DATABASE

The Cancer Drug Fraction of Metabolism Database

Liyan Hua1,†, Chien-Wei Chiang2,†, Wang Cong1, Jin Li1, Xueying Wang1, Lijun Cheng2, Weixing Feng1, Sara K. Quinney3,4,†, Lei Wang1,2,*,‡ and Lang Li2,‡

This study aims to create a database for quantifying the fraction of metabolism of cytochrome P450 isozymes for cancer drugs approved by the US Food and Drug Administration. A reproducible data collection protocol was developed to extract essential information, including both substrate-depletion and metabolite-formation data from publicly available in vitro selective cytochrome P450 enzyme inhibition studies. We estimated the fraction of metabolism from the curated data. To demonstrate the utility of this database, we conducted an in vitro drug interaction prediction for the 42 cancer drugs. In the drug–drug interaction prediction, we identified 31 drug pairs with at least one cancer drug in each pair that had predicted area under concentration ratios > 2. We further found clinical drug interaction pieces of evidence in the literature to support 20 of these 31 drug–drug interaction pairs.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
✔ The fractions of drug metabolism frequently were investigated, but currently there was a lack of fraction of metabolism database.

WHAT QUESTION DID THIS STUDY ADDRESS?
✔ The high-quality fraction of metabolism database of cancer drugs was created from published literature based on a well-characterized data-curating procedure for data sharing.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
✔ This study demonstrated that there were significant amounts of literature that contributed to fraction of metabolism estimation. In this article, we focused on the drug metabolism related to cytochrome P450 and established a well-characterized data-curating procedure for constructing a pharmacology database.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
✔ This database was mainly designed for public data sharing that would facilitate academic research.

Drugs are eliminated by excretion or metabolism after entering the human body.1 Drug metabolism refers to the process in which the chemical structure changes under the action of various drug metabolism enzymes (especially liver drug enzymes) in the body. The majority of small-molecule drugs are metabolized by cytochrome P450 (CYP450) enzymes, which are located in the hepatic endoplasmic reticulum.2,3 Many factors can alter hepatic drug metabolism, including genetic polymorphisms, disease, concomitant medications, and foods.4-6 Among these factors, concomitant medications are vital because of poly-pharmacy.7-11 Many enzymatic routes of elimination, including almost all of those via the CYP450 enzymes, can be inhibited or induced by concomitant medications. Notably, when the primary metabolic pathways of a drug are inhibited or induced by strong inhibitors or inducers, drug and metabolite concentrations in the blood and tissue can be significantly changed.12,13 The dramatically changed drug exposure may result in unwanted adverse reactions or reduced efficacy.14 A randomized, open-label, parallel-group study indicated that after coadministration of ketoconazole for 12 days, the AUC0–∞ of midazolam was about 6.56 times higher than midazolam alone (1,280 ng • hour/mL vs. 195 ng • hour /mL).15 Also, as a pharmacodynamic index, Symbol Digit Modalities Test score of midazolam was reduced to 34.7 from 48.3 when midazolam was co-administered with ketoconazole, which might indicate a significant increase in midazolam-related cognition impairment caused by drug interactions.16 These pieces of evidence showed that midazolam and ketoconazole

1College of Automation, Harbin Engineering University, Harbin, China; 2Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio, USA; 3The Center for Computational Biology and Bioinformatics, School of Medicine, Indiana University, Indianapolis, Indiana, USA; 4Department of Obstetrics and Gynecology, School of Medicine, Indiana University, Indianapolis, Indiana, USA. *Correspondence: Lei Wang (lei.wang2@osumc.edu)

Received: March 6, 2019; accepted: March 29, 2019. doi:10.1002/psp4.12417
have strong drug–drug interactions (DDIs) both in pharmacodynamics and pharmacokinetics (PK). As another example, in a randomized study, the coadministration of irbesartan (IRB) with hydrochlorothiazide (HCT; IRB/HCT 300 mg/25 mg) significantly decreased the HCT area under the curve (AUC) by 26.3% (1,373 ng • hour/mL vs. 1,087 ng • hour/mL HCT alone). At the same time, the effect of IRB on systolic blood pressure when administered with HCT was significantly different from those when IRB was administered alone. The average maximal reductions of systolic blood pressure and diastolic blood pressure when IRB was administered alone were 9.7 ± 5.1 mmHg and 6.1 ± 2.7 mmHg, respectively. Whereas in combination with HCT, these reductions of systolic blood pressure and diastolic blood pressure reached 11.1 ± 4.3 mmHg and 7.2 ± 3.9 mmHg, which suggested a synergistic blood pressure–lowering effect for the combination that was indicated as a result of pharmacodynamic DDI. 17

It is well recognized that both the precipitant’s inhibition potency (e.g., \( K_i \) or IC50 for inhibition, \( E_{\text{max}} \) and EC50 for induction) and the fraction of metabolism (Fm) of the victim drug for each CYP450 enzyme play critical roles for the effective prediction of PK DDI caused by the inhibition and induction of drug metabolism, especially through CYP450 enzymes.

