Predicting Chemical Toxicity Effects Based on Chemical-Chemical Interactions

Lei Chen²,³, Jing Lu³,⁴, Jian Zhang⁵, Kai-Rui Feng⁵, Ming-Yue Zheng³,⁶, Yu-Dong Cai²,⁶

¹ Institute of Systems Biology, Shanghai University, Shanghai, China, ² College of Information Engineering, Shanghai Maritime University, Shanghai, China, ³ Drug Discovery and Design Center (DDDC), Shanghai Institute of Materia Medica, Shanghai, China, ⁴ Department of Ophthalmology, Shanghai First People’s Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China, ⁵ Simcyp Limited, Blades Enterprise Centre, Sheffield, United Kingdom

Abstract

Toxicity is a major contributor to high attrition rates of new chemical entities in drug discoveries. In this study, an order-classifier was built to predict a series of toxic effects based on data concerning chemical-chemical interactions under the assumption that interactive compounds are more likely to share similar toxicity profiles. According to their interaction confidence scores, the order from the most likely toxicity to the least was obtained for each compound. Ten test groups, each of them containing one training dataset and one test dataset, were constructed from a benchmark dataset consisting of 17,233 compounds. By a Jackknife test on each of these test groups, the 1st order prediction accuracies of the training dataset and the test dataset were all approximately 79.50%, substantially higher than the rate of 25.43% achieved by random guesses. Encouraged by the promising results, we expect that our method will become a useful tool in screening out drugs with high toxicity.

Introduction

Toxicity is a key cause of late-stage failures in drug discovery. Even some approved drugs such as Phenacitin [1] and Troglitazone [2] have been withdrawn from the market because of unexpected toxicities that were not detected during Phase III clinical trials. Thus, early toxicity data on compounds are needed to reduce R&D costs. Evaluating toxicity and assessing risks of diverse chemicals require comprehensive experimental testing against a broad spectrum of toxicity end points. These tests can cost millions of dollars, involving several thousand animals, and take many years to complete. As a result, very few chemicals have undergone the degree of testing needed to support accurate health risk assessments or meet regulatory requirements for drug approval. In recent years, the number of synthetic compounds has surged with the advance of combinatorial chemistry, and accordingly large quantities of toxicity data are urgently demanded.

Recently, particular interest has been raised to apply fast and cost-effective in silico toxicological models to supplement those in vitro and in vivo testing. These models require high quality toxicity data for a large set of structurally diverse drug candidates. Accelrys Toxicity is a database of toxicity information compiled from the open scientific literature [3] and containing toxicological data for approximately 0.17 million chemicals. This database is of great value for investigating the pharmacokinetic properties, metabolism and potential toxicities of compounds. Six types of toxicity data are collected in the database: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; and (6) Multiple Dose Effects. It should be noted that these categories have multiple and overlapping mechanisms of toxic action and each category represents only specific types of experiments. The combination of these experimental results may help define the overall safety profile of a compound. However, this kind of databases only provides toxicological information for recorded compounds, not for new ones. It would be valuable to accurately predict toxicities of a new compound based on the information available for recorded compounds. In order to meet the demand, there is a drive to develop quick, reliable, and non-animal-involved prediction methods, e.g. using structure-activity relationships (SARs) to predict drugs toxicities.

Currently, most toxicological SAR models belong to binary classifiers, which only predict compounds to be toxic or non-toxic within a single toxicity class [4,5]. It is desired to modify the strategy to predict a series of toxicity effects. In this study, we chose to build a multiclass model to predict six categories of toxicity using the Accelrys Toxicity database instead of only one or two toxicity endpoints. However, the quadratic optimization problem...
in multiclass models is difficult to solve. Thus, many previous multiclass approaches tended to decompose a multiclass problem into multiple independent binary classifications. Investigators built a set of binary classifiers, such as the model of Dietterich et al [7], each classifier distinguishing only one of the classes from the others. Although this greatly simplifies the problem, such an approach cannot provide order prediction information for the query compounds. That is, it can only predict whether the query compound has some toxicity end points, but cannot determine which is the most likely toxicity, or even the order of toxicity end points by toxicity likelihoods.

