Computational and Structural Biotechnology Journal 21 (2023) 5807–5817

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Database Article

ccTCM: A quantitative component and compound platform for promoting the research of traditional Chinese medicine

Dongqing Yang [a], Zhu Zhu [b], Qi Yao [b], Cuihua Chen [c], Feiyan Chen [c], Ling Gu [c], Yucui Jiang [c], Lin Chen [d], Jingyuan Zhang [e], Juan Wu [a], Xingsu Gao [a], Junqin Wang [a], Guochun Li [a], Yunan Zhao [b,*]

[a] Department of Public Health, School of Medicine & Holistic Integrative Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China
[b] Department of Pathology and Pathophysiology, School of Medicine & Holistic Integrative Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China
[c] Research and Innovation Center, College of Traditional Chinese Medicine·Integrated Chinese and Western Medicine College, Nanjing University of Chinese Medicine, Nanjing 210023, China
[d] Department of Physiology, School of Medicine & Holistic Integrative Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China
[e] Department of Treatise on Febrile Diseases, School of Traditional Chinese Medicine & Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China

ARTICLE INFO

Keywords:
TCM (Traditional Chinese medicine)
Compound
Component
Content
Database

ABSTRACT

Traditional Chinese medicine (TCM) databases play a vital role in bridging the gap between TCM and modern medicine, as well as in promoting the popularity of TCM. Elucidating the bioactive ingredients of Chinese medicinal materials is key to TCM modernization and new drug discovery. However, one drawback of current TCM databases is the lack of quantitative data on the constituents of Chinese medicinal materials. Herein, we present ccTCM, a web-based platform designed to provide a component and compound-content-based resource on TCM and analysis services for medical experts. In terms of design features, ccTCM combines resource distribution, similarity analysis, and molecular-mechanism analysis to accelerate the discovery of bioactive ingredients in TCM. ccTCM contains 273 Chinese medicinal materials commonly used in clinical settings, covering 29 functional classifications. By searching and comparing, we finally adopted 2043 studies, from which we collected the compounds contained in each TCM with content greater than 0.001 %, and a total of 1449 were extracted. Subsequently, we collected 40,767 compound-target pairs by integrating multiple databases. Taken together, ccTCM is a versatile platform that can be used by TCM scientists to perform scientific and clinical TCM studies based on quantified ingredients of Chinese medicinal materials. ccTCM is freely accessible at http://www.cctcm.org.cn.

1. Introduction

As an alternative to modern medicine, traditional Chinese medicine (TCM) has been used to treat and prevent various diseases over thousands of years, playing an important role in improving the health of East Asian people [1]. In recent decades, great efforts have been exerted to study all aspects of TCM, such as clinical evaluation [2], chemical profiling [3], and bioactivities [4]. With the rapid increase in available TCM data, many web-based databases specializing in TCM have emerged, which in turn facilitate the scientific and clinical study of TCM.

First, the TCM Information Database [5], published in 2006, has been introduced earlier as a web resource to provide free-of-charge and comprehensive information about TCM, including herbs, prescriptions, herbal ingredients, structure and functional properties of compounds, as well as their therapeutic effects and clinical indications and applications. The database represents early efforts toward enhancing the ability to evaluate TCM herbs' beneficial and risk effects.

Second, are TCM Database@Taiwan [6] and SymMap [7], which emphasize phenotypic drug discovery (PDD) based on a large amount of information on natural products and their clinical applications. TCM Database@Taiwan, published in 2011, contains more than 20,000 natural compounds from 453 Chinese Materia Medica, including herbs, animal products, and minerals. The 2D and 3D formats of each compound in the database are available for virtual filtering or molecular

* Corresponding author.
E-mail address: zhaoyunan-js@163.com (Y. Zhao).

https://doi.org/10.1016/j.csbj.2023.11.030
Received 13 September 2023; Received in revised form 14 November 2023; Accepted 15 November 2023
Available online 20 November 2023

simulation. SymMap, an integrative database of TCM enhanced by symptom mapping, was published in 2019. SymMap presents the newly curated symptom-herb knowledge and connects symptoms and phenotypes to herbs and diseases, thereby providing both phenotypic changes and lead compounds for PDD screening efforts.

Finally, for TCMID [8], TCMSP [9], TCMIO[10], HERB [11], INPUT [12], LTM-TCM [13], and ITCM[14], these TCM databases focus on understanding of the action mechanisms underlying TCM through the concept and theory of network pharmacology. TCMID, a TCM integrative database for herb molecular mechanism analysis, was published in 2012. Based on predicted targets of compounds, the database displays herb-disease networks and compound-target networks, integrating TCM with modern science at the phenotypic and molecular levels. TCMSP, a database of system pharmacology for drug discovery from herbal medicines, was published in 2014. The database improves the network-pharmacology analysis of TCM with the help of absorption, distribution, metabolism, and excretion related properties of compounds. TCMIO is a comprehensive database of traditional Chinese medicine on immunology, which can be used to explore the molecular mechanisms by which traditional Chinese medicine regulates the immune microenvironment of cancer. Unlike TCMID and TCMSP, HERB, published in 2020, used gene targets guided by high-throughput transcriptomic screening experiments to identify herb-disease networks and compound-target networks. INPUT collects multiple resources and embeds a series of tools for network pharmacology analysis. LTM-TCM, published in 2022, is currently the most comprehensive TCM database. Using the LTM-TCM platform, the network-pharmacology analysis of TCM is enhanced by large amounts of data integration and high-quality normalization. The ITCM database, published in 2023, is based on a unified high-throughput experiment and constructs a drug transcriptomics platform containing 496 representative active ingredients.

Nowadays, these TCM databases play a crucial role in bridging the gap between TCM and modern medicine as well as in promoting the modernization and popularization of TCM [15]. However, some problems have emerged in the result reliability of PDD screening and network-pharmacology analysis. An obvious disadvantage of these TCM databases is that all of them do not provide quantitative data on ingredients in Chinese medicinal materials. Some ingredients of Chinese medicinal materials with very low content, even if they have better bioactivities, are not responsible for the therapeutic effects of TCM medicinal materials. If such ingredients are not discarded based on content data, the ranking of lead compounds and the construction of herb-compound-target-disease networks are bound to be seriously affected. Thus, adding quantitative data on ingredients into TCM databases contributes to upgrading PDD screening and network-pharmacology analysis.