The present study aims to focus on Fm data, which quantify the contribution of different metabolic pathways. There are multiple ways for Fm estimation through clinical PK studies or in vitro PK experiments. First, the contribution of a specific enzyme for a drug’s metabolism can be determined as the change in AUC or clearance (CL) in the absence and presence of a coadministered selective inhibitor according to an in vivo approach. 22,23 For example, Yeung et al. 24 used clinical drug interaction studies in which ketoconazole was used as the CYP3A4 index inhibitor and calculated a drug’s Fm in the CYP3A4 pathway using the following equation:

\[
\text{Fm}_{3A4} = 1 - \frac{\text{AUC}_{\text{control}}}{\text{AUC}_{\text{inhibited}}}
\]

where AUC is the area under the concentration-time curve of the victim drug. For example, bosutinib is metabolized primarily by CYP3A4. When coadministered with ketoconazole, its AUC ratio in plasma (AUC_inhibited/AUC_control) is 8.15. Thus, the Fm3A4 of cinacalcet is 0.88 according to the above equation.25

Second, Fm can be estimated via a pharmacogenetics study where it can be calculated from the fold change in the exposure of a victim drug in extensive metabolizers (EMs) when compared with poor metabolizers (PMs).26 In the study by Silas et al.,27 metoprolol metabolism was studied in a large population of patients. After a single 200 mg oral dose of metoprolol, the average AUC of metoprolol in the blood for the 24 hours of six CYP2D6 PMs was 7,250 ng • mL −1 • hour. Given that CL = Dose/AUC, thus, metoprolol’s Fm2D6 can be calculated from the following formula: Fm2D6 = 1 - CL_PM/CL_EM = 1 - AUC_EM/AUC_PM = 0.828.

Third, in vitro ADME studies in hepatocytes, liver microsomes, or cytosol using radiolabeled substrate (\(^ {14} \text{C} \) or \(^ {3} \text{H} \)), which measure the concentration of the unchanged radiolabeled drug and its metabolites in plasma, urine, and feces, also were regarded as a valuable clinical PK study to estimate the metabolic pathways of a drug.28,29 For example, faldaprevir was used in an ADME study to measure the formation rates of its metabolites by various recombinant human CYP450 isoforms or CYP450-selective chemical inhibitor by human liver microsomes (HLMs). The contribution of each CYP450 to the fraction of metabolites was determined by the rates of metabolite formation after normalization by relative liver content of each CYP450: Fm = (the rates of metabolite formation)/(relative liver content of CYP450). The results showed that the normalized contributions by rCYP3A4 were 0.94 and 0.97 for two kinds of metabolites.30

Several in vitro methods have been developed to determine the contribution of the individual enzyme in a drug’s metabolism. For example, substrate depletion in HLMs is one method that the drug was incubated with or without specific CYP450 selective inhibitors.31 Comparing the metabolism rate, \( V_{\text{max}}/K_m \), of the substrate without any inhibitor, the percent inhibition of a specific CYP450 pathway by the CYP450-selective chemical inhibitor reflects the contribution of this CYP450 toward the substrate’s metabolism. The chemical inhibitors should be potent, selective, and metabolically stable. In addition, a substrate-depletion experiment can be conducted with individual hepatic recombinant human enzymes isoforms. 32 This approach estimates the metabolism rate of the substrate in recombinant human CYP450 and scales the recombinant human CYP450 \( V_{\text{max}}/K_m \) to HLM CLint via an relative activity factors (RAF)/ intersystem extrapolation factors (ISEF) approach, assuming that the CYP450s are the only metabolism enzymes. The Fm was estimated from the percentage contribution of each CYP450 enzyme toward the total HLM CLint.

A metabolite formation study is another in vitro option. For example, after incubation, the mixtures of HLM and carbamazepine were analyzed by high-performance liquid chromatography/mass spectrometry. The rates of carbamazepine metabolites (two-hydroxycarbazepine and three-hydroxycarbazepine) formation were determined in microsomes and then compared with typical CYP450 enzyme activities. The data were analyzed by nonlinear regression (GraFit 5; Erithacus Software Ltd., Surrey, UK) and linear transformation (Eadie-Hofstee plots) to estimate the apparent PK values and enzyme models, respectively. The formation of two-hydroxylated and three-hydroxylated carbamazepine metabolites was evaluated in the presence or absence of known P450 inhibitors. 33 HLMs from high/low CYP450 activity donors were used to estimate the inhibition percentage of carbamazepine metabolite formation. In the meantime, cDNA-expressed isoforms were examined for the affinity of different metabolite formation.

