Article

In-Silico Drug Toxicity and Interaction Prediction for Plant Complexes Based on Virtual Screening and Text Mining

Feng Zhang [1,2], Kumar Ganesan [1,2], Yan Li [1,2] and Jianping Chen [1,2,*]

1   School of Chinese Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 10 Sassoon Road, Pokfulam, Hong Kong, China
2   Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen 518000, China
*   Correspondence: abchen@hku.hk; Tel.: +852-3917-6479

Abstract: Potential drug toxicities and drug interactions of redundant compounds of plant complexes may cause unexpected clinical responses or even severe adverse events. On the other hand, super-additivity of drug interactions between natural products and synthetic drugs may be utilized to gain better performance in disease management. Although without enough datasets for prediction model training, based on the SwissSimilarity and PubChem platforms, for the first time, a feasible workflow of prediction of both toxicity and drug interaction of plant complexes was built in this study. The optimal similarity score threshold for toxicity prediction of this system is 0.6171, based on an analysis of 20 different herbal medicines. From the PubChem database, 31 different sections of toxicity information such as "Acute Effects", "NIOSH Toxicity Data", "Interactions", "Hepatotoxicity", "Carcinogenicity", "Symptoms", and "Human Toxicity Values" sections have been retrieved, with dozens of active compounds predicted to exert potential toxicities. In *Spatholobus suberectus* Dunn (SSD), there are 9 out of 24 active compounds predicted to play synergistic effects on cancer management with various drugs or factors. The synergism between SSD, luteolin and docetaxel in the management of triple-negative breast cancer was proved by the combination index assay, synergy score detection assay, and xenograft model.

Keywords: herbal bioinformatics; in-silico toxicity prediction; drug-drug interaction; ligand-based virtual screening; synergism; triple-negative breast cancer

1. Introduction

For the treatment of some advanced cancer such as triple-negative breast cancer (TNBC), there are still seldom medications, but chemotherapeutic drugs can achieve moderate effects on patient overall survival according to the latest therapy guidelines and clinical trials [1–3]. More than 20 potential severe adverse events can be incurred during chemotherapy [4]. So, it's important to discover strategies to reduce the negative effects of chemotherapy to improve patients' quality of life. The synergistic effects of natural substances combined with chemotherapy medications may shed light on this [5–7]. Moreover, the paradigm shift from a "one-target, one-drug" mode to a "network-target, multiple-component-therapeutics" mode like network pharmacology will offer much more potential for cancer management. Nonetheless, conventional network pharmacology is based on simple additivity of the potencies and efficacies of individual active compounds of herbal medicines [8,9]. There are potentially super-additive (synergistic) effects and subadditivity (antagonism) in terms of drug combination [10], which may cause unexpected pharmacologic or clinical responses. Plus, conventional network pharmacology is lacking in toxicity analysis for plant complexes. Pharmacological and toxicological assessments for compound combinations by experiments would be exceedingly time- and cost-intensive. Similar compounds may be conferred by similar bioactivities [11]. This fundamental idea allows for the integration of chemical informatics and bioinformatics tools into hypotheses constructs for

drug discovery. Advancements in Big Data management have opened up access to massive public datasets and powerful tools, e.g., virtual screening. In-silico drug targets, toxicity, and drug-drug or drug-food interaction predictions based on chemical similarity through virtual screening and machine learning will offer us many preliminary outcomes efficiently for natural products research.

The concept and framework of network toxicity in Traditional Chinese Medicine (TCM) were first proposed by Fan et al. in 2011. They claimed to employ network pharmacology approaches to reconstruct the network of "compound-protein/gene-toxicity" to identify dangerous chemicals and anticipate the harmful side effects of existing compounds [12]. However, no successful practice has been reported based on this concept, which may because of the knowledge gap between compound-target and compound toxicity. And limited information about the toxicity and drug interactions of natural compounds can be found in the public databases for model training. But there are many expert systems (DEREK [13], AMBIT [14], DSSTox [15], Derek Nexus [16], Meteor [17], HazardExpert [18], PASS [19], cat-SAR [20], Toxmatch [21], VEGA [22], ChemIDplus [23]) built for toxicity prediction of synthetic drugs [24]. All of them are based on two different methodologies: The quantitative structure-activity relationship (QSAR) and molecular docking [25]. There are huge knowledge gaps between the molecular docking result and drug toxicity for the distinguished roles of one protein in systematic toxicity and the roles of compound-protein interactions in the activation or degradation of proteins are elusive. So, QSAR-based systems are the most employed computational approaches to predict drug toxicity. However, most QSAR-based systems employing models trained by datasets of FDA-approved synthetic drugs may show less confidence in the prediction of natural products. The performance of some expert systems for toxicity prediction such as TOPKAT, DEREK, and HazardExpert has been reported to be poor [26–29]. In addition, most toxicity prediction systems only offer limited endpoint alerts without any insight interpretation. Hence, this study aims to construct a prediction workflow (Figure 1) for natural products to alert comprehensive endpoints with detailed information for the first time. Although there are no qualified datasets for "one-step" alert model training for toxicity and drug interaction prediction of natural products. Many models for quantitative structure-activity relationship analysis are available. Firstly, based on one of the best QSAR models and platforms, the basic information about the similar bioactivity compounds of active compounds can be mined. Then the comprehensive toxicity and drug interaction information of similar compounds was collected as the indicator for active compounds. The toxicity and drug interaction prediction can be conducted based on a reasonable similarity score threshold.

PubChem (https://pubchem.ncbi.nlm.nih.gov/) (accessed on 31 December 2021) is the world's largest repository of publicly available chemical data at the National Institutes of Health. It provides detailed information on chemical and physical properties, biological activity, safety and toxicity, patents, and literature citations. It houses data on about 111 million chemicals, 295 million bioactivities, 34 million publications, and 42 million patents [30]. So, it is the most popular database for chemical-related data mining with detailed research protocols and insights. SwissSimilarity (http://www.swisssimilarity.ch/) (accessed on 30 December 2021) is run by the Molecular Modelling Group of the SIB Swiss Institute of Bioinformatics and the University of Lausanne. It is a user-friendly tool for ligand-based virtual screening from several libraries of small compounds using various methodologies. This platform can work based on various structure descriptors such as FP2 [31], ECFP4 [32], MHFP6 [33], Pharmacophore [34,35], ErG [36], Electroshape [37], and E3FP [38,39]. Even an in-house combined model on this platform can be available to show better performance for structure-activity prediction compared to an independent fingerprint similarity-based or shape similarity-based prediction model [40].
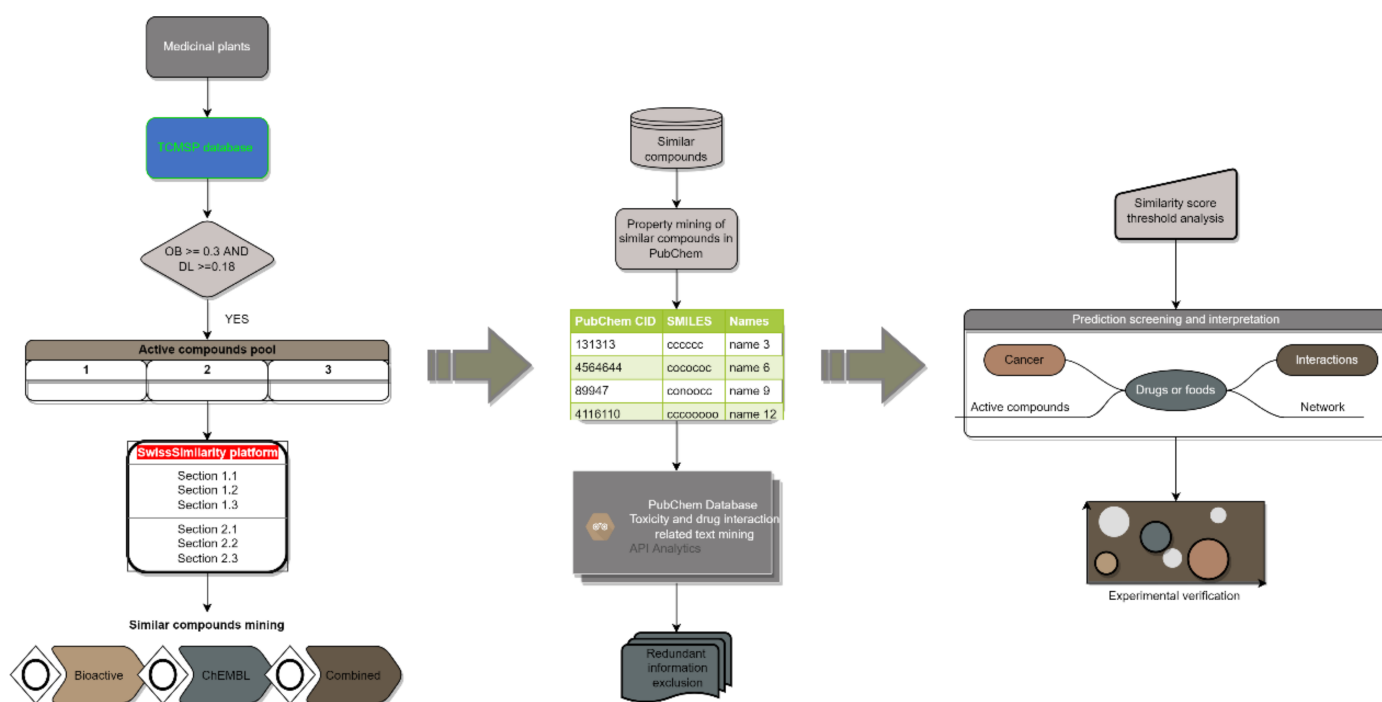
Figure 1. Workflow of in-silico toxicity and drug interaction analysis based on chemical similarity. Active compounds of herbal medicines are screened out based on some criteria. Then, as a query for mining similar compounds, the SMILES of active compounds will be collected. Based on the properties of similar compounds, the information on toxicity and drug interactions of similar compounds are retrieved. The final predictions and interpretations of active compounds will be made on a reasonable similarity score threshold. Experiments will be conducted to demonstrate the prediction results.

2. Results

2.1. Similarity Score Threshold Analysis

To build a workable and dependable in-silico prediction workflow, the similarity threshold setting plays an important role in the balance of prediction precision and prediction yield. Presumably, there is a positive correlation between the similarity score and a true prediction, which means a "strict" threshold will eliminate the false predictions. But it does not mean that a "strict" similarity threshold is better than a "low" similarity threshold in terms of prediction yield because there is a significantly negative correlation between the number of active compounds predicted and the similarity threshold (Figure 2). The similarity score in this workflow was regarded as a kind of descriptor of active compounds but not the final probability of a prediction model. Further analysis will be made based on the similarity threshold.