In recent years, the assessment of protein-protein interactions has been widely used to predict many attributes of proteins [8,9,10,11]. Furthermore, multiclass predictions of protein attributes have become more common [12,13,14]. These methods and their results show that interactive proteins tend to share the same functions with higher probability than do non-interactive ones. Likewise, it is reasonable to expect that interactive compounds are also more likely to share common functions as indicated by some pioneer studies [15,16]. Thus, toxicity, as part of the biological functions of compounds, should follow the same rule. Moreover, based on a previous work on the Anatomical Therapeutic Chemical (ATC) classification of drugs [16], compared to the SAR models based on physicochemical descriptors or structural alerts, a model based on chemical-chemical interactions can rank the order of the predictions more easily and yield better prediction results. In our study, we attempt to quantify chemical-chemical interactions for each pair of interactive compounds, and obtain the confidence scores of the interactions by which the toxicity end points were ordered. Briefly, compounds of seven categories including six categories of toxicity plus non-toxicity were collected. The interactive compounds of each query compound were identified utilizing STITCH (Search tool for interactions of chemicals) [17,18]. Then, the score of each class of the query compound was obtained from the confidence scores of interactions between the query compound and its interactive compounds using the toxicity profile of the interactive compounds. Finally, the prediction quality of the model was evaluated using the Jackknife test through ten test groups. Each of these was constructed from the benchmark dataset and contained one training dataset and one external test dataset. Details are described in the following sections.

**Materials and Methods**

**Benchmark Dataset**

We obtained a total of 171,266 compounds from the Accelrys Toxicity Database 2011.4 [19], which had at least one toxicity effect belonging to the following six categories: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; (6) Multiple Dose Effects. Based on compound toxicity, these compounds are allocated to the 6

<table>
<thead>
<tr>
<th>Tag</th>
<th>Toxicity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₁</td>
<td>Acute Toxicity</td>
<td>12,633</td>
</tr>
<tr>
<td>T₂</td>
<td>Mutagenicity</td>
<td>6,110</td>
</tr>
<tr>
<td>T₃</td>
<td>Tumorigenicity</td>
<td>2,293</td>
</tr>
<tr>
<td>T₄</td>
<td>Skin and Eye Irritation</td>
<td>2,353</td>
</tr>
<tr>
<td>T₅</td>
<td>Reproductive Effects</td>
<td>2,501</td>
</tr>
<tr>
<td>T₆</td>
<td>Multiple Dose Effects</td>
<td>4,198</td>
</tr>
<tr>
<td>T₇</td>
<td>Non-toxicity</td>
<td>646</td>
</tr>
<tr>
<td>Total</td>
<td>--</td>
<td>30,734</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0056517.t001
categories, allowing multiple assignments. In addition, 2,871 “nontoxic” compounds including FDA-approved drugs from DrugBank [20] and endogenic metabolites from the Human Metabolome database (HMDB) [21] were collected and labeled as a negative class. For convenience, the ‘n-toxic’ set is regarded as the 7th category of compound toxicity. Due to lack of chemical-chemical interaction information in STITCH [17,18], some compounds cannot be investigated by this approach. After excluding these compounds, a benchmark dataset \( S \) consisting of 17,293 compounds was retrieved, of which 16,587 were toxic and 646 were nontoxic. These compounds are classified into 7 categories of compound toxicity. Shown in Table 1 is the distribution of compounds in each category. The codes of 17,293 compounds and their toxicity information can be found in Table S1.

It is observed from Table 1 that the sum of the number of compounds in all the 7 categories is much larger than the number of compounds, indicating that some compounds are allocated to more than one category of toxicity. Of the 17,293 compounds in the benchmark dataset, 10,151 compounds belong to only one category of toxicity, 3,475 compounds belong to two categories of toxicity, while others belong to 3–5 categories of toxicity and no compounds belong to more than five categories of toxicity - refer to Figure 1 for a plot of the number of compounds against the number of categories of toxicity. Thus, prediction of compound toxicity is a multi-label classification problem. Like the case of processing proteins or compounds with multiple attributes [15,16,22], the proposed method would provide a series of candidate toxicities, ranging from the most to the least likely, instead of presenting only the most likely one.

