

## REVIEW

# A survey of web resources and tools for the study of TCM network pharmacology

Jing Zhao<sup>1,\*</sup>, Jian Yang<sup>2</sup>, Saisai Tian<sup>2</sup>, Weidong Zhang<sup>1,2,\*</sup>

<sup>1</sup> Institute of Interdisciplinary Complex Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

<sup>2</sup> School of Pharmacology, Second Military Medical University, Shanghai 200433, China

\* Correspondence: zhaojanne@gmail.com, wdzhangy@hotmail.com

Received June 8, 2018; Revised August 13, 2018; Accepted September 25, 2018

**Background:** Traditional Chinese medicine (TCM) treats diseases in a holistic manner, while TCM formulae are multi-component, multi-target agents at the molecular level. Thus there are many parallels between the key ideas of TCM pharmacology and network pharmacology. These years, TCM network pharmacology has developed as an interdisciplinary of TCM science and network pharmacology, which studies the mechanism of TCM at the molecular level and in the context of biological networks. It provides a new research paradigm that can use modern biomedical science to interpret the mechanism of TCM, which is promising to accelerate the modernization and internationalization of TCM.

**Results:** In this paper we introduce state-of-the-art free data sources, web servers and softwares that can be used in the TCM network pharmacology, including databases of TCM, drug targets and diseases, web servers for the prediction of drug targets, and tools for network and functional analysis.

**Conclusions:** This review could help experimental pharmacologists make better use of the existing data and methods in their study of TCM.

**Keywords:** TCM network pharmacology; molecular networks; signaling pathways; databases; web servers

**Author summary:** TCM network pharmacology studies the therapeutic mechanism of TCM formulae from a systems perspective and at the molecular level. Years of research in related fields has developed many databases and tools that are useful for the study of TCM network pharmacology. In this paper, we introduce some of such free resources.

## INTRODUCTION

These years have witnessed great progress in the application of systems biology to the research field of traditional Chinese medicine (TCM). Especially, TCM network pharmacology has arisen as a new interdisciplinary science, which integrates the study of TCM pharmacology with network science, systems biology, computational science and bioinformatics [1–3]. TCM treats diseases in a holistic way. At the molecular level, TCM formulae are multi-component and multi-target agents. Traditional reductionism method is difficult to reveal the complicated interplays between the multiple components and multiple targets of TCM formulae. TCM network pharmacology is an exciting new approach to

help examine the explicit targets of active TCM ingredients and define their function in the context of molecular networks.

One of the critical problems of TCM network pharmacology is to evaluate the synergistic effect of TCM's multiple targets in the disease-associated networks. According to the features of TCM that have complex components, unclear targets and treat diseases by holistic regulations, several network-based TCM research strategies have been established (see Figure 1 for an example) [1,2,4,5]. A series of databases [6–9] and algorithms [10–15] have been set up to provide data source and methodological support for TCM network pharmacology. Using these strategies, resources and methods, the therapeutic mechanisms of many TCM

formulae have been illustrated at the molecular level [16–20], which help facilitate the modernization of traditional Chinese medicine. However, it is still difficult for experimental researchers to apply the theories and algorithms of TCM network pharmacology. In this paper, we survey some data resources and tools that can be freely accessed online and easily used in the study of TCM network pharmacology.

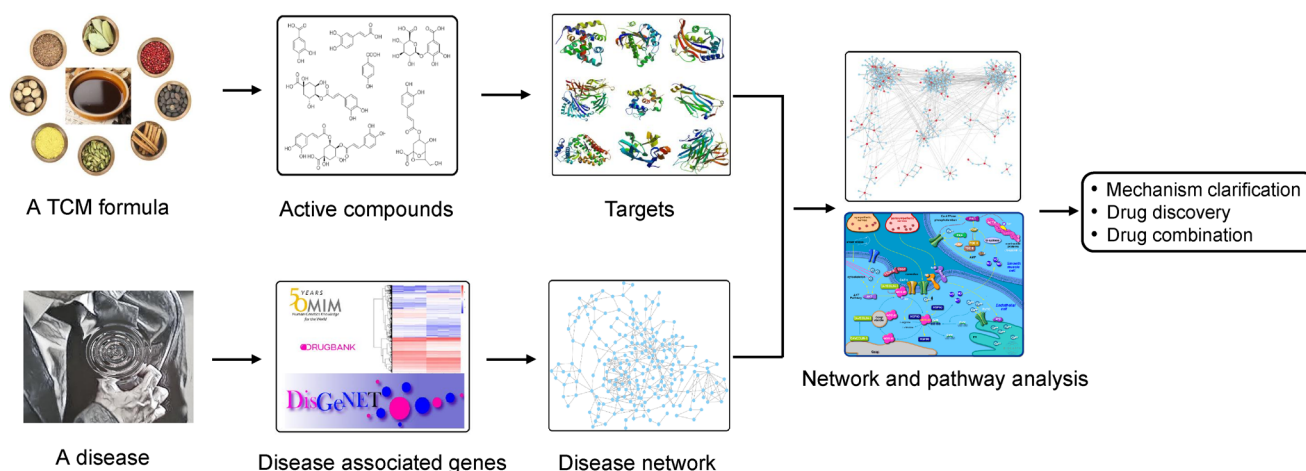
## TCM DATABASES

A TCM formula usually consists of many natural products such as herbs, animals and minerals, each of which contains large numbers of chemical molecules. The active compounds in the formula interact with dysfunctional proteins related to a specific disease so as to treat the disease. Therefore, the identification of active TCM compounds and their targets is essential to understand the underlying mechanisms of TCM formulae. Several databases have been constructed to provide information

concerning all aspects of TCM, including diseases, formulae, herbs or natural products, bioactive ingredients and targets. These databases bridge the gap between TCM and modern biomedical sciences and play important roles in the study of TCM pharmacology. This section introduces some free TCM databases. Table 1 lists a summary of contents of these databases.

## TCMSP

TCMSP (Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform) [8] consists of all the 499 Chinese herbs registered in the Chinese pharmacopoeia with 29,384 ingredients, 3,311 targets and 837 associated diseases. Drug-target mappings were obtained from HIT database [21] and the prediction algorithm SysDT [22]. The disease information was obtained from TTD database [23] and PharmGKB [24]. The information can be queried and downloaded. The very special value of this database is that for each



**Figure 1.** A general workflow for TCM network pharmacology in the study of a TCM formula. The workflow includes the identification of TCM's effective active compounds, targets of the compounds and disease associated genes. Based on these data, the construction and analysis of disease/drug-associated networks, and the identification of related signaling pathways help to clarify the mechanism of the TCM formula. Such studies can facilitate drug discovery from well-validated TCM remedies and the development of new drug combinations.