With the rapid development of quantitative analysis techniques including chemical and instrumental analysis methods [16], many studies have aimed to determine the contents of different ingredients of Chinese medicinal materials and link them to the biodiversity and quality evaluation of Chinese medicinal materials [17]. In particular, a few TCM quality-control studies have focused on detecting the content differences of multiple compounds in certain Chinese medicinal materials derived from different botanical origins [18], producing areas [19], cultivation years [20], harvesting seasons [21], and processing methods [22] through high-performance liquid chromatography (HPLC) coupled with different up-to-date detectors. The rapidly increasing number of quantified ingredients in Chinese medicinal materials provides us an opportunity to develop a component and compound-content-based database integrating comprehensive information on TCM (**ccTCM**).

Hence, in the present study, we obtained the quantified ingredients data of TCM based on comprehensive literature searches with focus on botanical origins, producing areas, harvesting seasons, and processing methods. The scope of our collection covered 29 categories of TCM, 273 Chinese medicinal materials (Supplementary Table S1), and 1499 compounds in total. The content-determination information of each TCM was obtained through manual literature retrieval, totaling 2043 articles, with an average of 7.48 literature supporting each TCM. For the convenience of use, the metadata of each TCM was collected from the Chinese Pharmacopoeia (2020 edition), and the metadata of each compound was collected from PubChem [23]. We also provided 40,767 pieces of target information for compounds. In brief, the data-based connections between TCM and modern medicine described in ccTCM provided reliable support for understanding the molecular mechanisms underlying TCM clinical therapy. Moreover, ccTCM provided similarity analysis of Chinese medicinal materials and resource-distribution analysis of components and compounds, thereby enabling the progress of the TCM industry and scientific research.

2. Materials and methods

2.1. Data sources of Chinese medicinal materials and compounds

The metadata of Chinese medicinal materials in ccTCM (including name, species, Latin name, medicinal part, basic characteristics, dosage, toxicity, main efficacy, identification, source information, trait, storage conditions, etc.) originated from the Chinese Pharmacopoeia (2015 edition) and were automatically translated into English through Google translation API.

The metadata of compounds was retrieved, using the compound name and their synonyms, from PubChem by using PubChemPy (v1.0.4) package via PubChem's PUG REST web service. The metadata of each compound includes molecular formula, molecular weight, complexity, classification, properties, synonyms, IUPAC, InChi, InChiKey, Canonical Smiles, Isomeric Smiles, exact mass, etc. The structures of compounds were searched from PubChem, ChEMBL [24], and ZINC [25]. For those compounds not found in these databases, their structures were drawn by using InDraw 5.2 software (https://www.integle.com/static/indraw).

2.2. Manual collection of quantified TCM ingredients

The schematic of document retrieval, quantitative data collection, and ingredient rating is shown in Fig. 1. The chemical profiling of Chinese medicinal materials was searched from the Chinese Academic Journal Network Publishing Database (CAJD) (https://www.cnki.net/) by using the combinations of the keywords ("name", "progress", and "chemistry") in the title. Original research on ingredient content analysis was also searched from CAJD by using the combinations of the keywords ("name", "content", and "determination") in the title. The 5777 pieces of literature involving 273 Chinese medicinal materials were adopted, as the botanical origin of Chinese herbs and the content unit of quantified ingredients was clearly clarified in the context. The detailed inclusion criteria for different content results of the same ingredients in the same TCM reported in different references can be referring to supplementary document 1. In brief, based on the content requirements in the Chinese Pharmacopoeia, the content of ingredients in ccTCM was shown as the range of quantitative data covering most of determination values reported in different references. Abnormal quantitative data that deviated so far from the rest of the data were not adopted. In addition, quantitative data were preferentially extracted from the articles focusing on the quality assessment of Chinese medicinal materials derived from different botanical origins, producing areas, cultivation years, harvesting seasons, or processing methods through the quantitative analysis of multi-components by a single marker or HPLC-based simultaneous detection. Specially, the relative content of compounds in essential oils varies greatly in different references. Considering the common characteristic of essential oils, the relative content of compounds in essential oils was shown as the achievable maximum determination values in the database.

We divided ingredients into 26 categories based on the structural characteristics of natural products, which were named major category in