To sufficiently evaluate the prediction method described in the following section, we constructed 10 test groups, denoted by \( TG_1, TG_2, \ldots, TG_{10} \), respectively. In each test group \( TG_i \) \( 1 \leq i \leq 10 \), there is one training dataset \( S_{T}^{0} \) and one test dataset \( S_{T}^{0} \), i.e., \( TG_i = \langle S_{T}^{0}, S_{T}^{0} \rangle \), where the test dataset consisted of 1,723 compounds which were randomly selected from \( S \) while the training dataset contained the remaining 15,510 samples in \( S \), i.e., \( S = S_{T}^{0} \cup S_{T}^{0} \) for each \( 1 \leq i \leq 10 \). It is necessary to point out that, in each test group, the portion of the data in each class of the test dataset is roughly the same as that of the training dataset. Shown in Table 2 is the distribution of compounds in training and test datasets of each test group.

**Chemical-chemical Interactions**

It is known that two proteins that can interact with each other are more likely to share common biological functions than non-interactive ones [8,9,10,11]. Likewise, two interactive compounds are also more likely to share similar biological functions [15,16]. Since toxicity is one of a compound’s properties and functions, utilizing chemical-chemical interactions to identify compound toxicity is deemed to be feasible.

The data for chemical-chemical interactions were retrieved from STITCH (chemical_chemical.links.detailed.v3.0.tsv.gz, http://stitch.embl.de/cgi/show_download_page.). [17], a well-known database including known and predicted interactions of chemicals and proteins collected from experiments, literature or other reliable sources. In the obtained file, the interaction unit contains two compounds and five kinds of scores with titles “Similarity”, “Experimental”, “Database”, “Textmining” and “Combined_score”. The last kind of score was used here to indicate the interactivity of two compounds, i.e., two compounds with “Combined_score” greater than zero were deemed interactive compounds, because the last kind of score integrates the information of the other kinds of scores. Thus, the considered interactive compounds in this study contain the following three categories: (1) those participating in the same reactions; (2) those sharing similar structures or activities and (3) those with literature associations [17]. It is known that these categories correspond to the following three facts: (I) compounds involved in the same reactions occupy the same biological pathways; (II) compounds with similar structures or activities are likely to share similar functions, thereby occupying the same pathways with high probability; (III) the co-occurrence of two compounds, as noted in many studies, indicates some direct or indirect relationships, suggesting that they have the potential to share the same pathways. On the other hand, compounds in the same biological pathways always induce similar side effects, thereby having similar toxicity effects. Accordingly, it is reasonable to suppose that interactive compounds tend to have similar toxicity effects.

The value of the “Combined_score” of two interactive compounds indicates the likelihood that they can interact, i.e., two interactive compounds with high “Combined_score” can interact with high probability. Thus, this score is also termed a confidence score in this study. For two compounds \( c_1 \) and \( c_2 \), let us denote the confidence score of an interaction between them by \( Q(c_1,c_2) \). Specifically, if there is no interaction information between \( c_1 \) and \( c_2 \) based on the current records in STITCH, their interaction confidence score is assigned zero, i.e., \( Q(c_1,c_2) = 0 \). In this study, 323,432 interaction units, i.e., 323,432 pairs of compounds with confidence scores greater than 0, were used to predict compound toxicity. The detailed information on these interaction units can be found in Table S2.

**Prediction Method**

As is mentioned in the above section, interactive compounds are more likely to have common toxicity. Accordingly, the toxicities of a query compound can be identified according to its interactive compounds.

For convenience, let \( T_1, T_2, \ldots, T_7 \) denote the seven categories of toxicity, where \( T_1 \) denotes “Acute Toxicity”, \( T_2 \) “Mutagenicity”, and so forth (see column 1 and 2 of Table 1). Suppose that there are \( n \) compounds in the training dataset, that is \( c_1, c_2, \ldots, c_n \), the toxicity of a compound \( c_i \) in the training dataset is formulated as

\[
T(c_i) = [t_{i,1}, t_{i,2}, \ldots, t_{i,7}] (i = 1, 2, \ldots, n)
\]

where

\[
t_{i,j} = \begin{cases} 
1 & \text{if } c_i \text{ has toxicity } T_j \\
0 & \text{otherwise}
\end{cases}
\]

Given a query compound \( c_0 \), its toxicity is predicted not only by its interactive compounds but also by the confidence scores of their interactions. The score indicating that the query compound \( c_0 \) has toxicity \( T_j \) is calculated by

\[
\Theta(c_0 \rightarrow T_j) = \sum_{i=1}^{n} Q(c_i, c_0) \cdot t_{i,j} = 1, 2, 3, 4, 5, 6, 7
\]

The high score \( \Theta(c_0 \rightarrow T_j) \) means that there are many interactive compounds of \( c_0 \) in the training dataset that have toxicity \( T_j \) or some interactions between \( c_0 \) and its interactive compounds having toxicity \( T_j \) are labeled by high confidence...
scores. In view of this, the greater the score $\Theta(c_{q_j} \rightarrow T_j)$, the more likely that the compound $c_q$ has toxicity $T_j$. In particular, if $\Theta(c_{q_j} \rightarrow T_j)$ for some $j$, it is indicated that the probability that the query $c_q$ having the $j$-th category of toxicity is zero because there are no interactive compounds of $c_q$ in the training dataset that have toxicity $T_j$.