Recently, as a result of the success of the cryopreservation of human hepatocytes, 34 a hepatocyte suspension model 35 became a new method to estimate Fm. Physiologically, the cryopreserved human hepatocyte is closer to human hepatic
metabolism than the other in vitro system. Desbans et al.\textsuperscript{36} used cryopreserved human hepatocytes from 12 donors to estimate \( \text{Fm}_{3A} \) for five prototypical CYP3A substrates with varying degrees of CYP3A-dependent in vivo CL using intrinsic metabolic stability measurements in the presence and absence of the CYP3A probe inhibitor ketoconazole. After hepatocytes were incubated with test compounds and/or the inhibitor, the intrinsic CL was estimated from the parent compound depletion profile. Then \( \text{Fm}_{3A} \) was calculated from the ratio between CL\textsubscript{int} in the absence and presence of ketoconazole as:

\[
\text{Fm}_{3A} = 1 - \frac{\text{CL}_{\text{int}} \text{ with ketoconazole}}{\text{CL}_{\text{int}} \text{ without ketoconazole}}
\]

Although there are some widely used databases for drug metabolism—such as (i) DrugBank\textsuperscript{37} which is a comprehensive database that combines detailed drug (e.g., chemical, pharmacological, and pharmaceutical) data with comprehensive drug target information (e.g., sequence, structure, and pathway); (ii) Transformer (the former Super CYP450),\textsuperscript{38} which integrates CYP450 enzyme interactions and some pharmacological information; and (iii) DIDB,\textsuperscript{39} which can evaluate the impact of DDI in the clinic by in vitro and in vivo DDI data—limited Fm data can be found in these databases.

In this article, we present our initial effort in developing the Fm database on CYP450. As a demonstration example, we primarily focus on small-molecule cancer drugs because of their broad applications. We further focus on the hepatic CYP450-based Fm calculation, as hepatic CYP450 enzymes are the most widely investigated. Therefore, in our data-curation study, Fm data are collected and estimated from in vitro inhibition studies of HLM, including metabolite-formation and substrate-depletion studies. The selection for this in vitro system was mainly because of its rich data sources. Other data types from clinical studies (e.g., pharmacogenetics (PGx) PK studies, or clinical PK drug interaction studies) or in vitro experiments (i.e., cytosol or hepatocyte studies) for Fm estimation were also investigated in our data-curation process. However, these data sources turned out to be highly sparse. They will be further explored in the future.

**MATERIALS AND METHODS**

The Fm database was curated from published articles in the PubMed index (http://www.ncbi.nlm.nih.gov/pubmed/). An overview of the data collection is available in Figure 1. In cancer drug selection, 237 cancer drugs approved by the US Food and Drug Administration (FDA; Table S1) were identified in DrugBank and the National Cancer Institute. We mainly focused on the 188 small-molecule cancer drugs. The next stage is the keyword search, including cancer drug names, “CYP450,” “HLMs,” and/or “metabolism.” If the drug has specific CYP450 metabolism information in DrugBank, the keyword may be composed of the drug name and the specific CYP450. Cancer drug generic names and their synonyms and brand names published in DrugBank were included in the search. Similarly, CYP450 enzyme names include their synonymous names in the HUGO Gene Nomenclature Committee.

**Figure 1** The flowchart of data-collection procedure. Fm, fraction of metabolism.

All of the identified PubMed papers were then filtered. In each abstract, we checked whether this was a drug metabolism study, whether the contribution of CYP450 to the metabolism were investigated, or whether the CYP450 enzyme inhibitors was discussed in the abstracts. If neither of these pieces of evidence was reported in the abstract, this paper was then removed. If an abstract passed the filtering step, its full text received a further examination. In the article Method section, substrate depletion and CYP450 enzyme inhibition in HLM study information or the information in a metabolite-formation study were examined. In particular, in the metabolite-formation studies, several experiment details were required. All incubations were performed at...
37°C; HLMs were incubated with sodium phosphate buffer and NADPH (usually NADP + glucose-6-phosphate dehydrogenase) before adding the substrates, and the drug was incubated with pooled HLMs in the presence or absence of P450 selective chemical inhibitors for several minutes; control samples containing no NADPH or substrates were also included. Incubations were carried out for a defined time, and at the final stage, the drug metabolites were evaluated using high-performance liquid chromatography/LC-MS/MS in the presence or absence of known CYP450 enzyme selective inhibitors. In the substrate-depletion studies, several experiment details were required to report. Preincubation and incubation procedures were the same as the metabolite-formation studies, and those that had no NADPH or substrate samples for control groups were also needed. After being vortexed, the concentration of the remaining parent drugs in the supernatants were measured using the high-performance liquid chromatography assay in the presence or absence of selective inhibitors. If this information was not reported in the Method section for either a substrate-depletion study or a metabolite-formation study, the paper was then removed.

The Results section was then reviewed carefully, including the narrative, figures and tables (including their legends), and the supplemental materials if available. We collected data from the paper and calculated the corresponding Fms. The following are Fm calculation steps for the metabolite-formation data:

1. The following data elements were extracted from the results: the metabolites for each drug and/or their relative contribution of the metabolisms, the percentage of inhibitions for each CYP450 enzyme, and their related CYP450 inhibitor. (Sometimes this percentage was directly reported. Otherwise, the metabolism rates under inhibition and control were reported.)
2. The formation of metabolites by HLMs estimated in the presence or absence (i.e., a control sample) of known P450 isoform-selective inhibitors. The relative proportion of the metabolite formation in the reaction mixtures with no inhibitor was set as 100%. If the inhibition percentage of the drug metabolite could not be collected directly from the paper, the proportion was estimated from the changes in the portion of control metabolite that was calculated relative to incubations with no inhibitor.
3. If multiple metabolites were found to be the predominant drug metabolic pathways (such as exemestane, tamoxifen, toremifene, lapatinib), the contribution of each metabolite in the total drug metabolism was calculated.
4. If there existed several competitive inhibitors for a specific CYP450 enzyme, the percent of inhibition for the inhibitor was calculated. The mean value was calculated and taken as the inhibition percentage when there were multiple inhibitors.
5. If the substrate concentration of the inhibition experiment varied in the paper, the rate of inhibition at each concentration was calculated. Then their mean value was calculated.
6. If there were several HLM samples in the experiment, the percentage of inhibition at each sample was calculated and the mean value was taken (such as fluorouracil).

