

Predicting new indications of compounds with a network pharmacology approach: Liuwei Dihuang Wan as a case study

Yin-Ying Wang^{1,2,3,*}, Hong Bai^{4,*}, Run-Zhi Zhang⁴, Hong Yan³, Kang Ning⁴ and Xing-Ming Zhao¹

¹Institute of Science and Technology for Brain-Inspired Intelligence (ISTBI), Fudan University, Shanghai 200433, China

²Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

³Department of Electronic Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong

⁴Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

*These authors have contributed equally to this work

Correspondence to: Xing-Ming Zhao, **email:** xmzhao@fudan.edu.cn

Keywords: network pharmacology; drug repurposing; TCMs; pathway profile; LDW

Received: February 13, 2017

Accepted: September 05, 2017

Published: September 30, 2017

Copyright: Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

With the ever increasing cost and time required for drug development, new strategies for drug development are highly demanded, whereas repurposing old drugs has attracted much attention in drug discovery. In this paper, we introduce a new network pharmacology approach, namely PINA, to predict potential novel indications of old drugs based on the molecular networks affected by drugs and associated with diseases. Benchmark results on FDA approved drugs have shown the superiority of PINA over traditional computational approaches in identifying new indications of old drugs. We further extend PINA to predict the novel indications of Traditional Chinese Medicines (TCMs) with Liuwei Dihuang Wan (LDW) as a case study. The predicted indications, including immune system disorders and tumor, are validated by expert knowledge and evidences from literature, demonstrating the effectiveness of our proposed computational approach.

INTRODUCTION

With the ever increasing cost and time-consuming process of drug development, new strategies for drug development are highly demanded. Drug repurposing, which aims for identifying novel indications for existing drugs, attracts a lot of attention since the toxicity of known drugs is already understood [1]. For example, Metformin has been widely used for more than 30 years for the treatment of type 2 diabetes, but extensive preclinical and clinical studies over the past decade have demonstrated the antitumor effects of the drug [2]. It has been reported that Metformin was able to lower the risk of cancer mortality and incidence in patients with diabetes [3]. Nowadays, drug repurposing has been considered as

an effective approach in drug development. However, identifying novel indications with drug repurposing is highly challenging since the novel indications of one drug may have no obvious relationship with its initial purpose.

During the past decade, much effort has been made to develop new computational approaches for the purpose of repositioning drugs and elucidating the molecular mechanisms of drugs. For example, Wang *et al* proposed a novel method to predict drug target proteins based on drug-domain interactions [4], and Zhang *et al* constructed a post-translational regulatory network to explore network motifs as potential drug targets which can help design multi-component or combinatorial drugs [5]. With the popular deep learning (DL) techniques, Kadurin *et al* proposed a DL-based model for screening

potential anti-cancer compounds [6]. Recently, the network pharmacology approaches have been widely employed for understanding the mechanisms of drug actions, resistance and side effects [7–9]. At the same time, some network pharmacology approaches have been proposed to predict the associations between drugs and diseases. For instance, Martinez *et al* developed DrugNet to prioritize drugs for certain diseases by integrating complex associations among disease, drugs and proteins [10]. Besides, Alaimo *et al* also introduced a method that can be used to integrate biological knowledge and bipartite interaction network to predict new indications of drugs [11].

As multi-target or multi-component therapies gain increasing attention recently, Traditional Chinese Medicines (TCMs) are being re-evaluated and becoming important resources for the discovery of alternative treatments for certain diseases, where various network pharmacology approaches have been proposed for this purpose [12–15]. For example, Qing Luo Yin (QLY) is an effective formula in the treatment of arthritis and antiangiogenic. With the target network of QLY, not only the diseases related key biological processes including angiogenesis, inflammatory and immune response were revealed, but also the active ingredients and synergistic combinations of this herbal formula were identified [16]. Another example is Liuwei Dihuang Wan (LDW), which shows potential for regulating the imbalance of hormones and metabolism [17]. Therefore, the network pharmacology approaches are capable of providing insights into the mechanisms of actions of known drugs and identifying new indications of those drugs [18–20]. However, current network pharmacology methods for repurposing drugs are mainly based on the target proteins of active compounds, whereas the target information may not be indicative of diseases that the drugs can be used for.

In this paper, we investigate the mechanisms of drug actions based on the pathways modulated by the drugs. By further integrating pathway profiles with chemical structures as well as disease phenotypes, we present a network pharmacology approach namely PINA (Predicting new Indications of compounds with a Network pharmacology Approach) as shown in Figure 1, to predict potential indications of old drugs. Benchmark results on FDA approved drugs have proven the superiority of our method over traditional network pharmacology approaches, as regard to revealing new associations between compounds and diseases. We further extend PINA to predict the novel indications of Traditional Chinese Medicines (TCMs) with Liuwei Dihuang Wan (LDW) as a case study. The predicted indications, including immune system disorders and tumor, are validated by expert knowledge and evidences from literature, demonstrating the effectiveness of our proposed computational approach.

RESULTS

Identification of the pathway profiles associated with diseases

In this work, given a disease, we assume that the compounds that can significantly affect the pathway profiles associated with the disease can be used for the disease. With 4,774 known drug-disease associations composed of 928 compounds and 608 diseases extracted from CTD database [21], we first identified the pathways that are dysfunctional in diseases. Assuming that diseases with similar pathway profiles should have similar mechanisms, based on the pathways we identified a disease-disease association network was constructed, where two diseases were linked if they shared at least one pathway. We further detected modules from the network with density-based MCODE [22] tool (Supplementary Figure 2). Table 1 listed the 14 modules and the corresponding average similarities among diseases within the module as well as the most enriched disease class. Supplementary Table 2 has shown the detailed information of the 14 modules. If the pathway profiles we identified are indeed associated with diseases, we expected that the diseases belonging to the same module should have similar mechanisms. It could be seen that the diseases grouped into the same module based on pathway profiles tended to have similar symptoms, where the disease similarity was calculated as described in [23]. Furthermore, the diseases can be grouped into 22 classes based on the physiological systems affected by the diseases as defined in [24]. By investigating the diseases belonging to same module, we found that the diseases in the same module tend to be in the same class as shown in Table 1, indicating that the diseases from the same module have similar mechanisms. In addition, by investigating the number of disease classes that the pathway profiles were associated with, we found that each of more than 79% pathways was associated with only one specific disease class, implying that each pathway profile is specifically associated with a certain type of diseases (Supplementary Figure 3).