Twenty herbal medicines, 495 active compounds, and 84,056 similar compounds were involved in similarity threshold analysis. The lower the similarity threshold is the more distinct retrieves can be collected. There is a significantly negative correlation between similar compounds retrieved and the similarity threshold (Figure 2a). There is a significantly negative correlation between the number of active compounds with similar compounds retrieved and the similarity threshold when the similarity score is higher than 0.56 (Figure 2b). Similarly, there is a linear (negative) correlation between the total toxicity information retrieved and the similarity threshold (Figure 3a). And there is a significantly negative correlation between the number of active compounds with toxicity prediction retrieved and the similarity threshold.

a

b
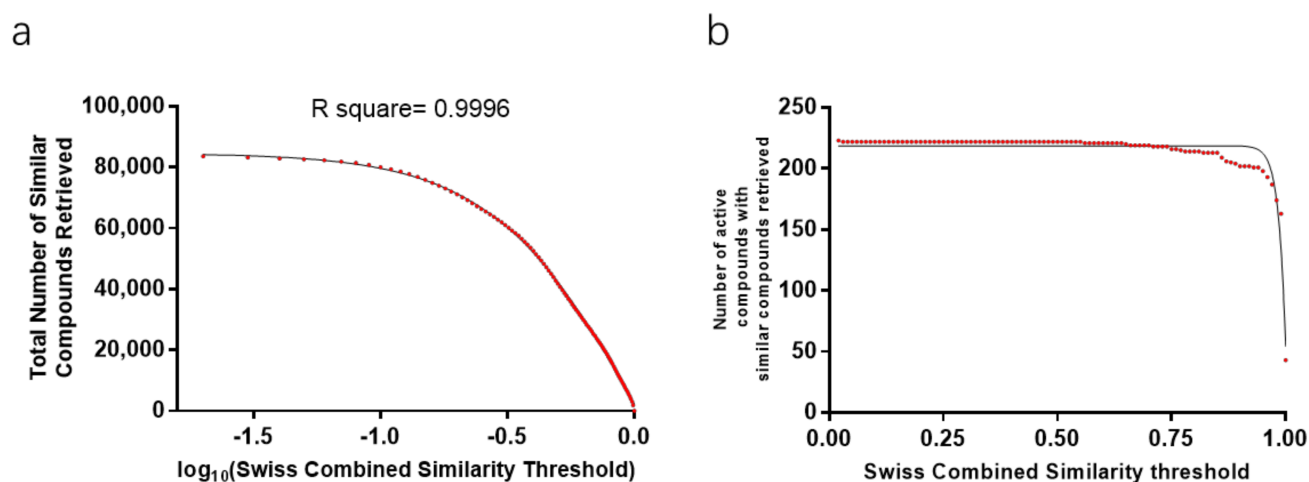
Figure 2. Similar compounds properties retrieved through computer programming for 20 herbal medicines. (**a**) Retrieve number curve of similar compounds properties dependent on similarity threshold. There is a significantly negative correlation between similar compounds retrieved and the similarity threshold. (**b**) Number curve of active compounds in herbal medicines for which the corresponding similar compounds were retrieved dependent on similarity threshold.

a

b
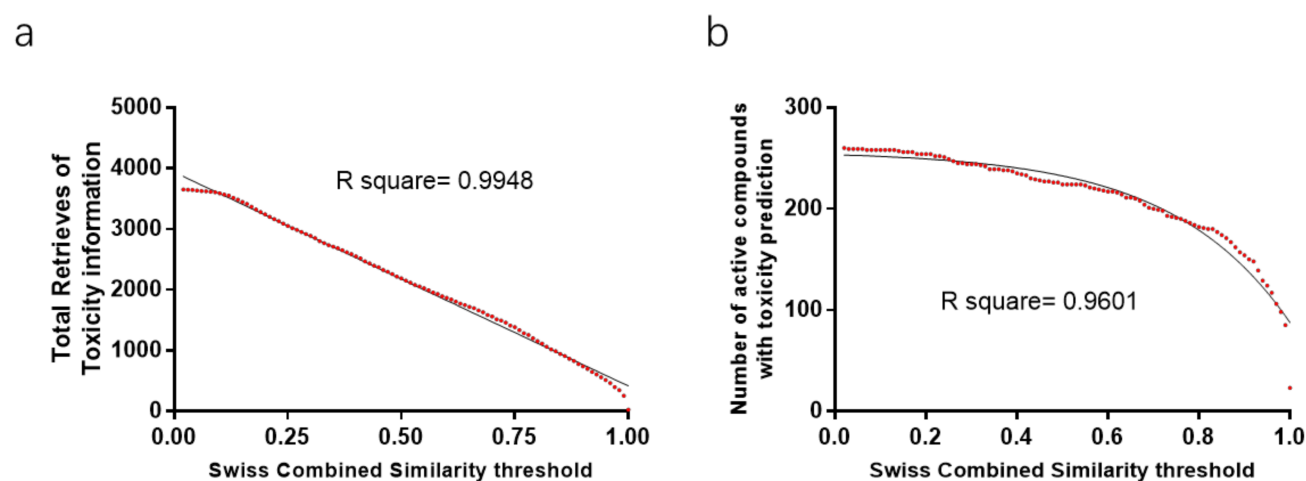
Figure 3. Toxicity information retrieved for 20 representative herbal medicines on different similarity thresholds. (**a**) Retrieve curve of toxicity information dependent on similarity threshold. There is a significantly negative correlation between the total toxicity information retrieved and the similarity threshold. (**b**) Number curve of active compounds in the 20 representative medicinal plants with toxicity information retrieved on different similarity thresholds. There is a negative correlation between the number of active compounds with toxicity information retrieved and the similarity threshold.

However, elusive information or even contrast information about the toxicities of the same compound can be retrieved due to a low similarity score. For example, in Table 1, the first elusive retrieve regarding hepatotoxicity of 3-Hydroxystigmast-5-en-7-one compared to the retrieve with the largest similarity score occurred when the similarity score decreased to 0.498, while the first contrast retrieves for beta-sitosterol and campesterol occurred when the similarity score decreased to 0.588 and 0.594, respectively. Herein, we introduced two different concepts: FEP-SS and FCP-SS (Defined in Box 1) which can be utilized to set an optimal similarity score threshold for predictions to balance the prediction precision and yield. The FCP-SS of one active compound is the threshold for retrieves without any controversial results, while the FEP-SS is the threshold for that all the retrieves are consistent. Based on two different toxicity aspects: hepatotoxicity and carcinogenicity, the optimal similarity score threshold analysis for consistent retrieves was conducted. The

representative FEP-SS and FCP-SS data of four different herbal medicines can be referred to in Table 2. The sum of all the FEP-SS and FCP-SS values of 20 herbal medicines can be referred to in Supplementary Table S1. Low FCP-SS and FEP-SS values probably were gained because of insufficient information from PubChem, herein, values of FCP-SS or FEP-SS ≤ 0.3 were excluded in the descriptive analysis of FCP-SS and FEP-SS values (Figure 4a). Generally, the FCP-SS is numerically less than FEP-SS (Figure 4b). The mean value of FCP-SSs and FEP-SSs are 0.6171 and 0.6181, respectively, and the third quartile value of FCP-SS is 0.759. Given the prediction yield and precision, the similarity score threshold of this prediction project was set at the concentrated value of FCP-SS, the representative value of FCP-SS, 0.6171.

Table 1. The toxicity information summary of similar compounds of representative active compounds.

Active Compound Name	Similarity Score	PubChem CID of Similar Compounds	Hepatotoxicity	Carcinogenicity
3-Hydroxystigmast-5-en-7-one	0.992	6010	0	N.A.
3-Hydroxystigmast-5-en-7-one	0.986	10631	0	N.A.
3-Hydroxystigmast-5-en-7-one	0.976	5997	N.A.	−1
3-Hydroxystigmast-5-en-7-one	0.917	6917715	0	N.A.
3-Hydroxystigmast-5-en-7-one	0.540	54454	0	N.A.
3-Hydroxystigmast-5-en-7-one	0.520	53232	0	N.A.
3-Hydroxystigmast-5-en-7-one	0.498	5280453	−1	N.A.
3-Hydroxystigmast-5-en-7-one	0.302	445354	1	N.A.
beta-sitosterol	0.999	5997	N.A.	−1
beta-sitosterol	0.804	5280453	−1	N.A.
beta-sitosterol	0.588	445354	1	N.A.
beta-sitosterol	0.588	445354	1	N.A.
campesterol	0.999	5997	N.A.	−1
campesterol	0.836	5280453	−1	N.A.
campesterol	0.594	445354	1	N.A.

1: Toxic; 0: Ambiguous; −1: Non-toxic or anti-toxic; N.A.: Not applicable.

Table 2. Similarity score corresponding to the first elusive or contrast prediction for certain toxicity of active compounds of four herbal medicines.

Active Compound Name	PubChem CID of Similar Compound	FEP-SS (Hepatotoxicity)	FCP-SS (Hepatotoxicity)	FEP-SS (Carcinogenicity)	FCP-SS (Carcinogenicity)
(+)-catechin	2369	0.472	-	-	-
(20S)-Dammar-24-ene-3beta,20-diol 3-acetate	5280453	0.31	-	-	-
18alpha-hydroxyglycyrrhetic acid	10133	0.331	-	-	-
3,22-Dihydroxy-11-oxo-delta(12)-oleanene-27-alpha-methoxycarbonyl-29-oic acid	5281004	0.265	-	-	-
3-Hydroxystigmast-5-en-7-one	5280453	0.498	-	-	-
DFV	4764	-	-	0.558	-
Glabranin	16078	0.703	-	-	-
Glabrene	3005573	-	-	-	0.353
Kanzonol F	441140	0.212	-	-	-
Medicarpin	441140	-	0.561	-	-
Olitoriside	54687/12560	0.337	0.449	-	-
Psi-Baptigenin	6237	-	0.176	-	-
Stigmasterol	445354	-	0.468	-	-
Aloe-emodin	42890/3059	0.463	0.413	-	-
Beta-sitosterol	445354	-	0.588	-	-
Campesterol	445354	-	0.594	-	-
Hederagenin	10133	0.486	-	-	-
Liquiritin	30323	-	-	0.379	-
Naringenin	5281576	-	-	0.477	-
Sitosterol	445354	-	0.588	-	-