**Table 1** The statistics of the TCM databases covered in this review

Databases	No. of formulae	No. of organisms	No. of ingredients	No. of protein targets
TCMID	46,929	8,129	43,413	None
TCMSP	None	499	29,384	3,311
TCM-ID	1197	1102	12,117	None
HIT	None	> 1300	586	1301
TCM@Taiwan	None	453	61,000	None
NPACT	None	None	1574	248
CancerHSP	None	2439	3575	832
NPASS	None	25,041	35,032	2,946

compound, it provides its drug pharmacokinetic information such as drug-likeness (DL), oral bioavailability (OB), human intestinal absorption (HIA), blood-brain barrier (BBB), intestinal epithelial permeability (Caco-2), ALogP, fractional negative surface area (FASA-) and number of H-bond donor/acceptor (Hdon/Hacc). Thus the users can choose compounds with good drug-likeness and ADME (absorption, distribution, metabolism, excretion) feature for further study. Wang *et al.* used TCMSP to perform ADME screen for active ingredients of Chinese patent medicine Xinnaixin Pill. They set the screening thresholds as  $OB \geq 30\%$ ,  $DL \geq 0.18$ ,  $HIA \geq 70\%$ , and  $BB > 0.3$  and generated 18 active compounds satisfying these criteria [25]. From this database they then obtained 218 putative targets for these compounds. Lee *et al.* collected 447 compounds included in Yijin-Tang from the TCMSP database and only 42 of them satisfy OB and DL screening criteria [26]. Besides the 42 compounds, they also included 4 other compounds below these criteria which were reported as bioactive compounds in literatures for further analysis. At last, they suggested that 10 compounds were bioactive compounds and key chemicals of Yijin-Tang because each of them was linked to more than four genes associated with hyperlipidaemia and atherosclerosis.

## TCMID

TCMID (Traditional Chinese Medicine Integrated Database) [9] provides information about TCM formulas, herbs, and compounds of herbs. It also includes information of diseases and western drugs. This database was constructed by integrating former databases and text-mining. Specifically, the TCM formulas were collected by text mining from literatures; herbs were extracted from TCM-ID database [7] and text-mining; herbal ingredients were got by text-mining methods and combining information from other TCM databases, including TCM@Taiwan [6], TCM-ID, and HIT. Information of diseases and their associated proteins and drugs was retrieved from DrugBank and OMIM. Previous version of TCMID includes target information, which were collected from databases HIT, STITCH [27], OMIM [28], DrugBank [29] and text-mining. Recently, target information cannot be found in this database. TCMID database allows querying and downloading.

Here we briefly introduce the three databases whose information has been included in TCMID. TCMID (Traditional Chinese Medicine Information Database) [7] collects information about TCM formulas, herbs and compounds. It provides query about these information but does not allow downloading the whole data. HIT (Herbal Ingredients' Targets Database Introduction) [21] is a manually curated database that contains information on

the protein targets of compounds found in Chinese herbs. Based on over 3,250 papers in the literature, HIT includes 1,301 protein targets affected by 586 herbal compounds found in more than 1,300 Chinese herbs. TCM@Taiwan [6] provides information about 61,000 compounds included in 453 TCM herbs collected from Chinese medical texts and scientific publications. Because its data are manually retrieved from literature, HIT is a very reliable database. We collected part of targets for Shexiang Baixin Pill's active compounds from this database for further network and functional analysis [16].

## NPACT

NPACT (Naturally occurring Plant based Anticancerous Compound-activity-Target database) [30] is a manually curated database of plant derived compounds that show anti-cancerous activity extracted from 762 papers. It contains 1,574 compounds exhibiting activity against 353 cancer cell lines, 284 cancer-related protein targets and 1,980 experimentally validated compound-target interactions. For each compound, it provides information on their structure, properties, cancer type, cell lines, inhibitory values (IC<sub>50</sub>, ED<sub>50</sub>, EC<sub>50</sub>, GI<sub>50</sub>), molecular targets, commercial suppliers and drug likeness of compounds. Users can query the database by compounds and browse the information of the database from its web.

## CancerHSP

CancerHSP (Anticancer Herbs database of Systems Pharmacology) [31] contains 3,575 anticancer compounds included in 2,439 anticancer herbs and 832 targets. The targets are got by text-mining or prediction. The anticancer activities of these compounds based on 492 different cancer cell lines are also provided. The database also shows the molecular structure and nine key ADME parameters of each compound. Users can query the database by herbs, compounds and targets. The data of the database can be downloaded.

## NPASS

The NPASS (the Natural Product Activity and Species Source Database) [32] provides detailed information of species sources and biological activities of natural products. This database contains 35,032 unique natural products (compounds) isolated from 25,041 source organisms, including 16,581 plant species, 1,675 bacteria species, 2,503 metazoa species and 2,107 fungi species. A total of 5,863 targets of the compounds are provided, 2,946 of which are proteins. It also includes information of other classes of targets, such as organisms and cell lines. The data were collected from existing resources and

manually annotated from literatures.

## DRUG TARGET DATABASES

From Table 1 we can see that target information in current free TCM databases is too scarce to satisfy the needs of TCM research. Thus it is necessary for us to utilize other sources. There are many general drug target databases or chemical-protein interaction databases that provide large-scale experimental or predicted drug/compound-protein interaction data for small molecules. These databases have been widely accepted and used by drug discovery community. They can also be used to search or predict protein targets for TCM active compounds. In fact, some of them have been used in the study of TCM network pharmacology. To get comprehensive information about TCM formulae's active ingredients and targets, it is better to combine the results obtained from multiple resources. For example, Fang *et al.* conducted a systems study for identifying potential natural products which target mutated genes across 15 cancer types or subtypes [33]. They collected natural products from 6 TCM databases, including TCMID, TCMSP and TCM@Taiwan. The targets of natural molecules were searched from 2 TCM databases (TCMID and TCMSP) and 3 general chemical-protein interaction databases (STITCH, ChEMBL and BindingDB). This study found 848 significant anti-cancer indications for 224 natural products. When we study TCM formula Shexiang Baixin Pill (SBP)'s effects on cardio-vascular diseases (CVD), we collected its targets from HIT and STITCH [16]. We found that SBP only targets a very small fraction of CVD disease genes, which is not like FDA approved CVD drugs that mainly target specific disease genes. However, from the perspective of network regulation, our study suggested that SBP achieves its efficacy on CVD treatment by regulating the disease network though interactions between genes, instead of directly acting on CVD disease genes.

In this section we survey drug-target and chemical-protein interaction databases. See Table 2 for statistics of some databases of this class.

### DrugBank

DrugBank provides comprehensive molecular information about drugs and their mechanisms, including their chemical, pharmacological, pharmaceutical, ADME, interaction information and their targets [29,34]. The first version of DrugBank was released in 2006. It has been updated frequently. Currently DrugBank 5.0 contains the information of 10,971 drugs and 4,900 protein targets. The drugs consist of 2,391 FDA approved small molecule drugs, 934 approved biotech drugs, 109 nutraceuticals and over 5,090 experimental drugs. This

database allows for web query and downloading. In the study of TCM pharmacology, target information of modern drugs can be used for the prediction of TCM's targets. In addition, modern drugs which treat the same disease can be used as contrast. Information of these drugs can be obtained from DrugBank. For example, in their study on Yin-Huang-Qing-Fei capsule (YHQFC)'s effects on chronic bronchitis, Yu *et al.* identified DrugBank's drugs having similar structures with YHQFC's active compounds and then regarded targets of these drugs as putative targets of YHQFC [41]. They also acquired the known therapeutic targets of drugs used to treat chronic bronchitis from DrugBank for further network analysis. In our study for Huang-Lian-Jie-Du-Tang (HLJDT)'s effects on rheumatoid arthritis (RA), we retrieved all the 32 FDA-approved anti-RA drugs and their 51 targets from DrugBank. We designed a network score to measure drugs' effects and found that HLJDT achieves a high effect score, which is in the same order as that of one class of anti-RA agents, anti-inflammatory agents [42].