By further investigating the pathway profiles that were associated with one disease class, we found that those pathways were indeed related to the disease class. For instance, the calcium signaling pathway played a crucial role in the control of neuronal functions and plasticity by regulating members of the neuronal calcium sensor (NCS) proteins [25]. It was reported that the deregulation of calcium signaling pathway was one of the key processes in the pathogenesis of neurodegenerative disorders [26]. In our study, the neurological class consists of 48 diseases while the calcium signaling pathway was predicted to be related with 18 out of them. Besides, the transmission across chemical synapses pathway we identified was related to more than 20% of neurological diseases, where the chemical synapses were specialized

junctions used for communications between neuron [27]. Furthermore, the GPCR ligand binding pathway was predicted to be associated with all psychiatric diseases, where the G protein-coupled receptors have been found to play important roles in major psychiatric disorders, such as depression and schizophrenia [28].

From the findings shown above, we can see that the pathway profiles identified here are indeed related to the corresponding diseases.

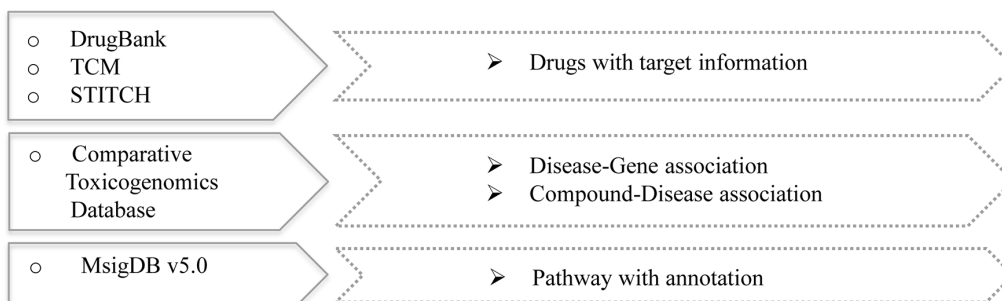
Prediction of potential indications for FDA approved drugs

With the pathway profiles identified above, the potential associations between compounds and diseases could be predicted. Based on the FDA approved drugs with target information and their known associations with diseases obtained from the CTD database, PIPP, NP_c

and NP_d were respectively applied to predict potential compound-disease associations.

By comparing the three approaches, we noticed that many of the predictions by pathway profile approach (PIPP) could be validated with those predicted by chemical structures and disease similarities based on the ‘guilt by association’ rule, where drugs with similar structures were assumed to be able to treat the same disease while similar diseases could be treated with the same drug. For example, the compound Nortriptyline (CID: 4543) was originally used as an anti-depressive agent [2], and it was predicted for the treatment of schizophrenia (OMIM: 603176) with a score of 0.9985 by PIPP. In fact, the drug Nortriptyline had similar structure with Amitriptyline (CID: 2160), which was used for schizophrenia [9], with a similarity score of 0.92. On the other hand, schizophrenia was similar with Attention Deficit Hyperactivity (ADH) disorder (OMIM: 143465), and Nortriptyline have already been reported for

A Data sources



B Predicting new Indications of compounds with a Network pharmacology Approach (PINA)

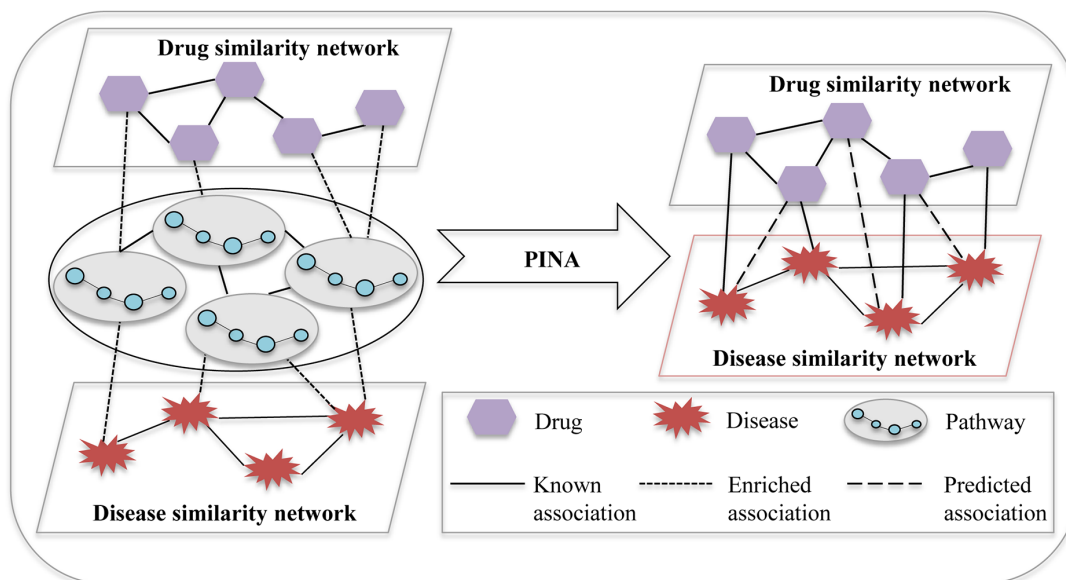


Figure 1: The pipeline of predicting new indications of compounds with a network pharmacology approach. (A) Data sources for network pharmacology analysis; **(B)** Predicting new indications of compounds with a network pharmacology approach. Node: irregular, disease; hexagon, compound; circle, gene. Line: solid line, known association; square dot, enriched association; long dash, predicted association.

Table 1: The modules detected by MCODE from the disease association network generated with disease related pathway profiles

Modules	Number of diseases	Average similarity [#]	Disease class (Coverage) [*]
1	16	0.1623	Psychiatric (0.625)
2	9	0.3210	Neurological (0.67)
3	7	0.2448	Ophthalmological (0.71)
4	5	0.3294	Connective tissue (0.6)
5	5	0.6367	Cardiovascular (1.0)
6	4	0.4166	Endocrine (1.0)
7	4	0.2861	Neurological (0.5), Metabolic (0.5)
8	4	0.2427	Neurological (0.5), Cancer (0.5)
9	3	0.5193	Neurological (1.0)
10	3	0.3635	Bone (1.0)
11	3	0.2281	Immunological (1.0)
12	3	0.4612	Metabolic (1.0)
13	3	0.4284	Multiple (1.0)
14	3	0.5407	Gastrointestinal (1.0)

[#]Average similarity means the average of similarities over all disease pairs in each module detected in disease associated network.

