Send Orders for Reprints to reprints@benthamscience.ae

RESEARCH ARTICLE

Title: Relating substructures and side effects of drugs with chemical-chemical interactions

Bo Zhou[a,*], Xian Zhao[b], Jing Lu[c], Zuntao Sun[d], Min Liu[b,e], Yilu Zhou[f], Rongzhi Liu[g], Yihua Wang[f],

[a]Shanghai University of Medicine & Health Sciences, Shanghai 201318, China; [b]College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China; [c]School of Pharmacy, Key Laboratory of Molecular Pharmacology and Drug Evaluation (Yantai University), Ministry of Education, Collaborative Innovation Center of Advanced Drug Delivery System and Biotech Drugs in Universities of Shandong, Yantai University, Yantai 264005, China; [d]Informatization Office, Shanghai Maritime University, Shanghai 201306, China; [e]Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China; [f]Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton SO17 1BJ, UK; United Kingdom; [g]Center for medical device evaluation, China Drug Administration, Sate Administration for Market Regulation, Beijing 100081, China.

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract: Background: Drugs are very important for human life because they can provide treatment, cure, prevention, or diagnosis of different diseases. However, they also bring side effects, which can give great risks for human bodies and pharmaceuticals companies. It is essential to identify drug side effects in drug discovery. To date, lots of computational methods have been proposed to predict the side effects of drugs and most of them used the fact that similar drugs always have similar side effects. However, previous studies did not analyze which substructures are highly related to which side effect.

Method: In this study, we did a computational investigation in this regard. To do that, we extracted a drug set for each side effect, which consisted of drugs having such effect side. Also, for each substructure, a set was constructed by picking up drugs owing such substructure. The relationship between one side effect and one substructure was evaluated based on linkages between drugs in their corresponding drug sets, resulting in an Es value. Then, the statistical significance of Es value was measured by a permutation test.

Results and Conclusion: Lots of highly related pairs of side effects and substructures were obtained and some were extensively analyzed to confirm the reliability of the results reported in this study.

Keywords: side effect, chemical substructure, chemical-chemical interaction, permutation test, statistical significance, tardive dyskinesia

1. INTRODUCTION

Drugs are always chemical substances that are discovered or designed for the treatment, cure, prevention, or diagnosis of different diseases. Clearly, drugs provide the guarantee for human health. However, they also produce unwanted effects, called side effects or adverse drug reactions (ADRs). According to the US Food and Drug Administration (FDA), several experimental drug compounds that have passed the clinical trials fail to gain FDA approvals because of their potential side effects, which may bring great risks for both human bodies and pharmaceuticals companies. Investments used in early stage become useless, which give financial burden for pharmaceuticals companies. Thus, it is urgent for drug developers to design effective methods for detecting all potential side effects of drugs and eliminate unacceptable drug candidates before costly human clinical trials.

The wet methods for identifying the side effects of a given drug candidate can provide solid results. However, these methods always need lots of time and are very expensive. In recent years, with the development of computer science and information techniques, it is an alternative way to determine drug side effects through designing computational methods. These methods can adopt different information of drugs and infer possible side effects that a given drug candidate may have. Although these methods cannot provide solid results, they can reduce the scope of side effects a drug may have, thereby speeding up the drug development and further reducing cost.

In recent years, several computational methods have been proposed to identify drug side effects. Pauwels et al.

*Address correspondence to this author at Shanghai University of Medicine & Health Sciences, Shanghai 201318, China; Tel/Fax: ++86-021-65882815, +86-021-65882570; E-mails: zhoub@sumhs.edu.cn

1386-2073 /19 $58.00+.00

employed sparse canonical correlation analysis on chemical structures and side effects of drugs to develop a novel method for predicting potential side-effects of drug candidate molecules [1]. Liu et al. proposed a machine-learning based method for ADR prediction using chemical, biological, and phenotypic properties of drugs [2]. Cheng et al. produced a phenotypic network inference model to identify drug side effects [3]. Yamanishi et al. developed a regression model, dealing with heterogeneous data sources, to identify potential side effects for drug candidate molecules [4]. Zhao et al. proposed a similarity-based method, incorporating several properties of drugs and random forest [5] as the classifier, to predict whether a given drug had a given side effect [6]. Chen et al. integrated the information of chemical-chemical interactions and protein-chemical interactions to construct a multi-label classifier for inferring which side effects a given drug may have [7]. Zhang et al. designed a feature selection based multi-label k-nearest neighbor (FS-MLKNN) method for ADR prediction [8]. Most previous methods adopted the classic property of drugs, that is, chemical structure and they believed that drugs with similar structures always share similar properties, including side effects. However, to our best knowledge, the associations between chemical substructures and side effects have not been reported. It is believed that some side effect may highly related to some substructures. If one can extract these relationships, it is helpful to infer the side effects of a give drug by checking its substructures.

In this study, we gave a computational investigation in this regard. First, we accessed the drug side effects from SIDER [9, 10] and constructed a drug set under each side effect. At the same time, the substructures of above drugs were extracted via RDKit [11] and a drug set was also constructed for each substructure. Then, the linkage of one side effect and one substructure was evaluated by considering all possible drug-drug pairs in two drug sets, resulting in a score, named Es value. Thereafter, a permutation test was performed to test the statistical significance of each Es value. Finally, we obtained lots of substructure-side effect pairs with statistical significance and selected some of them for detailed analyses.

