant systems or cell/tissue models. Clinical PK studies, on the other hand, evaluate whether one objective drug's exposure is changed due to the co-administrated precipitant drug. The molecular mechanisms of clinical PK DDIs are not necessarily known, unless the two

[1]Genentech Inc., San Francisco, California, USA; [2]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio, USA; [3]Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA; [4]School of Informatics, Computing & Engineering, Indiana University, Bloomington, Indiana, USA; [5]Instituto Gulbenkian de Ciência, Oeiras, Portugal; [6]Department of Computer and Information Sciences, University of Delaware, Newark, Delaware, USA; [7]Department of Obstetrics and Gynecology, Indiana University School of Medicine, Indianapolis, Indiana, USA; [8]Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio, USA.
*Correspondence: Aditi Shendre (aditi.shendre@osumc.edu), Lang Li (lang.li@osumc.edu), and Xia Ning (xia.ning@osumc.edu)

Drs. Heng-Yi Wu and Aditi Shendre have contributed equally to this work.

Received September 25, 2019; accepted November 25, 2019. doi:10.1002/cpt.1745

drugs are either known substrates/inhibitors/inducers of an enzyme. Clinical PD studies investigate whether the objective drug's efficacy or ADEs are changed because of the co-administrated precipitant.[7] *In vitro* PK experiments can be easily connected to pharmacogenetic (PG) studies because of their shared proteins and genes but this is not necessarily true in case of clinical PK or PD DDI studies.[8,9]

The goal of translational research in relation to DDI and PG studies is to achieve a comprehensive understanding of the PD, PK, and molecular mechanisms underlying drug effects in order to ultimately achieve clinical utility. However, it usually takes a long time to accomplish this overarching goal because of existing barriers between different scientific domains.[4] A salient example is tamoxifen, whose CYP2D6 metabolic pathway was initially discovered *in vitro* in 2004.[10] The genetic effects of *CYP2D6* on the exposure of tamoxifen and its active metabolites was later published in 2007.[11] The PK interactions between tamoxifen and selective serotonin reuptake inhibitors, such as paroxetine, were subsequently revealed in 2009.[12] Finally, the combined effect of the *CYP2D6* genotypes and drug inhibitors on tamoxifen efficacy and ADEs (hot flashes) was determined in 2012.[13] This example clearly demonstrates the association between DDI and drug-gene interaction (DGI) studies; however, it also shows the extended duration of DDI and PG research needed to achieve translational goals.

The translational landscape of drug interactions research has created an enormous opportunity for the field of informatics. The diverse and independent scientific disciplines involved in DDI and DGI research make it difficult to provide comprehensive evidence for all drugs.[4] Despite the existence of several databases, none of them have been successful in linking all the available information. DrugBank is probably the only database that comes close to identifying and including both DDI and DGI evidence.[14] However, it's PK and PD DDI evidence lack details on magnitude of drug exposure change and clinical phenotypes, respectively. DGI evidence in DrugBank include a drug's relationship with metabolic enzymes or transporter proteins but not the effect of PG on PK and PD effects. On the other hand, PharmGKB is designed to provide PG evidence on PK and PD outcomes, but no DDI evidence.[15] The Drug Interaction Database includes a collection of *in vitro* PK and clinical PK DDI evidence, but very limited PD DDI and PG evidence.[16] Therefore, it is of greatest translational research interest to consolidate these evidence in order to promote discovery of knowledge gaps between discordant DDI and DGI studies.[5,6]

Text mining, as an efficient knowledge discovery tool, has been extensively applied to mine drug interaction signals from the biomedical literature.[17–21] For example, Percha and Altman have developed a novel classification model to map all DGIs in MEDLINE abstracts, and discover new drug-gene relationships.[18] Previously, our group generated new DDI pairs by mining the PubMed literature using known cytochrome P450 (CYP450) probe substrates and inhibitors, and identifying all existing CYP450 substrates and inhibitors from *in vitro* experiments.[17] Recently, we also developed a DDI and DGI corpus with the goal of developing a new text mining algorithm and evaluating the performance of the text mining analyses separately for *in vitro* and clinical PK DDI evidence.[21] However, we did not investigate the overlapping or nonoverlapping evidence between the two. None of the existing informatics analyses have fully investigated the translational landscape of DDI and PG studies and the knowledge gaps that exist between them, nor have they differentiated between *in vitro* and clinical PK, and clinical PD evidence in the published DDI studies.

In this paper, a text mining approach was utilized to differentially screen *in vitro* PK DDI, clinical PK DDI, and clinical PD DDI, and DGI evidence, followed by an overlapping analysis. Our aim was to investigate and identify knowledge gaps among *in vitro* PK, clinical PK, and clinical PD studies, and translate the literature-based discovery evidence between DDI and DGI studies.

METHODS

A detailed description of the methods involved in the development of the text mining approach is presented in the **Supplemental File**. A brief description is included below.

Lexica construction

Lexica comprising of drug names, enzymes, action terms, and ADE terms were prepared. Based on the drug groups in DrugBank, the US Food and Drug Administration (FDA) approved and withdrawn drugs (2,403 generic names) were extracted for text mining. For drug enzyme terms, 94 symbol names and synonyms (350 terms in total) were collected from Gene ontology,[22] HUGO Gene Nomenclature Committee (HGNC),[23] and the Human Cytochrome P450 (CYP) Allele Nomenclature Database.[24] The action terms that describe the drug and enzyme relationships (i.e., inhibition or induction) were collected from our PK ontology[21] and the recent work by Percha and Altman.[18] The 19,550 preferred terms of adverse drug reactions were normalized from 70,177 lowest level terms in The Medical Dictionary for Regulatory Activity (MedDRA) database.[25]

Corpus construction

Two types of corpora, including information retrieval (IR) and information extraction (IE) were constructed for retrieving DDI and DGI abstracts and extracting DDI/DGI pairs, respectively. The IR corpus has 300 manually curated DDI abstracts in each one of the *in vitro* PK, clinical PK, and clinical PD studies. For PG studies, 3,429 DGI-relevant abstracts were collected from PharmGKB. The IE corpus consists of 210 *in vitro* PK DDI, 218 clinical PK DDI, 140 clinical PD DDI, and 395 DGI abstracts. In the IE corpus, terms such as drugs, enzymes, and relationships between drug-drug/drug-gene pairs were annotated. The details of the data collection (**Table S1**), text annotation, and annotation evaluation process are provided in the **Supplemental File**.

Text mining schemes

As shown in **Figure 1**, text mining for each type of DDI or DGI evidence was accomplished in two stages: IR and IE. In the IR stage, an optimal model that maximizes the recall rate of identifying relevant abstracts for each type of study was built using the IR corpus. The document-level classifier was trained upon N1-positive DDI abstracts and N2 randomly selected negative abstracts. In addition, the performance of the document-level classifier was evaluated using the testing dataset (N3 DDI abstracts and N4-negative abstracts). The data collection statistics for the IR models is shown in **Table 1**. After the optimal IR models were built, 25 million abstracts were screened, and relevant DDI and DGI abstracts were identified.