^{*}Coverage means the number of the diseases in the most enriched class divided by the number of disease in the module.

treating ADH in the CTD database, which validated that Nortriptyline could be used for schizophrenia. Moreover, we noticed that the pathway profile approach could successfully recover known associations that were missed by the chemical or disease similarity based approach. For example, the compound Retinoic Acid was used for femur head necrosis, which was successfully identified by our pathway profile method with a score of 1.0. However, the nearest profile approach based on chemical and disease similarity failed to identify this association with scores of 0.0 and 0.27, respectively.

The results shown above demonstrate that the pathway profile approach can complement with other approaches, e.g. chemical or disease similarity based ones, very well. Therefore, we further proposed an ensemble approach named as PINA that combines the pathway profile method with chemical and disease similarity based methods to predict potential compound-disease associations. The novel potential indications of all compounds are list in Supplementary Table 4. We also compared PINA with three existing methods from literature, including DrugNet [10], HGBI [29] and NBI [30]. DrugNet is a network-based drug repositioning method, which integrates the information of diseases, drugs and proteins to prioritize drug-disease associations. HGBI and NBI have been originally developed for predicting drug-protein interactions, and can also be used for the prediction of drug-disease associations. HGBI predicts the drug-disease associations with the guilt-by-association principle based on the drug-disease heterogeneous graph, while NBI can

predict new drug-disease associations based on a two-step diffusion model on a drug-disease bipartite graph. To evaluate the performance of our approach, PINA was compared with the other three approaches on the same benchmark drug-disease associations from the Comparative Toxicogenomics Database, where the same pre-process was used for all the four computational approaches. The chemical similarities between compounds were calculated based on their fingerprints by using the Single Linkage algorithm [31] while the disease similarities were defined as described in [23]. All the four approaches were evaluated with 5-fold cross-validations. Table 2 shows the performances of different methods, from which we can see that PINA has the highest AUC (0.8969) and F1 (0.3833) and significantly outperforms the other approaches.

Prediction of potential indications for LDW

In this part, we further extended PINA to predict the novel indications of Traditional Chinese Medicines (TCMs) with Liuwei Dihuang Wan (LDW) as a case study. With the known compound-disease associations from the CTD database, we built a model as described in Equation (4) and identified 59 diseases that LDW can be used for. Among the 156 compound components of LDW, only the eight compounds that can be found new indications with PINA were considered here, where the eight compounds were further required to be drug-like. Table 3 shows the detailed information about the eight compounds. By

Table 2: The performances of different methods which were obtained with 5-fold cross-validation

Method	AUC	Precision	Recall	F1 score
<i>PIPP</i>	0.8515	0.1517	0.4899	0.2313
<i>NP_c</i>	0.8132	0.0873	0.6511	0.1539
<i>NP_d</i>	0.8633	0.3005	0.4760	0.3684
<i>PINA</i>	0.8969	0.4325	0.3446	0.3833
<i>DrugNet</i>	0.8034	0.3411	0.3923	0.3568
<i>HGBI</i>	0.8125	0.3867	0.3639	0.3752
<i>NBI</i>	0.7983	0.3297	0.3321	0.3308

AUC - Area under ROC curve;

Precision - TP/(TP+FP), positive predictive value;

Recall - TP/(TP+FN), true positive rate;

F1 score - Harmonic mean of precision and recall.

Table 3: The detail information about eight compounds belonging to LDW

Compound ID	Name	FDA Status	Part of known indications obtained from CTD database
CID000445354	Retinol	Approved	Acne Vulgaris; Acute Kidney Injury; Adrenal Insufficiency; Carcinoma, Hepatocellular; Colonic Neoplasms; Diabetes Mellitus, Type 1; Fatty Liver; Hypertension, Portal; Hypertriglyceridemia; Liver Cirrhosis; Nephrosis;
CID000024360	Camptothecin	Experimental	Neoplasms; Leukemia, Lymphoid
CID027237R1936	Nicotinamide	Experimental	Diabetes Mellitus, Type 2; Hypercholesterolemia; Hyperglycemia; Hyperlipoproteinemias; Hypertension; Hypertriglyceridemia; Kidney Diseases; Nerve Degeneration; Ventricular Dysfunction, Left;
CID000006137	L-methionine	Approved	Carcinoma, Hepatocellular; Fatty Liver; Kidney Diseases; Memory Disorders;
CID005280343	Quercetin	Experimental	Acute Kidney Injury; Autoimmune Diseases; Breast Neoplasms; Cognition Disorders; Diabetes Mellitus; Fatty Liver; Hypertension; Kidney Diseases; Memory Disorders; Ovarian Neoplasms; Prostatic Neoplasms;
CID000005641	Urethane	Withdrawn	Arrhythmias, Cardiac; Hypertension; Liver Neoplasms; Ovarian Neoplasms;
CID027237R1305	Choline	Approved	Amnesia; Cognition Disorders; Fatty Liver; Memory Disorders;
CID027237R1681	Dopamine	Approved	Acute Kidney Injury; Arrhythmias, Cardiac; Central Nervous System Diseases; Heart Failure; Hypertension; Learning Disorders; Nerve Degeneration; Nervous System Diseases; Parkinson Disease;

investigating the indications of the eight compounds obtained from CTD, we found that LDW, as a mixture of multiple compounds, achieves its therapeutic effects through its individual components. For example, LDW was used for anti-aging, delayed development and blurred vision, whereas Retinol, also known as vitamin A, plays an essential role in anti-aging, promoting bone growth and the treatment of various eye conditions. Moreover, it was found that LDW was useful for decreasing blood sugar, suppressing blood pressure and improving the renal function. Another compound component Quercetin, an antioxidant, was reported to treat many LDW associated disease, such as acute kidney injury, diabetes mellitus

and hypertension. The combination of Nicotinamide and Retinol could be effective for acne treatment for which LDW has been used for [32].