FEP-SS: First elusive prediction-similarity score; FCP-SS: First contrast prediction-similarity score; -: Not applicable.
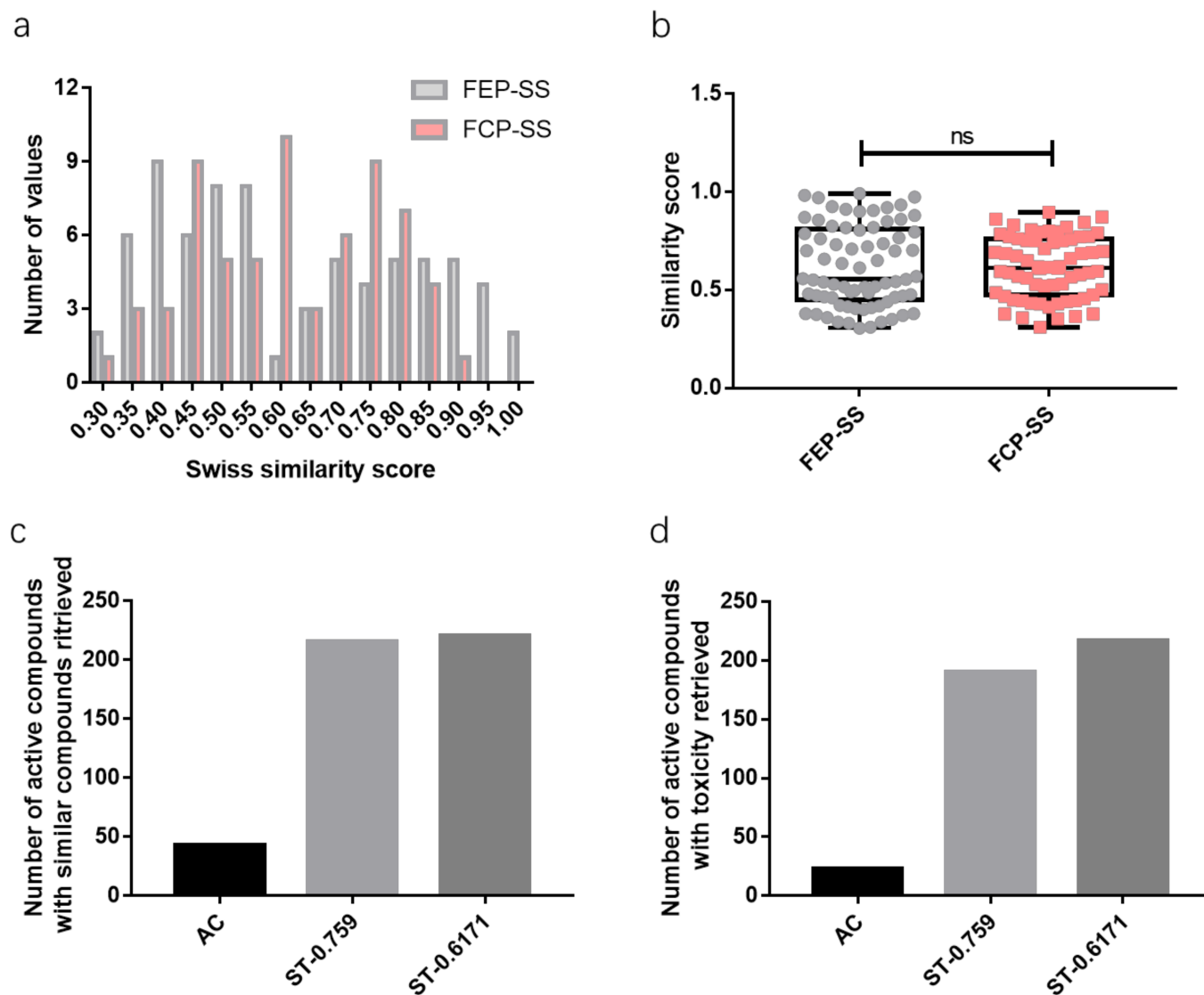
Figure 4. Statistical analysis of in-silico drug toxicity and interaction prediction for 20 different herbal medicines. (**a**) Histogram of FEP-SS and FCP-SS values of >0.3; (**b**) Box plots of FEP-SS and FCP-SS values of >0.3 for toxicity prediction; (**c,d**) The different amounts of active compounds with similar compounds or toxicity predictions based on different Swiss combined similarity score thresholds. AC: The group of toxicity information mining using active compounds of medicinal plants on PubChem platform, ST-0.759: Toxicity prediction group based on similarity score threshold of the third quartile value of FCP-SS (0.759), ST-0.6171: Toxicity prediction group based on a threshold of the mean value of P-SS (0.6171).

2.2. Toxicity Prediction Interpretation

After the construction of the prediction workflow and similarity threshold setting, the toxicity and drug interaction predictions were made for several herbal medicines. From the PubChem database, for the four representative herbal medicines, thirty-one different sections of toxicity information of active compounds such as "Acute Effects", "NIOSH Toxicity Data", "Interactions", "Hepatotoxicity", "Evidence for Carcinogenicity", "Symptoms", "Human Toxicity Values", and "TSCA Test Submissions" sections have been retrieved (See Supplementary Table S2, with 26 active compounds predicted to exert various potential toxicities.

(20S)-Dammar-24-ene-3beta,20-diol 3-acetate was predicted to possess potential hepatotoxicity and reproductive developmental toxicity. Cajinin, calycosin, formononetin, glyza-

glabrin, hederagenin, jaranol, kaempferol, luteolin, odoratin, olitoriside, psi-Baptigenin, quercetin, 3-Hydroxystigmast-5-en-7-one and so on may have potentially reproductive developmental toxicity. Olitoriside is an analog of digoxin with a relatively high similarity score of 0.992, which means it may also exert similar toxicity to digoxin. In addition, 3-Hydroxystigmast-5-en-7-one, glycyrrhiza flavonol A, isolicoflavonol, jaranol, kaempferol, luteolin, isoflavanone, and quercetin are probably genotoxic predicted by the toxicity information retrieve of their similar compounds, respectively. Presumably, reproductive and developmental toxicity should be paid enough attention to the consumption of all these four herbal medicines. It is also warranted that digoxin toxicity may occur when *Fructus ligustri* Lucidi is taken at toxic doses. Besides, the prediction results of acute toxicity, antidote and emergency treatment, protein binding, ecotoxicity values, ongoing test status, skin symptoms, eye symptoms, and target organs of some active compounds can be available in Supplementary Table S2.

2.3. Drug Interaction Prediction Interpretation

The prediction of drug interaction of plant complexes is another important role of this workflow. There are 41 different active compounds involved in the 4 representative herbal medicines, of which the detailed information can be referred to Supplementary Table S3. Many active compounds may have significant super-additive or sub-additive effects on drug pharmacokinetics, cancer management, cell survival, drug-induced reproductive developmental toxicity, antibacterial, anticoagulation, or/and cardiovascular function.

For cancer management, there are roughly seven different activities (Enhanced radiotherapy, metastasis inhibition, carcinogenesis inhibition, enhanced chemotherapy, enhanced genotoxicity, enhanced bioavailability, and weakened target therapy) influenced by potential drug interactions predicted based on the similarity score threshold of 0.6171. Cajinin, calycosin, formononetin, glyzaglabrin, isotrifoliol, luteolin, odoratin, and psi-Baptigenin may potentiate the sensitivity of cancer cells to ionizing radiation. Isorhamnetin, isotrifoliol, jaranol, kaempferol, luteolin, quercetin, and sitosterol were predicted with the same similar compound, apigenin or lupeol, to inhibit cancer metastasis. Hederagenin may inhibit carcinogenesis by 1,2-dimethyl-hydrazine, 12-O-tetradecanoylphorbol 13-acetate, or azoxymethane. Gadelaidic acid and icos-5-enoic acid were predicted to inhibit carcinogenesis caused by methyl nitrosourea. Glycyrrhiza flavonol A, 8-C-alpha-L-arabinosylluteolin, isolicoflavonol, isorhamnetin, jaranol, kaempferol, luteolin, and quercetin probably suppress UV-induced skin tumorigenesis. 8-C-alpha-L-arabinosylluteolin, Glycyrrhiza flavonol A, hederagenin, isorhamnetin, isotrifoliol, jaranol, licochalcone B, kaempferol, liquiritin, luteolin, mairin, olitoriside, quercetin, and sitosterol may have potential synergistic effects when treated in combination with many chemotherapeutic drugs on cancer management.

However, mairin may also be a promoter of N-Nitrobis(2-hydroxypropyl)amine and N-methyl-N′-nitro-nitrosoguanidine triggered cancer progression. There is a theoretical risk of enhanced genotoxicity using cisplatin with isorhamnetin, isotrifoliol, kaempferol, luteolin, quercetin, or 8-C-alpha-L-arabinosylluteoli. In addition, weakened target therapy of bortezomib may occur due to the combined treatment of 8-C-alpha-L-arabinosylluteolin, glycyrrhiza flavonol A, isorhamnetin, jaranol, kaempferol, luteolin, or quercetin.

2.4. Synergism Detection

Based on the drug interaction prediction, in *Spatholobus suberectus* Dunn (See Figure 5), 3-Hydroxystigmast-5-en-7-one, 8-C-alpha-L-arabinosylluteolin, beta-sitosterol, cajinin, calycosin, campesterol, formononetin, luteolin, and psi-Baptigenin (9 out of 24 screened active compounds) may play synergistic effects on cancer metastasis inhibition, carcinogenesis inhibition, chemotherapy, or/and radiotherapy with various chemotherapeutic drugs or factors. Cajinin, calycosin, luteolin, and psi-Baptigenin, similar to kaempferol, were predicted to enhance the chemotherapeutic drug bioavailability because of the inhibition effects on P-glycoprotein (P-gp) and cytochrome P450 (CYP) (See Supplementary Table S4). These two enzyme families play important role in the neutralization and effluxion of vari-