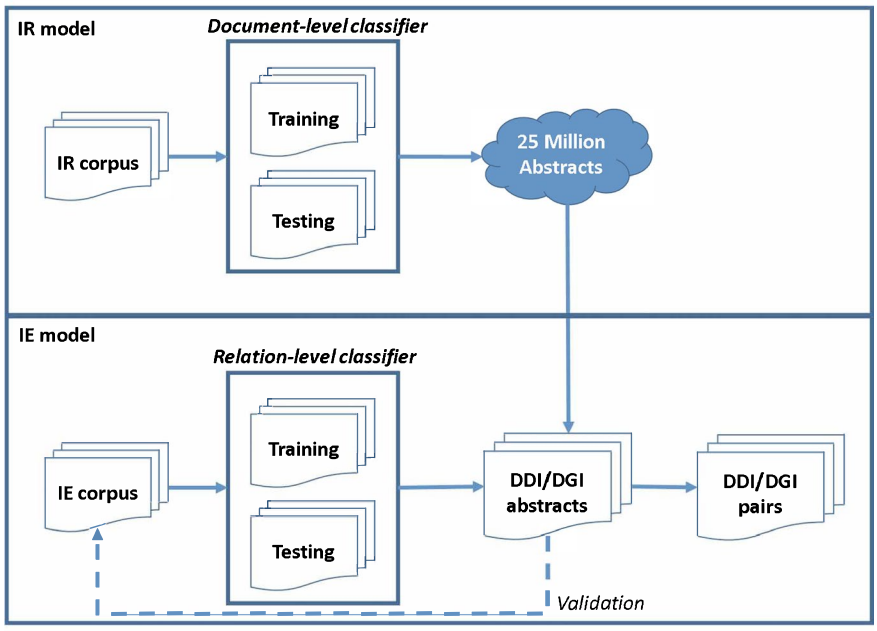

Figure 1 Text mining pipeline for the information retrieval and information extraction tasks. DDI, drug-drug interaction; DGI, drug-gene interaction; IE, information extraction; IR, information retrieval.

Table 1 Data collection statistics for IR models

	Training data		Testing data	
Study types	Positive (N1)	Negative (N2)	Positive (N3)	Negative (N4)
DDI				
In vitro PK	150[a]	10,000	150[a]	800[b]
Clinical PK				
Clinical PD				
DGI	1,700	1,700	1,729	8,300

DDI, drug–drug interaction; DGI, drug–gene interaction; IR, information retrieval; PD, pharmacodynamic; PK, pharmacokinetic.
[a]The 150 abstracts were different for training and testing as well as each of the *in vitro* PK, clinical PK, and clinical PD DDI studies.[b]Among 800 negative abstracts, 500 were single-drug or nutrition-related abstracts and 300 were randomly selected abstracts.

In the IE stage, an optimal model that maximizes F-measure of extracting relation pairs was built using the IE corpus. The DDI or DGI relationship classifiers were built upon 60% of the true entity relation pairs in the IE corpus (i.e., training data) and the remaining 40% were used for performance evaluation (i.e., testing data). Finally, using the optimal IE models, DDI and DGI pairs were extracted from their respective abstracts retrieved in the IR stage.

IR model development

IR was implemented in Weka.[26] String attributes in each abstract were converted into a set of attributes representing word occurrence information from the text using "StringToWordVector" module. Within the module, a set of word features converted from the normal text were extracted using IteratedLovinsStemmer, stopwordsHandler, NGramTokenizer (1–3), lowerCaseTokens, and wordsToKeep (1,000). The statistics for these word features, including term frequency-inverse document frequency and output word counts were prepared using TFTransform, IDFTransfrom, and outputWordCounts. Subject to the optimization of recall rate, Sequential Minimal Optimization was utilized for the text classification.

IE model development

IE of the DDI and DGI pairs was achieved in two steps: entity recognition and normalization, and relation pair extraction.

Entity Recognition and Normalization. The relevant entities, including drugs, enzymes, ADEs, and interaction terms, were tagged using name-entity recognition by string-matching against the lexica. Extracted drugs, enzymes, and ADEs were normalized to generic drug names, gene symbol names, and preferred terms in MeDRA, respectively. Interaction terms were normalized to their stemmed forms.

Relation pair extraction. The existing text mining methods recognize a piece of text that contains a semantic property of interest and extracts syntactic relations between entities in a single sentence using natural language processing.[19,27–30] Different from these works, we developed a feature-based approach to extract DDI/DGI pairs from context in an entire abstract. If N unique drug names are mentioned in an abstract, there are N*(N-1)/2 possible drug combinations that may or may not have interactions. Our IE model was built to predict the interaction relationship between each drug combination and optimize the F-measure.

In our DDI IE models, 16 features were created. These features capture syntactic, statistic, and scientific patterns from drug interactions present in the text. They were mainly derived from three types of information (entity location, entity statistics, and entity background knowledge). The location features provided location information for drug entities and interacting terms, or their co-occurrence (i.e., drug pairs co-occurring in the title sentence or the same sentence or the relative distance between drug pairs and interacting terms). The statistical features offer the frequency of drugs, drug pairs, and drug co-occurrence in a sentence or cross sentences. In addition, the knowledge features supply the background knowledge of

drug pair relations, such as enzyme-substrate/inhibitor relationships and anatomical therapeutic chemical (ATC) classification information.

To perform the DDI IE task, we customized 5 groups of feature sets from the 16 features using different strategies (**Tables S2 and S3**). Three manual (G1, G2, and G5) and two statistically (G3 and G4) determined group sets were adopted for the three types of studies. For G3 and G4, a stepwise regression model was used to determine statistically significant features, with G4 also involving two-way interaction terms. To maximize the F-measure for prediction, the optimal combination of five feature groups and seven popular classifiers (J48, Naïve Bayes, Sequential Minimal Optimization, Logistic Regression, Random Forest, Logistic Model Trees, and Iterative Classifier Optimizer) were explored for each study type. The details of the feature creation and selection are described in the section under "Experimental settings" in the **Supplemental File**.

To perform DGI IE, all descriptive structures for the drug-gene relationships were identified from PubMed abstracts.[18] Based on their findings, two important types of terms, including interacting verbs (e.g., inhibit) and mechanism terms (e.g., methylation), were included to characterize all dependency paths for DGI presentations. Four types of features (50 features) were created: (i) 44 features were scored based on the relative location, distance, and negation of each combination of 22 verb or 22 mechanism terms; (ii) co-occurrence frequency of each combination or each combination with verb terms; (iii) relative position of the bracket containing drugs or enzymes; and (iv) the order of the drug, gene, and verb/mechanism terms present in the sentences (**Table S4**). For this task, logistic regression was utilized for both feature selection and DGI prediction.

Hypotheses generation

By integrating the DDI and DGI evidence discovered through screening of the biomedical literature, and implementing a translational research method to discover knowledge gaps in drug interaction studies, we generated research hypotheses to: (i) understand the hazards of specific drugs given certain genetic polymorphisms and (ii) explore molecular mechanisms of drug interactions (**Figure 2**).