With the findings above, we assumed that the new indications we predicted for LDW can be validated with the indications of its component compounds. Figure 2 shows a compound-disease association network constructed with the associations between diseases and compounds we have predicted, where the 59 diseases that LDW has been predicted to be used for were linked to the eight individual compounds based on PINA. Among the 104 compound-disease associations shown in Figure 2, 21 of them have been reported in CTD. For example,

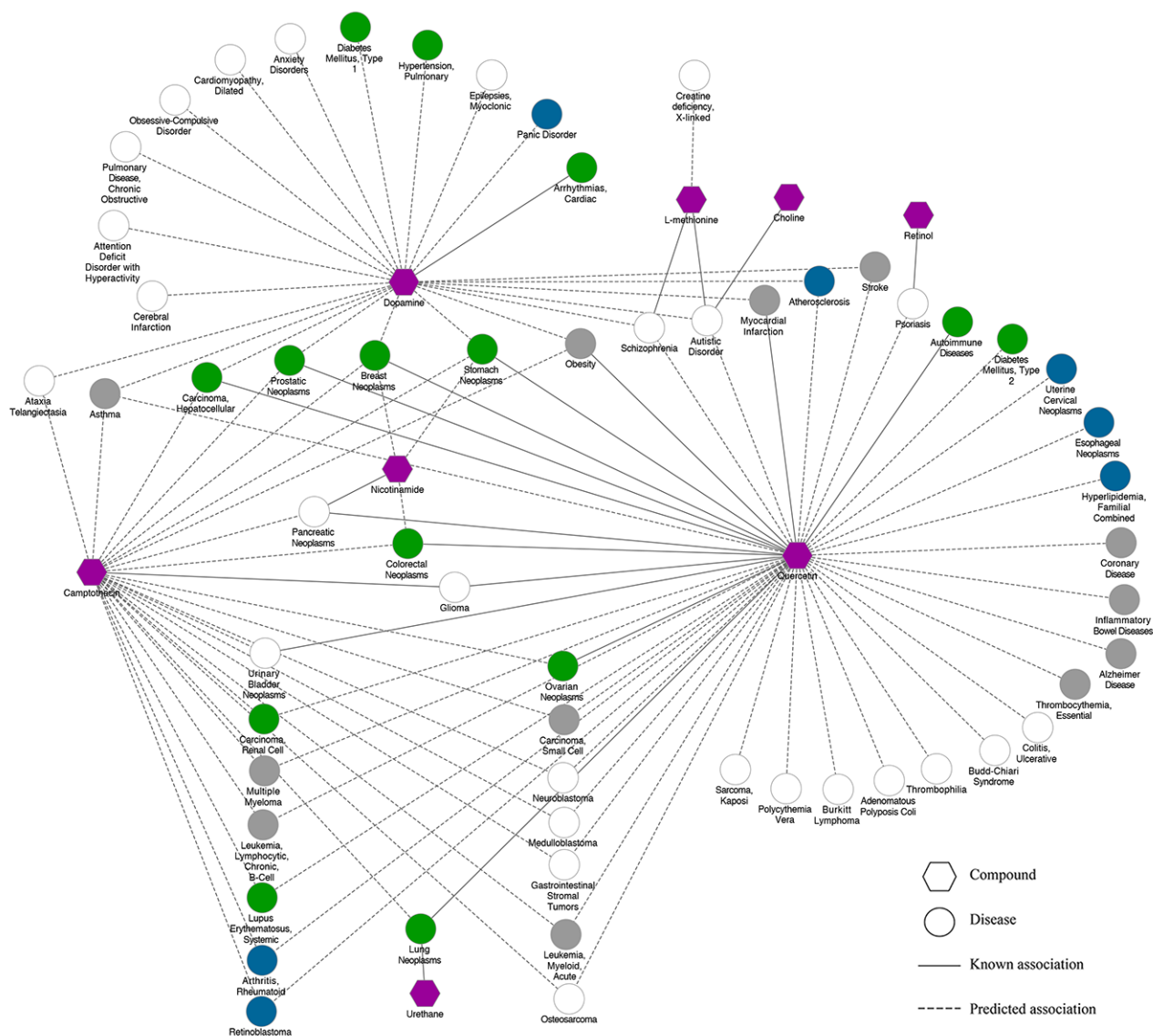


Figure 2: The LDW associated compound-disease network. Node: green circles, diseases for which LDW was known to be used for (Number=14); blue circles, diseases reported in Li *et al* that LDW can be used for (Number=7); grey circles, diseases reported in literature that LDW can be used for (Number=12); white circles, diseases predicted to be treated with LDW (Number=26); purple hexagons, compounds. Line: solid lines, known associations between compounds and diseases; dotted lines, predicted associations between compounds and diseases.

Quercetin could inhibit the growth of MCF-7 breast cancer cell line and promoted apoptosis by reducing G0/G1 phase arrest [33, 34]. Besides, it was widely accepted that the compound could be used to treat a certain disease if it targeted the disease related genes. With the known target and disease genes information, we also found there are another 27 predicted compound-disease associations which can be validated by sharing same genes. For instance, dopamine D2 receptor (DRD2) played an essential role in dopamine signaling which was strongly implicated in the etiology of schizophrenia (SZ) [35], and was also one of the targets of Dopamine. By targeting the gene DRD2, Dopamine may be used for schizophrenia. As a result, 48 compound-disease associations can be validated by difference evidence while the rest associations need further experimental validation.

By considering the indications (59 diseases) we predicted for LDW, we further investigated whether LDW has been reported to be effective for some of these diseases in literature by expert knowledge. As a result, 14 of them have been known to be associated with LDW as shown in Supplementary Table 3. For example, it was found that LDW could significantly inhibit the breast cancer tumor growth and progression, and promoted the recovery of breast ducts in mice [36, 37]. Likewise, LDW decoction could exert therapeutic effects on liver cancer in mice by affecting tumor cell cycle and down-regulate serum VEGF level [38]. Moreover, it was well accepted that LDW could counteract the adverse effect of steroid and immunosuppressive agents, significantly improving the therapeutic effectiveness in the treatment of Systemic Lupus Erythematosus (SLE) [39]. Besides, many other predicted diseases, i.e. diabetes mellitus, hypertension and so on [40–43], were also known to be associated with LDW.