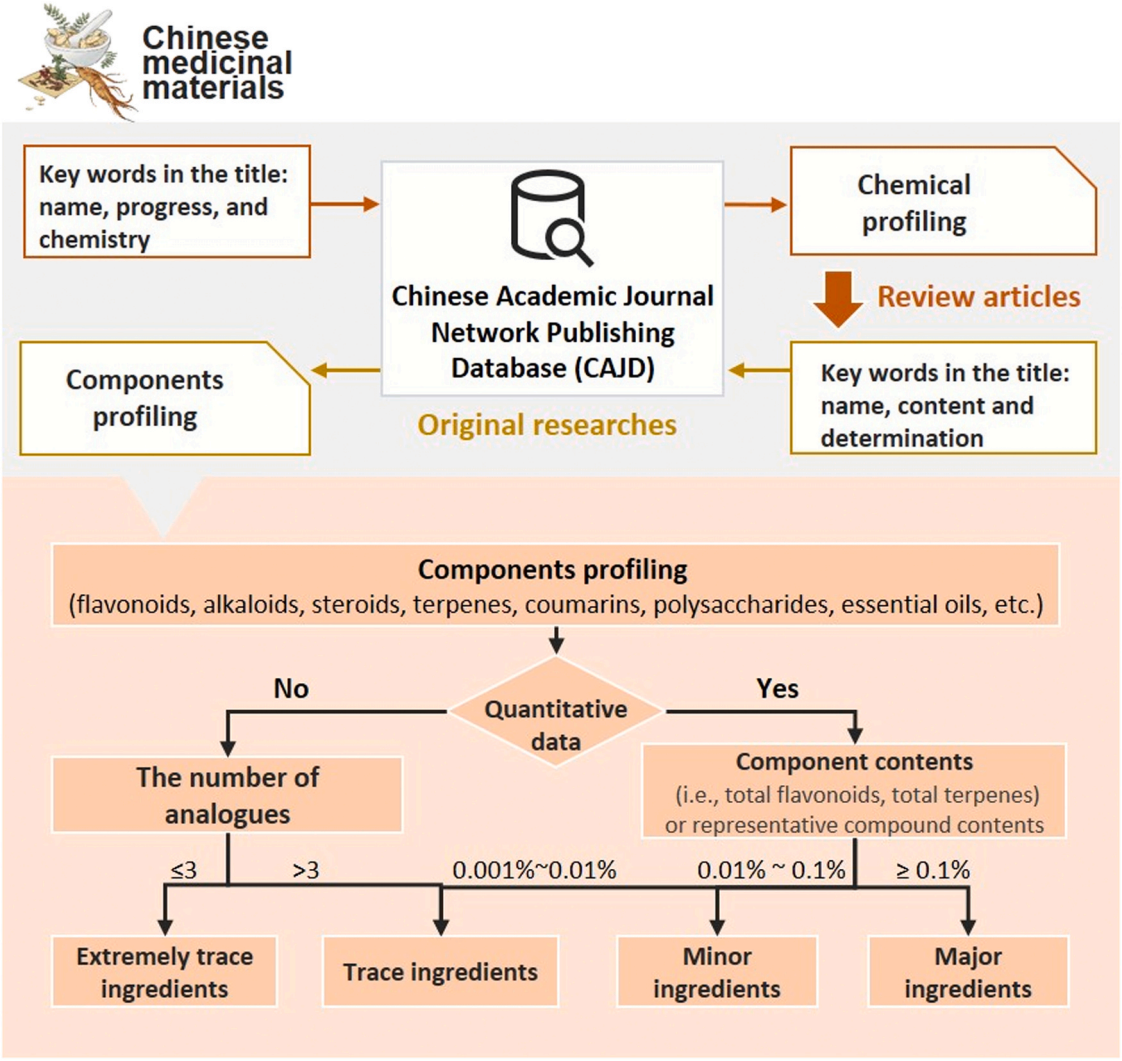


Fig. 1. Schematic of document retrieval, quantitative data collection, and ingredient rating. Related papers were collected through Chinese Academic Journal Network Publishing Database (CAJD). These articles were selected as candidates, in the context of which the botanical origin of Chinese herbs and the content unit of quantified ingredients is clearly clarified. The inclusion criteria of ingredient content data into ccTCM can be referring to supplementary document 3. The basis for the setting of major ingredients was as follows: if the patient was given 10 g of medicinal materials per day, the value of 0.1 % (g/g) indicated that the patient can take 10 mg of ingredients. In fact, most of drugs are orally used at a dosage of not less than 10–20 mg per day. Particularly, alkaloids were identified as major ingredients if the content of total alkaloids or a representative compound exceeded 0.01 %. When the content of polysaccharides, aliphatic organic acids or fatty oils in medicinal materials exceeded 10 %, they can be considered as major ingredients.

ccTCM, including aliphatic organic acids, alkaloids, benzyls, caffeoylquinic acids, chromones, coumarins, diarylheptanoids, essential oils, fatty oils, flavonoids, inorganic compounds, lignans, nucleosides, phenanthrenes, phenols, phenylethanols, phenylpropanoids, polyacetylenes, polypeptides, polysaccharides, quinones, steroids, stilbenes, tannins, terpenes, and others. Each major category contained some minor categories and subcategories, with a total of 115 minor categories and 132 subcategories (Supplementary Table S2). These ingredients were regarded as major ingredients when component contents (i.e., total flavonoids and total terpenes) or representative compound contents were equal to or greater than 0.1 % (g/g). Minor and trace ingredients were defined as component contents or representative compound contents of 0.01–0.1 % (g/g) and 0.001–0.01 % (g/g), respectively. If no quantitative data existed, such ingredients that possessed more than three analogs were regarded as trace ingredients. As regards the weight factor of ingredients in TCM, we ranked major ingredients, minor ingredients, and trace ingredients as 1, 0.3, and 0.1.

2.3. Compound–target relationships

We collected compound-target relationships primarily by integrating multiple reliable databases, such as Human Metabolome Database (HMDB, v5.0) [26], DrugBank v5.1.9 [27], Comparative Toxicogenomics Database (CTD, 2022–04) [28], Natural Product Activity and Species Source (NPASS v2022) [29], and Collective Molecular Activities of Useful Plants (CMAUP, v1.0) [30]. To avoid omission of information, we used compound names and their synonyms for matching. The literature links of compound primary targets were provided when the PubMed IDs were available. To obtain the targets for all compounds, first use PubChem's PUG REST web service to obtain synonyms for each compound, and then use the compound name and all its synonyms as the search item to search from the downloaded HMDB, DrugBank, CTD, NPASS, and CMAUP data files. For the same compound hitting the same target in multiple databases, merge and note its source. The processing details of compound-target relationships can be referring to supplementary document 2.

2.4. Implementation of ccTCM

The ccTCM database was developed on the PostgreSQL database (v14.0) and Django server framework (v3.2). Its web interfaces were built using the Vue3 framework, and ECharts was used for front-end visualization. The entire database was designed to enable the access of its entries by TCM and compounds by using multiple browse and search facilities. When applicable, the compound entries were cross-linked to the PubChem, CTD, and ZINC databases. The relevant pieces of literature on compound–target relation was provided by PubMed identifiers and cross-linked to PubMed. ccTCM is freely accessible at http://www.cctcm.org.cn without a need for user registration. The website is compatible with most major browsers. Enrichment analysis was conducted using the R package "clusterProfiler" (v4.2.2) (Yu et al., 2012), and networkx (https://networkx.org/, v2.6.3) was used for network-module analysis and net-properties calculations including diameter, clustering coefficient, closeness centrality, and betweenness centrality.

3. Results

3.1. Database statistics

ccTCM currently contains 273 Chinese medicinal materials containing 1449 unique compounds targeting 9880 proteins. We collected a total of 1248 records of TCM component or representative compound contents with 1073 supporting literature, a total of 2757 TCM-compound content pairs with 1126 supporting literature, and 40,767 compound–target pairs (Table 1).