ous chemotherapeutic drugs including docetaxel [41–43], the first-line chemotherapeutic drug in TNBC management [44]. Collectively, there may be potential synergism between SSD and docetaxel in TNBC therapy. Luteolin, extremely similar to apigenin, myricetin, genistein, and kaempferol which have been proved to play synergistic effects on cancer metastasis inhibition, carcinogenesis inhibition, chemotherapy, drug bioavailability enhancement, and radiotherapy with various drugs or factors, attracted our special interests. To prove the efficacies of SSD when treated in combination with docetaxel, a combination index assay, synergy score of matrix assay, and xenograft model were conducted.
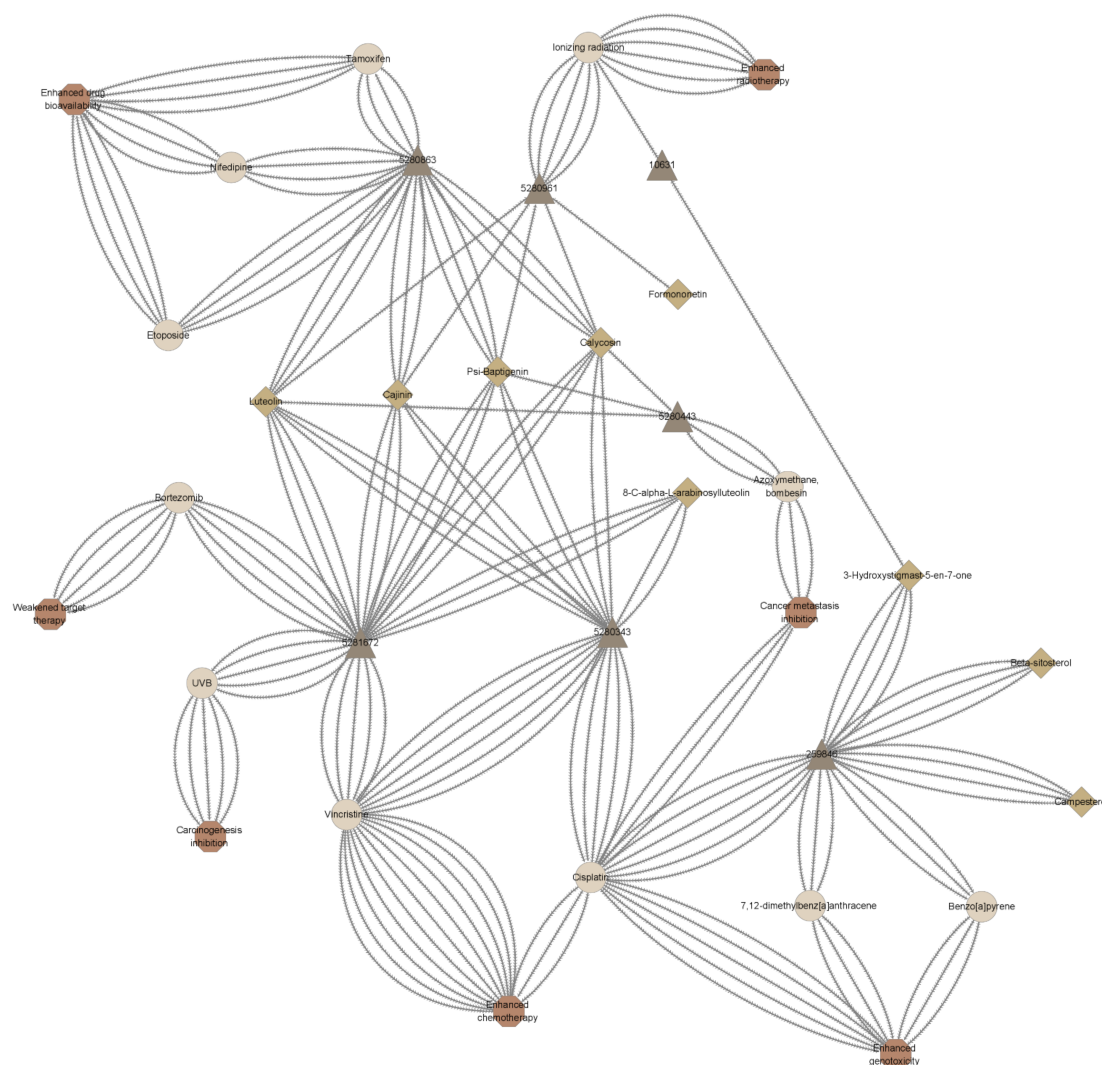

Figure 5. The network of cancer management-related interactions of active compounds in *Spatholobus suberectus* Dunn with drugs, ionizing radiation, or carcinogens, predicted through chemical similarity. Cajinin, calycosin, formononetin, luteolin, and psi-Baptigenin may potentiate the sensitivity of cancer cells to ionizing radiation. Luteolin enjoying similarity of 0.999 to apigenin was predicted to inhibit cancer metastasis. 8-C-alpha-L-arabinosylluteolin, and luteolin probably not only suppress UV-induced skin tumorigenesis but also have potential synergistic effects when treated in combination with many chemotherapeutic drugs. Cajinin, calycosin, luteolin, and psi-Baptigenin, similar to kaempferol, were predicted to enhance the chemotherapeutic drug bioavailability because of the inhibition effects on P-glycoprotein (P-gp) and cytochrome P450 (CYP). However, there is a theoretical risk of enhanced genotoxicity using cisplatin with luteolin or 8-C-alpha-L-arabinosylluteolin supplements. In addition, weakened target therapy of bortezomib may occur due to the combined treatment of 8-C-alpha-L-arabinosylluteolin or luteolin.

In the Combination Index assay (Figure 6), the IC_{50} values of SSP and docetaxel in anti-MDA-MB-231 cells are 70.48 µg/mL, and 1.85 nanomolar, respectively. When treated simultaneously, the IC_{50} values of SSP and docetaxel decreased to 4.73 µg/mL, and 1.18 nanomolar, respectively. The Combination Index is 0.70, which means there is a synergism of SSP and docetaxel in anti-MDA-MB-231 cells. The consistent results can be gained by the Synergy score detection assay where the mean value of the synergy score calculated by the ZIP method is 5.79, with the most synergistic area score of 20.68. In the combination of luteolin and docetaxel, the mean value of the synergy score is 7.217, and the synergy score of the most synergistic area (White rectangle) is 19.58 (Figure 7).
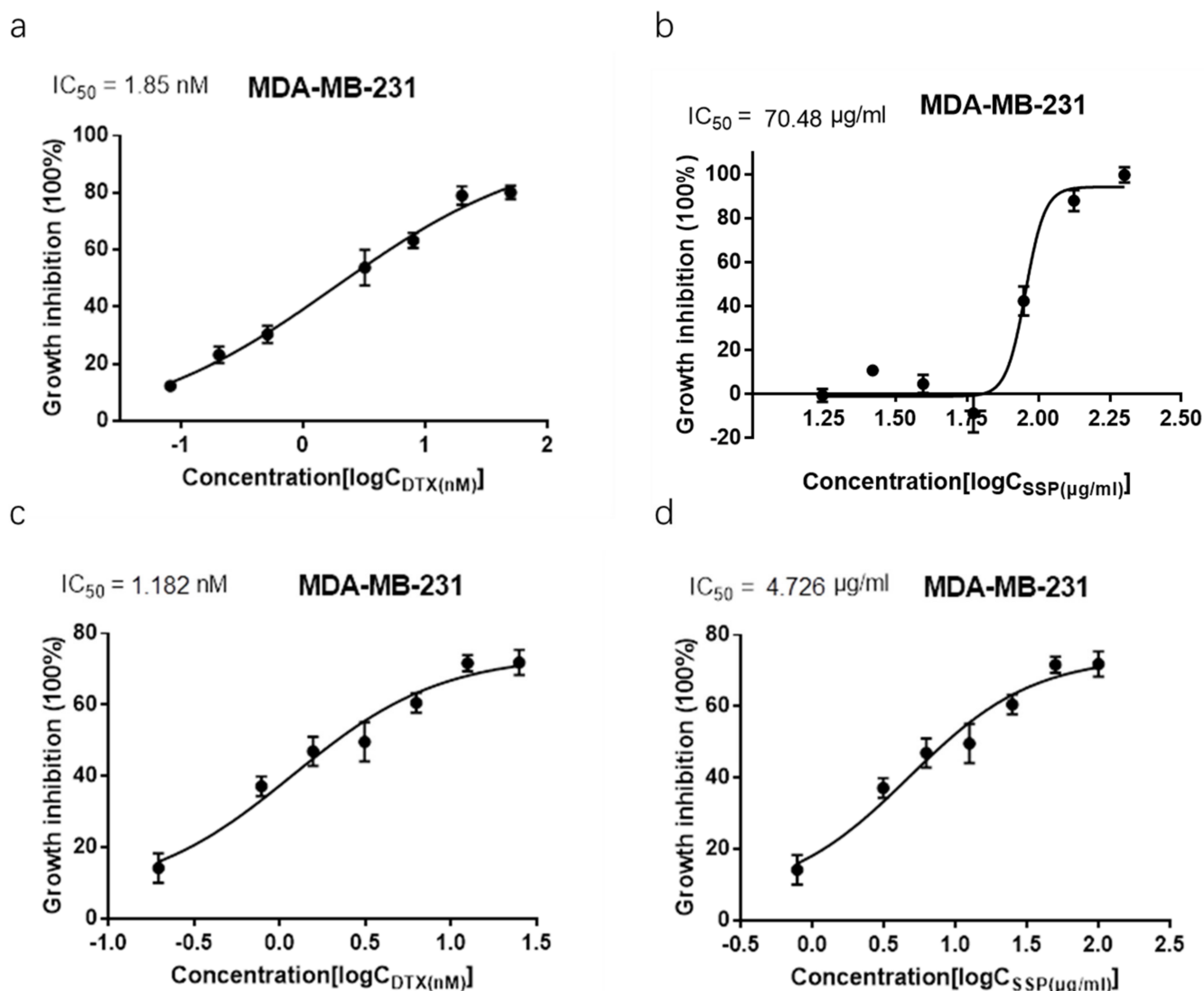