Since this is a multi-label classification problem, i.e., some compounds have more than one category of toxicity. A prediction method only providing the most likely toxicity is not an optimal choice. Thus, our method is valuable in that it can provide a series of candidate toxicities for a query compound, ranging from the most likely to the least likely. For example, if the results obtained from Eq. 3 are

$$\Theta(c_{q_j} \rightarrow T_3) \geq \Theta(c_{q_j} \rightarrow T_1) \geq \Theta(c_{q_j} \rightarrow T_6) > 0$$  \hspace{1cm} (4)$$

it can be interpreted to mean that there are three candidate toxicities for the query compound $c_q$ and the most likely toxicity for $c_q$ is $T_3$ ("Tumorigenicity", cf. Table 1), followed by $T_1$ ("Acute Toxicity") and $T_6$ ("Multiple Dose Effects"). In addition, $T_3$ is called the 1st order prediction, $T_1$ the 2nd order prediction, and so forth.

**Jackknife Test**

The Jackknife test [16] is often used to examine the performance of various predictors, because it can always provide a unique prediction result for a given dataset. It has been widely used by investigators to evaluate their predictors [23,24,25,26,27,28,29,30,31,32,33]. During the test, each sample in the training dataset is singled out one-by-one and tested by the predictor trained by the other samples. Thus, each sample is tested exactly once.

**Accuracy Measurement**

The $j$-th order prediction accuracy is calculated by the following formula [15,16]:

$$\Gamma_j = \frac{CT_j}{N}$$  \hspace{1cm} (5)$$

where $CT_j$ denotes the number of compounds whose $j$-th order prediction is one of its true toxicities, and $N$ denotes the total number of compounds in the dataset. If a prediction method can obtain high $\Gamma_j$ with small $j$ and low $\Gamma_j$ with large $j$, it implies that the method arranges the candidate toxicities well. Among them, the 1st order prediction accuracy is the most important indicator of good or bad performance.

Although the seven prediction accuracies can be obtained by Eq. 5, none of them provides the overall prediction accuracy. In view of this, we employ another measurement that calculates the proportion of true toxicities of the first $m$ predictions. It can be calculated as follows [16]:

$$\Delta_m = \frac{\sum_{i=1}^{N} S_{i,m}}{\sum_{i=1}^{N} N_i}$$  \hspace{1cm} (6)$$

where $S_{i,m}$ represents the number of the correct predictions of the $i$-th compound among its first $m$ predictions, and $N_i$ represents the number of toxicities of the $i$-th compound has. Since different compounds may have different numbers of toxicities, the parameter $m$ in Eq. 6 is usually taken as the smallest integer no less than the average number of toxicities in the dataset, which can
be computed by

\[ M = \frac{\sum_{i=1}^{N} N_i}{N} \]  

(7)

where \( m = \lceil M \rceil \). Obviously, a larger \( m \) implies better prediction performance by the method for the identification of compound toxicity.

### Results

As described in the Section “Benchmark dataset”, 10 test groups were constructed to evaluate the method described in Section “Prediction method”. In each test group, there were one training dataset consisting of 15,510 compounds and one test dataset containing 1,723 compounds. The predicted results for each test group obtained by the proposed method are as follows.