On the ccTCM main page, users can view the sunburst plot containing all Chinese medicinal materials and click the tick next to the TCM name to open the detail page. These Chinese medicinal materials were classified into 7 categories according to TCM function, and each category was further divided into 29 subcategories (Supplementary Table S1).

3.2. Browsing and searching Chinese medicinal materials, compounds, and literature

Users can view all Chinese medicinal materials, compounds, and literature through the resource browser. The TCM category filter can help users screen the list of Chinese medicinal materials. Similarly, compound browsing can also be filtered by a major category filter. Users can specify a range of years to view the list of available literature.

The resource browser also provides different angles for users to view the data contained in ccTCM. The component profile lists the weight factors of components included in each TCM by using the numbers 1, 0.3, and 0.1. In the component content page, users can view the content data of components or representative compounds in each TCM, and each record provided the corresponding literature. The compound-content page lists the quantified compounds in each TCM, whose average contents in Chinese medicinal materials are generally more than 0.01 % (g/g). Each compound was given a structural classification including major category, minor category, and subcategory.

The search page is convenient for users to search for wanted Chinese medicinal materials, compounds, and targets included in ccTCM. The search keywords can be the names of Chinese medicinal materials or compounds in English or Chinese.

On each TCM page, users can visit the metadata, origin picture, identification pictures, component profiling, compound contents, and corresponding targets on which the compounds act (Fig. 2). On each compound page, users can visit the molecular formula, molecular weight, complexity, classification, properties, cross-references and corresponding targets (Fig. 3).

3.3. Pot function, take Gegen Qinlian Tang as an example

Gegen Qinlian Tang is mostly used in diarrhea and diabetes clinically. This prescription contains 15 g of Puerariae Lobatae Radix, 9 g of Scutellariae Radix, 9 g of Coptidis Rhizoma, and 6 g of Glycyrrhizae Radix Et Rhizoma. ccTCM provides a Pot function like a shopping cart for users to customize the prescription on their own. On the Pot page, the prescription can be named, and the TCM quantity can also be modified or even deleted (Fig. 4A). At the bottom of the page, users can view all the compounds and their quantitative information contained in the current prescription (Fig. 4B). To demonstrate the reliability of the quantitative information provided by ccTCM, we compared it with the measurement data in the reference [31], and the comparison results are shown in Table 2. Spearman's rank correlation analysis showed that the quantitative data provided by ccTCM and the data measured by Li et al. had high consistency ($r = 0.943$, P value=0.005). The details of the formulation using Pot function and subsequent molecular mechanism analysis can be referring to supplementary document 3.

3.4. Resource distribution, similarity analysis, and molecular-mechanism analysis

The resource distribution of ingredients in Chinese medicinal materials can be viewed by selecting or specifying compounds or components. ccTCM uses cascade mode to facilitate users to select the component object (Fig. 5A). For example, the user wants to find the

Table 1
Overview of data from peering databases.

Items	Data type	ccTCM	TCMSP	TCMID 2.0	ETCM	HERB	HIT 2.0
Published year			2014	2018	2019	2020	2022
Chinese medicinal materials	The total number	273	499	8159	402	7263	1237
	Origin images	Yes	No	n.a.	No	No	No
	Identification images	Yes	No	n.a.	Yes	No	No
Compounds	The total number	1449	29,384	43,413	7284	49,258	1284
	Classification	Yes	No	n.a.	No	No	No
	Drug likeness	Yes	No	n.a.	Yes	No	No
	ADMET properties	Yes	Yes	n.a.	Yes	No	No
Quantification information	Quantified ingredients	1248	0	0	0	0	0
	TCM-compound content pairs	2757	0	0	0	0	0
Targets	The total number	9880	3311	82	2266	12,933	2208
	Source by literature supported	Yes	No	n.a.	No	Yes	Yes
	Compound-target pairs	40,767	84,260	n.a.	n.a.	4815	10,031
Literature	The total number	7027	1288	n.a.	n.a.	1966	7100
Online analysis	Resource distribution and similarity analysis	Yes	No	No	No	No	No
	Molecular-mechanism analysis (network analysis and enrichment analysis (KEGG and GO))	Yes	No	No	Yes	Yes	No

n.a.: not available.

A

Leonuri Herba

Yi Mǔ Cǎo (ccTCM000474)
益母草

Taxonomy: *Lamiaceae*

Origin: *Leonurus Japonicus Houtt.*

Use part: Aboveground Part

Feature: slightly cold; bitter, bitter

Usage: 9~30g; fresh product 12~40g.

Toxicity: Non-Toxic

1 − + Add

B

C

Component Profiling

Flavonoids Total Flavonoids (0.04~0.33%)

Alkaloids Total Alkaloids (0.04~0.5%)

Alkaloids Total Alkaloids (0.2~1.9%)

Polysaccharides Total Polysaccharides (~3%)

Polysaccharides Total Polysaccharides (~7%)

Essential oils Essential Oils (~0.08%)

D

Similar TCMs

Cyathulae Radix
Chuanxiong Rhizoma
Salviae Miltiorrhizae Radix Et Rhizoma
Curcumae Rhizoma
Carthami Flos
Spatholobi Caulis
Curcumae Longae Rhizoma
Myrrha

E

List of quantified compounds in **Leonuri Herba**

Major category	Minor category	Subcategory	Compound	Content (Essential oils)	Content (Chinese medicinal materials)	Literature
Alkaloids	Pyrrolidines	quaternary a...	stachydrine hydrochloride	NA	0.6~1.3%	ccRef01393
Alkaloids	Organic Ami...	guanidines	leonurine hydrochloride	NA	0.10~1.20%	ccRef01394
Alkaloids	Pyrrolidines	quaternary a...	stachydrine hydrochloride	NA	0.01~0.61%	ccRef01434
Flavonoids	Flavonols		rutin	NA	0.07~0.47%	ccRef01394
Alkaloids	Organic Ami...	guanidines	leonurine hydrochloride	NA	0.06~0.36%	ccRef01393
Flavonoids	Flavonols		quercetin	NA	0.05~0.31%	ccRef01842
Flavonoids	Flavonols		isoquercitrin	NA	0.04~0.31%	ccRef01394

Fig. 2. The metadata of Leonurus japonicas is taken as an example (A): taxonomy, origin, medicinal part, feature, usage, and toxicity. (B) The identification image of the morphology of L. japonicas. (C) Component profiling of L. japonicas. (D) Lists of similar TCMs belonging to the same functional category. (E) List of all the quantified compounds in L. japonicas with a content ratio greater than 0.001 %.