### TTD

TTD (Therapeutic target database) [23] provides information about drugs, targets, targeted diseases and pathways. The current version collects 34,019 drugs, including 2,544 proved drugs, 8,103 clinical trial drugs and 18,923 investigative drugs. Its 3,101 targets are classified as 445 successful targets, 1,121 clinical trial targets and 1,535 research targets. For each drug, its chemical structure, targets, targeted diseases and involved pathways are provided. Users can search the database by targets, drugs, diseases and biomarkers, as well as predict targets for compounds without target information using its drug similarity searching tool. The similarity searching is based on the Tanimoto similarity searching method. A query compound can be input by its MOL, SDF or SMILES format, and then the tool lists its similar compounds and corresponding Tanimoto similarity scores. The targets of the compounds with the highest scores can be predicted as the targets of the query compound. The database also allows for batch download. To study how the combination of herb pair *Euphorbia kansui* (GS) and *Glycyrrhiza* (GC) act on hepatocellular carcinoma ascites, Zhang *et al.* used structural similarity comparison in TTD to get 13 and 102 putative targets of GS and GC, respectively [43]. Their further topological analysis for the protein-protein network between these targets found that the ADRB1-PIK3CG interaction may play crucial role in connecting the other targets in the network. Considering that ADRB1 and PIK3CG were putative targets of GS and GC, respectively, and both had functional interactions with AVPR2, which is a known therapeutic target for ascites, they proposed that the GS-GC combination may exert

synergy effects on ascites through the ADRB1-PIK3CG-AVPR2 signal axis.

## STITCH

STITCH (Search Tool for Interacting Chemicals) [27] is a database of known and predicted interactions between chemicals and proteins. The current version covers interactions between 430,000 small molecules and 9,643,763 proteins from 2,031 organisms. STITCH shares protein space with the gene association database STRING, which is developed by the same team. The interactions are derived from high-throughput experiments, other primary databases, text-mining and computational prediction. The databases that STITCH collects data include manually curated databases such as DrugBank [34], GLIDA (GPCR-ligand database) [44], Mator [37], TTD (Therapeutic Targets Database) [45] and CTD (Comparative Toxicogenomics Database) [46]; pathway databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) [47], PID (Pathway Interaction Database) [48], Reactome [49] and BioCyc [50]; and experimental result databases such as ChEMBL [35], PDSP Ki [51], and PDB (Protein Data Bank) [52]. Each interaction in the database is assigned a score to indicate its probability or binding affinity. When a compound is searched via web of the database, its similar compounds and similarity scores will be listed. Like TTD, the database also supports the prediction of targets based on structural similarity. The STITCH database can be downloaded in full as well.

## ChEMBL

ChEMBL contains much more drug-like bioactive compounds than other drug target databases [35]. Currently, it contains 1,735,442 distinct compounds and 11,538 targets. The database provides structure, functional, target and ADMET information for the compounds. These data are manually collected from 67,722 peer-reviewed publications. The database can be searched by compounds or targets.

## BindingDB

BindingDB provides experimentally measured protein–ligand binding affinity data, including values of  $K_i$ ,  $K_d$ ,  $IC_{50}$  and  $EC_{50}$  [36]. It focuses chiefly on the interactions between drug-like small molecules and proteins considered to be drug-targets. The data are collected from US patents, scientific publications, and other databases such as the PubChem [53] (4,5), ChEMBL [35], PDSP Ki [51] and CSAR ([www.csardock.org](http://www.csardock.org)) [54]. The database launched on the web in the year of 2000 and is updated

continually. Currently BindingDB contains 1,427,022 binding data between 639,152 small molecules and 7,026 protein targets.

## ZINC

ZINC provides information about ligands and their purchasability, targets, clinical trials and so on [55]. Currently the 15th version of ZINC contains over 400 million purchasable “drug-like” compounds, which covers 204 commercial catalogs from 145 companies. The biological annotations of the compounds are derived from third party databases and libraries, such as HMDB [56], ChEMBL and DrugBank. ZINC normalizes ligand-target binding affinity measure  $pK_i$ ,  $IC_{50}$ ,  $EC_{50}$ ,  $AC_{50}$ , and  $pIC_{50}$  to a single standard  $pK_i$  value. The database also affords predicted targets, which were calculated by an algorithm combining the Similarity Ensemble Approach (SEA) with the maximum Tanimoto similarity. Applying this algorithm, over 171 million commercially available compounds were predicted targeting at a total of 2,629 targets with significance. When searching for a compound, the users can obtain information about its chemical characteristics, known and predicted targets, clinical trials and vendors who sell this compound.

## WEB SERVERS FOR TARGET PREDICTION

Some TCM databases and chemical-protein interaction databases, such as TCMSP, TTD, BATMAN-TCM [57] and STITCH, also provide target prediction services. That is, when we search these databases, some results are actually predicted by the algorithms of databases. There are some web servers and software that are specialized in target prediction, which can be used like these databases in the research of TCM pharmacology. For example, when exploring the anti-proliferative activity of herb *Pelargonium sidoides* DC, Pereira *et al.* used web servers PharmMapper [58] and DRAR-CPI [59] to predict its active compounds' targets [60]. They identified 142 potential protein targets of the herb, in which 90 targets were found to be related to cancer. Wei *et al.* used commercial software MetaDrug from GeneGo, Inc to predict putative targets of TCM formula Danhong injection (DHI) [61]. They found that 8 transcription factors (TFs), including pre-B-cell leukemia transcription factor 1 and cyclic AMP-dependent transcription factor 1, are putative target TFs for DHI-mediated protection against cerebral ischemia.

In this section we introduce some non-commercial web servers that provide service for the prediction of drug targets.

**Table 2** Statistics of some drug target databases

Databases	No. of compounds	No. of targets	Ref.
DrugBank	10,971	4,900	[29,34]
TTD	34,019	3,101	[23]
STITCH	430,000	9,643,763	[27]
ChEMBL	1,735,442	11,538	[35]
BindingDB	639,152	7,026	[36]
SuperTarget	195,770	6,219	[37]
MATADOR	801	2,901	[37]
CancerDR	148	116	[38]
DGIdb	32,106	88,533	[39]
SuperDRUG	4,587	3,000	[40]

## SuperPred

SuperPred is a web server for predicting targets and the ATC (Anatomical Therapeutic Chemical) codes of compounds [62]. Its predictions are based on the similar property principle, which includes the 3D structural similarity, the occurrence of fragments and the concordance of physico-chemical properties. The target prediction is performed by screening the input compound against a database containing about 341,000 compounds, 1,800 targets and 665,000 compound-target interactions. The database was created by extracting compound–target interaction data from SuperTarget, ChEMBL and BindingDB. The ATC code prediction is based on a similarity comparing pipeline including 2D, fragment and 3D similarity. The input compound is screened against 2,600 compounds having known ATC codes. The ATC codes of drugs are published by the World Health Organization (WHO) for the classification of drugs. The classification is based on therapeutic and chemical characteristics of the drugs. The query compound can be input by its name in PubChem or structure by SMILES or MOL format.