Previously, Li *et al* predicted new indications for LDW based on drug targets and disease genes [17]. We further investigated how many of our predictions could be validated by those reported in their work, and the new indications found by both works for LDW would be more convincing. Consequently, 7 of our predictions were also reported by Li *et al*, including atherosclerosis, retinoblastoma, rheumatoid arthritis, esophageal neoplasms, uterine cervical neoplasms, familial combined hyperlipidemia and panic disorder. For example, it was found that LDW had already been reported for treating esophageal neoplasms [44]. In addition, studies have shown that LDW pills could effectively inhibit the expression of IL-beta, MMP-1 and MMP-3, thus protecting and repairing the articular cartilage which had significant therapeutic effects on Osteoarthritis [45].

Moreover, we also performed text mining by querying the PubMed database to see whether LDW have been reported effective for the rest of our predictions. As a result, except the diseases mentioned above, 12 diseases have been reported to be treated by LDW in literature as listed in Supplementary Table 3. For instance, it was found

that LDW could simultaneously disturb the regulations of apoptosis and protein ubiquitination among biological processes, such as RPS6KA1, FHIT and AMFR, which may be the therapeutic targets of Alzheimer Disease [46, 47]. Moreover, traditional Chinese doctors have already used LDW to treat asthma patients based on the cytokine gene expression perturbed by LDW [48].

Taken together, 33 out of 59 diseases we predicted to be treated by LDW have been validated in different ways, where the known indications with direct evidences tend to rank top. These results demonstrate that LDW can really work for those diseases. The detailed results with corresponding evidences were presented in Supplementary Table 3.

DISCUSSION

Repurposing old drugs has drawn increasing attention, since they could serve as the effective and cost-saving strategies for drug discovery. In this study, we first introduce pathway profiles associated with diseases and affected by compounds. By integrating the pathway profiles with chemical structure as well as disease phenotype, we present PINA to predict new indications of compounds. Benchmark results on FDA approved drugs have demonstrated the predictive power of PINA. We further extended PINA to predict the potential indications of traditional Chinese medicine with LDW as a case study. The new indications we predicted for LDW have been validated with expert knowledge and evidences from literature.

We also noticed that improvement of our PINA approach is possible when predicting novel indications of TCMs. For example, a TCM formula is typically composed of multiple herbs or hundreds of chemical compounds. Here, the indications of a TCM formula were predicted with a Bayesian model, where the compound components were regarded to be independent with each other. Although the synergistic effects among compounds cannot be explicitly described in the Bayesian model, the good performance on LDW shows the effectiveness of the model. In the future, more efficient models should be developed to take into account the synergistic effects among compounds. Another concern is that many compound components of TCMs are not known while it is expensive and time-consuming to determine all bioactive compounds of TCMs, a comprehensive knowledgebase about compound components of TCMs is highly demanded.

MATERIALS AND METHODS

Data sources

The FDA approved human drugs used in our study were retrieved from the DrugBank database (Version 4.3) [49], of which we only focused on the 932 compounds

that had target information according to the DrugBank and STITCH databases (Version 4) [50] which provides a confidence score for each interaction. Here, a score of 700 was used as threshold to choose the high-confidence interactions [51]. Specifically, the interactions marked with prediction or text mining were removed to make sure high-quality interactions used in this paper. The LDW was composed of *Rehmannia glutinosa Libosch.*, *Cornus officinalis Sieb. et Zucc.*, *Paeonia suffruticosa Andr.*, *Dioscorea opposita Thunb.*, *Poria cocos (Schw.) Wolf* and *Alisma orientalis (Sam.) Juzep*. In our work, the chemical constituents of LDW were mainly obtained from the TCM Database@Taiwan [52] by searching the herb names. Meanwhile, the other constituents were also collected manually from published articles by text mining. Then we transformed all constituents into mol2 format with ChemDraw software (<http://www.cambridgesoft.com/software/ChemDraw/>), and the chemicals were then converted into the canonical SMILES format. We downloaded all known chemicals with each of them annotated with PubChem identity from STITCH database (version 4.0). By querying the known compounds with the chemical SMILES files, the chemical constituents of LDW can be identified. Here, we only picked up the chemicals that had target information according to DrugBank and STITCH databases. Consequently, 156 compounds of LDW (Supplementary Table 1) were finally collected.

The disease-gene associations were obtained from the Comparative Toxicogenomics Database (CTD) [21]. As a result, the associations between 4937 diseases and 8536 genes were collected. We further collected compound-disease associations from the CTD database, and the 4774 associations with direct evidence (therapeutic/maker) between 928 compounds and 608 diseases were used as the positive set while the other possible compound-disease associations were used as the negative set.

All predefined biological pathways used in this study were obtained from the Molecular Signatures Database (Version 5.0) [17], where the canonical pathways from the curated (c2) gene sets were adopted. The physical protein-protein interactions were obtained from HPRD [53], BioGRID [54], IntAct [55], MINT [56] and DIP [57] databases.

Predicting new indications of compounds with a network pharmacology approach

Predicting indications of compounds based on pathway profile

We assumed that the occurrence of a disease was due to the aberrant functions of certain pathways. Accordingly, to treat a disease, the drugs should affect the dysfunctional pathways that were associated with the disease. With this assumption, for each drug and its related disease, the pathways linking the pair of drug

and disease were firstly identified. For this, the pathway profiles associated with a drug and a disease were respectively identified, where the drug related pathways were enriched by its target proteins while the disease associated pathways were enriched by its related genes [58]. Given a pair of pathways respectively associated with a drug and a disease, we only considered the pathways that met one of the following conditions: (1) the two pathways are the same one (common pathway); (2) the two pathways share at least one gene (cross-talking pathways); (3) there are protein interactions between the two pathways (interacting pathways) (As shown in Supplementary Figure 1). To avoid possible false positives, the cross-talking or interacting pathways were required to have correlated activities based on the gene expression data obtained from 36 normal tissues [59]. Here, the pathway activity in a tissue was defined as the average expression value of all genes within the pathway and only the pathway pairs with a significant correlation coefficient (p-value <0.01) in 36 tissues were kept for further analysis.