#### Performance of the Method on the Training Dataset

For the 15,510 compounds in each training dataset \( S_{\text{tr}}(1 \leq i \leq 10) \), we conducted the prediction and evaluated its performance by the Jackknife test. Listed in the column with title \( S_{\text{tr}} \) of Table 3 are seven prediction accuracies, calculated by Eq. 5, from which we can see that the 1st order prediction accuracies were all around 79.50%, where the maximum was 79.57%, while the minimum was 79.23%; the 2nd order ones were all around 37.30%. It is indicated that the proposed method is very stable. It is also observed from the corresponding columns of Table 3 that the accuracies followed a descending trend when increasing the order number, indicating that the method sorted the candidate toxicities quite well for the compounds in each training dataset \( S_{\text{tr}}(1 \leq i \leq 10) \). The average

### Table 3. Prediction accuracies obtained by the method as applied to training and test datasets of each test group.

<table>
<thead>
<tr>
<th>Prediction Order</th>
<th>( TG_1 )</th>
<th>( TG_2 )</th>
<th>( TG_3 )</th>
<th>( TG_4 )</th>
<th>( TG_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79.40%</td>
<td>79.69%</td>
<td>79.45%</td>
<td>79.28%</td>
<td>79.23%</td>
</tr>
<tr>
<td>2</td>
<td>37.16%</td>
<td>38.42%</td>
<td>37.14%</td>
<td>38.24%</td>
<td>37.54%</td>
</tr>
<tr>
<td>3</td>
<td>22.18%</td>
<td>23.16%</td>
<td>22.20%</td>
<td>22.87%</td>
<td>22.32%</td>
</tr>
<tr>
<td>4</td>
<td>15.45%</td>
<td>16.66%</td>
<td>15.49%</td>
<td>16.77%</td>
<td>16.35%</td>
</tr>
<tr>
<td>5</td>
<td>11.06%</td>
<td>11.61%</td>
<td>11.04%</td>
<td>11.49%</td>
<td>11.00%</td>
</tr>
<tr>
<td>6</td>
<td>6.92%</td>
<td>7.25%</td>
<td>6.84%</td>
<td>7.89%</td>
<td>7.23%</td>
</tr>
<tr>
<td>7</td>
<td>1.21%</td>
<td>1.33%</td>
<td>1.22%</td>
<td>1.04%</td>
<td>1.27%</td>
</tr>
</tbody>
</table>

#### Table 4. Proportions of true toxicities covered by the first two predictions for training and test datasets of each test group.

<table>
<thead>
<tr>
<th>Test group</th>
<th>Training dataset</th>
<th>Test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>( TG_1 )</td>
<td>65.52%</td>
<td>64.69%</td>
</tr>
<tr>
<td>( TG_2 )</td>
<td>65.54%</td>
<td>64.52%</td>
</tr>
<tr>
<td>( TG_3 )</td>
<td>65.43%</td>
<td>66.49%</td>
</tr>
<tr>
<td>( TG_4 )</td>
<td>65.32%</td>
<td>65.83%</td>
</tr>
<tr>
<td>( TG_5 )</td>
<td>65.48%</td>
<td>64.36%</td>
</tr>
<tr>
<td>( TG_6 )</td>
<td>65.46%</td>
<td>65.71%</td>
</tr>
<tr>
<td>( TG_7 )</td>
<td>65.55%</td>
<td>65.21%</td>
</tr>
<tr>
<td>( TG_8 )</td>
<td>65.43%</td>
<td>65.82%</td>
</tr>
<tr>
<td>( TG_9 )</td>
<td>65.61%</td>
<td>64.07%</td>
</tr>
<tr>
<td>( TG_{10} )</td>
<td>65.61%</td>
<td>64.79%</td>
</tr>
</tbody>
</table>

Figure 2. The structures of the alkyl N-nitroso group and the primary aromatic amine group.

doi:10.1371/journal.pone.0056517.t002

doi:10.1371/journal.pone.0056517.t003
numbers of toxicities for compounds in each training dataset $S_{ti}^{(0)}$ were about 1.78 according to Eq. 7, i.e., $M = 1.78$. It is noteworthy that if one predicts compound toxicity by random guesses, the average success rate would be only 25.43% (1.78/7), which is much lower than each of the 1st order prediction accuracies by our method. To evaluate the prediction accuracy by the method more thoroughly, Eq. 6 was calculated by taking $m = 2$, i.e., we considered the first two predictions for each compound in $S_{ti}^{(0)} (1 \leq i \leq 10)$ to see the proportions of true toxicities covered by these predictions. These proportions are shown in column 2 of Table 4, from which we can see that they were all about 65.50%, where the maximum was 65.61% while the minimum was 65.32%. Thus, it is indicated once again that our method is reliable.