2. MATERIALS AND METHOD

2.1. Side effects of drugs

The side effects of drugs were retrieved from SIDER (http://sideeffects.embl.de/) [9, 10], a public database reporting the information of marked medicines and ADRs, which is collected from public documents and package inserts. From the downloaded file, we obtained 888 drugs and 1,385 different side effects. However, several side effects were quite rare, that is, few drugs own these side effects. The purpose of this study is to investigate which side effect and chemical substructure have strong associations using drugs have each side effect and substructure. And the proposed method is based on statistical theory. The results on these side effects would be quite sensitive. To strengthen the reliability of our results, we have to discard these side effects. On the other hand, the chemical-chemical interaction information described in Section 2.3 of some drugs was not available, these drugs were also excluded. In summary, we refined the obtained data as follows: (1) drugs without chemical-chemical interaction information were excluded; (2) side effects with less than five drugs were also discarded. Finally, we obtained 828 drugs and 820 side effects.

For 820 side effects, let us denote them as $s_1, s_2, \ldots, s_{820}$. For each side effect s_i, the set consisting of drugs having such side effect was represented by SD_i. The distribution of sizes of all SD_i $(i = 1,2,\ldots,820)$ is illustrated in **Figure 1(A)**. It can be observed that majority side effects have less than 100 drugs and some side effects have more than 500 drugs.

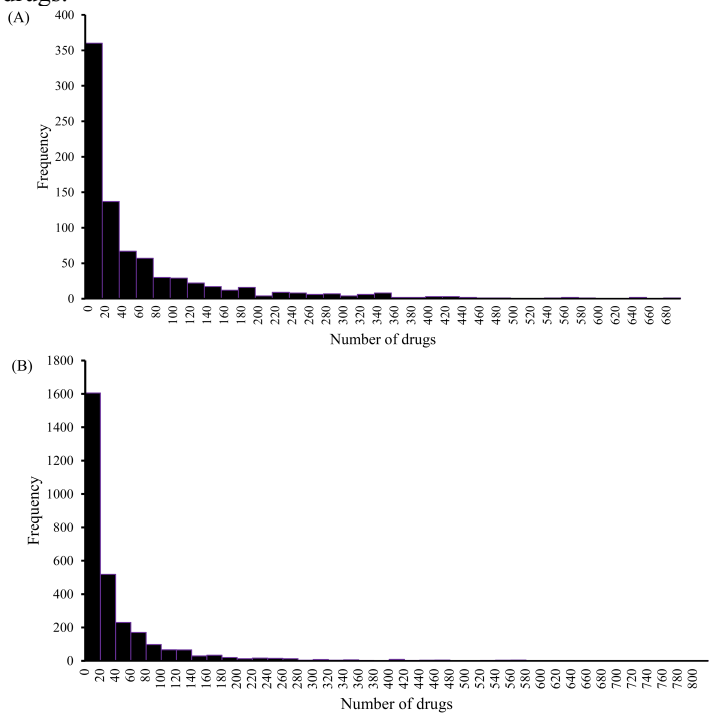

Figure 1. Bar charts to show the distribution of the numbers of drugs side effects or substructures have. (A) A bar chart for side effect; (B) A bar chart for substructure.

2.2. Substructures of drugs

As mentioned in Section 2.1, we obtained 828 drugs and their side effects. For accessing the substructures of these 828 drugs, the RDKit [11], a widely used source chemistry informatics and machine learning toolkit, was adopted to generate their Morgan fingerprints [12], resulting in 6,441 substructures. Similar to the side effect, some sub-structures own very few drugs and thus should be excluded. With the same manner, we excluded substructures with less than five drugs, obtaining 2,966 substructures. For latter formulation, we denoted obtained 2,966 substructures as $ss_1, ss_2, \ldots, ss_{2966}$ and the sets containing corresponding drugs by $SSD_1, SSD_2, \ldots, SSD_{2966}$. Also, **Figure 1(B)** shows the distribution of sizes of SSD_i $(i = 1,2,\ldots,2966)$, from which we can see that the distribution of the sizes of SSD_i $(i = 1,2,\ldots,2966)$ is quite similar to that of SD_i $(i = 1,2,\ldots,820)$.

2.3. Chemical-chemical interactions

For each substructure or side effect, several drugs own such substructure or side effect. Thus, the relationship

between one substructure and one side effect can be measured by their corresponding drug sets. Here, we designed a scheme to evaluate the association between two drug sets via chemical-chemical interaction information.

The chemical-chemical interaction information was retrieved from STITCH (http://stitch.embl.de/, version 4.0) [13, 14], a well-known online database exploring known and predicted interactions of chemicals and proteins. The chemical-chemical interactions are collected in a file, named "chemical_chemical.links.v4.0.tsv.gz", in which each interaction contains two chemicals, represented by PubChem IDs, and one score, denoted by "combined_score" in the obtained file, ranging from 1 to 999. These interactions were derived from experiments, databases and the literature, and the score integrates the associations between two chemicals from several aspects, such as structures, reactions, activities and text descriptions. Thus, they can widely measure the associations between two chemicals. To date, several studies have used this information to investigate different chemical or drug-related problems [6, 7, 15-29]. For formulation, let us denote the score between two chemicals c_1 and c_2 by $S(c_1,c_2)$. Specially, if c_1 and c_2 were identical, $S(c_1,c_2)$ was set to 1000; and if chemicals c_1 and c_2 cannot interact with each other, $S(c_1,c_2)$ was set to be zero.