Here, the pathway profile method named PIPP (predicting indications based on pathways profile) were proposed. Given one disease and related drugs as well as the pathways associated with any pair of drug and disease as defined above, the score of a pathway pair associating a drug with the disease it could be used for was defined as follows:

$$P(p_i - D) = \frac{N(C | p_i)}{N(C' | D)} \quad (1)$$

where $N(C | p_i)$ is the number of compounds treating disease D and the pathway pair p_i is the one that occurs commonly between compound set C and the disease D , and $N(C' | D)$ is the number of all compounds used for disease D . If $P(p_i - D)$ is above a certain threshold, the pathway pair p_i will be regarded as the pathway profiles for associating a disease with the drugs treating the disease.

Given a new drug, the score of the drug used for treating the disease D can be defined as below:

$$P(C_i - D) = 1 - \sum_{m=1}^3 \alpha_m \prod (1 - P(p_m - D)) \quad (2)$$

where m represents one of the three types of pathway profiles, i.e. common, cross-talking and interacting pathway(s), $P(p_m - D)$ is the score of the m th type of pathway profiles associated with disease D , and α_m is the weight for the m th type of pathway profiles. To determine the weights for the three types of pathway profiles, the 5-fold cross validation was employed and the AUC was used to choose the proper values. As a result, the weights in Equation (2) were determined as: $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.2$, where the best results were obtained.

Predicting new indications of compounds based on the nearest neighbor profile

It has been found that similar drugs tend to have similar mechanism and can be used to treat similar diseases, and *vice versa* [24]. Therefore, given a new drug, the new indications of the drug can be predicted based on its similarity with other drugs. Here, the nearest neighbor profile approach proposed by Yamanishi *et al* [26], *i.e.* nearest profile based on chemical similarity and nearest profile based on disease similarity we named as NP_{-C} and NP_{-D}, was adopted to predict whether a new drug could be used for a certain disease. The chemical similarity between compounds is calculated based on their fingerprints by using the Single Linkage algorithm [31]. The disease similarities are defined in [23], where the similarity was calculated based on disease descriptions from the OMIM database [60].

Predicting new indications of compounds based on an ensemble method

The three independent methods mentioned above, *i.e.* PIPP, NP_{-C} and NP_{-D}, showed different performance on different datasets. Here, we further proposed an ensemble approach named PINA to predict the compound-disease associations by integrating the pathway profile, chemical similarity and disease similarity. In particular, a weight was set for each method based on its performance on a benchmark dataset, and the ensemble learner was constructed as follows:

$$P(C_D) = \sum_{i=1}^n W_i \cdot M_i \quad (3)$$

where w_i is the weight for each method, and M_i is the output of the i th method. Here, the weight for each method is set to the AUC (area under the curve) score of a receiver operating characteristic (ROC) curve. For a given compound, we can use the ensemble approach to predict whether the drug can be used for the disease.

Predicting new indications of LDW

To evaluate the performance of our proposed approach, the PINA method was applied to infer the therapeutic indications of TCM and investigate the curative effect between TCM and its individual components. To this end, we chose the LDW as a case study since its chemical constituents and indications were well known. Subsequently, we proposed PINA to the 156 chemical constituents to predict compound-disease associations, where a score was calculated based on Equation (3) as the confidence score of the prediction. To determine the threshold above which a prediction is regarded as positive, the 5-fold cross-validation was employed on the known drug-disease associations, *i.e.* training set. Especially, the threshold that can lead to the highest F1 score was chosen, where the F1 score can evaluate the overall performance of the learner and is a tradeoff between Precision and Recall.

Here, the threshold of 0.6 that can lead to the highest F1 score in the cross-validation was chosen. Then we defined an efficacy score for LDW to a certain disease by considering the synergistic effect of all compounds based on Bayesian models. The efficacy score could be described as follows:

$$P(LWDH_D) = 1 - \prod_{C_i \in LWDHW} (1 - P(C_i_D)) \quad (4)$$

where C_i is the component of LDW. $P(C_i_D)$ is the association score between compound C_i and disease D calculated with Equation (3).

Author contributions

Y.Y.W. and X.M.Z. conceived and designed the study. Y.Y.W. and R.Z. conducted the data analysis. Y.Y.W. and K.N. drafted the manuscript. X.M.Z., H.Y. and H.B. revised the manuscript. All authors contributed to writing and finalizing the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was partly supported by National Nature Science Foundation of China (61772368, 61572363, 91530321, 61602347, 81573702 and 31671374), Ministry of Science and Technology's high-tech (863) grant (2014AA021502), Sino-German Research Center grant (GZ878), and China Postdoctoral Science Foundation Funded Project (2016M601647), and City University of Hong Kong (Project 7004707).

REFERENCES

1. Jin G, Wong ST. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today*. 2014; 19:637–44. <https://doi.org/10.1016/j.drudis.2013.11.005>.
2. Leone A, Di Gennaro E, Bruzzese F, Avallone A, Budillon A. New perspective for an old antidiabetic drug: metformin as anticancer agent. *Cancer Treat Res*. 2014; 159:355–76. https://doi.org/10.1007/978-3-642-38007-5_21.
3. Landman GW, Kleefstra N, van Hateren KJ, Groenier KH, Gans RO, Bilo HJ. Metformin associated with lower cancer mortality in type 2 diabetes: ZODIAC-16. *Diabetes Care*. 2010; 33:322–26. <https://doi.org/10.2337/dc09-1380>.
4. Wang YY, Nacher JC, Zhao XM. Predicting drug targets based on protein domains. *Mol Biosyst*. 2012; 8:1528–34. <https://doi.org/10.1039/c2mb05450g>.
5. Zhang XD, Song J, Bork P, Zhao XM. The exploration of network motifs as potential drug targets from