Translate DDI signals into PG hypotheses. A knowledge discovery method was used to translate DDI signals into PG hypotheses. The process included the examination of evidence to determine whether drug D1 changed drug D2 efficacy or ADEs (i.e., PD DDI),

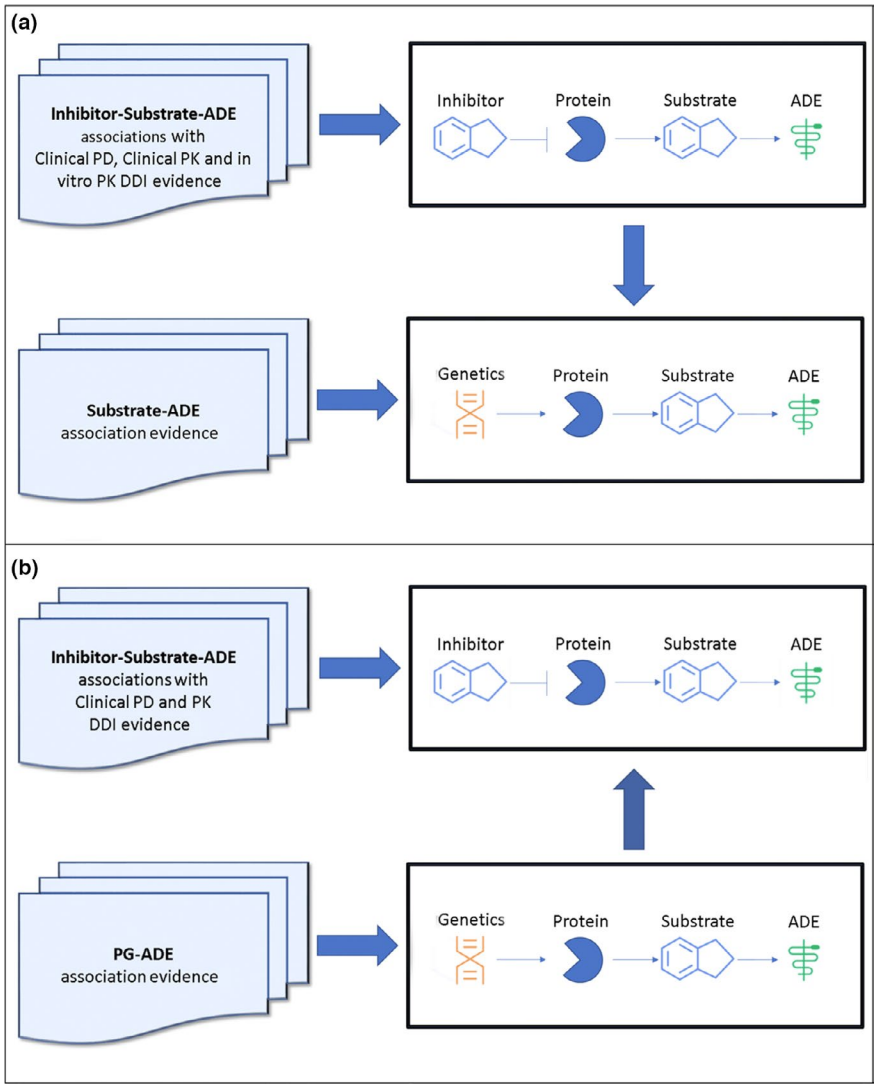

Figure 2 Hypotheses generation. (**a**) Translate drug-drug interaction (DDI) signals to predict genetic effects related to adverse drug events (ADEs) and (**b**) translate drug-gene interaction signals to predict molecular mechanisms of DDI. PD, pharmacodynamic; PG, pharmacogenetic; PK, pharmacokinetic.

D2 exposure (i.e., PK DDI), or inhibited D2 metabolic enzyme E (*in vitro* PK). If these DDI effects were noted, we then hypothesized that the functional genetic polymorphisms of E may be associated with D2 efficacy or ADEs.

Translate DGI signals into DDI mechanistic hypotheses. To explore unknown mechanisms involving drug interactions with only clinical PK and PD evidence, a discovery method was proposed to translate DGI signals into DDI mechanisms. The process included the evaluation of both drugs (D1 and D2) to discern their shared target genes and ADEs. If D1 and D2 were reported to interact and had common interacting genes, we hypothesized that their interaction may be synergistic or antagonistic for a given ADE.

RESULTS

The number of abstracts from each type of study, as well as the recall, F-measure, and validity-related statistics, are presented below. **Figure 3** presents the number of DDI and PG abstracts retrieved and the DDI and DGI pairs extracted from each type of study. The Venn diagram in **Figure 3** shows the overlap of drug pairs from the three types of DDI studies to help identify potential knowledge gaps between DDI and DGI evidence. The data related to the Venn Diagram and the DGI associated ADEs are presented in the **Supplemental Excel Files** "Venn diagram data and statistics.xlsx" and "DGI-ADE information.xlsx," respectively.

IR: Identifying DDI and DGI-relevant abstracts from MEDLINE

Using our recently developed corpus, the optimal IR models were built for each study type. The F-measures for the performance of the IR models were 0.94, 0.84, 0.70, and 0.78, respectively; and the recall rates were 0.98, 0.99, 0.86, and 0.97 for *in vitro* PK DDI, clinical PK DDI, clinical PD DDI, and PG, respectively (**Table S5**). Using these optimally trained models, a large-scale IR analysis of 25 million MEDLINE abstracts (1975–2015) was conducted. Studies involving animal models were removed using MeSH terms under the tree "B01.050" (Animal). We retrieved 5,199 *in vitro* PK, 17,048 clinical PK, 80,246 clinical PD DDI, and 479,865 PG abstracts (**Figure 3**). To further demonstrate the performance of these IR models, studies in the IE corpora (210, 218, 140, and 395 abstracts for *in vitro* PK, clinical PK, clinical PD, and PG studies) were used because there is no overlap between the IR and IE corpora. Recall rates for these IE studies were determined to be 1.00, 0.96, 0.99, and 0.94, respectively.

IE: Identifying DDI and DGI pairs from the MEDLINE Abstracts

To extract DDI and DGI pairs from the MEDLINE abstracts identified in the IR step, IE models were customized and optimized. The DDI extraction performances for each of the *in vitro* PK, clinical PK, and clinical PD studies were compared across five feature sets (G1–G5) and seven classifiers and are presented in **Tables S6**, **S7**, and **S8**, respectively. The optimal F-measure for *in vitro* PK studies was 0.83 using feature group 5 (G5) and the Naïve Bayes classifier; the optimal F-measure for clinical PK studies was 0.85 using feature group 1 (G1) and the Iterative Classifier Optimizer; and the optimal F-measure for clinical PD studies was 0.73 using feature group 1 (G1) and the Naïve Bayes classifier. For the DGI IE model, 50 features were trained over a