Performance of the Method on the Test Dataset

For the 1,723 compounds in each test dataset $S_{ti}^{(0)} (1 \leq i \leq 10)$, the toxicities of these compounds were predicted by the proposed method described in Section “Prediction method” based on the numbers of toxicities for compounds in each training dataset $S_{ti}^{(0)}$. After processing by Eq. 5, seven prediction accuracies for each test dataset $S_{ti}^{(0)}$ were obtained and were listed in the column with title m in Table 4, from which we can see that they were all about 65.50%, where the maximum was 65.61% while the minimum was 65.32%. Thus, it is indicated once again that our method is reliable.

Understanding of the Toxicity Prediction Results

It is observed from Table 3 that the performance of the method on ten test groups is similar. Thus, the first test group (i.e., $TG_{1}$) is used as an example to show how to interpret the toxicity predicting results in detail.

Our multiclass model achieved a quite promising performance using the chemical-chemical interactions data on test group $TG_{1}$ (see Table 3 for details). For example, the compound 1-[N-methyl-N-nitrosamino]-1-[(3-pyridyl)-1-butanol (CID000047289, NNK) shows positive results for five toxicity endpoints: $T_{1}$, $T_{2}$, $T_{3}$, $T_{5}$, and $T_{6}$. Our model accurately predicted these five kinds of endpoints, and provided the order predictions as $T_{1} > T_{2} > T_{3} > T_{4} > T_{5} > T_{6}$. The 7th label representing ‘non-toxic’ was ranked as the last, suggesting that this compound is very likely to have toxic effects. As stated in the Section “Chemical-chemical interactions”, the interactive compounds derived from STITCH tend to have the same toxicity categories. 4-(Methylnitrosamino)-1-[(3-pyridyl)-1-butanol (CID000104856, NNAL), an interactive compound of NNK, has toxicities $T_{2}$ and $T_{3}$, which are also shared by NNK. The alkyl N-nitroso group (see Figure 2) of these two compounds associates with the formation of DNA adducts, and induces lung cancer in laboratory animals [34,35,36]. Another example is trimethoprim (CID000005578), which is positive for five toxicity endpoints: $T_{1}$, $T_{2}$, $T_{3}$, $T_{4}$, and $T_{6}$. The prediction order of our model was $T_{1} > T_{4} > T_{5} > T_{2} > T_{3} > T_{6}$. This compound was considered to be a carcinogen according to chemical-chemical interactions, but the Accelrys Toxicity database [19] labeled this compound only as a mutagen. However, it is reasonable to assume this compound as a carcinogen because it has a genotoxic toxicrophore-aromatic amine (see Figure 2) [5,37,38]. Typically, mutation is one of the first steps in the development of cancer [39].
Tasosartan (CID000060919) is an angiotensin II (AngII) receptor blocker [40], which is labeled as a relatively “non-toxic” compound in the dataset. Using our model, the order prediction of this compound was \(T_7 > T_6 > T_5 > T_2\). The 1st order prediction is “non-toxic”, consistent with the experimental data available. Among seven interactive compounds in the training dataset retrieved from STITCH (see Table 5), the top five interactive compounds are “non-toxic”, and their confidence scores are relatively high. However, the latter two interactive compounds are toxic, so tasosartan is predicted to have some toxicity effects in our model. However, the possibility of its possessing these toxicities is less than that of its not possessing toxicity (i.e., “non-toxic”).

The predictions for NNK, trimethoprim, and tasosartan and the prediction accuracies of the method indicate that interactive compounds can share common toxicity with high probability, which assessment conforms to the results of predicting other attributes of compounds [15,16]. The confidence scores of chemical-chemical interactions contribute significantly to the prediction of compound toxicity. As shown in Table 5, the interactive compounds of tasosartan with high confidence scores dominantly have the same toxicity as tasosartan. On the other hand, the predicted results for NNK, trimethoprim, and tasosartan reflect a limitation of our model: the judgment of “toxic” or “non-toxic” is based on a collective set of compounds with interactive information. However, some compounds with low confidence scores exist and they may contribute to the input of promiscuous interaction information to the final classification model. To address this issue, a future endeavor should introduce a threshold to the interaction confidence score and exclude “noisy” information to obtain a more accurate prediction.

Moreover, many more compounds are without chemical-chemical interactions in the original Accelrys Toxicity database. It is expected that the problem of predicting compound toxicity can be solved more favorably by the method as increasing amounts of chemical-chemical interaction information become available.