2.4. Es values of substructure-side effect pairs

For side effect s_i, drugs having such side effect were collected in set SD_i and SSD_j contained the drugs owning substructure ss_j. The associations between s_i and ss_j can be measured by drugs in SD_i and drugs in SSD_j as follows:

$$Es(s_i,ss_j)=\sum_{d_x\in SD_i}\sum_{d_y\in SSD_j}S(d_x,d_y) \qquad (1)$$

Clearly, such score of s_i and ss_j was the sum of all possible drug pairs in SD_i and SSD_j. We named this score as Es value. A high Es value suggests a potential strong association between the corresponding side effect and substructure.

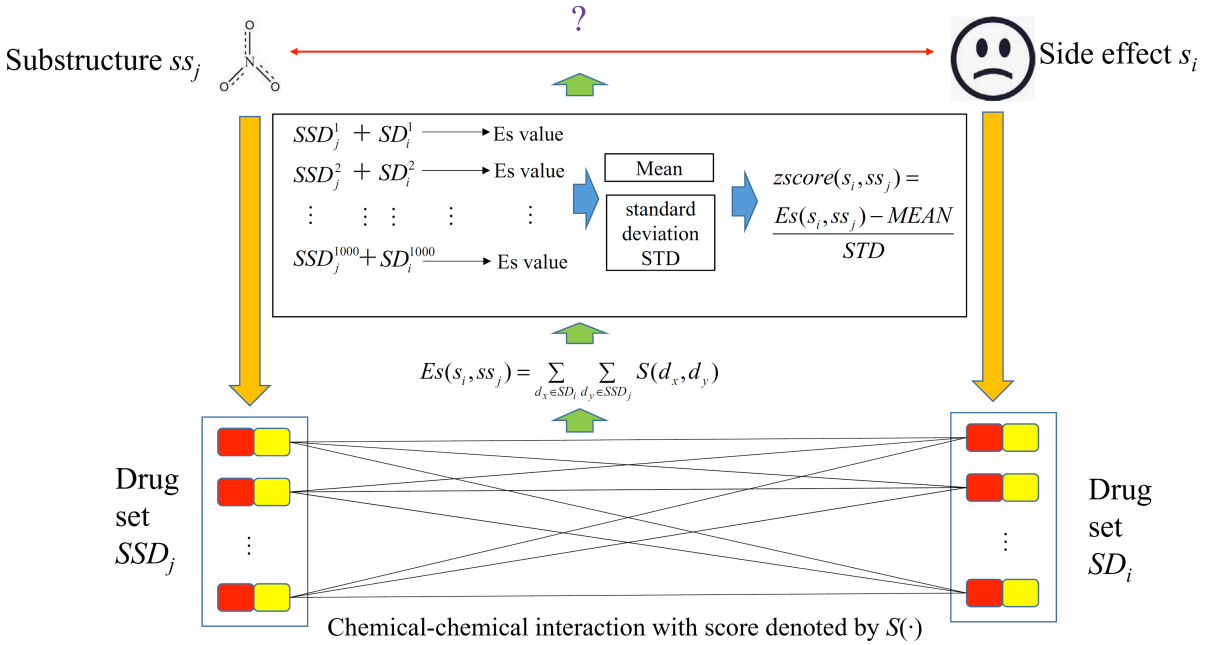

Figure 2. A flow chart to illustrate the procedures of evaluating the associations between one side effect s_i and one substructure ss_i. First, drugs having the side effect s_i comprised a set SD_i and drugs owing the substructure ss_i constituted another drug set SSD_j. Second, calculate Es value by checking the chemical-chemical interactions between any drug in SD_i and any drug in SSD_j. Third, 1000 drug sets $SD_i^1, SD_i^2, \ldots, SD_i^{1000}$ with the same size of SD_i were randomly constructed and 1000 sets $SSD_j^1, SSD_j^2, \ldots, SSD_j^{1000}$ with the same size of SSD_j were also built. These sets were adopted to calculate 1000 Es values. Finally, calculate the zscore according to the true Es value and 1000 Es values on randomly produced sets, thereby determining the relationship between the side effect and substructure.

2.5. Zscores of substructure-side effect pairs

In Section 2.4, we evaluated the associations between one side effect and one substructure via an Es value. However, this score highly relied on the sizes of SD_i and SSD_j and it was difficult to determine how high of this score was significant. Thus, a permutation test was performed to measure the statistical significance of each Es value. In detail, we constructed 1,000 drug sets with the same size of SD_i, denoted by $SD_i^1, SD_i^2, \ldots, SD_i^{1000}$, which were generated by randomly selecting drugs from 828 drugs. For SSD_j, we also constructed 1,000 drug sets in a similar way, denoted by $SSD_j^1, SSD_j^2, \ldots, SSD_j^{1000}$, each of which

had the same size of SSD_j. Then, we computed 1,000 Es values between SD_i^a and SSD_j^a $(a = 1,2,\ldots,1000)$ using **Eq. 1**, and further calculate the zscore of $Es(s_i, ss_j)$ by

$$zscore(s_i, ss_j) = \frac{Es(s_i, ss_j) - MEAN}{STD}, \qquad (2)$$

where MEAN represented the average of 1,000 Es values yielded by randomly produced drug sets and STD represented the standard deviation of these Es values. By statistical theory, if the zscore is larger than 1.96, the side effect and substructure would have a statistically significant association.