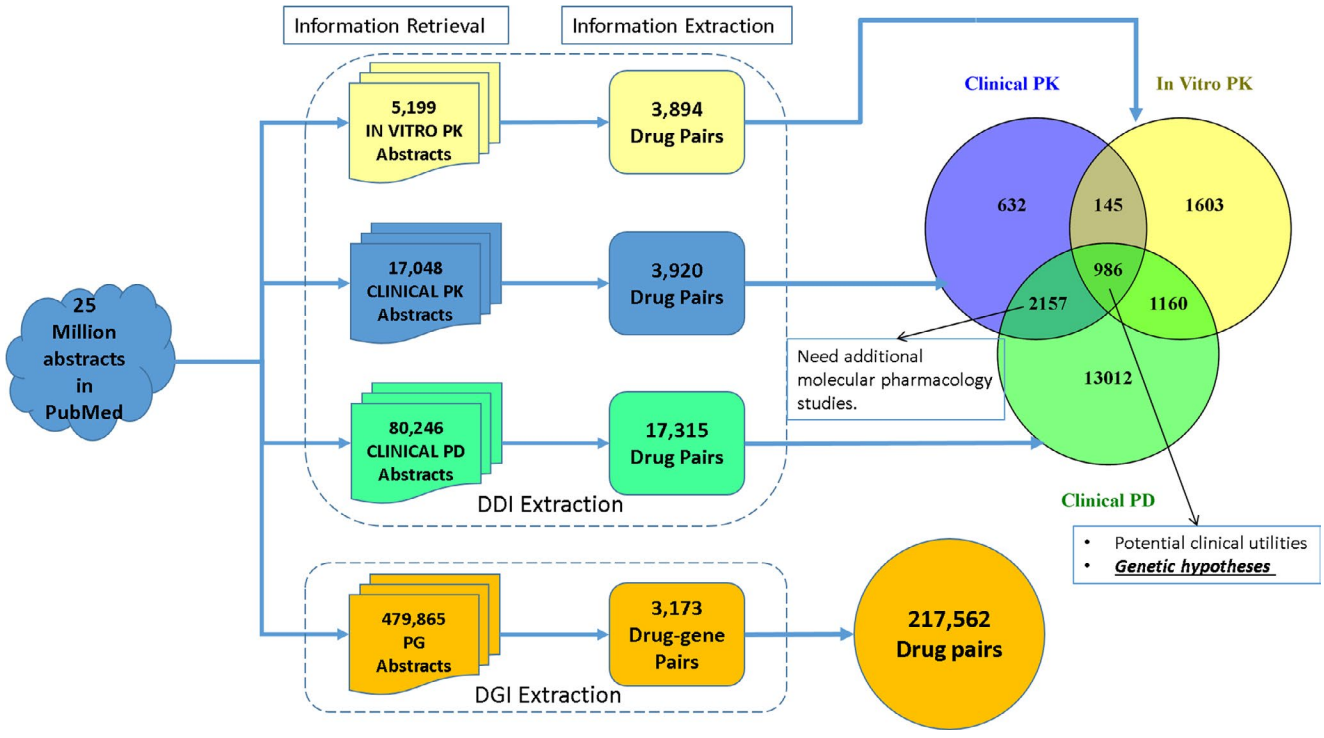

Figure 3 Results from the information retrieval and information extraction stages accompanied by a Venn diagram illustrating the overlap between the different DDI studies. DDI, drug-drug interaction; DGI, drug-gene interaction; PD, pharmacodynamic; PG, pharmacogenetic; PK, pharmacokinetic.

logistic linear regression classifier to reach the optimal F-measure of 0.82 (**Table S9**). All four optimized IE models were then applied to the relevant DDI/DGI abstracts retrieved from the previous IR stage. The IE analysis focused on FDA approved and withdrawn drugs, and identified 3,894, 3,920, and 17,315 unique *in vitro* PK, clinical PK, and clinical PD DDI pairs, respectively, and 3,173 unique DGI pairs (**Figure 3**). Using 3,173 retrieved DGI pairs, 217,562 drug pairs were further generated when both drugs shared enzyme relationships.

The overlap and knowledge gap among DDI evidence

With the drug pairs extracted in the IE stage, the Venn diagram shown in **Figure 3** was constructed to present the overlapping DDIs. A total of 986 unique drug pairs were found to overlap between all three study types. Another 2,157 DDIs represented the overlap between clinical PK and clinical PD studies. Last, 13,012 DDI pairs were found to only have clinical PD evidence.

Another overlapping analysis was performed to compare "extracted DDIs" from DDI IE to "predicted DDIs" from DGI IE. In **Table 2**, 94.8% of the 986 DDI pairs shared by all 3 DDI study types were predicted by DGI results. Other types of DDI evidence that overlap well with DGI predicted DDIs include: "clinical PK – *in vitro* PK" (95.9%), "clinical PD – *in vitro* PK" (86.6%), and "*in vitro* PK" (85.2%). For the remaining DDIs without *in vitro* PK evidence, DGI does not predict DDIs well, and the overlapping percentages are below 70%. Only 42.7% of the clinical PD DDIs were predicted from DGI results.

Comparing DDI text mining evidence to DDI data in the DrugBank database

DDI text mining performance was also evaluated by comparing the results with DDI data from the DrugBank database. For our comparison analysis, we only focused on the FDA approved and withdrawn drugs. Between 222,409 DrugBank DDIs and 19,695 text-mined DDIs, 9,587 DDIs overlapped. We compared the overlapping DDIs under the subgroups defined by the three DDI evidence types. In **Table 2**, DDI pairs with all three types of evidence from our text mining analysis overlapped the most with DrugBank DDIs (~ 88%), whereas DDI pairs with only clinical PD evidence had the lowest overlapping rate (~ 40%).

To demonstrate the validity of our text mined DDI evidence, the top 20 DDI pairs in each study type were evaluated manually (**Table 2**). DDI pairs were ranked by their reporting frequencies in different PubMed abstracts. Among the top 20 DDIs from *in vitro* PK, clinical PK, and clinical PD studies, 17, 20, and 19 pairs were manually validated as true DDIs, respectively. However, only 9, 16, and 13 of these DDIs were found to be reported in the DrugBank.

Table 2 Overlapping analysis for DDI studies

Venn diagram area	No. of DDIs in LS result	No. of DDIs predicted by DGI (%)	No. of DDIs found in DrugBank (%)	Top 20 DDIs (found in DrugBank/validated DDIs)
Full *in vitro* PK area	3,894	3,443 (88.4)	2,594 (66.6)	9/17
Full clinical PK area	3,920	2,980 (76.0)	2,734 (69.8)	16/20
Full clinical PD area	17,315	8,991 (51.9)	8,296 (47.9)	13/19
Overlap of clinical PD – clinical PK – *in vitro* PK	986	935 (94.8)	867 (87.9)	19/20
Overlap of clinical PK – *in vitro* PK	145	139 (95.9)	112 (77.2)	19/20
Overlap of clinical PD – *in vitro* PK	1,160	1,004 (86.6)	785 (67.7)	19/20
Overlap of clinical PD – clinical PK	2,157	1,494 (69.3)	1,406 (65.2)	13/19
Only clinical PK	632	412 (65.2)	349 (55.2)	13/14
Only *in vitro* PK	1,603	1,365 (85.2)	830 (51.8)	11/11
Only clinical PD	13,012	5,558 (42.7)	5,238 (40.3)	8/18
Venn diagram area	**No. of DDIs in Venn diagram area**		**No. of DDIs found in DrugBank (%)**	**Type of evidence**
Clinical PD – clinical PK – *in vitro* PK	6,683		3,271 (48.9)	PK evidence
Clinical PK – *in vitro* PK				
Clinical PD – *in vitro* PK				
Clinical PD – clinical PK				
Clinical PK				
In vitro PK				
Clinical PD – clinical PK – *in vitro* PK	17,315		4,016 (23.2)	PD evidence
Clinical PK – *in vitro* PK				
Clinical PD – *in vitro* PK				
Clinical PD – clinical PK				
Clinical PD				

DDI, drug–drug interaction; DGI, drug–gene interaction; LS, literature search; PD, pharmacodynamic; PK, pharmacokinetic.

Additionally, for the top 20 DDI pairs in the overlapping areas among 2 or 3 evidence types, almost all them were validated in our manual review but only a few of these DDIs were reported in DrugBank. DDI pairs that did not overlap with DrugBank data were also manually reviewed for validity. Among the top 20 DDIs from our 119 three-way overlapped DDIs, 17 were found to have confirmed DDI evidence in the literature. Similarly, all of the top 20 DDIs with overlapping clinical PD and clinical PK evidence were confirmed to have DDI evidence in the literature.

Translate DDI signals into genetics hypotheses

The 986 DDI pairs shared among 3 types of DDI studies were translated into genetic hypotheses with respect to their ADEs. Among these 986 DDIs, 865 (87.8%), 481 (48.8%), 193 (19.6%), 419 (42.5%), and 365 (37%) were associated with CYP3A, *CYP2D6*, *CYP2C8*, *CYP2C9*, and *CYP2C19*, respectively, with some DDIs involving more than one CYP450 enzyme. In our following genetic hypothesis generation analysis, we focused on *CYP3A* and *CYP2D6* as they were responsible for 88% and 49% of the 986 DDIs, respectively.