## SwissTargetPrediction

SwissTargetPrediction is a web server to predict the targets of compounds based on a combination of 2D and 3D similarity measures with known compounds [63]. Predictions can be carried out in five different organisms, including human, mouse, rat, cow and horse. The dataset of known compound-protein interactions for prediction was retrieved from the 16th version of ChEMBL database. It consists of 280,381 small molecules interacting with 2,686 targets, with the majority of targets (66%) found in human. SwissTargetPrediction provides a score for each predicted target to assess the likelihood of the predictions to be correct. It also maps predictions by

homology within and between different species and provides the likelihood score. Query compounds can be input by structure as SMILES string or MOL file.

## TargetNet

The web server TargetNet [64] predicts the targets of compounds based on QSAR (quantitative structure activity relationships) model, which mathematically relates specific chemical features of molecules to their bioactivities. Seven distinct molecular fingerprints are employed to describe chemical characteristics of molecules. The prediction uses a machine learning approach called Naïve Bayes classifier. The database BindingDB is used as training datasets. After a filtering process, a total of 109,061 compounds and 623 human target proteins in this database are used for model building. When the user inputs a compound, the prediction scores of this molecule toward all the 623 in-house proteins are calculated via 623 established QSAR models with selected performance statistics. A molecular structure or a molecular file with SMILES format is required for input.

## PharmMapper

The pharmacophore means a molecular framework that carries the essential features for a molecule to interact with a specific target protein. A pharmacophore model is a group of steric and electronic features of a molecule that is necessary to ensure its optimal interactions with a specific target protein for triggering or blocking its biological activity. Ligand-based pharmacophore modeling has been widely used in drug target identification. The Pharm Mapper is a web server that predicts protein targets of compounds based on pharmacophore model [58]. The prediction is performed by matching the pharmacophore of the query compound against an in-house pharmacophore model database.

Currently, the PharmMapper's in-house pharmacophore database includes 53,184 unique pharmacophore models. PharmMapper accepts a file with a compound stored in mol or SDF format.

### SystemsDock

Molecular virtual docking is a kind of computational method to evaluate the binding potential of a small molecule to a target protein. It has been widely applied in drug discovery. The web server SystemsDock [65] utilizes a molecular docking algorithm called dock-IN to assess protein–ligand binding potential and provides a score to present the potential [66]. Users need to input chemical molecules and their potential target proteins they want to validate. Each time at most five molecules and a group of proteins are allowed to input to the web server. The input for compounds can be a structure file in formats of 2D/3D SDF, Mol2 or SMILES. Proteins can be input by their gene symbols or PDB IDs. Users can also predict which proteins in a specific pathway are potential targets of their compounds. In this case the molecular pathway should be uploaded in SBML format with CellDesigner™ tags. It is noted that molecular virtual docking cannot be applied to proteins whose 3-D structures are unknown. The SystemsDock can be used to evaluate the binding affinity between small molecules and their putative targets. For example, Zhang *et al.* applied a similar software eHiTS [67] to evaluate the direct binding efficiencies of the main chemical components of *Euphorbia kansui* and *Glycyrrhiza* with ADRB1 and PIK3CG, respectively [43]. These two targets were predicted by TTD using structural similarity comparison, and their interaction is likely to play crucial role in connecting the other targets in the PPI network between the targets.

## DISEASE DATABASES

Clinical practices have proved the efficacy of TCM formulae to the treatment of many chronic complex diseases. Complex diseases are multi-factorial disorders that are influenced by multiple genes in combination with lifestyle and environmental factors. There are many databases that collect disease-related gene information for systems-level understanding of human diseases. They are important sources for the research of TCM network pharmacology. In this section, we introduce three representative disease databases.

### OMIM

The Online Mendelian Inheritance in Man (OMIM) database [28] is a curated database that provides information about the genetic component and relevant

genes for all the known mendelian disorders. It is a comprehensive, authoritative compendium of human genes, genetic phenotypes and the relationships between them. The OMIM contains over 15,000 genes associated with all known mendelian disorders. The information of OMIM comes from the published peer-reviewed biomedical literature. The database is updated daily.

### DisGeNET

DisGeNET is a comprehensive gene-disease association (GDA) database [68] that provides the current knowledge of human genetic diseases including Mendelian, complex and environmental diseases. The information in DisGeNET is integrated from expert curated repositories, GWAS catalogues, animal models and the scientific literature. Specifically, its data sources include the CTD™ (Comparative Toxicogenomics Database™) [69], UniProt/SwissProt [70], ClinVar [71], Orphanet [72], the NHGRI-EBI GWAS Catalog [73], GAD (Genetic Association Database) [74], MGD (the Mouse Genome Database) [75], RGD (the Rat Genome Database) [76], PsyGeNET [77], the Human Phenotype Ontology [78], LHGDN (the Literature Human Gene Derived Network) [79], and the BeFree [80]. The DisGeNET gives each GDA three scores range from 0 to 1, the confidence score, the Disease Specificity Index (DSI), and the Disease Pleiotropy Index (DPI). The confidence score reflects the reliability of the GDA by its recurrence across all data sources. The DSI is inversely proportional to the number of diseases associated to a particular gene, while the DPI is proportional to the number of different (MeSH) disease classes the gene is associated to. Thus a GDA with high DSI suggests that the gene is more specific for that disease, and a GDA with low DPI means that the gene is more specific for the disease class which the disease belongs to. The current version of DisGeNET (v5.0) contains 561,119 gene-disease associations between 17,074 genes and 20,370 diseases, disorders, traits, and clinical or abnormal human phenotypes.

### MalaCards

The MalaCards human disease database [81] is an integrated compendium of human diseases and their annotations. Currently, it includes information of 19,592 diseases consolidated from 72 sources. For each disease, the database displays a web card with diverse annotative information about this disease, such as disease classifications, disease summary, related diseases, associated genes, therapeutic drugs and publications. The “Genes” section of the web card provides the list of the genes found to be associated with the disease. The MalaCards assigns each disease-gene association a priority score,

which is a weighted sum of individual scores derived from eight sources, *i.e.*, OMIM [28], ClinVar [71], Orphanet [72], Humsavar of SwissProt [70], GeneTests [82], DISEASES [83], Novoseek [84] and GeneCards [85]. For each disease, the MalaCards defines a set of ‘elite’ genes, whose associations with this disease are supported by manually curated and reliable sources. Generally, the average number of associated genes per disease in the MalaCards is larger than that in the OMIM and smaller than that in the DisGeNET.