7. The total inhibition percentage of one metabolite was normalized. The fraction of metabolite for the $i$th CYP450 enzyme in the $j$th metabolite is:

$$F_{m_{ij}} = \frac{\text{inhibition}_{i}}{\sum \text{inhibition}_{j}} \times \text{percentage of metabolite/j}$$

where inhibition$_i$ refers to the percentage of inhibition for the $i$th enzyme. The sum of $F_{m_{ij}}$ over all metabolites is regarded as the fraction of metabolized for enzyme $i$.

For example, two metabolite pathways of tramadol were reported in ref. 40 with CYP450 information. According to the data collection pipeline we proposed previously, we can collect a proportion of the contribution of each pathway and normalize:

| Pathway | % Contribution | Normalized |
|---------|----------------|------------|
| M1      | 32.881         | 33.532     |
| M2      | 65.18          | 66.471     |

Then CYP450 inhibition information can be collected from each metabolite pathway and normalized by metabolite pathway contribution:

Finally, Fm is estimated as:

| Drug    | Pathway | 1A2 | 2C9 | 2C19 | 2D6 | 2E1 | 3A4 |
|---------|---------|-----|-----|------|-----|-----|-----|
| Tramadol| M2      | 7.381| 2   | 2.2  | 16.171| 2.592| 3.193|
|         | M1      | 6.701| 6.701| 6.701| 7.732| 7.732| 31.431|
| Total   |         | 14.082| 8.18| 8.9  | 23.903| 10.324| 34.634|

On the other hand, Fm can be calculated from the substrate-depletion study:

1. The inhibition percentage or the remaining proportions of substrates was evaluated in the presence or absence (i.e., a control sample) of known CYP450 isoform-selective inhibitors. It was assumed that the sum of inhibition percentage of a specific substrate and its remaining proportion equaled to 100%. If the
inhibition percentage of the drug was not collected directly in the paper, the ratio was estimated from the changes of a control sample or the remaining percentage of the substrate.

2. If a CYP450 enzyme was inhibited by several inhibitors simultaneously, the mean percent of inhibition for all inhibitors was chosen.

3. If there were two or more substrate concentrations of the inhibition experiment, we took the mean percentage of inhibition at all concentrations for estimation.

4. The metabolized fraction for the $i$th CYP450 enzyme is:

$$\text{Fm}_i = \frac{\text{inhibition}_i}{\sum \text{inhibition}_i}$$

where inhibition$_i$ refers to the percentage of inhibition for the $i$th enzyme.

To ensure data integrity, two curators with biology backgrounds conducted the data-curating process. Dr. Sara K. Quinney, who has an extensive pharmacology training background further checked any differentially annotated abstracts. Then the data extraction from the full text was carried out by those two annotators again. Among the disagreed data collection between these two annotators, a group review was conducted by Drs. Wang, Quinney, and Li to reach the final agreement.

The data-validation step was conducted, and two sets of drugs were used for validation. The part of the drugs with Fm data was selected randomly as the positive set, and their Fm data were reevaluated. Some cancer-related drugs without Fm data were selected randomly as the negative set and reevaluated. During this validation process, two independent annotators went through the entire Fm data-curating process for these 16 drugs. These two validation annotators have masters or doctorate degrees in computational biology and/or biology. Drs. Quinney and Li, who have pharmacology backgrounds, further evaluated the concordance among these three sets of annotations. The consistency of the Fm data is reported later.

RESULTS

Among 237 cancer-related drugs (Table S1), there are 42 drugs that have identified and curated Fm data (Table S2) from the PubMed literature. During the validation process, among the eight negative cancer-related drugs (i.e., no Fm data detected from the first annotator), the other two validation annotators also did not find their Fm data. Therefore, the validation accuracy rate is 100% for the negative drug set. Among those eight positive cancer drugs, the first validation annotator managed to find Fm data for seven of eight drugs, and the second validation annotator found the Fm data for all eight drugs. The overlap rate is 93.8% for the positive drug set. However, the consistency of the Fm value was not good. The first validation annotator had consistent Fm calculations for six of eight drugs, 75%. The inconsistency was primarily attributed to one completely missed drug and another one that was miscalculated. The second validation annotator had only 62.5% of Fm values that were in concordance with the original annotator because he only considered the top one or two major CYP450 metabolic pathways and ignored the other minor pathways. Finally, among the discordant Fm values for the positive cancer drugs, their full-text papers and Fm values were further evaluated by two additional pharmacologists. They found the original annotator had the highest accuracy in calculating the Fm. She was right on 90% of the discordant values among the three annotators. The validation results are shown in Figure 2.