Table 1. Top ten substructure-side effect pairs with highest zscores

Substructure (Morgan fingerprints)	Side effects	Zscore
c(c[cH])(c[cH])N(c)[CH2]	tardive dyskinesia	40.722
N(C[CH2])(c(c)[cH])c(c)[cH]	tardive dyskinesia	39.033
N(c)(c)[CH2]	tardive dyskinesia	38.845
c(cc)(c[cH])N(c)[CH2]	tardive dyskinesia	38.743
c(c[cH])(c[cH])N(c)[CH2]	testicular swelling	37.083
N(c)(c)C	tardive dyskinesia	36.870
c(c[cH])(Sc)c([cH])N	tardive dyskinesia	36.423
N(c(c)[cH])c(c)[cH]	tardive dyskinesia	35.683
N(c)(c)C	testicular swelling	35.518
c(cc)(c[cH])N(c)[CH2]	testicular swelling	34.526

Table 2. Ten side effects with most highly related substructures and their most related substructures

Side effect	Number of highly related substructures (Morgan fingerprints)	Most related substructures (Morgan fingerprints)	Zscore
arterial insufficiency	1138	c(c[cH])(O[CH2])c(C)[cH]	12.145
heart block	997	c(cc)(c[cH])N(c)[CH2]	18.155
hyperprolactinemia	980	c(c(c)N)c([cH])Cl	20.418
galactorrhea	940	c(cc)(c[cH])N(c)[CH2]	33.505
tardive dyskinesia	919	c(c[cH])(c[cH])N(c)[CH2]	40.722
Parkinson	919	N(c)(c)C	32.450
Diabetic ketoacidosis	915	c(c(c)N)c([cH])Cl	28.085
testicular swelling	910	c(c[cH])(c[cH])N(c)[CH2]	37.083
dysarthria	882	c(c[cH])(c[cH])C(c)=N	18.867
nightmares	877	N(c)(c)[CH2]	18.455

Table 3. Ten substructures with most highly related side effects and their most related side effects

Substructure (Morgan fingerprints)	Number of highly related side effects	Most related side effect	zscore
c(cc)(Nc)c([cH])C	664	testicular swelling	25.908
c(c[cH])(Nc)c(c)C	641	testicular swelling	26.987
c(c[cH])(Nc)c(C)[cH]	640	testicular swelling	25.392
c(c[cH])(c(C)[cH])N(c)C	634	testicular swelling	34.489
c(c[cH])(Nc)c([cH])[CH2]	630	testicular swelling	30.905
N(c)c	627	tardive dyskinesia	29.509
c(cc)(c[cH])Nc	627	tardive dyskinesia	29.965
c(c(c)N)c([cH])Cl	620	tardive dyskinesia	34.306
N(c)(c)C	597	tardive dyskinesia	36.870
c(c[cH])(c[cH])C(c)=N	596	somnambulism	20.178

3. RESULTS

In this study, we gave a computational investigation on associations between side effects and substructures. The purpose of this study was to extract highly related substructure-side effect pair. The whole procedures are shown in **Figure 2**. This section gave a detailed description on the results.

As mentioned in Sections 2.1 and 2.2, 820 side effects and 2,966 different substructures were investigated in this study, meaning there were 2,432,120 (820 × 2966) substructure-side effect pairs. For each pair, we calculated its Es value according to Eq. (1). However, Es value was highly related to the sizes of drug sets of the substructure and side effect. Then, a permutation test was performed to evaluate whether each Es value was statistical significance, resulting in a zscore for each substructure-side effect pair. The distribution of obtained 2,432,120 zscores is illustrated in a bar chart, as shown in **Figure 3**. It can be observed that most zscores gathered between -5 and 5. By setting the confidence level as 95%, we selected the substructure-side effect pairs with zscores larger than 1.96, obtaining 193,298 (193,298/2,432,120=7.948%) pairs. In statistics, the Es values of these pairs were statistically significant high, meaning corresponding substructure and side effect were highly related. These 193,298 substructure-side effect pairs together with their zscores are provided in Supplementary Material I. **Table 1** lists the top ten substructure-side effect pairs with highest zscores. It can be observed that side effects "tardive dyskinesia" and "testicular swelling" are highly related to some substructures with high zscores.

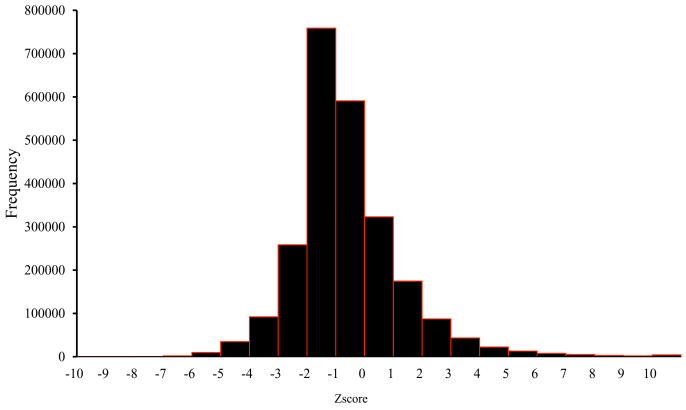

Figure 3. A bar chart to illustrate the distribution of zscores for all substructure-side effect pairs.

For 193,298 highly related substructure-side effect pairs, 792 side effects were identified to have at least one highly related substructures and 1,875 substructures were predicted to be highly related to at least one side effects. The number of highly related substructures/side effects for each side effect/substructure is provided in Supplementary Material II/III. **Figure 4** shows the number of highly related substructures/side effects for each of these side effects/substructures, from which we can see that most side effects had less than 600 highly related substructures and most substructures own less than 400 highly related side effects. We list the top ten side effects with most highly related substructures in **Table 2**, indicating that the side effect "arterial insufficiency" had most related substructures, followed by "heart block", "hyperprolactinemia" and so forth. While in **Table 3**, top ten substructures with most highly related side effects are listed. Substructure "c(cc)(Nc)c([cH])C" has most related side effects, followed by "c(c[cH])(Nc)c(c)C", "c(c[cH])(Nc)c(C)[cH]" and so forth.