Figure 6. Combination index analysis for the synergistic effects of SSP and docetaxel. (**a**) MDA-MB-231 cells treated with different doses of docetaxel independently for 48 h, IC_{50} (DTX) = 1.85 nM, R square = 0.9805. (**b**) MDA-MB-231 cells treated with different doses of SSP independently for 48 h, IC_{50} (SSP) = 70.48 µg/mL, R square = 0.9854. (**c,d**) MDA-MB-231 cells treated with different doses of SSP (µg/mL)/docetaxel (nM) (100/25; 50/12.5; 25/6.25; 12.5/3.13; 6.25/1.56; 3.13/0.78; 0.78/0.20) for 48 h, IC_{50} (DTX) = 1.18 nM, IC_{50} (SSP) = 4.726 µg/mL, R square = 0.9424. CI ≈ 0.70, which means there is a synergism between SSP and docetaxel.
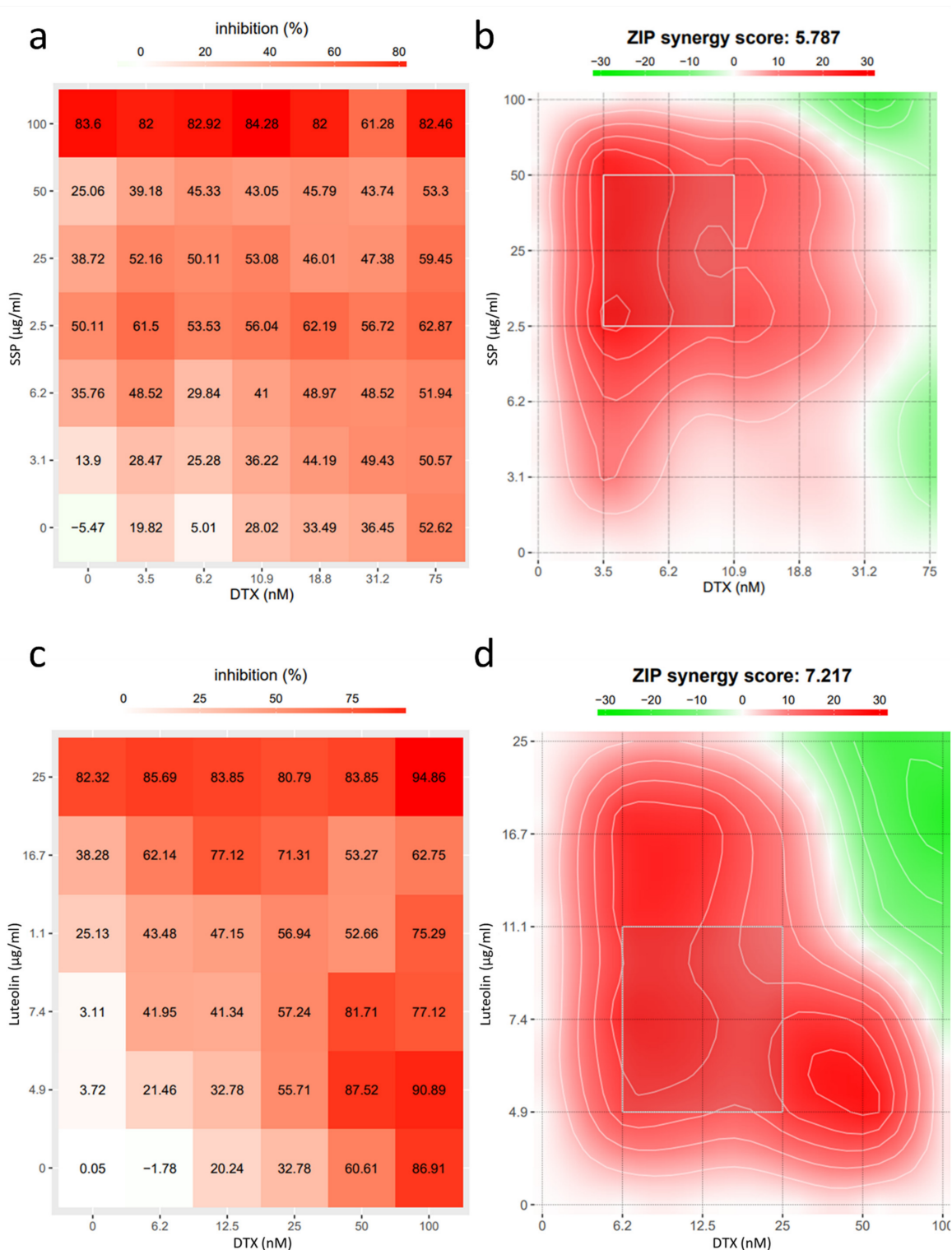
Figure 7. Quantitative analysis of the synergistic effect of SSP or luteolin combined with docetaxel through the ZIP method on the growth inhibition of MDA-MB-231 cells. (**a**,**c**) 2-D heat map of the dose-response matrix (Inhibition ratio) of drugs. (**b**,**d**) 2-D heat map of synergy score. In the combination of SSP and docetaxel, the mean value of the synergy score is 5.79, and the synergy score of the most synergistic area (White rectangle) is 20.68; In the combination of luteolin and docetaxel, the mean value of the synergy score is 7.217, the synergy score of the most synergistic area (White rectangle) is 19.58. The data were gained from two independent experiments.

In-vivo assay (Figure 8), there are no significant differences among the Vehicle control group, DTX group, and SSP-L group in terms of tumor volume. However, the tumor volume of the combination group of docetaxel treatment at low dose plus the SSP treatment at low dose was significantly less than that of the Vehicle control group, which means there is supper-additivity between SSP and docetaxel in anti-TNBC.
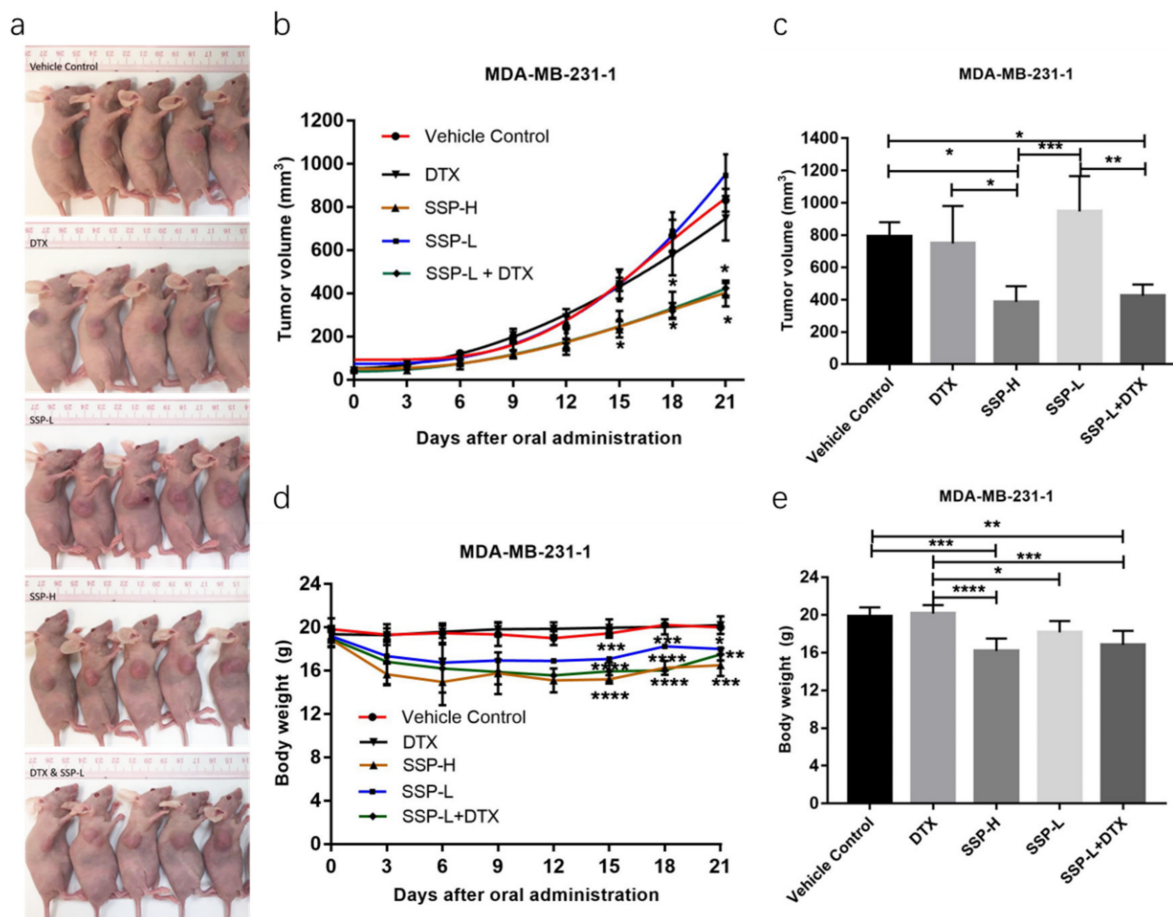
Figure 8. Qualitative analysis of the synergistic effects of SSP and docetaxel in anti-TNBC. (**a**) Representative pictures of mice xenograft with different treatments for 21 days. (**b**) Tumor volume curve. The Vehicle Control group received oral administration of Milli-Q water; The SSP-L group received oral administrations of SSP (0.4 g/kg/day); The SSP-H group received oral administrations of SSP (0.8 g/kg/day). The DTX group received administration of docetaxel (i.p., 2.5 mg/kg/week). The DTX & SSP-L group received oral administrations of SSP (0.4 g/kg/day) and docetaxel (i.p., 2.5 mg/kg/week). Data are shown as mean $\pm$ SEM ($n = 6$) with two independent experiments. (**c**) Statistical analysis of tumor volume at the endpoint of different mice with different treatments aforementioned. Data are shown as mean $\pm$ SEM ($n = 6$). * $p < 0.05$ (Vehicle control vs. SSP-H; Vehicle control vs. SSP-L + DTX; DTX vs. SSP-H), ** $p < 0.01$ (SSP-L vs. SSP-L + DTX) and *** $p < 0.001$ (SSP-L vs. SSP-H). (**d**) Bodyweight detection of the xenograft model experiment. Data are shown as mean $\pm$ SD ($n = 6$) with two independent experiments. (**e**) Statistical analysis of mice body weight at the endpoint of different groups. Data are shown as mean $\pm$ SD ($n = 6$). * $p < 0.05$ (DTX vs. SSP-L), ** $p < 0.01$ (Vehicle control vs. SSP-L + DTX), *** $p < 0.001$ (Vehicle control vs. SSP-H; DTX vs. SSP-L + DTX), **** $p < 0.0001$ (DTX vs. SSP-H).

3. Discussion

For the first time, a workflow (Figure 1) for both toxicity and drug interaction prediction of herbal medicine based on virtual screening and text mining [45,46] was constructed. For studies on drug toxicity, drug-drug interactions, and drug-food interactions, with

detailed related information retrieved, this workflow is beneficial for hypothesis construction and insight interpretation. Moreover, it has many superiorities over fixed prediction models. First, no prediction model can predict the toxicity and drug interactions like this workflow at the same time, which is important for a comprehensive safety assessment of complex drug mixture. Second, drug-drug interaction prediction models can only be utilized for drug pairs of interest, where the name of the interested drug pair should be offered in advance. However, this workflow can show some insights into drug combinations of >2 compounds without any pre-purpose needed. Even, the interaction between active compounds and some other factors such as ionizing radiation or carcinogens can be indicated by this workflow. Third, no detailed insights or interpretations of any toxicity prediction model are available. But this text mining-based procedure will include data resources, clinical trial details, and even experiment protocols for a result assessment in addition to the endpoint alert. Forth, as more and more information is documented in the chemical databases, a flexible workflow show much more potential and comprehensive assessment of the compound toxicity and interactions compared to a fixed prediction model which is only trained for specific toxicities prediction and utilized in limited scenarios. Fifth, this workflow is based on a combined SwissSimmlarity score, which has been proved to show better performance compared to fingerprint as the unique structure descriptor in activity prediction.