All of the extracted and validated Fm data are presented in Table S2. In this data set, the key data elements included the drug names, PubMed unique identifier, and Fms for various CYP450 and their corresponding inhibitors. Especially note that column C annotates whether the data were from the substrate-depletion study or the metabolite formation study, and columns D–P display Fm data for various CYP450 pathways. The numbers represent the percentiles, and a blank indicates that the corresponding CYP450 pathway was not being investigated. Please keep in mind that in some experiments, there existed more than one inhibitor for one enzyme, and we annotate them all in the database.

One important utility of the Fm database was to predict drug interactions. These data characterized all of the hepatic CYP450 metabolic pathways and their contributions in predicting drug interactions. We therefore further explored and predicted the drug interactions between these 42 cancer drugs (i.e., substrates) and 178 drugs (i.e., inhibitors), in which their inhibitions on CYP450 were known, in the following case study.

AUC ratio (AUCR) is the key parameter to measure drug interaction. The predicted AUCR estimate as follows:

$$\text{AUCR} = \frac{\text{AUC}_i}{\text{AUC}} = \frac {1} {(1-f_e) \sum_{i=1}^{n} \text{Fm}_i + \sum_{j=1}^{m} \frac {1} {K_i} + f_e}$$

where AUC/AUC is the ratio of the area under the plasma concentration-time profile of the substrate drug in the presence (AUC) and absence (AUC) of the inhibitor drug; Fm is the fraction of the total hepatic metabolism mediated through a CYP450 enzyme (from our Fm database), $f_e$ is the fraction of the unchanged drug via urine. $I_i$ is the unbound inhibitor concentration, $K_{iu}$ is the unbound inhibition constant. In this paper, $I_i$ equals $C_{\text{max}} \times f_u$, where $C_{\text{max}}$ is the maximum concentration, and $f_u$ is the fraction of unbound drug in plasma. In this paper, we have 42 cancer drugs with Fm and $f_u$, there are 178 drugs (16 cancer drugs and 162 noncancer drugs where there was nonoverlap with 237 cancer drugs) with $K_i$ and $I_i$, with $K_i$ obtained from published literature and $I_i$ collected from published clinical studies, and comprehensive pharmacoinformatics databases such as DIDB and Goodman and Gilman’s The Pharmacological Basis of Therapeutics, where most of the data were collected through the PubMed literature review.

Following the FDA DDI guideline and expert experience, an AUCR > 1.25 is regarded as PK DDI evidence. Balancing recommendations provided by regulatory agencies and the scientific community, we define AUCR > 2 as evidence of significant PK DDI. Based on our steady-state model-based...
DDI predictions, we find 31 drug pairs (the substrate is a cancer drug, and the inhibitor is a noncancer drug) with an AUCR > 2. After having been validated in DrugBank, Drugs.com and PubMed, 15 drug pairs have supporting clear clinical PK DDI evidence in these databases. There are only four inhibitors (acetaminophen, amprenavir, ritonavir, and indinavir) in Table 1 because the $K_i$ on the CYP3A4 of these four drugs (1.855 μM, 0.79 μM, 0.18 μM, and 3.51 μM) were less than 10 μM, a value considered in this study as a cutoff that indicated potential inhibition to the CYP3A4 substrate.

**DISCUSSION**

The primary goal of this paper is to provide a database of Fm. In this cancer drug Fm database, the relative contributions of the CYP450 isozyme were curated and calculated from drug metabolism studies using the HLMs. First, relevant papers were identified through the keyword search in PubMed. Second, only the articles with relevant in vitro experiment information were selected. These in vitro experiments included both metabolite formation and substrate-depletion studies. In this database, there are four drugs ( exemestane, tamoxifen, toremifene, lapatinib) whose Fms were calculated from their metabolite-formation experiments, and the other 38 drugs were estimated from substrate-depletion studies. A predefined data-curating protocol was established to assure data quality and data reproducibility. Multiple annotators were employed in the data-filtering, data-curating, and data-validation stages. Two independent validation annotators reevaluated the randomly selected drugs from positive
Table 1 The validation results of drug–drug interaction prediction based on steady-state model for AUC