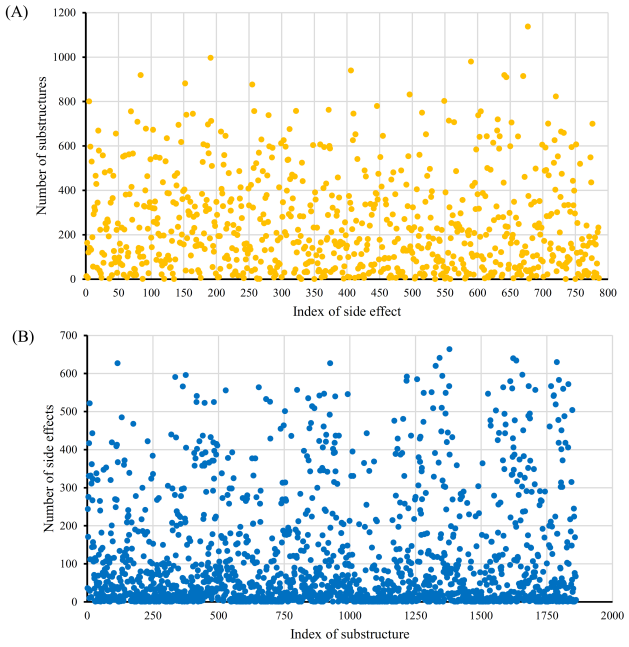

Figure 4. Scatter diagrams to show the number of highly related substructures/side effects for each side effect/substructure. (A) The scatter diagram for side effect; (B) The scatter diagram for substructure.

4. DISCUSSION

In this study, we extracted lots of highly related substructure-side effect pairs (Supplementary Material I) via a computational analysis. Very frequent side effects that often occur in many drugs, such as "headache" or "nausea", hardly appear in **Tables 1-3**, which is consistent with the fact that they are common reactions. However, we found more specific side effects which are related to special types of substructures. Here, we analyzed some of them to confirm that our results were reliable. For rest pairs, we only listed them in Supplementary Material I. readers can give further investigations based on these materials.

Table 4. Distribution of the top ten related substructures for "tardive dyskinesia" in six drugs

Rank	Substructure (Morgan fingerprints)	Zscore	Typical APDs		Atypical APDs	Antidepressants	Antiemetics	
			Chlorpromazine	Haloperidol	Clozapine	Clomipramine	Metoclopramide	Prochlorperazine
1	c(c[cH])(c[cH])N(c)[CH2]	40.722	+	-	-	+	-	+
2	N(C[CH2])(c(c)[cH])c(c)[cH]	39.033	+	-	-	+	-	+
3	N(c)(c)[CH2]	38.845	+	-	-	+	-	+
4	c(cc)(c[cH])N(c)[CH2]	38.743	+	-	-	+	-	+
5	N(c)(c)C	36.870	+	-	-	+	-	+
6	c(c[cH])(Sc)c([cH])N	36.423	+	-	-	-	-	+
7	N(c(c)[cH])c(c)[cH]	35.683	+	-	-	+	-	+
8	c(c(c)N)c([cH])Cl	34.306	+	-	+	+	-	+
9	c(c[cH])(c([cH])S)N(c)[CH2]	33.948	+	-	-	-	-	+
10	S(c(c)[cH])c(c)[cH]	33.878	+	-	-	-	-	+

+/-: these symbols indicate whether the substructure is contained in the corresponding drug.

As listed in **Table 1**, "tardive dyskinesia" (TD) was highly related to some substructures with quite high zscores. It is a disorder that results in involuntary, repetitive body movements, and most generally occur in patients treated with long-term neuroleptic drugs. Neuroleptic drugs, also known as antipsychotic drugs (APDs), are used to treat and manage symptoms of many psychiatric disorders. Sometimes they are prescribed for gastrointestinal (GI) disorders. APDs can be divided into two classes: "typical" antipsychotics or first-generation and "atypical" antipsychotics or second-generation. In our results, 919 substructures were identified to be highly related to this side effects (see **Table 2**). According to our results, ten substructures, listed in **Table 4**, were related to this side effect with highest zscores. Six drugs (chlorpromazine, haloperidol, clozapine, clomipramine, metoclopramide and prochlorperazine) were selected to detect the distribution of these substructures.

Chlorpromazine and haloperidol are the best known typical APDs. Typical APDs do have a high risk of TD. The exact mechanism by which typical APDs result in the development of TD remains largely uncertain. However, the most compelling line of evidence suggests that chronic dopamine blockade caused by dopamine D2 receptor antagonists or APDs could result in an upregulation of dopamine receptor responsiveness, resulting in a compensatory supersensitivity of the receptors in the nigrostriatal pathway [30, 31]. As shown in **Table 4**, chlorpromazine contains all of top ten substructures. It is also reflected in other phenothiazines (thioridazine, fluphenazine, mesoridazine, trifluoperazine, perphenazine, pipotiazine), and they contain almost all of top ten substructures. These substructures in phenothiazines may contribute to greater affinity for the D2 binding site, which is associated with high risk for TD. As the first synthetic butyrylbenzene drug, haloperidol, its pharmacological action and TD incidence are similar to those of phenothiazine drugs. However, its chemical structure is completely different from that of phenothiazine, and therefore does not contain any substructures listed in **Table 4**. Evidence suggests that haloperidol has been shown to induce TD through the influx of proinflammatory cytokines and neurotransmitters [32], unlike phenothiazines that cause TD by blocking D2 receptors. From the viewpoint of all substructures of haloperidol, 92% (174/189) of the substructures were significantly associated with TD. Its 174 substructures associated with TD may be related to the TD induction mechanism.