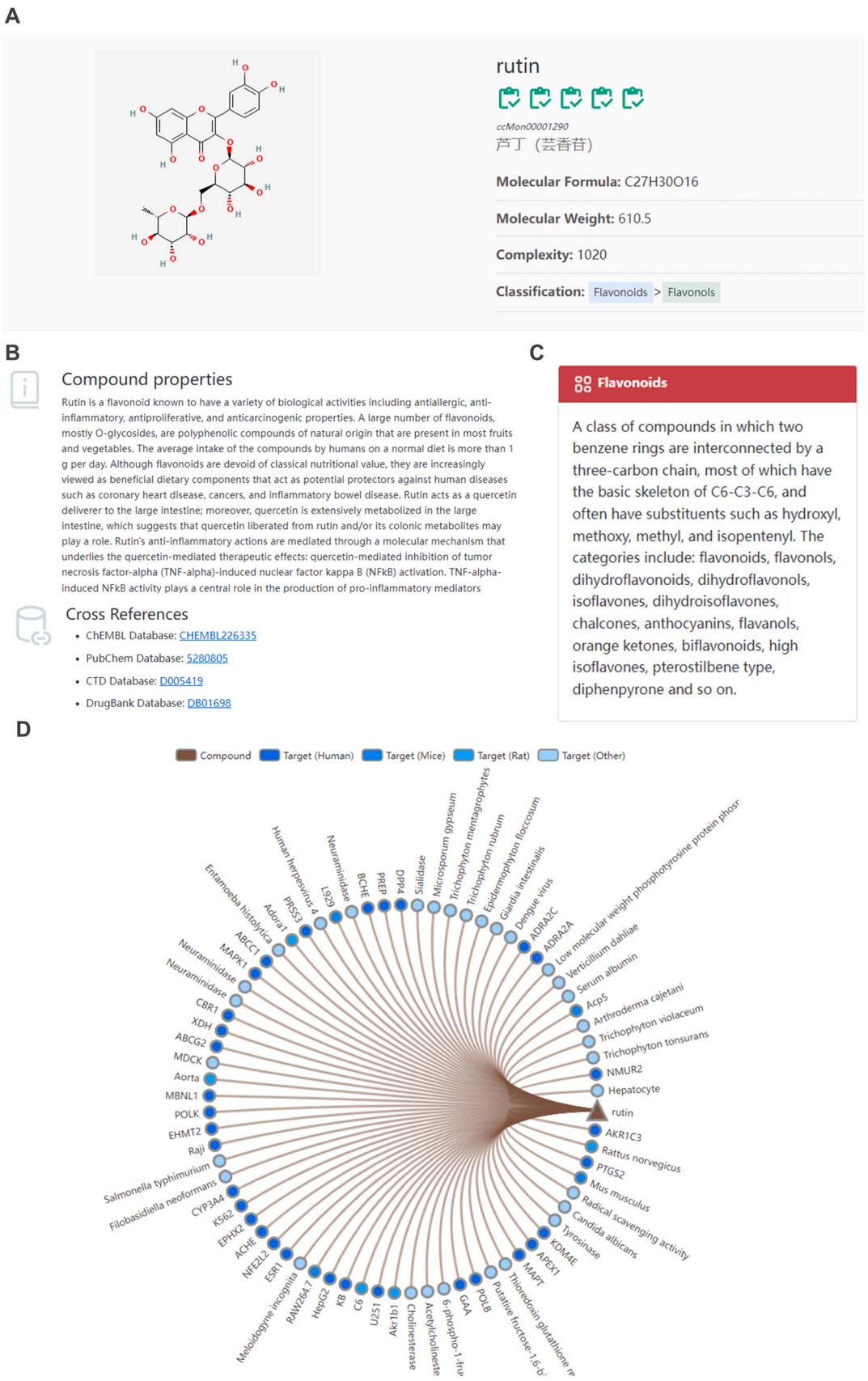
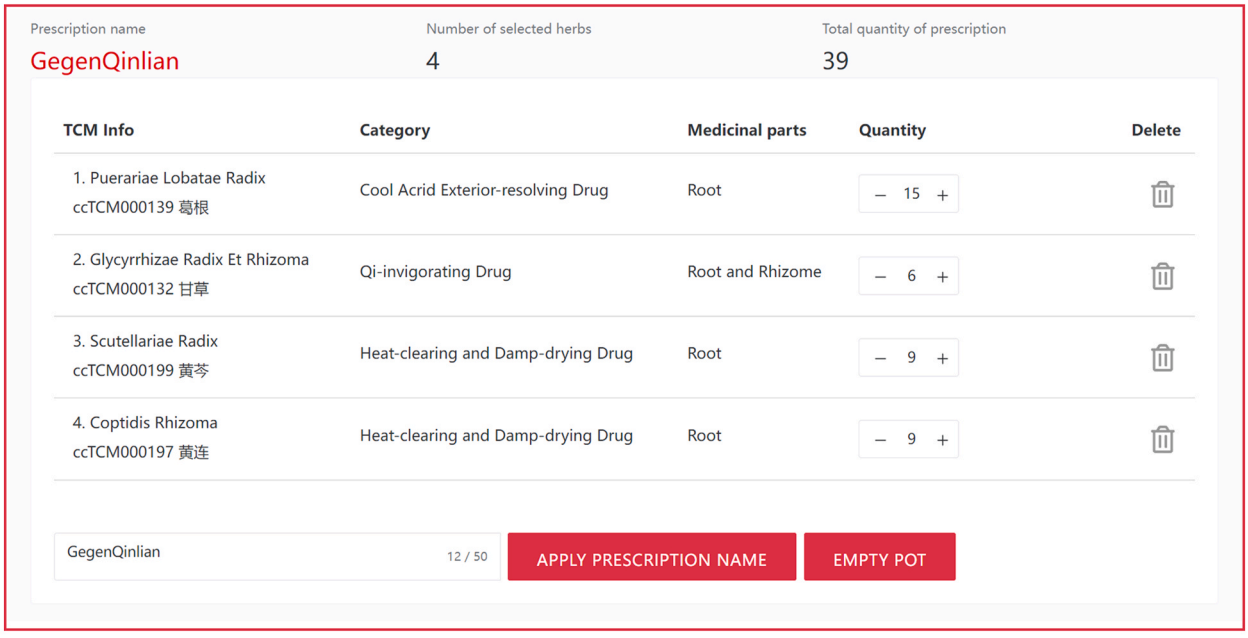


Fig. 3. Information of a specific compound. (A) The metadata of rutin is taken as an example: molecular formula, molecular weight, complexity, and classification. (B) Description of compound properties and available cross-references. (C) Functional description of the main category to which the compound rutin belongs. (D) Network of relation between rutin and targets in humans, mice, rats, and other organisms.