<table>
<thead>
<tr>
<th>Nongeneric SAs</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>thiocarbonyl</td>
<td>diphenylthiohydantoin (CID000854150) rubane acid (CID002777982)</td>
</tr>
<tr>
<td>halogenated benzene</td>
<td>4-chlorothiophenol (CID000007815) 4-bromophenol (CID000007808)</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>halogenated PAH</td>
<td>4-hydroxy-2',4',6'-trichlorobiphenyl (CID000105036) 2,2',5,5'-tetrachlorobenzidine (CID000027465)</td>
</tr>
<tr>
<td>halogenated dibenzodioxins</td>
<td>1,2,3,4,6,7,8-heptachlorobenzodioxin (CID000037270) 2,3,7,8-tetrabromobenzodioxin 4-dioxin (CID000039729)</td>
</tr>
</tbody>
</table>

Figure 3. Nongeneric SAs (Benigni) and some carcinogens matching these SAs. doi:10.1371/journal.pone.0056517.g003

Analysis of the Relationship between Different Chemical Toxicity Effects

In the Accelrys Toxicity Database, there are 3,607 compounds with more than two types of toxicity effects and 3,475 compounds with exact two effects (refer to Figure 1). We analyzed the number of common compounds belonging to two categories, and the ratio of the number of common compounds to the number of non-overlapping compounds of the two categories (see Table 6). It can be found that the intersection of \(T_5\) (“Reproductive Effects”), cf. Table 1 and \(T_6\) (“Multiple Dose Effects”) is the largest, sharing 26.6% of common compounds. The overlapping compounds suggest that there may be a causal relationship between the two categories. Specifically, the reproductive effects may cause multiple dose effects, i.e., reproductive toxicities may be cumulative, and hence be regarded as showing multiple dose effects in the meantime. The followed instances of correspondence between two categories are \(T_2\) (“Mutagenicity”) vs. \(T_3\) (“Tumorigenicity”) and \(T_1\) (“Acute Toxicity”) vs. \(T_6\) (“Multiple Dose Effects”). Since, in many cases, mutation is one of the first steps in the development of cancer [39], we took \(T_2\) (“Mutagenicity”) vs. \(T_3\) (“Tumorigenicity”) as an example to study the relationship between the two toxic categories.

From the viewpoint of mechanism of action, carcinogens can be classified into genotoxic or epigenetic carcinogens. Genotoxic carcinogens can bind covalently to DNA, and many known mutagens belong to this category. In the dataset, there are 1,720 common compounds with simultaneous toxicity \(T_2\) (“Mutagenicity”) and \(T_3\) (“Tumorigenicity”) and \(T_1\) (“Acute Toxicity”) vs. \(T_6\) (“Multiple Dose Effects”). The Structural alerts (SAs) provided by Benigni [37], which are molecular functional groups associated with a specific toxicity end point [38], were used here to gain insights into the correspondence of the two toxic effects. As summarized in Table S3, we illustrated a few examples for each of the matched SAs.

As previously mentioned, not all of the mutagens are carcinogens. For example, \(\omega_2\beta\)-unsaturated carbonyl compounds can interact with DNA by Michael addition, then lead to mutagenic and carcinogenic responses [37], e.g. acrylamide (CID000006579) and 2-butenal (CID000447466). However, if an
α,β-unsaturated carbonyl compound has conformational constraints or alkyl groups at the site of nucleophilic attack, the compound would be prone to reaction via Schiff base formation [41]. This change may only generate the DNA-adducts, but not carcinogens, not mutagens (see Figure 3). Thus, some compounds that can match nongeneric SAs [37] are only behave negatively in the standard mutagenicity assay [42]. Hence, our approach suggests that the multi-classification scheme is feasible and effective for in silico chemical toxicity prediction.

Supporting Information

Table S1 List of 17,233 compounds investigated in this study and their toxicological information.

(PDF)

Table S2 List of 323,432 interaction units used to predict compound toxicity in this study.

(PDF)

Table S3 List of SAs (Benigni) and examples matching SAs in our dataset.

(PDF)

Author Contributions

Conceived and designed the experiments: LC, JZ, MYZ, YDC. Performed the experiments: LC, JZ, KRF. Analyzed the data: JZ, MYZ. Contributed reagents/materials/analysis tools: LC, JZ, MYZ. Wrote the paper: LC, JZ, KRF.