### DigSee

DigSee [86] is a text mining search engine which provides evidence sentences describing genes involved in the development of disease through biological events. Currently, DigSee supports all disease types and important biological events, such as point mutation, gene expression, regulation (positive regulation, negative regulation), protein catabolism, phosphorylation, localization, binding, transcription, catalysis, acetylation, hydroxylation, ubiquitination, methylation, glycosylation, and DNA methylation. All human genes can be searched, and 14,608 genes are indexed in current system version (v2.01). With the input of disease, genes or events, users can obtain Medline abstracts with highlighted evidence sentences.

## WEB SERVERS AND TOOLS FOR NETWORK AND FUNCTIONAL ANALYSIS

Since TCM formulae are actually multi-component, multi-target agents, annotating their targets in the context of networks could facilitate revealing the mode of action. Using network methods we can identify signaling pathways regulated by active ingredients in TCM formulae and examine the interactions between target proteins and the role of target proteins within networks. This helps illustrate the potential impact of TCM-based therapy and better elucidate its mechanisms of action. Here we introduce some web servers and software tools that can be used in network-based study of TCM pharmacology.

### BATMAN-TCM

BATMAN-TCM (Bioinformatics Analysis Tool for Molecular mechANism of TCM) [57] is an online bioinformatics analysis tool for studying the molecular mechanism of TCM. Users can input TCM formulae, herbs, or compounds for analysis. The tool will predict targets for the compounds in the formulae or herbs, perform functional analysis of the predicted targets, and construct and visualize the ingredient-target-pathway/disease association network. Functional analyses of

targets output the enrichment of targets in biological pathways, Gene Ontology functional terms and diseases. Formula-herb-compound association data used in this tool are extracted from the TCMID database. Potential targets of TCM ingredients are predicted by a similarity-based algorithm, which ranks potential drug-target interactions based on their similarity to known drug-target interactions in DrugBank, TTD and KEGG databases. This tool provides a rapid one-stop service for analyzing TCM formulae.

### DAVID

Pathway enrichment analysis identifies signaling pathways significantly enriched with a group of genes or proteins. Usually, drug target enriched pathways are considered most likely to be regulated by the drug. Thus this method has been widely used in the study of TCM network pharmacology [17,18,26,87]. Enrichment analysis can be easily performed using the online bioinformatics platform DAVID (the Database for Annotation, Visualization and Integrated Discovery) [88]. For example, Yu *et al.* used DAVID to conduct pathway enrichment analysis for the hubs in the PPI network between the putative targets of Yin-Huang-Qing-Fei capsule (YHQFC) and known therapeutic targets of the chronic bronchitis [41]. It was found that the hubs were significantly associated with various biological processes associated with chronic bronchitis, such as asthma pathway, cytokine–cytokine receptor interactions, T cell receptor signaling pathway, vascular smooth muscle contraction, and neuroactive ligand–receptor interaction. These results suggest that YHQFC’s effects on chronic bronchitis could be realized by regulating these pathways.

At the platform of DAVID, the users can upload a gene/protein list and then use the Functional Annotation Tool of DAVID to annotate the genes. This tool provides enrichment analysis for the gene list from different biological aspects, including biological pathways, GO terms, protein-protein interactions, protein functional domains, disease associations, gene tissue expressions, literatures, and so on. Based on the co-association of genes in different annotation terms, the DAVID Functional Annotation Clustering tool can group similar, redundant, and heterogeneous annotation terms into annotation groups. This service condenses a long list of annotation terms into fewer gene functional groups, thus it can help the users better interpret the functions of their gene list.

### CMap

The CMap (Connectivity Map) [89,90] is an online platform for finding disease–gene–drug relationships

based on the similarity of their gene-expression signatures. The CMap database contains huge number of gene-expression profiles of cultured human cells which are treated with drugs or bioactive compounds. Using the online software tool provided by CMap, users can compare gene expression signatures derived from their own research, such as a disease or a drug-treatment condition, with all perturbational signatures in the database. CMap gives a score ranging from +1 to -1 (+100 to -100 in the new version CMap-L1000v1) to measure the similarity between a query gene signature and a reference profile in the database. A positive score represents the degree of positive correlation and a negative score denotes a negative correlation between the query signature and the reference profile derived from a specific chemical perturbation. Thus high positive correlation of two chemical-treatment expression profiles suggests that the two chemicals may confer related physiological effects on the cell.

The original version of CMap (build 02) contains gene expression profiles of the treatment of 1,309 small molecules on 5 human cell lines [90]. Last year, the next generation CMAP called CMap-L1000v1 was released, which was expanded to 19,811 small molecule compounds and 9 human cell lines. Besides small molecule compounds, the new version also includes the treatment of other perturbagens (18,493 shRNAs, 3,462 cDNAs and 314 biologics) on cell lines [89].

Wen *et al.* used CMap to examine TCM formula Si-Wu-Tang (SWT) induced changes in gene expression [91]. They found that the gene expression signature of SWT-treated cells showed a high similarity with the CMAP profiles of estradiol-treated MCF-7 cells, which indicates SWT's phytoestrogenic effect. This result is consistent with SWT's widely application in the treatment of women's diseases. Our team established the gene expression profiles of MCF7 cells in response to 102 distinct TCM compounds. Then we used CMap to perform a case study to show how our data can be applied to reveal synergistic mechanism of TCM components [92]. Recently, Yoo *et al.* reanalyzed our data using the new version of CMap, they found that the 102 TCM compounds can be clustered into 4 groups based on their mechanism of actions (MoAs), while the MoAs could be clustered into 7 groups. Their further analysis suggests that TCM components clustered in the same MoAs are associated with similar disease indications [93]. These studies show that the CMap is a useful platform for elucidating molecular mechanisms of TCM when combined with microarray study.

## Cytoscape

Cytoscape [94] is free software for integrating, visualiz-

ing, and analyzing data in the context of networks. Cytoscape's software Core provides basic functionality to construct, visualize, layout and analyze the network; to visually integrate the network with expression profiles, phenotypes, and other molecular states; and to link the network to databases of functional annotations. Many works used Cytoscape for network visualization [33]. Network topological analysis is an important task in the research of TCM network pharmacology, which can locate hub nodes playing key roles in the network. When studying the pharmacological mechanisms of Wu-Tou-Dang (WTD) for the treatment of RA, Zhang *et al.* constructed the interaction network of WTD's putative targets and known RA-related targets. Then they used four topological measures (degree, node betweenness, closeness and *k* value) to find out hub proteins for further analysis [87]. In our research for Huang-Lian-Jie-Du-Tang (HLJDT)'s anti-rheumatic effects, we found that most targets of HLJDT are hubs of RA-associated network, suggesting that HLJDT may interfere with RA by acting on proteins in the central locations of the disease network with multiple components [14]. Network topological analysis can be easily done using Cytoscape's tool called "NetworkAnalyzer".

The Core of Cytoscape is extensible through a series of APPs, which can be installed and applied from a menu interface of the software. The users first need to download this software from the website of Cytoscape and then install it on their computer. For example, an APP called "clusterMaker" provides a series of methods, such as *k*-means, MCL, Glay community detection algorithm, to split a network into topological clusters. It has been known that topological clusters in biological networks may indicate relatively independent biological functions [95–97]. Thus the decomposition of networks can help better understand the interplays between multiple targets of TCM formula at the scale of modules. In our previous study on the anti-rheumatic mechanism of Tibetan Medicated-Bath Therapy using Wuwei-Ganlu-Yaoyu-Keli (WGYK), we first constructed a gene association network affected by adjuvant arthritis (AA) and WGYK and then decomposed this network into modules. We found that different modules participate in different aspects or stages of immune process regulation, including pathogen recognition, proinflammatory response and inflammatory signaling in innate immune defenses, innate immune response and adaptive immune response [17]. This result suggests that WGYK performs its therapeutic effect on RA by regulating multiple stages of immune process. These analyses can be performed by Cytoscape and DAVID.