| Substrate     | Inhibitor       | AUCR  | DrugBank | Drugs.com | Pubmed (clinical) |
|---------------|-----------------|-------|----------|-----------|-------------------|
| Fenretinide   | Acetaminophen   | 11.19 | —        | —         | The fenretinide plasma level increased sharply after ceftriaxone and acetaminophen were initiated. Kang et al.45 |
| Aprepitant    | Acetaminophen   | 5.75  | —        | —         | —                 |
| Bosutinib     | Acetaminophen   | 4.68  | —        | —         | —                 |
| Vinblastine   | Amprenavir      | 3.06  | —        | —         | Combining these medications may significantly increase the blood levels and effects of vinblastine. |
| Trabectedin   | Acetaminophen   | 3     | —        | —         | Trabectedin may cause liver problems, and using it with other medications that can also affect the liver such as acetaminophen may increase that risk. |
| Everolimus    | Acetaminophen   | 2.39  | —        | —         | —                 |
| Aprepitant    | Amprenavir      | 2.3   | —        | —         | Amprenavir may increase the blood levels and effects of aprepitant. |
| Vinblastine   | Ritonavir       | 2.24  | —        | —         | Combining these medications may significantly increase the blood levels and effects of vinblastine. | There are multiple case reports describing serious infectious complications in patients receiving vinblastine and ritonavir. Ezzat et al.46 |
| Gefitinib     | Amprenavir      | 2.21  | —        | —         | Amprenavir may increase the blood levels and effects of gefitinib. |
| Bosutinib     | Amprenavir      | 2.17  | —        | —         | Amprenavir may significantly increase the blood levels of bosutinib. |
| Vincristine   | Acetaminophen   | 2.14  | —        | —         | —                 |
| Axitinib      | Amprenavir      | 2.11  | —        | —         | Amprenavir may significantly increase the blood levels of axitinib. This may increase side effects such as high blood pressure, diarrhea, nausea, vomiting, constipation, decreased appetite, weight loss, and rash, itching, or peeling of skin on the hands and feet. |
| Vinblastine   | Indinavir       | 2.09  | —        | —         | Combining these medications may significantly increase the blood levels and effects of vinblastine. |
| Tamoxifen     | Acetaminophen   | 2.07  | —        | —         | —                 |
| Vinorelbine   | Amprenavir      | 2.06  | —        | —         | Combining these medications may significantly increase the blood levels and effects of vinorelbine. |

AUCR, AUC ratio.
*Information is from a case report.
and negative sets. The overlap rates were 93.8% and 100% for the positive and negative sets, respectively. After being evaluated by two additional pharmacologists, the original annotation has the highest accuracy in collecting and calculating the Fm (Figure 2).

This paper, for the first time, demonstrates a well-characterized data-curating procedure in constructing a pharmacology database. Although there are many pharmacology databases, there is limited published evidence on how data annotation and validation are conducted, and most importantly, where the potential errors come from and how these errors are identified and corrected during the data-curating process. To our knowledge, the best practice is a well predefined data-curating guideline. The more accurate and detailed it is, the less likely it is the annotators will make mistakes. A double-annotation process is also a great way to identify potential errors. We believe this is one of the primary contributions in this paper.

In our initial effort to curate and calculate Fm data from the literature, we chose to focus on primarily the Fm data from the HLM inhibition experiments. These experiments are suitable for studying the relative contribution of CYP450 isozymes to drug metabolism. If we want to evaluate the contribution of other enzymes (e.g., uridine 5’-diphospho-glucuronosyltransferase) to drug metabolism, we will need to explore hepatocyte or cytosol studies. Although we focused on HLM studies, which can provide us with abundant data, their Fms are limited to CYP450 metabolism enzymes. In the future, we will further expand Fm data collections to the hepatocyte or cytosol studies.

Publication bias may influence the quantity and quality of the available data. We only search literature from PubMed, not considering other resources. All of the drugs did not have in vitro data. We also recognize that there is no selective inhibitor available in vitro for CYP2C19 and CYP2B6. Also, there is a lack of discussion for the fraction of the renal CL with the unchanged drug and the inducer. Our Fm estimates may not be entirely accurate. Furthermore, in our Fm estimation, we assumed a total inhibition of the in vitro experiment, which is also questionable in reality. It may lead to biased Fm estimates.

There are additional data, such as pharmacogenetics and drug interaction clinical PK studies, for Fm estimation. These studies could provide more accurate data than in vitro studies. According to PharmGKB, there are two drugs among 42 cancer drugs—docetaxel and paclitaxel—where Fm can be estimated from genetic polymorphisms. The related literature about the CYP450 variant can find in PharmGKB, and we can calculate the contribution of the CYP450 enzyme via the pharmacogenetics research. Using the informatics approach, we have successfully identified Fm data for 42 of 237 cancer drugs. In this paper, we focus on the drug metabolism data related to CYP450, and these data are generated from either metabolite formation studies or substrate-depletion studies. To create a more comprehensive Fm database, we will further consider alternative metabolism pathways, such as uridine 5’-diphospho-glucuronosyltransferases; alternative in vitro experiments, such as hepatocytes; and alternative clinical data, such as pharmacogenetic PK studies or clinical DDI studies. In addition, we will develop annotation guidelines for each task and conduct literature reviews and validation analyses afterward.

**Supporting Information.** Supplementary information accompanies this paper on the CPT: Pharmacometrics & Systems Pharmacology website (www.psp-journal.com).

**Table S1.** Drug list for 237 cancer drugs.

**Table S2.** Fraction of drug metabolism.

Application: AUCR.xlsx: The predicted AUCR results, fe.txt: The fraction of unchanged drug via urine, fm.txt: The fraction of drug metabolism, kui.txt: The unbound inhibition constant of inhibitors, Model code.docx: The code was used for AUCR prediction.

**Funding.** This work has been supported by several National Institutes of Health grants (DK102694, GM10448301-A1, R01 GM117206, and R01LM011945) and National Science Foundation Grant NSF1622526. It has also been supported by the Natural Science Foundation of Heilongjiang Province, China (QC2016086) and the Fundamental Research Funds for the Central Universities, China (HEUCF160420).