CYP3A-related DDIs had 68 distinctive substrates, and CYP2D6-related DDIs had 25 different substrates. Based on these CYP3A and CYP2D6 substrates, 552 and 192 ADE terms were found to co-occur in their clinical PD DDI abstracts, respectively. Similarly, 199 and 57 ADE terms related with the 68 CYP3A and 25 CYP2D6 substrates from DGI abstracts were retrieved. The common ADE terms from both DDI and DGI abstracts were considered as potential *CYP3A* or *CYP2D6* gene-related ADEs. These common DDI and DGI ADEs were further evaluated through manual review. Overall, 150 and 31 genetic hypotheses were generated from the 68 CYP3A substrates and 25 CYP2D6 substrates, respectively. Of these, 31 CYP3A-related and 13 CYP2D6-related PG evidence were reported in published PG studies (**Table 3**). As a result, 119 and 18 new PG hypotheses were generated for *CYP3A* and *CYP2D6*, respectively.

Translate DGI signals into DDI molecular mechanistic hypotheses

Among the 2,157 DDIs shared between clinical PD and PK DDI evidence, 1,497 DDI pairs shared the same metabolic enzymes (i.e., CYPs and uridine 5′-diphosphate glucuronosyltransferase) in their DGIs. Therefore, these 1,497 DDI pairs potentially have a PK drug interaction mechanism. Among the remaining 660 DDIs, 68 DDI pairs were found to share the same molecular pathways, 38 DDI pairs shared common genes, and 12 DDI pairs shared common genetic variants. The 38 DDI pairs with shared genes were reviewed further to determine if they elicit similar responses, whether the shared responses were modified by these DDIs, and whether any *in vitro* cell culture studies had investigated their DDI mechanisms. After manual review, 7 of 38 DDI pairs were validated to have DDI effects, and 3 had additional DDI evidence from *in vitro* experiments (**Table 4**).

DISCUSSION

ADEs caused by drug interactions are a critical issue for prescriptions. In clinical practice, prescription decision support typically stems from *in vivo* and clinical evidence. However, there is high variability in drug responses, which are affected by both genetic and environmental factors. Therefore, studying genetic or molecular mechanisms underlying DDIs is essential to help: (i) understand the hazards of specific drugs given certain genetic polymorphisms, and (ii) explore molecular mechanisms of such interactions. Today, more than 2,000 genetic tests are currently available, but not every drug is covered and the tests can be expensive.[31] To address these challenges, we introduced a translational research method to discover knowledge gaps in drug interaction studies. Utilizing the results from our large-scale screening, two sets of hypotheses were generated by (i) translating DDI signals into genetic information for ADEs and (ii) translating DGI signals into molecular mechanistic hypotheses.

To demonstrate the process of PG hypotheses generation, evaluation, and validation, we use the example of tacrolimus, a CYP3A substrate. Among 87 clinical PD DDI abstracts showing the interaction evidence between tacrolimus and CYP3A inhibitors, such as ketoconazole, clarithromycin, cyclosporine, or ritonavir, 141 ADE terms were identified and extracted. From these results, we assume that 141 genetic hypotheses can be generated for tacrolimus. Another 153 ADEs were extracted from DGI abstracts related to tacrolimus and *CYP3A*. A total of 25 ADE terms were common between the DDI and DGI abstracts. From these, three ADEs (nephrotoxicity, hepatotoxicity, and hyperglycemia) were validated and found to be associated with the *CYP3A5* polymorphism, rs776746 (PharmGKB level 2A or level 3 evidence).[32–35] Thus, we were able to validate our tacrolimus ADE-related genetic hypotheses, underscoring the accuracy of our text mining algorithm. More importantly, the generation of 119 new PG hypotheses highlights the significance of our translational research method in enabling the discovery of potentially new genetic mechanisms that may otherwise not have been explored through conventional DDI and PG research methods. An example of this is simvastatin-induced rhabdomyolysis. Even though there are reported evidence for the association between *CYP3A4* and *CYP3A5* genetic polymorphisms and simvastatin-induced mild myopathy symptoms,[36] there is no reported study on the association between *CYP3A* and the more severe form of myopathy (i.e., rhabdomyolysis).

Similar to the PG hypotheses, molecular mechanistic hypotheses were validated or determined to be new through our manual review process. For example, the combination of cisplatin and pemetrexed showed improved response rate in mesothelioma patients in a randomized phase III trial.[37] Both drugs share the *ABCC2, MTHFR*, and *SLC19A1* target genes among them (PharmGKB level 4 evidence). Patients with *ABCC2* rs2273697 (AA or AG genotype) were reported to have improved overall and progression-free survival when treated with cisplatin and pemetrexed compared with patients with the *ABCC2* GG genotype.[38] The synergistic PD DDI between cisplatin and pemetrexed has been demonstrated in *in vitro* studies over multiple cancer cell lines (MCF7, A549, and PA1 cells).[39] In particular, when MCF7 cells were incubated with pemetrexed for 24 hours followed by cisplatin for 24 hours, synergistic inhibition of cell proliferation was noted. Similar synergistic effects were also observed in the A549 and PA1 cell lines. Another example is the interaction between etanercept and methotrexate, which results in improved response in patients with rheumatoid arthritis especially with the *ATP5E* rs1059150 (GG), *HLA-E* rs1264457 (AA), or *KLRC1* rs7301582 (CT or TT) variants (PharmGKB level 3 evidence). However, no *in vitro* experiments have been reported that illustrate their DDI mechanisms. This etanercept/methotrexate

Table 3 Novel or tested pharmacogenetics hypotheses related to *CYP3A* and *CYP2D6* genes

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
CYP3A	Alprazolam	Nefazodone/Fluoxetine/Ritonavir	N/A	Sedation	Yes	–
	Amprenavir	Ritonavir	1,200 mg/200 mg q.d., 600 mg/100 mg b.i.d.	Diarrhea	Yes	–
				Hyperlipidemia	Yes	–
				Nausea	Yes	–
				Oral paresthesias	Yes	–
				Rash	Yes	–
				Vomiting	Yes	–
	Atazanavir	Ritonavir	300 mg/100 mg q.d., 400 mg/100 mg q.d. if given with other HIV drugs	Cholelithiasis	Yes	–
				Diarrhea	Yes	–
				Hyperlipidemia	Yes	–
				Increase in total and subcutaneous fat	Yes	–
				Metabolic syndrome	Yes	–
				Upper respiratory tract infection	Yes	–
				Vomiting	Yes	–
				Hyperbilirubinemia	No	28599374
				Nephrolithiasis	No	25151207
	Atorvastatin	Itraconazole/Ticagrelor/Ritonavir	> 20 mg atorvastatin to be given with caution; no adjustment with Ticagrelor; minimum required dose with ritonavir	Rhabdomyolysis	Yes	
	Atorvastatin	Itraconazole/Ritonavir	> 20 mg atorvastatin to be given with caution; minimum required dose with ritonavir	Myopathy	No	28812116, 17289397, 15900215
	Carbamazepine	Ritonavir	N/A	Ataxia	Yes	–
	Cerivastatin	Cyclosporine	N/A	Myopathy	Yes	–
				Rhabdomyolysis	Yes	–
	Cisapride	Ketoconazole/Clarithromycin/ Diltiazem	Contraindicated; N/A for diltiazem	QTc prolongation	Yes	–
	Cisapride	Clarithromycin/Diltiazem	Contraindicated; N/A for diltiazem	Syncope	Yes	–
	Cisapride	Ketoconazole/Clarithromycin	Contraindicated	Ventricular arrhythmias	Yes	–
	Colchicine	Clarithromycin	0.3 or 0.6 mg colchicine; dose and frequency depend on the condition treated	Death	Yes	–
				Gi toxicity	Yes	–
				Multiorgan failure	Yes	–
				Myelosuppression	Yes	–
				Pancytopenia	Yes	–