The Cytoscape software is user-friendly and very useful for the study of TCM network pharmacology.

## PERSPECTIVE

By the combination of various chemical components, TCM formulae could achieve similar therapeutic efficacy as mono-ingredient agents at much lower doses of separate compounds. The low-dose and multi-target characteristics of TCM formulae may suppress drug resistance and side effects. Thus TCM has unique advantage in the treatment of complex chronic diseases that require long-term medication. Currently, the paradigm of “multi-target drug discovery” based on molecular network is becoming a new trend, and drug discovery tends to get resources and inspiration from traditional drugs and folk medicines. Through studying the main ingredients of TCM formulae and the role of the targets in the context of cellular networks, TCM network pharmacology helps us better understand the holistic, coordinated and complementary features of TCM. The web resources and tools introduced here may help experimental pharmacologists making better use of existing data and methods in the study of TCM network pharmacology. Current TCM network pharmacology investigates the interactions between multiple components of TCM mainly through the interactions between their targets. Such research can be thought as protein/gene-centered study, which focuses on protein-protein or gene-gene interactions. Recently, Kim *et al.* proposed a compound-centered opinion, suggesting that structural similarities of natural compounds to human metabolites can facilitate the understanding of the mode of action of traditional medicines [98]. We hope this inspirational work can bring new insights and approaches to TCM study. We expect that TCM network pharmacology could provide useful inspiration and reference for the discovery of new multi-component, multi-target drugs and combination medication.

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 81520108030, 21472238, 61372194 and 81260672), Professor of Chang Jiang Scholars Program, Shanghai Engineering Research Center for the Preparation of Bioactive Natural Products (No. 16DZ2280200), the Scientific Foundation of Shanghai China (Nos. 13401900103 and 13401900101), the National Key Research and Development Program of China (No. 2017YFC1700200), the Natural Science Foundation of Chongqing (No. cstc2018jcyjAX0090) and Chongqing Education Reform Project of Graduate (No. yjgl52017). The funders had no role in study design, data collection, analysis, decision to publish and preparation of the manuscript.

## COMPLIANCE WITH ETHICAL GUIDELINES

The authors Jing Zhao, Jian Yang, Saisai Tian and Weidong Zhang declare that they have no conflict of interest.

This paper is a review and does not contain any studies with human or

animal subjects performed by any of the authors.

## REFERENCES

1. Li, S. and Zhang, B. (2013) Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chin. J. Nat. Med.*, 11, 110–120
2. Zhao, J., Jiang, P. and Zhang, W. (2010) Molecular networks for the study of TCM pharmacology. *Brief. Bioinform.*, 11, 417–430
3. Li, S., Fan T.-P., Jia, W., Lu, A. and Zhang, W. (2014) Network pharmacology in traditional Chinese medicine, evidence-based complementary and alternative medicine. Article ID 138460 <https://doi.org/10.1155/2014/138460>
4. Li, P., Chen, J., Wang, J., Zhou, W., Wang, X., Li, B., Tao, W., Wang, W., Wang, Y. and Yang, L. (2014) Systems pharmacology strategies for drug discovery and combination with applications to cardiovascular diseases. *J. Ethnopharmacol.*, 151, 93–107
5. Huang, C., Zheng, C., Li, Y., Wang, Y., Lu, A. and Yang, L. (2014) Systems pharmacology in drug discovery and therapeutic insight for herbal medicines. *Brief. Bioinform.*, 15, 710–733
6. Chen, C. Y.-C. (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS One*, 6, e15939
7. Chen, X., Zhou, H., Liu, Y. B., Wang, J. F., Li, H., Ung, C. Y., Han, L. Y., Cao, Z. W. and Chen, Y. Z. (2006) Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation. *Br. J. Pharmacol.*, 149, 1092–1103
8. Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., Li, P., Guo, Z., Tao, W., Yang, Y., *et al.* (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.*, 6, 13
9. Xue, R., Fang, Z., Zhang, M., Yi, Z., Wen, C. and Shi, T. (2013) TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res.*, 41, D1089–D1095
10. Li, S., Zhang, B. and Zhang, N. (2011) Network target for screening synergistic drug combinations with application to traditional Chinese medicine. *BMC Syst. Biol.*, 5, S10
11. Lin, L., Yang, T., Fang, L., Yang, J., Yang, F. and Zhao, J. (2017) Gene gravity-like algorithm for disease gene prediction based on phenotype-specific network. *BMC Syst. Biol.*, 11, 121
12. Sun, Y., Sheng, Z., Ma, C., Tang, K., Zhu, R., Wu, Z., Shen, R., Feng, J., Wu, D., Huang, D., *et al.* (2015) Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nat. Commun.*, 6, 8481
13. Yang, K., Bai, H., Ouyang, Q., Lai, L. and Tang, C. (2008) Finding multiple target optimal intervention in disease-related molecular network. *Mol. Syst. Biol.*, 4, 228
14. Fang, H., Wang, Y., Yang T., Ga, Y., Zhang, Y., Liu, R., Zhang, W. and Zhao, J. (2013) Bioinformatics analysis for the antirheumatic effects of Huang-Lian-Jie-Du-Tang from a network perspective. *Evid-Based Compl. Alt.*, Article ID 245357, <http://dx.doi.org/10.1155/2013/245357>
15. Le, D. H. and Le, L. (2016) Systems pharmacology: a unified framework for prediction of drug-target interactions. *Curr. Pharm.*