Atypical APDs bind weakly to dopamine D2 receptors and seem to be associated with a decreased prevalence of TD. As the first atypical APD, clozapine is considered to have significantly less TD compared to typical APDs [33], even to other atypical APDs [34]. For the top ten substructures in **Table 4**, clozapine contains only one substructure significantly associated with TD. The results indicate that clozapine contains less substructures closely related to TD than phenothiazine. On the other hand, this also provides a possible explanation for the lower risk of TD in clozapine from substructure analysis. Unfortunately, the advantage of clozapine in reducing TD is only a special case in atypical APDs. Later well-designed trials failed to show the anticipated decrease in TD with other atypical APDs [35-37]. Comparing the distributions of 919 TD-related substructures in typical APDs and atypical APDs, the results show typical APDs have more substructures with higher zscores than atypical APDs (see **Figure 5**). Whether true or not, it is clear that TD continues to be a major problem associated with the long term use of almost all APDs.

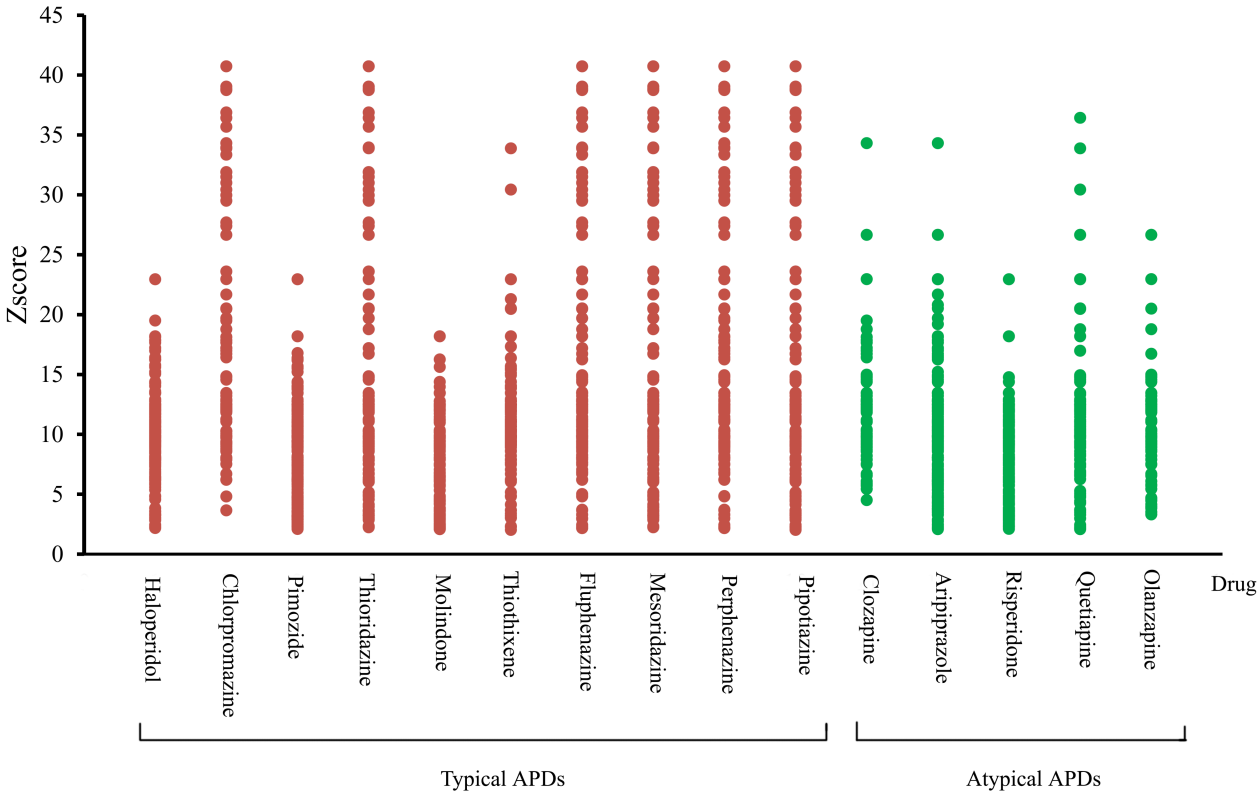

Figure 5. The distribution of zscores on the 919 TD-related substructures of typical/atypical APDs and TD. TD represents the side effect "tardive dyskinesia".

Clomipramine is a tricyclic antidepressant that acts by reducing the re-uptake of norepinephrine and serotonin in the central nervous system. TD induced by antidepressants is less prevalent than TD induced by APDs. Similarly, there are only a few reports showing cases of TD caused by clomipramine [38]. However, as shown in **Table 4**, clomipramine contains some substructures that are closely related to TD, suggesting that monitoring of TD should be strengthened in clinical long-term use.

Metoclopramide, a dopamine receptor antagonist, is, unsurprisingly, a well-documented cause of TD [39-41]. However, due to its being used for shorter periods of time than psychiatric medications, risk of TD seems to be lower than previously thought [42]. From the results of the substructure analysis in **Table 4**, compared with metoclopramide, prochlorperazine, another antiemetic drug, contains more substructures significantly associated with TD. Prochlorperazine may yield a higher frequency of TD than metoclopramide, as evidenced by some clinical trials [43].