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Cyclosporine	Telaprevir	N/A	Cytomegalovirus reactivation	Yes	–
				Acute rejection	No	27653228, 18444945, 29043387
	Cyclosporine	Ketoconazole/Nefazodone	N/A	Nephrotoxicity	No	22388796
	Darunavir	Ritonavir	800 mg/100 mg q.d., 600 mg/100 mg b.i.d.	Diarrhea	Yes	–
				Headache	Yes	–
				Hyperlipidemia	Yes	–
				Increase in alt	Yes	–
				Increase in total and subcutaneous fat	Yes	–
				Lipoma	Yes	–
				Metabolic syndrome	Yes	–
				Nasopharyngitis	Yes	–
				Nausea	Yes	–
				Upper respiratory tract infection	Yes	–
	Diazepam	Cimetidine	N/A	Sedation	Yes	–
	Docetaxel	Ketoconazole	N/A	Less febrile neutropenia	Yes	–
				Less neutrophil suppression	Yes	–
				Anemia	No	25495407, 17410042
				Diarrhea	No	25495407, 17410042
				Edema	No	19332043
				Fatigue	No	21468756, 17410042
				Increase in liver function tests	No	25495407
				Myelosuppression	No	25495407, 17545536, 19332043, 16765145, 17410042
				Nausea	No	25495407, 17410042
				Neuropathy	No	27574448, 17410042

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Everolimus	Voriconazole	Avoid voriconazole	Pneumonia	Yes	
	Fentanyl	Fluconazole/Voriconazole	N/A	Respiratory depression	No	21223952
	Imatinib	Docetaxel	N/A	Diarrhea	Yes	–
				Nausea	Yes	–
				Vomiting	Yes	–
	Indinavir	Ritonavir	800 mg/100 mg b.i.d., 800 mg/200 mg b.i.d.	Alopecia	Yes	–
				Flank pain	Yes	–
				Hematuria	Yes	–
				Hyperlipidemia	Yes	–
				Nausea	Yes	–
				Nephrolithiasis	Yes	–
				Vomiting	Yes	–
	Levomethadyl acetate	Ketoconazole	N/A	Prolongation of miosis	Yes	–
	Lopinavir	Ritonavir	800 mg/200 mg q.d., 400 mg/100 mg b.i.d., 600 mg/150 mg if given with other HIV drugs	Asthenia	Yes	–
				Diarrhea	Yes	–
				Gastritis	Yes	–
				Headache	Yes	–
				Hyperglycemia	Yes	–
				Hyperinsulinemia	Yes	–
				Hyperlipidemia	Yes	–
				Lipoma	Yes	–
				Rash	Yes	–
	Lovastatin	Ritonavir/Telaprevir/Boceprevir	Contraindicated	Myalgia	Yes	–
	Lovastatin	Itraconazole/Ritonavir	Contraindicated	Myopathy	Yes	–
	Lovastatin	Itraconazole/Cyclosporine/ Ritonavir	Contraindicated; avoid cyclosporine	Rhabdomyolysis	Yes	–
	Methadone	Fluvoxamine	N/A	Opioid withdrawal	No	21902501
	Midazolam	Ketoconazole/Saquinavir	Avoid ketoconazole; N/A for saquinavir	Cognitive impairment	Yes	–
	Midazolam	Posaconazole	Avoid posaconazole	Diarrhea	Yes	–
	Midazolam	Ketoconazole/Ritonavir/Diltiazem/ Verapamil/Ranitidine/Erythromycin	Avoid ketoconazole/diltiazem/verapamil/ erythromycin; N/A for ranitidine; ritonavir contraindicated	Sedation	No	17786417

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Nelfinavir	Ritonavir	N/A	Diarrhea	Yes	–
				Nausea	Yes	–
				Rash	Yes	–
	Paclitaxel	Cyclosporine	N/A	Nail changes	Yes	–
				Neurotoxicity	No	23640974, 20212519
				Vomiting	No	15901749
	Paclitaxel	Cyclosporine/Pazopanib	N/A	Diarrhea	Yes	–
				Nausea	No	15901749
	Paclitaxel	Cyclosporine/Pazopanib/Verapamil	N/A	Neutropenia	No	12454106, 26179145, 15901749
	Paclitaxel	Pazopanib	N/A	Abscess	Yes	–
				Dysgeusia	Yes	–
				Fatigue	Yes	–
				Gastrointestinal perforations	Yes	–
				Hair color changes	Yes	–
				Hepatotoxicity	Yes	–
				Hyperbilirubinemia	Yes	–
				Hypertension	Yes	–
				Myalgia	Yes	–
				Alopecia	No	24025145
				Myelosuppression	No	26179145, 21702053, 15901749
				Neuropathy	No	25398452
	Paclitaxel	Pazopanib/Verapamil	N/A	Thrombocytopenia	No	15901749
	Pimozide	Clarithromycin	Contraindicated	QTc prolongation	Yes	–
	Quetiapine	Ritonavir/Atazanavir	1/6th the usual dose	Delirium	Yes	–
				Sedation	Yes	–
				Weight gain	Yes	–
	Quinidine	Cimetidine	N/A	QT, QTc, QRS, and RR prolongation	Yes	–

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Saquinavir	Ritonavir/Nelfinavir	1,000 mg/100 mg b.i.d.; N/A for nelfinavir	Abdominal discomfort	Yes	–
				Diarrhea	Yes	–
				Dysphagia	Yes	–
				Headache	Yes	–
				Nausea	Yes	–
	Saquinavir	Ritonavir	1,000 mg/100 mg b.i.d.,	Asthenia	Yes	–
				Hyperlipidemia	Yes	–
				Increase in liver function tests	Yes	–
				Paresthesias	Yes	–
				QTc prolongation	Yes	–
	Simvastatin	Amiodarone	< 20 mg/day	Acute renal failure	Yes	–
				Azotemia	Yes	–
	Simvastatin	Diltiazem	N/A	Hepatitis	Yes	–
	Simvastatin	Amiodarone/Diltiazem	< 20 mg/day; N/A for diltiazem	Hepatotoxicity	Yes	–
	Simvastatin	Clarithromycin/Amiodarone/ Ritonavir/Itraconazole/ Cyclosporine/Verapamil	Avoid clarithromycin/itraconazole; < 20 mg/day with amiodarone and verapamil; < 10 mg/day with cyclosporine; ritonavir contraindicated	Rhabdomyolysis	Yes	–
	Simvastatin	Ritonavir/Telaprevir/Boceprevir	Contraindicated	Myalgia	No	16321621
	Simvastatin	Diltiazem/Clarithromycin/Ritonavir/ Amiodarone/Itraconazole/ Cyclosporine/Verapamil	Avoid clarithromycin/itraconazole; < 20 mg/day with amiodarone and verapamil; < 10 mg/day with cyclosporine; N/A for diltiazem; ritonavir contraindicated	Myopathy	No	17289397, 22122820
	Sirolimus	Cyclosporine	2 mg or 5 mg per day maintenance based on immunologic risk	Hypertension	Yes	–
				Hypertriglyceridemia	Yes	–
				Myelosuppression	Yes	–
				Nephrotoxicity	Yes	–
				Diabetes mellitus	No	28245187
				Hypercholesterolemia	No	21441846