There is limited information about toxicity and drug interactions of natural active compounds (Table 3) documented in public databases. In PubChem, the most powerful database of chemical information regarding toxicity, only 23 out of 495 active compounds of 20 representative herbs can be retrieved. So, it is not feasible to make a safety assessment on herbal medicines by searching through public databases. Although there are many expert systems constructed for drug toxicity prediction based on QSAR or molecular docking, most QSAR-based systems employing models trained by FDA-approved drugs may show less confidence in the prediction of natural products. Plus, there are huge knowledge gaps between the molecular docking result and drug toxicity for the distinguished roles of one protein in systematic toxicity and the roles of compound-protein interactions in the activation of proteins are elusive. To solve this problem, this workflow tries to predict the properties of the unknown compounds by their similar bioactive compounds based on optimal QSAR on the SwissSimilarity platform.

Table 3. Toxicity and drug interaction data mining in PubChem by active compounds of Fructus ligustri Lucidi, Spatholobus suberectus Dunn, Hedysarum multijugum Maxim, and Licorice.

PubChem CID	Active Compound Name	AE	Is	AET	HTE	NHTE	CC	PSR	HT	EC	NHTV	OTS	NTPS
5280448	Calycosin	-	+	+	+	+	-	-	-	-	-	-	-
9064	(+)-catechin	+	-	-	-	-	-	-	-	-	-	-	-
222284	Beta-sitosterol	+	-	-	-	-	-	-	-	-	-	-	-
73299	Hederagenin	+	-	-	-	-	-	-	-	-	-	-	-
5280863	Kaempferol	-	+	+	+	+	+	+	-	-	-	-	-
5280445	Luteolin	+	-	-	-	-	-	-	-	-	-	-	-
5280343	Quercetin	+	+	+	+	+	+	+	+	+	+	+	+
5280794	Stigmasterol	-	-	+	+	-	-	-	-	-	-	-	-
439533	Taxifolin	+	-	-	-	-	-	-	-	-	-	-	-

AE: Acute Effects; Is: Interactions; AET: Antidote and Emergency Treatment; HTE: Human Toxicity Excerpts; NHTE: Non-Human Toxicity Excerpts; CC: Carcinogen Classification; PSR: Populations at Special Risk; HT: Hepatotoxicity; EC: Evidence for Carcinogenicity; NHTV: Non-Human Toxicity Values; OTS: Ongoing Test Status; NTPS: National Toxicology Program Studies. +: with retrieves; -: no retrieves.

There are no available models for toxicity or drug-drug interaction prediction for a mixture of more than two different compounds because no dataset can be gained for this kind of model training. All drug interaction prediction models are just trained for drug pairs of interest [47–49]. Although there is some interaction information of three-

compound combinations retrieved in this study, even based on the most powerful database for chemical information-PubChem, most drug interaction retrieves are also documented for drug pairs. So, there is still a knowledge gap between the predictions of this workflow and the final clinical performance of plant complexes. All the predictions should be proved by experiments. The prediction result should be treated as preliminary hypotheses.

For a logistic regression model evaluation, metrics of sensitivity and specificity are the most introduced. Receiver operating characteristic curves (ROC), graphs of the specificity vs. the sensitivity, dependent on different thresholds, can show the performance of various models trained by the same dataset. Moreover, the area under the ROC curve can be used to compare different models trained by various algorithms and strategies. However, for the toxicity and drug interaction prediction of natural compounds, there are insufficient data for a logistic regression model training. Here we employed a combined system as beforementioned. To evaluate this system, two new concepts FEP-SS and FCP-SS (Defined as Box 1) were introduced to find the similarity range corresponding to consistent predictions or uncontroversial predictions compared to the retrieve with the highest similarity, respectively. More similar compounds and the toxicity and drug interaction information can be retrieved for active compounds with a relatively large similarity range because there is a significantly negative correlation between information retrieved and the similarity threshold (Figure 2a). The higher the similarity threshold is set, the more precise the prediction of this system is, but the less active compounds can be predicted (Figures 3b and 4c,d). Ambiguous predictions are acceptable for the toxicity prediction of natural compounds, so the mean value of FCP-SS was set as the similarity score threshold. As more and more data are documented in the PubChem database, theoretically, a "stricter" similarity threshold for the prediction of most active compounds can be set in the future.

There are no distinct conclusions about the relationship between the content of an active compound and the weight of the compound on the activity of plant complexes. Some compounds accounting for a small proportion of the total herbal medicine may still exert remarkable activities, while some compounds accounting for a large proportion of total extracts may show little bioactivities. Given the drug interactions, situations will be much more complex, that is where the significance of this manuscript comes from. It filled a vacancy in conventional network pharmacology which lacks drug toxicity and interactions analysis in a complex system.

In QSAR analysis, chemical similarities such as fingerprints [50,51] and shape similarity [37] are the most popular descriptors for the structure of small molecules. In general, fingerprint similarity performs better than shape similarity in terms of bioactivity prediction [40]. That is why most toxicity prediction-expert systems take fingerprints of molecules as the structure descriptors. But shape similarity, independent of fingerprint similarity, can bring some extra information for indicating the chemical structure of drugs. So, a combined model trained by machine learning plays better performance in terms of structure-activity relationship prediction [40]. Herein, we tended to employ the combined model offered by the SwissSimilarity platform, a user-friendly platform with the "Bioactive"-compound class and several compound libraries for natural compound-data mining, to find the similar compounds with similar bioactivities [38]. For the stage of endpoint alert, most expert systems show limited information, which is difficult for prediction and insight interpretation. To solve this problem, ligand-based virtual screening outperforms an ambitious prediction model. However, this workflow was much more time-consuming for manual prediction interpretation compared to conventional expert systems. And it needs professional knowledge to interpret toxicity information to avoid an interpretation error. Text-classification and interpretation models trained via machine learning may solve this problem someday.

There are synergistic effects of crude extracts of SSD combined with docetaxel in anti-TNBC. Luteolin in concomitant use of docetaxel was also proved to show super-additive effects in anti-TNBC cells at certain doses (Figure 7). These experimental results are consistent with the prediction results. Till now, there are no effective medications but chemotherapeutic drugs for the management of triple-negative breast cancer [52,53].

However, most chemotherapy will incur more than 20 different severe adverse events such as anemia, diarrhea, fatigue, nausea, vomiting, and hair changes [4]. Based on the potential synergism of SSD and docetaxel, with less toxicity, lower therapeutic doses of combination treatment of docetaxel and SSD may accomplish the counterpart even better efficacies compared to the independent treatment of docetaxel or SSP.

4. Conclusions

For the first time, a workable and dependable workflow of in-silico drug toxicity and interaction prediction for plant complexes was built. From the PubChem database, 31 different sections of toxicity information such as "Acute Effects", "NIOSH Toxicity Data", "Interactions", "Hepato-toxicity", "Carcinogenicity", "Symptoms", and "Human Toxicity Values" sections have been retrieved, with dozens of active compounds predicted to exert potential toxicities. In *Spatholobus suberectus* Dunn (SSD), there are 9 out of 24 active compounds predicted to play synergistic effects on cancer management with various drugs or factors, which is consistent with the experimental data.

5. Materials and Methods

5.1. Dataset Assembly

A dataset containing the active compounds of 20 herbal medicines was gathered from the TCMSP database [54], based on the ADME criteria (("Oral bioavailability" ≥ 0.3 and "Drug-likeness" ≥ 0.18), given all the herbal medicines are presumed to be administrated orally. Finally, the dataset "active_comp_pool_tcmsp.csv" contained 561 active compounds in total (495 distinct active compounds), 13 active compounds of Fructus ligustri Lucidi, and 24 active compounds of Spatholobus suberectus Dunn, 20 active compounds of Hedysarum multijugum Maxim, and 92 active compounds of Licorice, respectively.

5.2. Similar Compound Data Mining

From the PubChem database, the mining of properties of active compounds was conducted firstly through a script coded in Python 3 (version 3.8.10) called "compound_properties_mining.py" using pubchempy (version 1.0.4) and pandas (version 1.2.5) packages. This script iterates over the "active_comp_pool_tcmsp.csv" dataset, specifically, the "Molecule Name" column, while fetching one "Molecule Name" at a time. The gathered property data of active compounds were written to a CSV file named "active_comp_proper_pubchem.csv". After duplicate values deletion, the mining of similar compounds of active compounds was done through the web scraper script called "similar_comp_crawler.py". This script iterated the "Active_compound_name" column and the "isomeric_smiles" column of the dataset storing the properties of active compounds. The isomeric SMILES code is posted as a query to the SwissSimilarity website (updated version issued in Dec. 2021), selecting "Bioactive" compound class, choosing "ChEMBL (actives only)" natural product library [55], based on combined methods [40]. All the data of similar compounds were stored in the file named "similar_comp_pool_swiss.csv".

5.3. Toxicity and Drug Interaction Information Mining

Before toxicity and drug interaction information mining, using a script called "similar_compound_properties_mining.py", the properties of similar compounds were collected with a similar method as the mining of properties of active compounds beforementioned and were stored in the file named "similar_comp_properties_sum.csv". Then the toxicity and drug interaction information mining was conducted through the web scraper script called "toxicity_mining_pubchem.py". After redundant-value deletion, all the toxicity and drug interaction information were stored in the file named "Toxi_infor_sum.csv". Screened from the "Toxi_infor_sum.csv" file, the drug interaction information retrieved was separated and split into one "interaction" retrieve per row using a script named "drug_interactions_split.py" for further manual interpretation. The split data was stored in the file named "drug_interaction_pred_0.6171.csv".

5.4. Prediction Interpretation

The final prediction of the toxicity or drug interactions of active compounds of 4 representative herbal medicines was interpreted manually, based on the toxicity information and drug interaction data of its similar compounds, with a reasonable similarity score threshold. The definitions of "Active compound", "Prediction yield", "Drug interaction", "First elusive prediction-similarity score", and "First contrast prediction-similarity score" can be referred to Box 1. Every row in the dataset "Toxi_infor_sum.csv" was regarded as one retrieve. In a retrieve, for certain toxicity annotation, the similar compound was annotated as toxic if there was at least one in 32 sections clarifying the certain toxicity of similar compounds, or regarded as ambiguously toxic if there were controversial insights about the certain toxicity, or regarded as non-toxic or anti-toxic if all the available information indicating it was non-toxic or anti-toxic, or documented as "N.A." if there were no related insights or evidence in all sections. The prediction results of active compounds were indicated by the indicators (Similar compounds with a similarity score above the similarity score threshold). For certain toxicity predictions, the active compound was regarded as toxic if major indicators were annotated toxic, or regarded as ambiguously toxic if there were controversial insights about indicators, or regarded as non-toxic or anti-toxic if all the indicators were non-toxic or anti-toxic, or documented as "N.A." if there were no related insights or evidence about all the indicators of one active compound. The basic statistical analysis of this prediction study was done through the scripts named "parameter_similar_comp_properties.py", and "parameter_comp_toxicity.py". The predicted drug interaction network of active compounds in Spatholobus suberectus Dunn for cancer management was made by Cytoscape (Version 3.8.2) [56].