Drug-induced TD is a complex and unique neurologic disorder. By substructure analysis, some substructures closely related to TD can be identified, which is helpful for the detection of TD occurrence of existing APDs, and also has predictive significance for new drugs.

CONCLUSION

This contribution gave a computational investigation on the relationships between side effects and chemical substructures. The association between one side effect and one substructure was evaluated by the linkage between drugs in their corresponding drug sets. The statistical significance of each substructure-side effect pair was further measured by a permutation test. Lots of highly related substructure-side effect pairs were accessed. And we in detail analyzed the relationship between one side effect "tardive dyskinesia" and its top ten related substructures, partly proving the reliability of our results. It is hopeful that results reported in this study can give new insights for identification of drug side effects.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by Natural Science Foundation of Shanghai (17ZR1412500) and Science and Technology

Commission of Shanghai Municipality (STCSM) (18dz2271000).

SUPPORTIVE/SUPPLEMENTARY MATERIAL

Supplementary Material I: Highly related substructures and side effects; Supplementary Material II: The Number of highly related substructures for each side effect; Supplementary Material III. The number of highly related side effects for each substructure.

REFERENCES

[1] Pauwels, E.; Stoven, V.; Yamanishi, Y. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinformatics* **2011,** *12*, 169.

[2] Liu, M.; Wu, Y.; Chen, Y.; Sun, J.; Zhao, Z.; Chen, X. W.; Matheny, M. E.; Xu, H. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* **2012,** *19* (e1), e28-35.

[3] Cheng, F.; Li, W.; Wang, X.; Zhou, Y.; Wu, Z.; Shen, J.; Tang, Y. Adverse drug events: database construction and in silico prediction. *J Chem Inf Model* **2013,** *53* (4), 744-752.

[4] Yamanishi, Y.; Pauwels, E.; Kotera, M. Drug side-effect prediction based on the integration of chemical and biological spaces. *J Chem Inf Model* **2012,** *52* (12), 3284-3292.

[5] Breiman, L. Random forests. *Mach Learn* **2001,** *45* (1), 5-32.

[6] Zhao, X.; Chen, L.; Lu, J. A similarity-based method for prediction of drug side effects with heterogeneous information. *Math Biosci* **2018,** *306*, 136-144.

[7] Chen, L.; Huang, T.; Zhang, J.; Zheng, M. Y.; Feng, K. Y.; Cai, Y. D.; Chou, K. C. Predicting drugs side effects based on chemical-chemical interactions and protein-chemical interactions. *Biomed Res Int* **2013,** *2013*, 485034.

[8] Zhang, W.; Liu, F.; Luo, L.; Zhang, J. Predicting drug side effects by multi-label learning and ensemble learning. *BMC Bioinformatics* **2015,** *16* (1), 365.

[9] Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L. J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* **2010,** *6*, 343.

[10] Kuhn, M.; Letunic, I.; Jensen, L. J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res* **2016,** *44* (D1), D1075-1079.

[11] RDKit: Open-source cheminformatics, http://www.rdkit.org.

[12] Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J Chem Inf Model* **2010,** *50* (5), 742-754.

[13] Kuhn, M.; Von, M. C.; Campillos, M.; Jensen, L. J.; Bork, P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Research* **2008,** *36* (Database issue), 684-688.

[14] Kuhn, M.; Szklarczyk, D.; Pletscherfrankild, S.; Blicher, T. H.; Von, C. M.; Jensen, L. J.; Bork, P. STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Res* **2014,** *42* (Database issue), 401-407.

[15] Chen, L.; Liu, T.; Zhao, X. Inferring anatomical therapeutic chemical (ATC) class of drugs using shortest path and random walk with restart algorithms. *BBA - Mol Basis Dis* **2018,** *1864* (6, Part B), 2228-2240.

[16] Liu, T.; Chen, L.; Pan, X. An integrated multi-label classifier with chemical-chemical interactions for prediction of chemical toxicity effects. *Comb Chem High T Scr* **2018,** *21*(6), 403-410.

[17] Chen, L.; Zeng, W. M.; Cai, Y.D.; Feng, K.Y.; Chou, K. C. Predicting anatomical therapeutic chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. *PLoS ONE* **2012,** *7* (4), e35254.

[18] Cheng, X.; Zhao, S.-G.; Xiao, X.; Chou, K.C. iATC-mISF: a multi-label classifier for predicting the classes of anatomical therapeutic chemicals. *Bioinformatics* **2016,** *33* (3), 341-346.

[19] Hu, L.L.; Chen, C.; Huang, T.; Cai, Y.D.; Chou, K.C. Predicting Biological Functions of Compounds Based on Chemical-Chemical Interactions. *PLoS ONE* **2011,** *6* (12), e29491.

[20] Chen, L.; Lu, J.; Zhang, N.; Huang, T.; Cai, Y.D. A hybrid method for prediction and repositioning of drug Anatomical Therapeutic Chemical classes. *Mol BioSyst* **2014,** *10* (4), 868-877.

[21] Gao, YF.; Chen, L.; Cai, Y. D.; Feng, K. Y.; Huang, T.; Jiang, Y. Predicting Metabolic Pathways of Small Molecules and Enzymes Based on Interaction Information of Chemicals and Proteins. *PLoS ONE* **2012,** *7* (9), e45944.

[22] Chen, L.; Chu, C.; Lu, J.; Kong, X.; Huang, T.; Cai, Y.D. A computational method for the identification of new candidate carcinogenic and non-carcinogenic chemicals. *Mol BioSyst* **2015,** *11* (9), 2541-2550.