- Des., 22, 3569–3575
16. Fang, H.-Y., Zeng, H.-W., Lin, L.-M., Chen, X., Shen, X.-N., Fu, P., Lv, C., Liu, Q., Liu, R.-H., Zhang, W.-D., *et al.* (2017) A network-based method for mechanistic investigation of Shexiang Baixin Pill's treatment of cardiovascular diseases. *Sci. Rep.*, 7, 43632
17. Wang, T., Yang, J., Chen, X., Zhao, K., Wang, J., Zhang, Y., Zhao, J. and Ga, Y. (2017) Systems study on the antirheumatic mechanism of Tibetan medicated-bath therapy using Wuwei-Ganlu-Yaoyu-Keli. *BioMed Res. Int.*, 2017, 2320932
18. Liang, X., Li, H. and Li, S. (2014) A novel network pharmacology approach to analyse traditional herbal formulae: the Liu-Wei-Di-Huang pill as a case study. *Mol. Biosyst.*, 10, 1014–1022
19. Zhang, W., Tao, Q., Guo, Z., Fu, Y., Chen, X., Shar, P. A., Shahan, M., Zhu, J., Xue, J., Bai, Y., *et al.* (2016) Systems pharmacology dissection of the integrated treatment for cardiovascular and gastrointestinal disorders by traditional Chinese medicine. *Sci. Rep.*, 6, 32400
20. Zhou, W., Cheng, X. and Zhang, Y. (2016) Effect of Liuwei Dihuang decoction, a traditional Chinese medicinal prescription, on the neuroendocrine immunomodulation network. *Pharmacol. Ther.*, 162, 170–178
21. Ye, H., Ye, L., Kang, H., Zhang, D., Tao, L., Tang K., Liu, X., Zhu, R., Liu, Q., Chen, Y. Z. *et al.* (2011) HIT: linking herbal active ingredients to targets. *Nucleic Acids Res.*, 39 (suppl\_1), D1055–D1059
22. Yu, H., Chen, J., Xu, X., Li, Y., Zhao, H., Fang, Y., Li, X., Zhou, W., Wang, W. and Wang, Y. (2012) A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PLoS One*, 7, e37608
23. Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., *et al.* (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.*, 46, D1121–D1127
24. Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., Altman, R. B. and Klein, T. E. (2012) Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.*, 92, 414–417
25. Huang, C., Yang, Y., Chen, X., Wang, C., Li, Y., Zheng, C. and Wang, Y. (2017) Large-scale cross-species chemogenomic platform proposes a new drug discovery strategy of veterinary drug from herbal medicines. *PLoS One*, 12, e0184880
26. Lee, A. Y., Park, W., Kang, T.-W., Cha, M. H. and Chun, J. M. (2018) Network pharmacology-based prediction of active compounds and molecular targets in Yijin-Tang acting on hyperlipidaemia and atherosclerosis. *J. Ethnopharmacol.*, 221, 151–159
27. Kuhn, M., von Mering, M., Campillos, M., Jensen, L.J. and Bork, P. (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, 36(suppl\_1), D684–688
28. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res.*, 33(suppl\_1), D514–517
29. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.*, 34, D668–D672
30. Mangal, M., Sagar, P., Singh, H., Raghava, G. P. S. and Agarwal, S. M. (2013) NPACT: naturally occurring plant-based anti-cancer compound-activity-target database. *Nucleic Acids Res.*, 41, D1124–D1129
31. Tao, W., Li, B., Gao, S., Bai, Y., Shar, P. A., Zhang, W., Guo, Z., Sun, K., Fu, Y., Huang, C., *et al.* (2015) CancerHSP: anticancer herbs database of systems pharmacology. *Sci. Rep.*, 5, 11481
32. Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., Wang, Y., Tan, Y., Gao, D., Wang, B., *et al.* (2018) NPASS: natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res.*, 46, D1217–D1222
33. Fang, J., Cai, C., Wang, Q., Lin, P., Zhao, Z. and Cheng, F. (2017) Systems pharmacology-based discovery of natural products for precision oncology through targeting cancer mutated genes. *CPT Pharmacometrics Syst. Pharmacol.*, 6, 177–187
34. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, 46, D1074–D1082
35. Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, 42, D1083–D1090
36. Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L. and Chong, J. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, 44, D1045–D1053
37. Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Urdiales, E. G., Gewiss, A., Jensen, L. J., *et al.* (2008) SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, 36, D919–D922
38. Kumar, R., Chaudhary, K., Gupta, S., Singh, H., Kumar, S., Gautam, A., Kapoor, P., Raghava, G. P. S. and Cancer, D. R. (2013) CancerDR: cancer drug resistance database. *Sci. Rep.*, 3, 1445
39. Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., Wollam, A., Spies, N. C., Griffith, O. L. and Griffith, M. (2018) DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.*, 46, D1068–D1073
40. Siramshetty, V. B., Eckert, O. A., Gohlke, B.-O., Goede, A., Chen, Q., Devarakonda, P., Preissner, S. and Preissner, R. (2018) SuperDRUG2: a one stop resource for approved/ marketed drugs. *Nucleic Acids Res.*, 46, D1137–D1143
41. Yu, G., Zhang, Y., Ren, W., Dong, L., Li, J., Geng, Y., Zhang, Y., Li, D., Xu, H. and Yang, H. (2016) Network pharmacology-based identification of key pharmacological pathways of Yin-Huang-Qing-Fei capsule acting on chronic bronchitis. *Int. J. Chron. Obstruct. Pulmon. Dis.*, 12, 85–94
42. Fang, H., Wang, Y., Yang, T., Ga, Y., Zhang, Y., Liu, R., Zhang, W. and Zhao, J. (2013) Bioinformatics analysis for the antirheumatic

- effects of Huang-Lian-Jie-Du-Tang from a network perspective. *Evid. Based Complement. Alternat. Med.*, 2013, 245357
43. Zhang, Y., Lin, Y., Zhao, H., Guo, Q., Yan, C. and Lin, N. (2016) Revealing the effects of the herbal pair of *Euphorbia kansui* and *Glycyrrhiza* on hepatocellular carcinoma ascites with integrating network target analysis and experimental validation. *Int. J. Biol. Sci.*, 12, 594–606
  44. Okuno, Y., Tamon, A., Yabuuchi, H., Nijima, S., Minowa, Y., Tonomura, K., Kunimoto, R. and Feng, C. (2008) GLIDA: GPCR—ligand database for chemical genomics drug discovery—database and tools update. *Nucleic Acids Res.*, 36(suppl\_1), D907–D912
  45. Chen, X., Ji, Z. L. and Chen, Y. Z. (2002) TTD: Therapeutic Target Database. *Nucleic Acids Res.*, 30, 412–415
  46. Davis, A. P., Grondin, C. J., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., King, B. L., Wiegers, T. C. and Mattingly, C. J. (2015) The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.*, 43, D914–D920
  47. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28, 27–30
  48. Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T. and Buetow, K. H. (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.*, 37, D674–D679
  49. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., *et al.* (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, 44, D481–D487
  50. Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, 44, D471–D480
  51. Roth, B. L., Lopez, E., Patel, S. and Kroeze, W. K. (2000) The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist*, 6, 252–262
  52. Rose, P. W., Prlić, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, 43, D345–D356
  53. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Res.*, 44, D1202–D1213
  54. Carlson, H. A., Smith, R. D., Damm-Ganamet, K. L., Stuckey, J. A., Ahmed, A., Convery, M. A., Somers, D. O., Kranz, M., Elkins, P. A., Cui, G., *et al.* (2016) CSAR 2014: a benchmark exercise using unpublished data from pharma. *J. Chem. Inf. Model.*, 56, 1063–1077.
  55. Sterling, T. and Irwin, J. J. (2015) ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.*, 55, 2324–2337
  56. Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, 46, D608–D617
  57. Liu, Z., Guo, F., Wang, Y., Li, C., Zhang, X., Li, H., Diao, L., Gu, J., Wang, W., Li, D., *et al.* (2016) BATMAN-TCM: a bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Sci. Rep.*, 6, 21146
  58. Wang, X., Shen, Y., Wang, S., Li, S., Zhang, W., Liu, X., Lai, L., Pei, J. and Li, H. (2017) PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res.*, 45, W356–W360
  59. Luo, H., Chen, J., Shi, L., Mikailov, M., Zhu, H., Wang, K., He, L., and Yang, L. (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical–protein interactome. *Nucleic Acids Res.*, 39(suppl\_2), W492–W498
  60. Pereira, A. S. P., Bester, M. J. and Apostolides, Z. (2017) Exploring the anti-proliferative activity of *Pelargonium sidoides* DC with *in silico* target identification and network pharmacology. *Mol. Divers.*, 21, 809–820
  61. Wei, J., Zhang, Y., Jia, Q., Liu, M., Li, D., Zhang, Y., Song, L., Hu, Y., Xian, M., Yang, H., *et al.* (2016) Systematic investigation of transcription factors critical in the protection against cerebral ischemia by Danhong injection. *Sci. Rep.*, 6, 29823
  62. Nickel, J., Gohlke, B.-O., Ereman, J., Banerjee, P., Rong, W. W., Goede, A., Dunkel, M. and Preissner, R. (2014) SuperPred: update on drug classification and target prediction. *Nucleic Acids Res.*, 42, W26–W31
  63. Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O. and Zoete, V. (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.*, 42, W32–W38
  64. Yao, Z.-J., Dong, J., Che, Y.-J., Zhu, M.-F., Wen, M., Wang, N.-N., Wang, S., Lu, A.-P. and Cao, D.-S. (2016) TargetNet: a web service for predicting potential drug-target interaction profiling via multi-target SAR models. *J. Comput. Aided Mol. Des.*, 30, 413–424
  65. Hsin, K.-Y., Matsuo, Y., Asai, Y., Kamiyoshi, K., Watanabe, T., Kawaoka, Y. and Kitano, H. (2016) systemsDock: a web server for network pharmacology-based prediction and analysis. *Nucleic Acids Res.*, 44, W507–W513
  66. Hsin, K.-Y., Ghosh, S. and Kitano, H. (2013) Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PLoS One*, 8, e83922
  67. Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S. and Johnson, A. P. (2006) eHiTS: an innovative approach to the docking and scoring function problems. *Curr. Protein Pept. Sci.*, 7, 421–435.
  68. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F. and Furlong, L. I. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, 45, D833–D839
  69. Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., King, B. L., McMorran, R., Wiegers, J., Wiegers, T. C. and Mattingly, C. J. (2017) The Comparative Toxicogenomics Database: update 2017.