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Tacrolimus	Fluconazole/Voriconazole	1/3rd the usual dose followed by adjustments based on blood concentration levels	Hepatotoxicity	No	24438215, 25966085
				Hyperglycemia	No	25966085
				Nephrotoxicity	No	17495880, 20526235, 23574377, 19644155, 19067682, 21677300, 25201288, 26856709, 27977332, 27217047, 29539600
	Terfenadine	Nefazodone/Ketoconazole/ Itraconazole	N/A	QTc prolongation	Yes	–
				Torsades de pointes	Yes	–
	Tipranavir	Ritonavir	500 mg/200 mg b.i.d.	Diarrhea	Yes	–
				Headache	Yes	–
				Hepatotoxicity	Yes	–
				Hyperlipidemia	Yes	–
				Nausea	Yes	–
	Trazodone	Ritonavir	N/A	Dizziness	Yes	–
				Fatigue	Yes	–
				Hypotension	Yes	–
				Nausea	Yes	–
				Sedation	Yes	–
				Syncope	Yes	–
	Triazolam	Nefazodone/Ritonavir	Contraindicated	Sedation	Yes	–
	Verapamil	Cimetidine	N/A	Increase in negative dromotropic effect	Yes	–
CYP2D6	Codeine	Quinidine	N/A	Psychomotor effects	Yes	–
				Pupillary effects	Yes	–
				Respiratory depression	No	22492761, 18713907
	Dextromethorphan	Quinidine	20 mg/10 mg q.d. for 7 days and then b.i.d.	Diarrhea	Yes	–
				Falls	Yes	–
				QTc prolongation	Yes	–
				Urinary tract infection	Yes	–

(Continued)

Table 3 (Continued)

Gene	Substrate (S)	Inhibitors (I)	DDI dosing recommendations in drug labels (S/I)	ADEs	Novel genetic hypotheses	PMIDs for tested hypotheses
	Imipramine	Fluoxetine	N/A	QTc prolongation	Yes	–
	Lidocaine	Propafenone	N/A	Negative inotropic effect	Yes	–
				Central nervous system side effects	Yes	–
	Metoprolol	Cimetidine	N/A	Fatigue	No	2049246, 24637943
				Anxiety	Yes	–
				Asthenia	Yes	–
				Sweating	Yes	–
	Metoprolol	Diphenhydramine	N/A	Systolic blood pressure	No	24637943
	Metoprolol	Diphenhydramine/Verapamil	N/A	Heart rate reduction	No	19037197, 18784654, 24637943, 24193112, 29095089
	Metoprolol	Propafenone	N/A	Nightmares	Yes	–
				Left ventricular failure	Yes	–
	Metoprolol	Verapamil	N/A	Atrioventricular block	No	3437726
				Systemic arterial pressure	No	19037197, 24637943
	Mexiletine	Quinidine	N/A	Prolongation of refractory period and conduction	Yes	–
	Propafenone	Quinidine	Avoid quinidine	Heart rate reduction	No	7595187
	Propranolol	Cimetidine	N/A	Heart rate reduction	No	9399616, 12728976
	Risperidone	Ritonavir	N/A	Dystonia	No	20599499
				Parkinsonism	No	20599499
				Coma	Yes	–
				Akathisia	Yes	–
	Timolol	Paroxetine	N/A	Heart rate reduction	No	7474246, 20925579
	Venlafaxine	Fluoxetine	N/A	Serotonin syndrome	No	19822698, 23799451, 16958828, 10780263
	Zuclopenthixol	Fluoxetine	N/A	Extrapyramidal side effects	No	12107620
				Dystonia	Yes	–

ADEs, adverse drug events; DDI, drug-drug interactions; N/A, not available; PMID, PubMed IDs for journal articles.

Table 4 Novel or tested molecular mechanistic hypotheses

Drug 1	Drug 2	DDI dosing recommendations in drug labels (D1/D2)	DDI effects	Shared genes	Novel molecular mechanistic hypotheses	PMIDs for tested hypotheses
Ramipril	Hydrochlorothiazide	N/A	Improved blood pressure reduction	*ACE* rs4359 or rs4344	Yes	–
Azathioprine	Mercaptopurine	N/A	An increased risk of pancreatitis	*HLA-DQA1* (*02:01 allele), *HLA-DRB1* (*07:01 allele)	Yes	–
Etanercept	Methotrexate	50 mg etanercept once weekly	Increased efficacy in patients with rheumatoid arthritis	*ATP5F1E* rs1059150, *HLA-E* rs1264457, *KLRC1* rs7301582	Yes	–
Amlodipine	Benazepril	2.5 mg/10 mg q.d., 10 mg/40 mg q.d. maintenance	Improved blood pressure reduction	*ACE* rs1799752	Yes	–
Raltitrexed	Irinotecan	N/A	Increased incidence of asthenia	*TYMS* rs45445694	No	9607593
Cisplatin	Pemetrexed	500 mg/m^2 i.v. and 75 mg/m^2 i.v.	Improved response rate in mesothelioma	*ABCC2* rs2273697, *MTHFR* rs1801133, *SLC19A1* rs1051298	No	16898269, 22562354
Carbamazepine	Levetiracetam	N/A	Leads to beneficial anticonvulsant PD interactions	*SCN1A rs2298771*	No	24211788

DDI, drug-drug interactions; N/A, not available; PD, pharmacodynamic; PMID, PubMed IDs for journal articles.

example demonstrates that our translational discovery method can also generate novel PD DDI mechanisms.

Our study has both strengths and limitations. Our text mining algorithm enabled us to screen ~ 25 million MEDLINE abstracts in order to retrieve DDI and DGI evidence. We were also able to extract a substantial number of DDI and DGI pairs from the literature, and distinguish them based on the type of study involved. However, we only focused on drug pairs and the CYP450-related enzymes and genes. Therefore, we could not evaluate DDIs or ADEs associated with high dimensional drug combinations or their interactions with other enzymes and drug transporters. Our algorithm was also designed to identify co-occurrence of drugs and ADE terms, as such, a manual review process was required to verify and confirm these associations. Additionally, our algorithm does not collect information on drug dosage or sample size from individual studies at the moment but we are planning to add this information along with the information on drug transporters and other enzymes in the future. Despite these limitations, our study provides a tremendous amount of information on DDIs, DGIs, and ADEs and allowed us to generate several novel genetic hypotheses.

In conclusion, a text mining pipeline was developed to extract DDI evidence from the biomedical literature in the current study. Initially, golden standard corpora for DDIs and DGIs were created to facilitate the text mining development. Subsequently, a large-scale analysis was conducted to identify knowledge gaps in DDI and DGI research, which were then used to generate hypotheses in order to identify novel genetic mechanisms involving drug interactions and predict potential molecular mechanistic DDI mechanisms.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

Supplementary text.
Table S1. Inclusion-exclusion criteria for the abstract validation process.
Table S2. FDA's probe features.
Table S3. Feature selections for the three DDI study types.
Table S4. Categorization of drug-gene-verb presentation order.
Table S5. Performance evaluation of the IR task.
Table S6. Performance evaluation of IE for *in vitro* PK DDI studies.
Table S7. Performance evaluation of IE for clinical PK DDI studies.
Table S8. Performance evaluation of IE for clinical PD DDI studies.
Table S9. Performance evaluation of IE for DGI studies.
Figure S1. Examples of the calculation for the average angle of a drug pair to interaction verbs ($\angle_{i,j,k,s}$).
Excel files. 1) Venn diagram data and statistics.xlsx and 2) DGI-ADE information.xlsx

FUNDING

This work is supported by grants from the National Library of Medicine (R01LM011945), and the National Institute of General Medical Sciences (R01GM104483).