**Conflict of Interest.** The authors declared no competing interests for this work. As an Associate Editor for CPT: Pharmacometrics & Systems Pharmacology, Lang Li was not involved in the review or decision process for this article.

**Author Contributions.** L.H., S.K.Q., L.C. and L.W. wrote the manuscript. L.L., W.C. and C.-W.C. designed the research. L.H., F.W. and C.-W.C. performed the research. L.H., X.W., J.L. and L.W. analyzed the data.
1. Rowland, M. & Tozer, T.N. Clinical Pharmacokinetics/Pharmacodynamics. 4th edn. (Lippincott Williams and Wilkins, Philadelphia, PA, 2005).

2. Shen, D.D., Kunze, K.L. & Thummel, K.E. Enzyme-catalyzed processes of first-pass hepatic and intestinal drug extraction. *Adv. Drug Deliv. Rev.* 27, 99–127 (1997).

3. Renwick, A. The metabolism of antihistamines and drug interactions: the role of cytochrome P450 enzymes. *Clin. Exp. Allergy* 29, 116–124 (1999).

4. Ereshefsky, L. Drug–drug interactions involving antidepressants: focus on venlafaxine. *J. Clin. Psychopharmacol.* 16, 375–385 (1996).

5. Eichelbaum, M., Ingelman-Sundberg, M. & Evans, W.E. Pharmacogenomics and individualized drug therapy. *Annu. Rev. Med.* 57, 119–137 (2006).

6. Shah, R.R. & Smith, R.L. Inflammation-induced phenocconversion of polymorphic drug metabolizing enzymes: hypothesis with implications for personalized medicine. *Drug Metab. Dispos.* 43, 400–410 (2015).

7. Fitzgerald, R.J. Medication errors: the importance of an accurate drug history. *Br. J. Clin. Pharmacol.* 67, 671–675 (2009).

8. Admassie, E., Melese, T., Mequaneut, W., Hailu, W. & Srikanth, B.A. Extent of polymorphism, occurrence and associated factors of drug–drug interaction and potential adverse drug reactions in Sondar Teaching Referral Hospital, North West Ethiopia. *J. Adv. Pharm. Technol. Res.* 4, 183–189 (2013).

9. Sasaki, T. Enzyme- and transporter-mediated drug interactions. *Clin. Pharmacol. Ther.* 83, 151–154 (2008).

10. Rosholm, J.U., Bjerrum, L., Hallas, J. & Gram, L.F. Polypharmacy and the risk of drug–drug interactions among Danish elderly. *A prescription database study. Dan. Med. Bull.* 45, 210–213 (1998).

11. Vanol, J.T. et al. A randomized, controlled trial of a clinical pharmacist intervention to improve inappropriate prescribing in elderly outpatients with polypharmacy. *Am. J. Med.* 100, 426–437 (1996).

12. Vose, C. & Ings, R. Drug Metabolism. (Royal Society of Chemistry, London, 2014).

13. Yu, J., Ritchie, T.K., Mulgaonkar, A. & Ragueneau-Majilessi, I. Drug disposition and inter-individual variability in pharmacokinetics using virtual human populations and physiologically-based pharmacokinetic modeling. *AAPS J.* 17, 216–231 (2015).

14. James, M., Dickinson, G.L. & Rostami-Hodjegan, A. A framework for assessing the in vitro relevance of human cytochrome P450 enzymes in the human liver. *J. Hepatol.* 56, 809–814 (2012).

15. Shoaf, S.E., Bricmont, P. & Mallikaarjun, S. Effects of CYP3A4 induction and inhibition on the pharmacokinetics and pharmacodynamics of tolvaptan, a non-peptide AVP antagonist in healthy subjects. *Br. J. Clin. Pharmacol.* 73, 579–587 (2012).

16. Lam, Y., Alfaro, C.L., Ereshefsky, L. & Miller, M. Pharmacokinetic and pharmacodynamic interactions of oral midazolam with ketoconazole, fluvoxamine, fexofenadine, and nefazodone. *J. Clin. Pharmacol.* 43, 1274–1282 (2003).

17. Hedaya, M.A. & Helmy, S.A. Modeling of the pharmacokinetic/pharmacodynamic interaction between irbesartan and hydrochlorothiazide in normotensive subjects. *Biopharm. Drug Dispos.* 36, 216–231 (2015).

18. Jamei, M., Dickinson, G.L. & Rostami-Hodjegan, A. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of ‘bottom-up’ vs ‘top-down’ recognition of covariates. *Drug Metab. Pharmacokinet.* 24, 53–75 (2009).

19. Rostami-Hodjegan, A. & Tucker, G.T. Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nat. Rev. Drug Discov.* 1, 140–148 (2002).

20. Rostami-Hodjegan, A. & Tucker, G. “In silico” simulations to assess the “in vivo” consequences of “in vitro” metabolic drug–drug interactions. *Drug Discov. Today. Technol.* 1, 441–448 (2004).