A



B

List of quantified compounds

Compound	TCM(s)	Major Category	Minior Category	Subcategory	Content (%, g/g Prescr...
baicalin	Scutellariae Radix(黄芩)	Flavonoids	Flavone-Type		3.16154
berberine	Coptidis Rhizoma(黄连)	Alkaloids	Isoquinolines	berberine-type	1.51154
glycyrrhizic acid	Glycyrrhizae Radix Et Rhiz...	Terpenes	Pentacyclic Triterpenoids	oleanane-type	1.09231
puerarin	Puerariae Lobatae Radix(...	Flavonoids	Isoflavones		0.90385
3'-methoxy puerarin	Puerariae Lobatae Radix(...	Flavonoids	Isoflavones		0.78846
wogonoside	Scutellariae Radix(黄芩)	Flavonoids	Flavone-Type		0.49615
coptisine	Coptidis Rhizoma(黄连)	Alkaloids	Isoquinolines	berberine-type	0.48462
baicalein	Scutellariae Radix(黄芩)	Flavonoids	Flavone-Type		0.37846
palmatine	Coptidis Rhizoma(黄连)	Alkaloids	Isoquinolines	berberine-type	0.36923
3'-hydroxy puerarin	Puerariae Lobatae Radix(...	Flavonoids	Isoflavones		0.32692
daidzin	Puerariae Lobatae Radix(...	Flavonoids	Isoflavones		0.32308
epiberberine	Coptidis Rhizoma(黄连)	Alkaloids	Isoquinolines	berberine-type	0.3
liquiritin	Glycyrrhizae Radix Et Rhiz...	Flavonoids	Flavanones		0.29231
oroxyloside	Scutellariae Radix(黄芩)	Flavonoids	Flavone-Type		0.28846
luteolin	Scutellariae Radix(黄芩)	Flavonoids	Flavone-Type		0.26538
puerarinapioside	Puerariae Lobatae Radix(...	Flavonoids	Isoflavones		0.21923
liquiritin apioside	Glycyrrhizae Radix Et Rhiz...	Flavonoids	Flavanones		0.19615

Fig. 4. Pot function. (A) In the pot function, users can customize the quantity of various Chinese medicinal materials. (B) All TCM compounds and their quantitative information contained in the current prescription.

Table 2
Comparison of quantitative information of compounds in GQT.

Compound	**Source**	Content in GQT extract (%) [28]	Median content (ccTCM) (%, g/g Prescription)	Content (ccTCM) (%, g/g Prescription)
Puerarin	*Pueraria lobata (Willd) Ohwi.*	2.87	0.90385	0.115–1.692
Daidzein	*Pueraria lobata (Willd) Ohwi.*	0.77	0.01538	0.0038–0.0269
Liquiritin	*Glycyrrihiza uralensis Fisch*	1.42	0.29231	0.05–0.54
Baicalin	*Scutellaria baicalensis Georgi.*	28.84	3.16154	2.17–4.15
Baicalein	*Scutellaria baicalensis Georgi.*	9.72	0.37846	0.04–0.72
Berberine	*Coptis chinensis Franch.*	18.93	1.51154	1.2–1.8

distribution data of oxindole-type alkaloids in Chinese medicinal materials. First, alkaloids in the drop-down box of the major category are selected, and then indoles in the drop-down box of minor category are selected. Finally, the oxindole-type alkaloid in the drop-down box of subcategory is selected. Users can directly type the name of the compound in the dialog box to view the distribution in Chinese medicinal materials (Fig. 5B).

The TCM similarity analysis service provides a comparison of Chinese medicinal materials from three aspects: major category, minor category, and compound. By using the Pot function, users add TCM to the Pot and specify the quantity. The analysis method uses Spearman's rank correlation coefficient, and the analysis results are displayed in the form of heat map (Fig. 6A).

Molecular-mechanism analysis refers to network analysis and enrichment analysis (Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling-pathway enrichment analysis and gene ontology (GO) functional-module enrichment analysis) according to the quantified compounds in Chinese medicinal materials and the targets they act on. The currently accepted species are *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*. ccTCM provides three types of networks (Compound Target Network, Weighted Compound Target Network, and Module Identified Network).

As for the Compound Target Network, the box represents TCM, the triangle represents compound, and the circle represents gene. Different colors represent different classifications of Chinese medicinal materials or compounds. The TCM node size corresponds with its quantity in the prescription, and the compound node size corresponds with its quantity in the TCM multiplied by the TCM quantity in the prescription. As regards the Weighted Compound Target Network, the nodes were resized according to their degrees, which was an update of the previous network. A higher content of compounds in the prescription corresponded with a larger size of the compound nodes and corresponding gene nodes connected to them (Fig. 6B). The Module Identified Network is analyzed according to the network-module identification algorithm [32], and different colors represent different possible modules (Fig. 6C). All analysis results are available for user download.

3.5. Special subject of COVID-19

For the treatment of COVID-19, China has accumulated a considerable clinical experience in the aspect of TCM therapy and has proposed many effective prescriptions. The ccTCM platform provided the three prescriptions (Qingfei Paidu Decoction, Huashi Baidu Prescription, and Xuanfei Baidu Prescription) suggested by the State Administration of Traditional Chinese Medicine (http://www.satcm.gov.cn/xinxifabu/meitibaodao/2020–04–17/14712.html, Supplementary Table S3). Users can easily view the contents of the three prescriptions from the home page and carry out molecular-mechanism analysis. We also marked effective traditional Chinese materials for COVID-19 treatment on TCM pages.