- Nucleic Acids Res., 45, D972–D978
70. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., *et al.* (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.*, 32, D115–D119
71. Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, 44, D862–D868
72. Aymé, S. and Schmidtke, J. (2007) Networking for rare diseases: a necessity for Europe. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 50, 1477–1483, in German
73. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, 45, D896–D901
74. Becker, K. G., Barnes, K. C., Bright, T. J. and Wang, S. A. (2004) The Genetic Association Database. *Nature Genet.*, 36 431–432
75. Blake, J. A., Richardson, J. E., Bult, C. J., Kadin, J. A. and Eppig, J. T., and the Mouse Genome Database Group. (2003) MGD: the Mouse Genome Database. *Nucleic Acids Res.*, 31, 193–195
76. Twigger, S., Lu, J., Shimoyama, M., Chen, D., Pasko, D., Long, H., Ginster, J., Chen, C.-F., Nigam, R., Kwitek, A., *et al.* (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res.*, 30, 125–128
77. Gutiérrez-Sacristán, A., Grosdidier, S., Valverde, O., Torrens, M., Bravo, À., Piñero, J., Sanz, F. and Furlong, L. I. (2015) PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics*, 31, 3075–3077
78. Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D. and Mundlos, S. (2008) The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.*, 83, 610–615
79. Bundschuh, M., Dejori, M., Stetter, M., Tresp, V. and Kriegl, H.-P. (2008) Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics*, 9, 207
80. Bravo, À., Piñero, J., Queralt-Rosinach, N., Rautschka, M. and Furlong, L. I. (2015) Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC Bioinformatics*, 16, 55
81. Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., Gershoni, M., Morrey, C. P., Safran, M. and Lancet, D. (2017) MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.*, 45, D877–D887
82. Roberta, A. (2007) GeneTests: integrating genetic services into patient care. *Am. J. Hum. Genet.*, 81, 658–659
83. Pletscher-Frankild, S., Paljeja, A., Tsafou, K., Binder, J. X. and Jensen, L. J. (2015) DISEASES: text mining and data integration of disease-gene associations. *Methods*, 74, 83–89
84. Allende, R. A. (2009) Accelerating searches of research grants and scientific literature with novo|seek<sup>SM</sup>. *Nat. Methods*, 6, 394
85. Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., *et al.* (2010) GeneCards Version 3: the human gene integrator. Database (Oxford), 2010, baq020
86. Kim, J., So, S., Lee, H.-J., Park, J. C., Kim, J. J. and Lee, H. (2013) DigSee: disease gene search engine with evidence sentences (version cancer). *Nucleic Acids Res.*, 41, W510–W517
87. Zhang, Y., Bai, M., Zhang, B., Liu, C., Guo, Q., Sun, Y., Wang, D., Wang, C., Jiang, Y., Lin, N., *et al.* (2015) Uncovering pharmacological mechanisms of Wu-tou decoction acting on rheumatoid arthritis through systems approaches: drug-target prediction, network analysis and experimental validation. *Sci. Rep.*, 5, 9463
88. Huang, W., Sherman, B. T. and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, 4, 44–57
89. Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A. and Asiedu, J. K. (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, *Cell* 171, 1437–1452. e17
90. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A. and Ross, K.N. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313,1929–1935
91. Wen, Z., Wang, Z., Wang, S., Ravula, R., Yang, L., Xu, J., Wang, C., Zuo, Z., Chow, M. S., Shi, L., *et al.* (2011) Discovery of molecular mechanisms of traditional Chinese medicinal formula Si-Wu-Tang using gene expression microarray and connectivity map. *PLoS One*, 6, e18278
92. Lv, C., Wu, X., Wang, X., Su, J., Zeng, H., Zhao, J., Lin, S., Liu, R., Li, H., Li, X., *et al.* (2017) The gene expression profiles in response to 102 traditional Chinese medicine (TCM) components: a general template for research on TCMs. *Sci. Rep.*, 7, 352
93. Yoo, M., Shin, J., Kim, H., Kim, J., Kang, J. and Tan, A. C. (2018) Exploring the molecular mechanisms of traditional Chinese medicine components using gene expression signatures and connectivity map. *Comput. Methods Programs Biomed.*,
94. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13, 2498–2504
95. Vennix, P. P., Kuijpers, W., Tonnaer, E. L., Peters, T. A. and Ramaekers, F. C. (1990) Cytokeratins in induced epidermoid formations and cholesteatoma lesions. *Arch. Otolaryngol. Head Neck Surg.*, 116, 560–565
96. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A.-L. (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297,1551–1555
97. Padmanabhan, K., Wang, K. and Samatova, N. F. (2012) Functional annotation of hierarchical modularity. *PLoS One*, 7, e33744
98. Kim, H. U., Ryu, J. Y., Lee, J. O. and Lee, S. Y. (2015) A systems approach to traditional oriental medicine. *Nat. Biotechnol.*, 33, 264–268