CONFLICT OF INTEREST

All authors declare no competing interests for this work.

AUTHOR CONTRIBUTIONS

H.Y.W., A.S., and L.L. wrote the manuscript. H.Y.W., D.Z., L.M.R., H.S., S.K.Q., and L.L. designed the research. H.Y.W., A.S., S.Z., P.Z., L.W., X.N., and L.L. performed the research. H.Y.W., A.S., S.Z., and L.W. analyzed the data. H.Y.W. contributed to new analytical tools.

1. Hall, M.J., DeFrances, C.J., Williams, S.N., Golosinskiy, A. & Schwartzman, A. National Hospital Discharge Survey: 2007 summary. *Natl. Health Stat. Report* **29**, 1–20 (2010).
2. Niska, R., Bhuiya, F. & Xu, J. National Hospital Ambulatory Medical Care Survey: 2007 emergency department summary. *Natl. Health Stat. Report* **26**, 1–31 (2010).
3. Hajjar, E.R., Cafiero, A.C. & Hanlon, J.T. Polypharmacy in elderly patients. *Am. J. Geriatr. Pharmacother.* **5**, 345–351 (2007).
4. Hennessy, S. & Flockhart, D.A. The need for translational research on drug-drug interactions. *Clin. Pharmacol. Ther.* **91**, 771–773 (2012).
5. Boyce, R., Collins, C., Horn, J. & Kalet, I. Computing with evidence part II: an evidential approach to predicting metabolic drug-drug interactions. *J. Biomed. Inform.* **42**, 990–1003 (2009).
6. Boyce, R., Collins, C., Horn, J. & Kalet, I. Computing with evidence part I: a drug-mechanism evidence taxonomy oriented toward confidence assignment. *J. Biomed. Inform.* **42**, 979–989 (2009).
7. Prueksaritanont, T. *et al*. Drug-drug interaction studies: regulatory guidance and an industry perspective. *AAPS J.* **15**, 629–645 (2013).
8. Crews, K.R. *et al*. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. *Clin. Pharmacol. Ther.* **91**, 321–326 (2012).
9. Wilke, R.A. *et al*. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin. Pharmacol. Ther.* **92**, 112–117 (2012).
10. Desta, Z., Ward, B.A., Soukhova, N.V. & Flockhart, D.A. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *J. Pharmacol. Exp. Ther.* **310**, 1062–1075 (2004).
11. Goetz, M.P. *et al*. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.* **101**, 113–121 (2007).
12. In brief: Tamoxifen and SSRI interaction. *Med. Lett. Drugs Ther.* **51**, 45 (2009).
13. Hertz, D.L., McLeod, H.L. & Irvin, W.J. Jr. Tamoxifen and CYP2D6: a contradiction of data. *Oncologist* **17**, 620–630 (2012).
14. Wishart, D.S. *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
15. Whirl-Carrillo, M. *et al*. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **92**, 414–417 (2012).
16. Drug Interaction Database. DIDB has been specifically designed to support pharmaceutical scientists in their assessment of human PK-based drug interactions and drug safety <https://www.druginteractioninfo.org/solutions/drug-interaction-database/>. Accessed March 12, 2015.
17. Duke, J.D. *et al*. Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interactions. *PLoS Comput. Biol.* **8**, e1002614 (2012).
18. Percha, B. & Altman, R.B. Learning the structure of biomedical relationships from unstructured text. *PLoS Comput. Biol.* **11**, e1004216 (2015).
19. Percha, B., Garten, Y. & Altman, R.B. Discovery and explanation of drug-drug interactions via text mining. *Pac. Symp. Biocomput.* 410–421 (2012). https://doi.org/10.1142/9789814366496_0040
20. Tari, L., Anwar, S., Liang, S., Cai, J. & Baral, C. Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. *Bioinformatics* **26**, i547–i553 (2010).

21. Wu, H.Y. *et al*. An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinform.* **14**, 35 (2013).
22. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
23. HUGO Gene Nomenclature Committee at the European Bioinformatics Institute <http://www.genenames.org>. Accessed July 17, 2015.
24. Sim, S.C. & Ingelman-Sundberg, M. The Human Cytochrome P450 (CYP) Allele Nomenclature website: a peer-reviewed database of CYP variants and their associated effects. *Hum Genomics* **4**, 278–281 (2010).
25. Brown, E.G., Wood, L. & Wood, S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* **20**, 109–117 (1999).
26. Eibe, F., Hall, M.A. & Witten, I.H. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th edn. (Morgan Kaufmann, Burlington, MA, 2016).
27. Segura-Bedmar, I., Martinez, P. & de Pablo-Sanchez, C. Using a shallow linguistic kernel for drug-drug interaction extraction. *J. Biomed. Inform.* **44**, 789–804 (2011).
28. Segura-Bedmar, I. & Martinez, P. & de Pablo-Sánchez, C. A linguistic rule-based approach to extract drug-drug interactions from pharmacological documents. *BMC Bioinform*. **12** (suppl. 2), S1 (2011).
29. Segura-Bedmar, I., Martinez, P. & Herrero-Zazo, M. Lessons learnt from the DDIExtraction-2013 Shared Task. *J. Biomed. Inform.* **51**, 152–164 (2014).
30. Segura-Bedmar, I., Martınez, P. & Sánchez-Cisneros, D.The 1st DDIExtraction-2011 challenge task: extraction of drug-drug interactions from biomedical texts. Challenge Task on Drug-Drug Interaction Extraction 2011, 1–9 (2011).
31. Genetic Testing: MedlinePlus <https://medlineplus.gov/genetictesting.html>. Accessed September 01, 2019.
32. Quaglia, M., Terrazzino, S., Boldorini, R., Stratta, P. & Genazzani, A.A. Severe acute nephrotoxicity in a kidney transplant patient despite low tacrolimus levels: a possible interaction between donor and recipient genetic polymorphisms. *J. Clin. Pharm. Ther.* **38**, 333–336 (2013).
33. Wang, L., Li, N., Wang, M.X. & Lu, S.C. Benefits of minimizing immunosuppressive dosage according to cytochrome P450 3A5 genotype in liver transplant patients: findings from a single-center study. *Genet. Mol. Res.* **14**, 3191–3199 (2015).
34. Xue, F. *et al*. CYP3A5 genotypes affect tacrolimus pharmacokinetics and infectious complications in Chinese pediatric liver transplant patients. *Pediatr. Transplant* **18**, 166–176 (2014).
35. Kuypers, D.R., Naesens, M., de Jonge, H., Lerut, E., Verbeke, K. & Vanrenterghem, Y. Tacrolimus dose requirements and CYP3A5 genotype and the development of calcineurin inhibitor-associated nephrotoxicity in renal allograft recipients. *Ther. Drug Monit.* **32**, 394–404 (2010).
36. Yang, W.H. *et al*. Simvastatin-induced myopathy with concomitant use of cyclosporine: case report. *Int. J. Clin. Pharmacol. Ther.* **49**, 772–777 (2011).
37. Janne, P.A. *et al*. Pemetrexed alone or in combination with cisplatin in previously treated malignant pleural mesothelioma: outcomes from a phase IIIB expanded access program. *J. Thorac. Oncol.* **1**, 506–512 (2006).
38. Goricar, K., Kovac, V. & Dolzan, V. Clinical-pharmacogenetic models for personalized cancer treatment: application to malignant mesothelioma. *Sci. Rep.* **7**, 46537 (2017).
39. Kano, Y. *et al*. Schedule-dependent interactions between pemetrexed and cisplatin in human carcinoma cell lines in vitro. *Oncol. Res.* **16**, 85–95 (2006).