4. Discussion

Since 2006, there have been a dozen academic databases of TCM. Although different databases have different focuses, they are generally organized around Chinese medicinal materials, prescriptions, compounds, targets, symptoms, and diseases [9,33]. Among these TCM databases, TCMID 2.0 included 8159 Chinese medicinal materials and 43,413 compounds [33]. TCMSP and ETCM focused on collecting TCM with clear theoretical guidance, including 499 and 402 Chinese medicinal materials with 29,384 and 7284 compounds, respectively [9,34]. Compared to above databases, ccTCM has the lowest number of Chinese medicinal materials. The reason is that the information on TCM components and compounds in ccTCM was not imported from online databases but was collected manually. The quantitative data of TCM ingredients were obtained by searching in the CAJD literature database. Due to the variety and origin of TCM, there is often inconsistency in the quantitative data of ingredients measured in different laboratories. To ensure the reliability of the data, it is necessary to conduct multiple searches and verifications (Refer to supplementary document 1). The number of collected ingredient quantification literature has reached 5777. On average, we have reviewed 21 pieces of literature for each TCM. Such a large workload has led us to prioritize 273 commonly used Chinese medicinal materials. At the beginning of ccTCM establishment, the construction principle was that every piece of data had a reference. The total number of file data referenced by ccTCM has reached 7027, similar with that referenced by HIT 2.0 [35]. Additionally, ccTCM also provided the source and identification images of each TCM.

Among these published TCM databases, HERB has the highest number of compound-target relationships supported by literature, reaching 12,933 [11]. To provide reliable compound-target relationships, ccTCM adopted highly reliable and academically open databases (i.e., HMDB, DrugBank, CTD, NPASS, and CMAUP) as our data sources. CTD was first published in 2004, aiming to provide manually verified chemical-gene/protein interactions, chemical–disease and gene–disease relationships. Till August 2022, CTD has included over 3.4 million evidence-based manually curated interactions. Another valuable database is NPASS, which provides 222,092 natural product-target pairs with 446,552 quantitative activity records (i.e., IC50, Ki, EC50, GI50, or MIC).

Each major database integrates bioinformatics analysis functions to accelerate the discovery of new drugs. For example, ETCM provides enrichment analysis for each collected formula [34], while HERB [11] and ITCM [14] provides transcriptomic data query functions, aiming to improve the effective screening of active compounds in TCM. ccTCM also provides users with network analysis and enrichment analysis. Moreover, users are allowed to quickly browse a list of all quantitative compounds in a single TCM or prescription before conducting analysis. Accurate quantitative information is well known to play a crucial role in expediting the discovery of effective ingredients in TCM, thereby promoting the development of novel drugs. To maximize the utilization of the quantitative ingredients in ccTCM, users can easily search for the distribution of a certain compound or a class of compounds (resource distribution analysis). It is also possible to do similarity analysis of different Chinese medicinal materials based on component major categories, compound subclasses, or compounds.