Box 1. Definition of the basic concept in the prediction system.

---

1. Active compound
   The active compounds of herbal medicine defined here, are the natural products documented in the TSCSP database for a certain herbal medicine, screened out based on the criteria ("Oral bioavailability" ≥ 0.3 and "Drug-likeness" ≥ 0.18).
2. Drug interactions
   Drug interactions, in such a prediction system, include drug-food interactions, drug-drug interactions, and interactions of drugs with other factors such as carcinogens and ionizing radiation.
3. Prediction yield
   In such a prediction system, the prediction yield is defined as the number of active compounds of herbal medicines with at least one kind of toxicity or drug interaction information predicted.
4. First elusive prediction-similarity score (FEP-SS)
   Among all the information retrieved of similar compounds, for a certain toxicity prediction of an active compound based on such a system, as the similarity score decreases, the first elusive prediction-similarity score is the similarity score corresponding to the first elusive, arguable, or equivocal toxicity information retrieve compared to the toxicity information retrieve with the largest similarity score.
5. First contrast prediction-similarity score (FCP-SS)
   Among all the information retrieves of similar compounds, for a certain toxicity prediction of an active compound based on such a system, as the similarity score decreases, the first contrast prediction-similarity score is the similarity score corresponding to the first contrast toxicity information retrieve compared to the toxicity information retrieve with the largest similarity score.

5.5. Preparation of Spatholobus Suberectus Dunn-Percolation (SSP) Extract

SSP was prepared and made a quality control as before described, its chemical profile can be referred to in previous studies [57]. Dried SSD stems were ground into coarse powder, then it was extracted using a percolating device with 10 times volumes (v/w) of 60% ethanol. The filtrate was then concentrated under reduced pressure by a rotary

evaporator. The concentrated percolation extracts were then freeze-dried (Percent yield 20%) and stored at 4 °C for further use.

5.6. Cell Culture and Treatment

MDA-MB-231 cells were obtained from American Type Culture Collection (Manassas, VA, USA). All cells were maintained in glucose-containing (4.5 g/L) Dulbecco's modified Eagle medium (Gibco, Grand Island, NY, USA), supplemented with fetal bovine serum (10% *v/v*, Gibco, Grand Island, NY, USA), penicillin (Sigma-Aldrich, St. Louis, MO, USA, 100 U/mL), and streptomycin (Sigma-Aldrich, St. Louis, MO, USA, 100 μg/mL) in a humidified atmosphere of 5% CO2 at 37°C. Cells were seeded onto 96-well plates at the density of $3–5 \times 10^3$/well. After undergoing serum starvation for 24 h, they were treated with different concentrations of SSP, luteolin (DIECKMANN (HK) CHEMICAL INDUSTRY COMPANY LTD, Hong Kong, China), or docetaxel (Beijing Aosaikang Pharmaceutical Co., Ltd., Beijing, China). The tumor cell growth inhibitory effects of drugs were detected by CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay containing 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2- (4-sulfophenyl)- 2H-tetrazolium) (MTS) kit (Promega, Wisconsin, DA, USA) as per the manufacturer's protocol. The IC_{50} values of drugs were calculated by linear or nonlinear regression. The Combination Index was calculated after 40 h- of drug treatment using the formula [58]: Combination Index = (D)1/(Dx)1 + (D)2/(Dx)2, where (Dx)1, (Dx)2 are the concentrations of the tested substance 1 and the tested substance 2 used in the single treatment that was required to decrease the cell viability by x%, and (D)1, (D)2 are the concentrations of the tested substance 1 in combination with the concentration of the tested substance 2 that together decreased the cell viability by x%. The synergy score was calculated on the SynergyFinder platform (http://www.synergyfinder.org/) [59] with "Matrix" format and inhibition-Phenotypic Response, using the ZIP method [60] after 24 h-drug treatment.

5.7. Xenograft Model

The xenograft model was constructed as before described [57]. Female (BALB/c) nude mice (6–7 weeks old) were purchased from Harlan Laboratories, Indianapolis, IN, USA that were housed and maintained in the Laboratory Animal Unit, the University of Hong Kong, a specific pathogen-free and climate-controlled room (22 ± 2 °C, $50 \pm 10\%$ relative humidity) with a 12-h light/dark cycle and provided with diet and water ad libitum. MDA-MB-231 cells (2×10^6/site) were implanted subcutaneously into the bilateral flank of each mouse. Palpable and measurable tumors were initially found 10 days after cell injection. Then, the animals were randomly assigned into five groups that were received the following treatments: the Vehicle control group (*n* = 6) received Milli-Q water; the SSP-L group (*n* = 6) received SSP (0.4 g/kg/p.o, daily); the SSP-H (*n* = 6) group received SSP (0.8 g/kg/p.o, daily); the DTX group (*n* = 6) received docetaxel (2.5 mg/kg/i.p. week); the combination group (DTX & SSP-L) (*n* = 6) received docetaxel (2.5 mg/kg/i.p. week) plus SSP (0.4 g/kg/p.o, daily). The tumor size was calculated using the formula: $0.5 \times$ lengths $\times$ width2. All experiments were approved by the Institutional guidelines of Laboratory Animal Care and Committee on the Use of Live Animals in Teaching and Research (CULATR No.: 4484-17).

5.8. Statistical Analysis

Linear or non-linear regression was operated with GraphPad Prism 7 (GraphPad Software, San Diego, CA, USA) choosing log(inhibitor) vs. response-Variable slope (four parameters) as the equation. All data were expressed as Mean $\pm$ SD or Mean $\pm$ SEM. One-way ANOVA was employed to make a difference analysis for multiple groups' comparation. The difference between two groups was analyzed by a two-tailed Student's *t*-test. Significance was established at $p < 0.05$.

References

1. Elghazaly, H.; Rugo, H.S.; Azim, H.A.; Swain, S.M.; Arun, B.; Aapro, M.; Perez, E.A.; Anderson, B.O.; Penault-Llorca, F.; Conte, P.; et al. Breast-Gynaecological & Immuno-Oncology International Cancer Conference (BGICC) Consensus and Recommendations for the Management of Triple-Negative Breast Cancer. *Cancers* **2021**, *13*, 2262.
2. Miles, D.; Gligorov, J.; Andre, F.; Cameron, D.; Schneeweiss, A.; Barrios, C.; Xu, B.; Wardley, A.; Kaen, D.; Andrade, L.; et al. Primary results from IMpassion131, a double-blind, placebo-controlled, randomised phase III trial of first-line paclitaxel with or without atezolizumab for unresectable locally advanced/metastatic triple-negative breast cancer. *Ann. Oncol.* **2021**, *32*, 994–1004. [CrossRef] [PubMed]
3. Schmid, P.; Rugo, H.S.; Adams, S.; Schneeweiss, A.; Barrios, C.H.; Iwata, H.; Diéras, V.; Henschel, V.; Molinero, L.; Chui, S.Y.; et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): Updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2020**, *21*, 44–59. [CrossRef]
4. Borghaei, H.; Gettinger, S.; Vokes, E.E.; Chow, L.Q.M.; Burgio, M.A.; de Castro Carpeno, J.; Pluzanski, A.; Arrieta, O.; Frontera, O.A.; Chiari, R.; et al. Five-Year Outcomes from the Randomized, Phase III Trials CheckMate 017 and 057, Nivolumab Versus Docetaxel in Previously Treated Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2021**, *39*, 723–733. [CrossRef]
5. Zhao, M.Y.; Liu, P.; Sun, C.; Pei, L.J.; Huang, Y.G. Propofol Augments Paclitaxel-Induced Cervical Cancer Cell Ferroptosis In Vitro. *Front. Pharmacol.* **2022**, *13*, 816432. [CrossRef]
6. Li, R.T.; Zhu, Y.D.; Li, W.Y.; Hou, Y.K.; Zou, Y.M.; Zhao, Y.H.; Zou, Q.; Zhang, W.H.; Chen, J.X. Synergistic photothermal-photodynamic-chemotherapy toward breast cancer based on a liposome-coated core-shell AuNS@NMOFs nanocomposite encapsulated with gambogic acid. *J. Nanobiotechnol.* **2022**, *20*, 212. [CrossRef]
7. Date, T.; Kuche, K.; Chaudhari, D.; Ghadi, R.; Sahel, D.K.; Chitkara, D.; Jain, S. Hitting Multiple Cellular Targets in Triple-Negative Breast Cancer Using Dual-Action Cisplatin(IV) Prodrugs for Safer Synergistic Chemotherapy. *ACS Biomater. Sci. Eng.* **2022**, *8*, 2349–2362. [CrossRef]
8. Wang, C.; Gao, H.; Huang, L.; Wang, Z.; Ding, X. Network Pharmacological Analysis and Experimental Study of the Antipharyngitis Mechanism of the Chaiqin Qingning Capsule. *Biomed Res. Int.* **2022**, *2022*, 5616942. [CrossRef]
9. Li, X.; Ma, J.; Guo, L.; Dong, C.; Zhu, G.; Hong, W.; Chen, C.; Wang, H.; Wu, X. Identification of Bioactive Compounds and Potential Mechanisms of Kuntai Capsule in the Treatment of Polycystic Ovary Syndrome by Integrating Network Pharmacology and Bioinformatics. *Oxid. Med. Cell. Longev.* **2022**, *2022*, 3145938. [CrossRef]
10. Crowther, N.R.; Holbrook, A.M.; Kenwright, R.; Kenwright, M. Drug interactions among commonly used medications. Chart simplifies data from critical literature review. *Can. Fam. Physician* **1997**, *43*, 1972–1976, 1979–1981.
11. Martin, Y.C.; Kofron, J.L.; Traphagen, L.M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **2002**, *45*, 4350–4358. [CrossRef]
12. Fan, X.; Zhao, X.; Jin, Y.; Shen, X.; Liu, C. Network toxicology and its application to traditional Chinese medicine. *Zhongguo Zhong Yao Za Zhi* **2011**, *36*, 2920–2922.
13. Ridings, J.E.; Barratt, M.D.; Cary, R.; Earnshaw, C.G.; Eggington, C.E.; Ellis, M.K.; Judson, P.N.; Langowski, J.J.; Marchant, C.A.; Payne, M.P.; et al. Computer prediction of possible toxic action from chemical structure: An update on the DEREK system. *Toxicology* **1996**, *106*, 267–279. [CrossRef]

14. Jeliazkova, N.; Jeliazkov, V. AMBIT RESTful web services: An implementation of the OpenTox application programming interface. *J. Cheminform.* **2011**, *3*, 18. [CrossRef]