- post-translational regulatory networks. *Sci Rep.* 2016; 6:20558. <https://doi.org/10.1038/srep20558>.
6. Kadurin A, Aliper A, Kazennov A, Mamoshina P, Vanhaelen Q, Khrabrov K, Zhavoronkov A. The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget.* 2017; 8:10883–90. <https://doi.org/10.18632/oncotarget.14073>.
 7. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol.* 2012; 52:505–21. <https://doi.org/10.1146/annurev-pharmtox-010611-134520>.
 8. Qin G, Zhao XM. A survey on computational approaches to identifying disease biomarkers based on molecular networks. *J Theor Biol.* 2014; 362:9–16. <https://doi.org/10.1016/j.jtbi.2014.06.007>.
 9. Wist AD, Berger SI, Iyengar R. Systems pharmacology and genome medicine: a future perspective. *Genome Med.* 2009; 1:11. <https://doi.org/10.1186/gm11>.
 10. Martínez V, Navarro C, Cano C, Fajardo W, Blanco A. DrugNet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif Intell Med.* 2015; 63:41–49. <https://doi.org/10.1016/j.artmed.2014.11.003>.
 11. Alaimo S, Pulvirenti A, Giugno R, Ferro A. Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics.* 2013; 29:2004–08. <https://doi.org/10.1093/bioinformatics/btt307>.
 12. Li S, Zhang B, Zhang NB. Network target for screening synergistic drug combinations with application to traditional Chinese medicine. *BMC Syst Biol.* 2011; 5 Suppl 1:S10. <https://doi.org/10.1186/1752-0509-5-S1-S10>.
 13. Li S, Zhang B. Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chin J Nat Med.* 2013; 11:110–20. [https://doi.org/10.1016/S1875-5364\(13\)60037-0](https://doi.org/10.1016/S1875-5364(13)60037-0).
 14. Li S, Fan TP, Jia W, Lu A, Zhang W. Network pharmacology in traditional chinese medicine. *Evid Based Complement Alternat Med.* 2014; 2014:138460. <http://dx.doi.org/10.1155/2014/138460>.
 15. Zhao J, Jiang P, Zhang W. Molecular networks for the study of TCM pharmacology. *Brief Bioinform.* 2010; 11:417–30. <https://doi.org/10.1093/bib/bbp063>.
 16. Zhang B, Wang X, Li S. An Integrative Platform of TCM Network Pharmacology and Its Application on a Herbal Formula, Qing-Luo-Yin. *Evid Based Complement Alternat Med.* 2013; 2013:456747. <https://doi.org/10.1155/2013/456747>.
 17. Liang X, Li H, Li S. A novel network pharmacology approach to analyse traditional herbal formulae: the Liu-Wei-Di-Huang pill as a case study. *Mol Biosyst.* 2014; 10:1014–22. <https://doi.org/10.1039/C3MB70507B>.
 18. Wei S, Niu M, Wang J, Wang J, Su H, Luo S, Zhang X, Guo Y, Liu L, Liu F, Zhao Q, Chen H, Xiao X, et al. A network pharmacology approach to discover active compounds and action mechanisms of San-Cao Granule for treatment of liver fibrosis. *Drug Des Devel Ther.* 2016; 10:733–43. <https://doi.org/10.2147/DDDT.S96964>.
 19. Li X, Wu L, Liu W, Jin Y, Chen Q, Wang L, Fan X, Li Z, Cheng Y. A network pharmacology study of Chinese medicine QiShenYiQi to reveal its underlying multi-compound, multi-target, multi-pathway mode of action. *PLoS One.* 2014; 9:e95004. <https://doi.org/10.1371/journal.pone.0095004>.
 20. Lee JH, Zhao XM, Yoon I, Lee JY, Kwon NH, Wang YY, Lee KM, Lee MJ, Kim J, Moon HG, In Y, Hao JK, Park KM, et al. Integrative analysis of mutational and transcriptional profiles reveals driver mutations of metastatic breast cancers. *Cell Discov.* 2016; 2:16025. <https://doi.org/10.1038/celldisc.2016.25>.
 21. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Wieggers TC, Mattingly CJ. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.* 2015; 43:D914–20. <https://doi.org/10.1093/nar/gku935>.
 22. Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J Cardiovasc Magn Reson.* 2012; 14:83. <https://doi.org/10.1186/1532-429X-14-83>.
 23. van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome. *Eur J Hum Genet.* 2006; 14:535–42. <https://doi.org/10.1038/sj.ejhg.5201585>.
 24. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci USA.* 2007; 104:8685–90. <https://doi.org/10.1073/pnas.0701361104>.
 25. Arrell DK, Terzic A. Network systems biology for drug discovery. *Clin Pharmacol Ther.* 2010; 88:120–25. <https://doi.org/10.1038/clpt.2010.91>.
 26. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics.* 2008; 24:i232–40. <https://doi.org/10.1093/bioinformatics/btn162>.
 27. Roerig B, Feller MB. Neurotransmitters and gap junctions in developing neural circuits. *Brain Res Brain Res Rev.* 2000; 32:86–114. [https://doi.org/10.1016/S0165-0173\(99\)00069-7](https://doi.org/10.1016/S0165-0173(99)00069-7).
 28. Catapano LA, Manji HK. G protein-coupled receptors in major psychiatric disorders. *Biochim Biophys Acta.* 2007; 1768:976–93. <https://doi.org/10.1016/j.bbamem.2006.09.025>.
 29. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics.* 2014; 30:2923–30. <https://doi.org/10.1093/bioinformatics/btu403>.

30. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLOS Comput Biol.* 2012; 8:e1002503. <https://doi.org/10.1371/journal.pcbi.1002503>.
31. Loewenstein Y, Portugaly E, Fromer M, Linial M. Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. *Bioinformatics.* 2008; 24:i41–49. <https://doi.org/10.1093/bioinformatics/btn174>.
32. Emanuele E, Bertona M, Altabas K, Altabas V, Alessandrini G. Anti-inflammatory effects of a topical preparation containing nicotinamide, retinol, and 7-dehydrocholesterol in patients with acne: a gene expression study. *Clin Cosmet Investig Dermatol.* 2012; 5:33–37. <https://doi.org/10.2147/CCID.S29537>.
33. Deng XH, Song HY, Zhou YF, Yuan GY, Zheng FJ. Effects of quercetin on the proliferation of breast cancer cells and expression of survivin *in vitro*. *Exp Ther Med.* 2013; 6:1155–58. <https://doi.org/10.3892/etm.2013.1285>.
34. Jeong JH, An JY, Kwon YT, Rhee JG, Lee YJ. Effects of low dose quercetin: cancer cell-specific inhibition of cell cycle progression. *J Cell Biochem.* 2009; 106:73–82. <https://doi.org/10.1002/jcb.21977>.
35. Fan H, Zhang F, Xu Y, Huang X, Sun G, Song Y, Long H, Liu P. An association study of DRD2 gene polymorphisms with schizophrenia in a Chinese Han population. *Neurosci Lett.* 2010; 477:53–56. <https://doi.org/10.1016/j.neulet.2009.11.017>.
36. Zheng L, Liu H, Gong Y, Meng X, Jiang R, Wang X, Wang Q, Wang Y. Effect of Liuweidihuang pill and Jinkuishenqi pill on inhibition of spontaneous breast carcinoma growth in mice. *J Tradit Chin Med.* 2015; 35:453–59. [https://doi.org/10.1016/S0254-6272\(15\)30124-2](https://doi.org/10.1016/S0254-6272(15)30124-2).
37. Wu MY, Li P. [A survey on the pharmacological action and clinical application of liuwei dihuang pill]. [Article in Chinese]. *World Journal of Integrated Traditional and Western Medicine.* 2014; 9:1023–25.
38. Tao Q, Sun MY, Feng Q. [Syndrome identification of CC14 induced liver fibrosis model rats based on syndrome detecting from recipe used]. [Article in Chinese]. *Zhongguo Zhong Xi Yi Jie He Za Zhi.* 2009; 29:246–50.
39. Liao YN, Liu CS, Tsai TR, Hung YC, Chang SJ, Lin HL, Chen YC, Lai HM, Yu SF, Chen CJ. Preliminary study of a traditional Chinese medicine formula in systemic lupus erythematosus patients to taper steroid dose and prevent disease flare-up. *Kaohsiung J Med Sci.* 2011; 27:251–57. <https://doi.org/10.1016/j.kjms.2011.03.001>.
40. Pijl H, Ohashi S, Matsuda M, Miyazaki Y, Mahankali A, Kumar V, Pipek R, Iozzo P, Lancaster JL, Cincotta AH, DeFronzo RA. Bromocriptine: a novel approach to the treatment of type 2 diabetes. *Diabetes Care.* 2000; 23:1154–61. <https://doi.org/10.2337/diacare.23.8.1154>.
41. Dai W, Chen J, Lu P, Gao Y, Chen L, Liu X, Song J, Xu H, Chen D, Yang Y, Yang H, Huang L. Pathway Pattern-based prediction of active drug components and gene targets from H1N1 influenza's treatment with maxingshiganyinqiaosan formula. *Mol Biosyst.* 2013; 9:375–85. <https://doi.org/10.1039/c2mb25372k>.
42. Wang Y, Liu JW, Wang Q. [The pharmacological research and clinical application of liuwei dihuang pill]. [Article in Chinese]. *Chin J of Clinical Rational Drug Use.* 2011; 4:45–46.
43. Hu XY. [The clinical application of liuwei dihuang pill]. [Article in Chinese]. *Journal of Practical Traditional Chinese Medicine.* 2012; 28:244–45.
44. Jiang TL, Yan SC, Zhao LF. Preventing effect of “liuwei dihuang decoction” on esophageal carcinoma. *Gan To Kagaku Ryoho.* 1989; 16:1511–18.
45. Fang J, Zhang YX, Ru XB, Wei XL. [Effect of liuwei dihuang decoction, on the cytokine expression in splenocytes in AA rats]. [Article in Chinese]. *Zhongguo Zhong Yao Za Zhi.* 2001; 26:128–31.
46. Cheng XR, Cui XL, Zheng Y, Zhang GR, Li P, Huang H, Zhao YY, Bo XC, Wang SQ, Zhou WX, Zhang YX. A Co-Module Regulated by Therapeutic Drugs in a Molecular Subnetwork of Alzheimer's Disease Identified on the Basis of Traditional Chinese Medicine and SAMP8 Mice. *Curr Alzheimer Res.* 2015; 12:870–85. <https://doi.org/10.2174/1567205012666150710111858>.
47. Tian J, Shi J, Zhang X, Wang Y. Herbal therapy: a new pathway for the treatment of Alzheimer's disease. *Alzheimers Res Ther.* 2010; 2:30. <https://doi.org/10.1186/alzrt54>
48. Shen JJ, Lin CJ, Huang JL, Hsieh KH, Kuo ML. The effect of liu-wei-di-huang wan on cytokine gene expression from human peripheral blood lymphocytes. *Am J Chin Med.* 2003; 31:247–57. <https://doi.org/10.1142/S0192415X03000886>.
49. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014; 42:D1091–97. <https://doi.org/10.1093/nar/gkt1068>.
50. Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ, Bork P. STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res.* 2014; 42:D401–07. <https://doi.org/10.1093/nar/gkt1207>.
51. Zhao XM, Iskar M, Zeller G, Kuhn M, van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLOS Comput Biol.* 2011; 7:e1002323. <https://doi.org/10.1371/journal.pcbi.1002323>.
52. Chen CY. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One.* 2011; 6:e15939. <https://doi.org/10.1371/journal.pone.0015939>.
53. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen

- B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, et al. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* 2009; 37:D767–72. <https://doi.org/10.1093/nar/gkn892>.
54. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 2015; 43:D470–78. <https://doi.org/10.1093/nar/gku1204>.
55. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, et al. IntAct: an open source molecular interaction database. *Nucleic Acids Res.* 2004; 32:D452–55. <https://doi.org/10.1093/nar/gkh052>.
56. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. MINT: a Molecular INTeraction database. *FEBS Lett.* 2002; 513:135–40. [https://doi.org/10.1016/S0014-5793\(01\)03293-8](https://doi.org/10.1016/S0014-5793(01)03293-8).
57. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 2002; 30:303–05. <https://doi.org/10.1093/nar/30.1.303>.
58. Liu KQ, Liu ZP, Hao JK, Chen L, Zhao XM. Identifying dysregulated pathways in cancers from pathway interaction networks. *BMC Bioinformatics.* 2012; 13:126. <https://doi.org/10.1186/1471-2105-13-126>.
59. Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics.* 2005; 86:127–41. <https://doi.org/10.1016/j.ygeno.2005.04.008>.
60. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005; 33:D514–17. <https://doi.org/10.1093/nar/gki033>.