21. Zhang, D., Zhu, M. & Humphreys, W.G. Drug metabolism in Drug Design and Development (John Wiley & Sons, Hoboken, NJ, 2007).

22. Le Coutre, P. et al. Nilotinib (formerly AMN107), a highly selective BCR-ABL tyrosine kinase inhibitor, is active in patients with imatinib-resistant or intolerant accelerated-phase chronic myelogenous leukemia. *Blood* 111, 1834–1839 (2008).

23. Creighton, C.J. et al. Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res.* 68, 7493–7501 (2008).

24. Yeung, C.K. et al. Organelle impairment–drug–drug interaction database: a tool for evaluating the impact of renal or hepatic impairment and pharmacokinetic inhibition on the systemic exposure of drugs. *CPT: Pharmacometrics Syst. Pharmacol.* 4, 489–494 (2015).

25. Harris, R.Z., Safi, M., Sullivan, J.T. & Padhi, D. Pharmacokinetics of cinacalcet hydrochloride when administered with ketoconazole. *Clin. Pharmacol. Ther.* 46, 495–501 (2007).

26. Ito, K., Hallifax, D., Obach, R.S. & Houston, J.B. Impact of parallel pathways of drug elimination and multiple cytochrome P450 involvement on drug–drug interactions: CYP2D6 paradigm. *Drug Metab. Dispos.* 33, 837–844 (2005).

27. Silas, J.H., McSourty, J.C., Lennard, M.S., Tucker, G.T. & Woods, H.F. Polymorphic metabolism of metoprolol: clinical studies. *Eur. J. Clin. Pharmacol.* 28, 85–90 (1985).

28. Bohnert, T. et al. Evaluation of a new molecular entity as a victim of metabolic drug–drug interactions—a small-world industry? *Drug Metab. Dispos.* 44, 1399–1423 (2016).

29. Rodrigues, A.D., Winchell, G.A. & Dobrianska, M.R. Use of in vitro drug metabolism data to evaluate metabolic drug–drug interactions in man: the need for quantitative databases. *J. Clin. Pharmacol.* 41, 368–373 (2001).

30. Li, Y. et al. Enzyme-transporter interplay in the formation and clearance of abundant metabolites of faldaprevir found in excreta but not in circulation. *Drug Metab. Dispos.* 42, 384–393 (2014).

31. Hukey, S.W., Dean, D.C., Miller, R.R., Rasmussen, G.H. & Chu, S.H. Identification of human cytochrome P450 isoforms responsible for the in vitro oxidative metabolism of finasteride. *Drug Metab. Dispos.* 23, 1126–1135 (1995).

32. Li, Z.M., Guo, L.H. & Ren, X.M. Biotransformation of 8-fluorotoluol alcohol by recombinant human cytochrome P450s, human liver microsomes and human liver cytosol. *Environ. Sci. Proc. Imp.* 18, 536–546 (2016).

33. Pearce, R.E., Vakkalagadda, G.R. & Leeder, J.S. Pathways of carbamazepine bioactivation in vitro. Characterization of human cytochromes P450 responsible for the formation of 2- and 3-hydroxylated metabolites. *Drug Metab. Dispos.* 30, 1170–1179 (2002).

34. Stephens, X., Najimi, M. & Sokal, E.M. Hepatocyte cryopreservation: is it time to change the strategy? *World J. Gastroenterol.* 16, 1–14 (2010).

35. Mao, J., Mohutsky, M.A., Harrellson, J.P., Wrighton, S.A. & Hall, S.D. Prediction of CYP3A-mediated drug-drug interactions using human hepatocytes suspended in human plasma. *Drug Metab. Dispos.* 39, 591–602 (2011).

36. Desbans, C. et al. Prediction of fraction metabolized via CYP3A in humans utilizing cryopreserved human hepatocytes from a set of 12 single donors. *Xenobiotica* 44, 17–27 (2014).

37. Wishart, D.S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34(Database issue), D668–D672 (2006).

38. Peters, S.A. Physiologically-Based Pharmacokinetic (PBPK) Modeling and Simulations: Principles, Methods, and Applications in the Pharmaceutical Industry (John Wiley, Hoboken, NJ, 2012).

39. Benet, L.Z., Broccatelli, O. & Oprea, T.I. BDSS applied to over 900 drugs. *AAPS J.* 13, 519–547 (2011).

40. Laurence, B., Bruce, A. & Bjorn, K. Goodman and Gilman’s The Pharmacological Basis of Therapeutics 12th edn. (Pergamon Press, New York, NY, 1990).

41. US Food and Drug Administration: Drug interaction studies—study design, data analysis, implications for dosing, and labeling recommendations (draft guidance) (US Food and Drug Administration, Silver Spring, MD, 2012).

42. Kang, M.H. et al. Probable fatal drug interaction between intravenous fentanil, ceftriaxone, and acetaminophen: a case report from a New Approaches to Cancer Therapy (NANT) Phase I study. *J. Hematol. Thromb.* 2, 258 (2014).

43. Ezzat, H.M. et al. Incidence, predictors and significance of severe toxicity in patients with human immunodeficiency virus-associated Hodgkin lymphoma. *Leuk. Lymphoma* 53, 2390–2396 (2012).