15. Grulke, C.M.; Williams, A.J.; Thillanadarajah, I.; Richard, A.M. EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. *Comput. Toxicol.* **2019**, *12*, 100096. [CrossRef]

16. Langton, K.; Patlewicz, G.Y.; Long, A.; Marchant, C.A.; Basketter, D.A. Structure-activity relationships for skin sensitization: Recent improvements to Derek for Windows. *Contact Dermat.* **2006**, *55*, 342–347. [CrossRef]

17. Greene, N.; Judson, P.N.; Langowski, J.J.; Marchant, C.A. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. *SAR QSAR Environ. Res.* **1999**, *10*, 299–314. [CrossRef]

18. Benfenati, E.; Gini, G. Computational predictive programs (expert systems) in toxicology. *Toxicology* **1997**, *119*, 213–225. [CrossRef]

19. Poroikov, V.; Filimonov, D.; Lagunin, A.; Gloriozova, T.; Zakharov, A. PASS: Identification of probable targets and mechanisms of toxicity. *SAR QSAR Environ. Res.* **2007**, *18*, 101–110. [CrossRef]

20. Cunningham, A.R.; Carrasquer, C.A.; Mattison, D.R. A categorical structure-activity relationship analysis of the developmental toxicity of antithyroid drugs. *Int. J. Pediatr. Endocrinol.* **2009**, *2009*, 936154. [CrossRef]

21. Gallegos-Saliner, A.; Poater, A.; Jeliazkova, N.; Patlewicz, G.; Worth, A.P. Toxmatch—A chemical classification and activity prediction tool based on similarity measures. *Regul. Toxicol. Pharmacol.* **2008**, *52*, 77–84. [CrossRef] [PubMed]

22. Pedretti, A.; Mazzolari, A.; Gervasoni, S.; Fumagalli, L.; Vistoli, G. The VEGA suite of programs: An versatile platform for cheminformatics and drug design projects. *Bioinformatics* **2021**, *37*, 1174–1175. [CrossRef] [PubMed]

23. Tomasulo, P. ChemIDplus-super source for chemical and drug information. *Med. Ref. Serv. Q.* **2002**, *21*, 53–59. [CrossRef] [PubMed]

24. Roy, K.; Kar, S. In Silico Models for Ecotoxity of Pharmaceuticals. *Methods Mol. Biol.* **2016**, *1425*, 237–304.

25. Kar, S.; Leszczynski, J. Exploration of Computational Approaches to Predict the Toxicity of Chemical Mixtures. *Toxics* **2019**, *7*, 15. [CrossRef]

26. Prival, M.J. Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. *Environ. Mol. Mutagen.* **2001**, *37*, 55–69. [CrossRef]

27. Mombelli, E. An evaluation of the predictive ability of the QSAR software packages, DEREK, HAZARDEXPERT and TOPKAT, to describe chemically-induced skin irritation. *Altern. Lab. Anim.* **2008**, *36*, 15–24. [CrossRef]

28. Lewi, D.F.; Bird, M.G.; Jacobs, M.N. Human carcinogens: An evaluation study via the COMPACT and HazardExpert procedures. *Hum. Exp. Toxicol.* **2002**, *21*, 115–122.

29. Morita, T.; Shigeta, Y.; Kawamura, T.; Fujita, Y.; Honda, H.; Honma, M. In silico prediction of chromosome damage: Comparison of three (Q)SAR models. *Mutagenesis* **2019**, *34*, 91–100. [CrossRef]

30. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021, new data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef]

31. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]

32. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef]

33. Probst, D.; Reymond, J.L. A probabilistic molecular fingerprint for big data settings. *J. Cheminform.* **2018**, *10*, 66. [CrossRef]

34. Brown, B.P.; Vu, O.; Geanes, A.R.; Kothiwale, S.; Butkiewicz, M.; Lowe, E.W., Jr.; Mueller, R.; Pape, R.; Mendenhall, J.; Meiler, J. Introduction to the BioChemical Library (BCL): An Application-Based Open-Source Toolkit for Integrated Cheminformatics and Machine Learning in Computer-Aided Drug Discovery. *Front. Pharmacol.* **2022**, *13*, 833099. [CrossRef]

35. Gobbi, A.; Poppinger, D. Genetic optimization of combinatorial libraries. *Biotechnol. Bioeng.* **1998**, *61*, 47–54. [CrossRef]

36. Stiefl, N.; Watson, I.A.; Baumann, K.; Zaliani, A. ErG: 2D pharmacophore descriptions for scaffold hopping. *J. Chem. Inf. Model.* **2006**, *46*, 208–220. [CrossRef]

37. Armstrong, M.S.; Finn, P.W.; Morris, G.M.; Richards, W.G. Improving the accuracy of ultrafast ligand-based screening: Incorporating lipophilicity into ElectroShape as an extra dimension. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 785–790. [CrossRef]

38. Bragina, M.E.; Daina, A.; Perez, M.A.S.; Michielin, O.; Zoete, V. The SwissSimilarity 2021 Web Tool: Novel Chemical Libraries and Additional Methods for an Enhanced Ligand-Based Virtual Screening Experience. *Int. J. Mol. Sci.* **2022**, *23*, 811. [CrossRef]

39. Axen, S.D.; Huang, X.P.; Caceres, E.L.; Gendelev, L.; Roth, B.L.; Keiser, M.J. A Simple Representation of Three-Dimensional Molecular Structure. *J. Med. Chem.* **2017**, *60*, 7393–7409. [CrossRef]

40. Gfeller, D.; Michielin, O.; Zoete, V. Shaping the interaction landscape of bioactive molecules. *Bioinformatics* **2013**, *29*, 3073–3079. [CrossRef]

41. Choi, Y.H.; Suh, J.H.; Lee, J.H.; Cho, I.H.; Lee, M.G. Effects of tesmilifene, a substrate of CYP3A and an inhibitor of P-glycoprotein, on the pharmacokinetics of intravenous and oral docetaxel in rats. *J. Pharm. Pharmacol.* **2010**, *62*, 1084–1088. [CrossRef]

42. Van Waterschoot, R.A.; Lagas, J.S.; Wagenaar, E.; van der Kruijssen, C.M.; van Herwaarden, A.E.; Song, J.Y.; Rooswinkel, R.W.; van Tellingen, O.; Rosing, H.; Beijnen, J.H.; et al. Absence of both cytochrome P450 3A and P-glycoprotein dramatically increases docetaxel oral bioavailability and risk of intestinal toxicity. *Cancer Res.* **2009**, *69*, 8996–9002. [CrossRef]

43. Van Zuylen, L.; Verweij, J.; Nooter, K.; Brouwer, E.; Stoter, G.; Sparreboom, A. Role of intestinal P-glycoprotein in the plasma and fecal disposition of docetaxel in humans. *Clin. Cancer Res.* **2000**, *6*, 2598–2603.

44. National Comprehensive Cancer Network. NCCN Guidelines Version 4.2022 Breast Cancer. Available online: https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf (accessed on 29 June 2022).

45. Mayr, F.; Moller, G.; Garscha, U.; Fischer, J.; Rodriguez Castano, P.; Inderbinen, S.G.; Temml, V.; Waltenberger, B.; Schwaiger, S.; Hartmann, R.W.; et al. Finding New Molecular Targets of Familiar Natural Products Using In Silico Target Prediction. *Int. J. Mol. Sci.* **2020**, *21*, 7102. [CrossRef]

46. Bender, A.; Glen, R.C. Molecular similarity: A key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218. [CrossRef]

47. Yadav, J.; Paragas, E.; Korzekwa, K.; Nagar, S. Time-dependent enzyme inactivation: Numerical analyses of in vitro data and prediction of drug-drug interactions. *Pharmacol. Ther.* **2020**, *206*, 107449. [CrossRef]

48. Nozaki, Y.; Izumi, S. Recent advances in preclinical in vitro approaches towards quantitative prediction of hepatic clearance and drug-drug interactions involving organic anion transporting polypeptide (OATP) 1B transporters. *Drug Metab. Pharmacokinet.* **2020**, *35*, 56–70. [CrossRef]

49. Ryu, J.Y.; Kim, H.U.; Lee, S.Y. Deep learning improves prediction of drug-drug and drug-food interactions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4304–E4311. [CrossRef]

50. Nickel, J.; Gohlke, B.O.; Erehman, J.; Banerjee, P.; Rong, W.W.; Goede, A.; Dunkel, M.; Preissner, R. SuperPred: Update on drug classification and target prediction. *Nucleic Acids Res.* **2014**, *42*, W26–W31. [CrossRef]

51. Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner, C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijer, M.B.; Matos, R.C.; Tran, T.B.; et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181. [CrossRef]

52. D'Abreo, N.; Adams, S. Immune-checkpoint inhibition for metastatic triple-negative breast cancer: Safety first? *Nat. Rev. Clin. Oncol.* **2019**, *16*, 399–400. [CrossRef] [PubMed]

53. Sternschuss, M.; Yerushalmi, R.; Saleh, R.R.; Amir, E.; Goldvaser, H. Efficacy and safety of neoadjuvant immune checkpoint inhibitors in early-stage triple-negative breast cancer: A systematic review and meta-analysis. *J. Cancer Res. Clin. Oncol.* **2021**, *147*, 3369–3379. [CrossRef] [PubMed]

54. Ru, J.; Li, P.; Wang, J.; Zhou, W.; Li, B.; Huang, C.; Li, P.; Guo, Z.; Tao, W.; Yang, Y.; et al. TCMSP: A database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* **2014**, *6*, 13. [CrossRef] [PubMed]

55. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef]

56. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]

57. Zhang, F.; Liu, Q.; Ganesan, K.; Kewu, Z.; Shen, J.; Gang, F.; Luo, X.; Chen, J. The Antitriple Negative Breast cancer Efficacy of Spatholobus suberectus Dunn on ROS-Induced Noncanonical Inflammasome Pyroptotic Pathway. *Oxid. Med. Cell. Longev.* **2021**, *2021*, 5187569. [CrossRef]

58. Chou, T.-C. Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacol. Rev.* **2006**, *58*, 621–681. [CrossRef]

59. Zheng, S.; Wang, W.; Aldahdooh, J.; Malyutina, A.; Shadbahr, T.; Tanoli, Z.; Pessia, A.; Tang, J. SynergyFinder Plus: Toward Better Interpretation and Annotation of Drug Combination Screening Datasets. *Genom. Proteom. Bioinform.* **2022**, *in press*. [CrossRef]

60. Yadav, B.; Wennerberg, K.; Aittokallio, T.; Tang, J. Searching for Drug Synergy in Complex Dose-Response Landscapes Using an Interaction Potency Model. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 504–513. [CrossRef]