[23] Huang, G.; Lu, Y.; Lu, C.; Zheng, M.; Cai, Y.D. Prediction of Drug Indications Based on Chemical Interactions and Chemical Similarities. *BioMed Res Int* **2015,** *2015*, 584546.

[24] Chen, L.; Lu, J.; Huang, T.; Yin, J.; Wei, L.; Cai, Y.D. Finding Candidate Drugs for Hepatitis C Based on Chemical-Chemical and Chemical-Protein Interactions. *PLoS ONE* **2014,** *9* (9), e107767.

[25] Lu, J.; Huang, G.; Li, H.P.; Feng, K.Y.; Chen, L.; Zheng, M.-Y.; Cai, Y.D. Prediction of Cancer Drugs by Chemical-Chemical Interactions. *PLoS ONE* **2014,** *9* (2), e87791.

[26] Liu, L.; Chen, L.; Zhang, Y.H.; Wei, L.; Cheng, S.; Kong, X.-Y.; Zheng, M.; Huang, T.; Cai, Y.D. Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection. *J Biomol Struct Dyn* **2017,** *35* (2), 312-329.

[27] Guo, Z.H.; Chen, L.; Zhao, X. A network integration method for deciphering the types of metabolic pathway of chemicals with heterogeneous information. *Comb Chem High T Scr* **2018**, 21(9), 670-680.

[28] Zhao, X.; Chen, L.; Guo, Z.-H.; Liu, T., Predicting drug side effects with compact integration of heterogeneous networks. *Curr Bioinform* **2019**. doi: 10.2174/1574893614666190220114644

[29] Wang, T.; Chen, L.; Zhao, X., Prediction of drug combinations with a network embedding method. *Comb Chem High T Scr* **2018,** *21* (10), 789-797.

[30] Waln, O.; Jankovic, J. An update on tardive dyskinesia: from phenomenology to treatment. *Tremor Other Hyperkinet Mov (N Y)* **2013,** *3*.

[31] Lockwood, J.T.; Remington, G. Emerging drugs for antipsychotic-induced tardive dyskinesia: investigational drugs in Phase II and Phase III clinical trials. *Expert Opin Emerg Drugs* **2015,** *20* (3), 407-421.

[32] Galili, R.; Mosberg; Gil-Ad, I.; Weizman, A.; Melamed, E.; Offen, D. Haloperidol-induced neurotoxicity--possible implications for tardive dyskinesia. *J Neural Transm (Vienna)* **2000,** *107* (4), 479-490.

[33] Correll, C.U.; Schenk, E.M., Tardive dyskinesia and new antipsychotics. *Curr Opin Psychiatry* **2008,** *21* (2), 151-156.

[34] Kane, J.M.; Woerner, M.G.; Pollack, S.; Safferman, A. Z.; Lieberman, J. A. Does clozapine cause tardive dyskinesia? *J Clin Psychiatry* **1993,** *54* (9), 327-330.

[35] Woods, S.W.; Morgenstern, H.; Saksa, J.R.; Walsh, B.C.; Sullivan, M.C.; Money, R.; Hawkins, K.A.; Gueorguieva, R.V.; Glazer, W.M. Incidence of tardive dyskinesia with atypical versus conventional antipsychotic medications: a prospective cohort study. *J Clin Psychiatry* **2010,** *71* (4), 463-474.

[36] Peluso, M.J.; Lewis, S.W.; Barnes, T.R.; Jones, P.B. Extrapyramidal motor side-effects of first- and second-generation antipsychotic drugs. *Br J Psychiatry* **2012,** *200* (5), 387-392.

[37] Miller, D.D.; Caroff, S.N.; Davis, S.M.; Rosenheck, R.A.; McEvoy, J.P.; Saltz, B.L.; Riggio, S.; Chakos, M.H.; Swartz, M.S.; Keefe, R.S.; Stroup, T.S.; Lieberman, J.A. Clinical Antipsychotic Trials of Intervention Effectiveness, I., Extrapyramidal side-effects of antipsychotics in a randomised trial. *Br J Psychiatry* **2008,** *193* (4), 279-288.

[38] Clayton, A.H. Antidepressant-induced tardive dyskinesia: review and case report. *Psychopharmacol Bull* **1995,** *31* (2), 259-264.

[39] Ganzini, L.; Casey, D. E.; Hoffman, W.F.; McCall, A. L. The prevalence of metoclopramide-induced tardive dyskinesia and acute extrapyramidal movement disorders. *Arch Intern Med* **1993,** *153* (12), 1469-1475.

[40] Kenney, C.; Hunter, C.; Davidson, A.; Jankovic, J. Metoclopramide, an increasingly recognized cause of tardive dyskinesia. *J Clin Pharmacol* **2008,** *48* (3), 379-384.

[41] Lau Moon Lin, M.; Robinson, P.D.; Flank, J.; Sung, L.; Dupuis, L.L. The Safety of Metoclopramide in Children: A Systematic Review and Meta-Analysis. *Drug Saf* **2016,** *39* (7), 675-687.

[42] Rao, A.S.; Camilleri, M. Review article: metoclopramide and tardive dyskinesia. *Aliment Pharmacol Ther* **2010,** *31* (1), 11-19.

[43] Wijemanne, S.; Jankovic, J.; Evans, R. W. Movement Disorders From the Use of Metoclopramide and Other Antiemetics in the Treatment of Migraine. *Headache* **2016,** *56* (1), 153-161.

1386-2073 /19 $58.00+.00