5. Conclusion

The novelty of the ccTCM database includes the following: (i) ccTCM

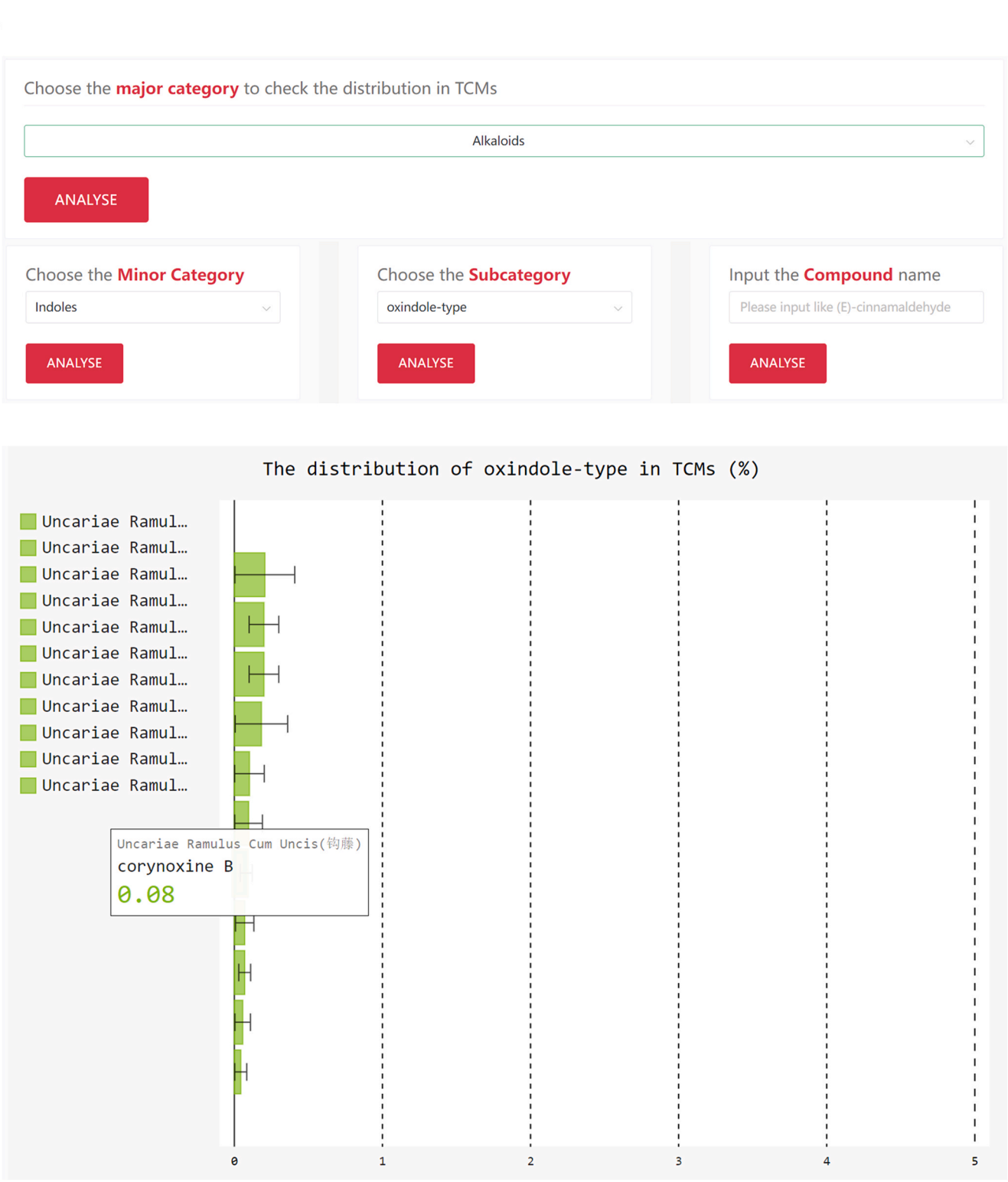


Fig. 5. Presentation of resource-distribution analysis. (A) The drop-down menu provides users with resource analysis at different aspects (major category, minor category, and subcategory) in a cascading manner. (B) Bar plot of distribution of Bisepoxylignans in ccTCM.

is the first available database containing quantitative component and compound data in Chinese medicinal materials; (ii) ccTCM integrates the Pot function for the user-defined analysis of molecular mechanism of TCM, visualized-distribution profiles of components and compounds in Chinese medicinal materials, and similarity analysis of different Chinese medicinal materials from three aspects (major category, minor category, and compounds). (iii) ccTCM is the first available database providing structural classification of natural compounds. The current version of ccTCM contains a total of 273 Chinese medicinal materials and covers almost all functional classifications of TCM. Nevertheless, some Chinese medicinal materials have not been collected yet. We plan to add more Chinese medicinal materials into the ccTCM database in the future. We will also try to integrate TCM theory into ccTCM and provide a more comprehensive and useful TCM database.

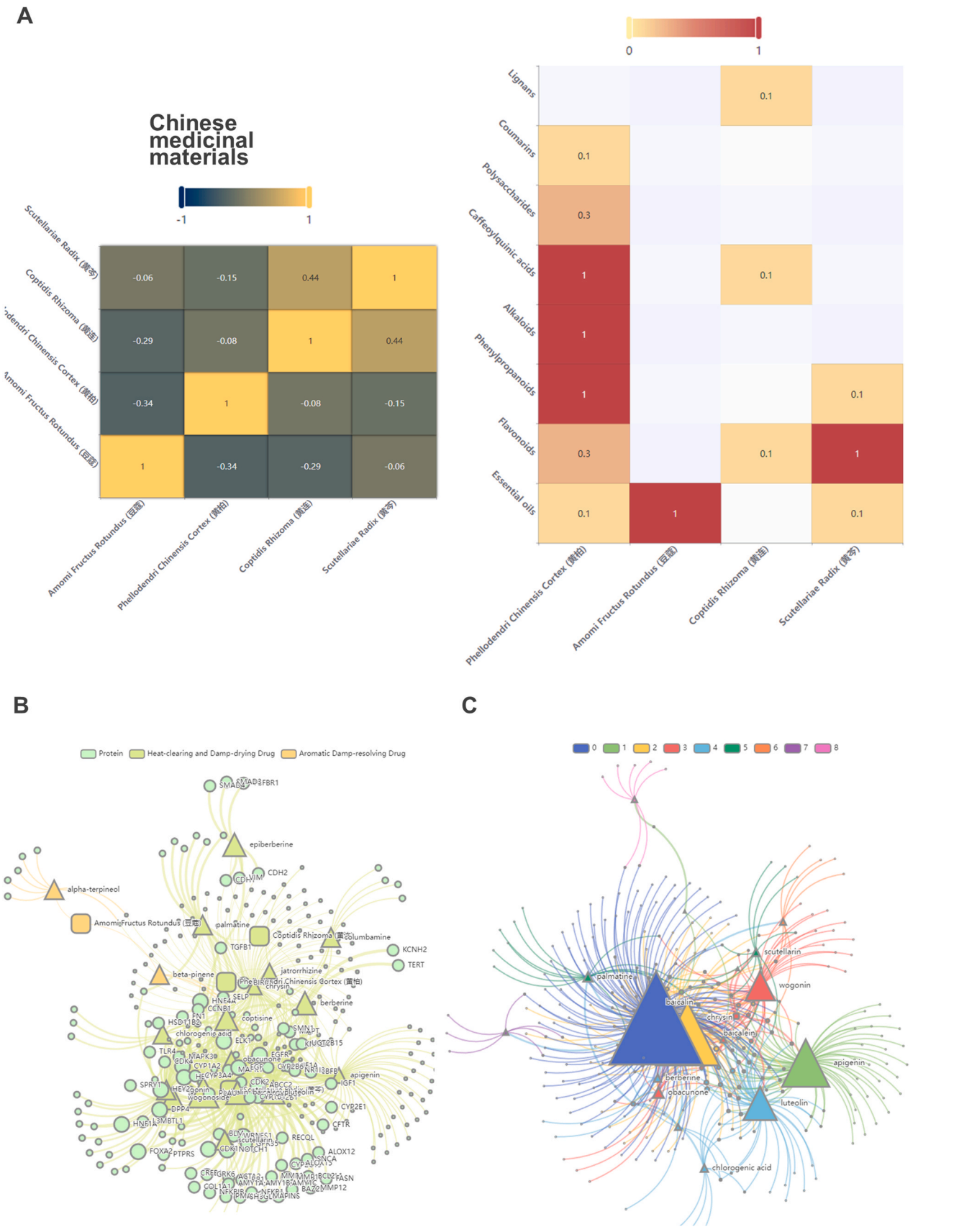


Fig. 6. Presentation of enrichment-analysis results. (A) Similarity analysis uses a similarity matrix to reflect the Spearman correlation among different Chinese medicinal materials. A heatmap showing the contents of compounds in each TCM. (B) Weighted compound-target network, with the nodes resized according to their degrees. (C) Module-identified network is analyzed according to the network module-identification algorithm, and different colors represent different possible modules. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

CRediT authorship contribution statement

Yunan Zhao and **Dongqing Yang:** Conceptualization, Writing – original draft. **Zhu Zhu**, **Qi Yao**, **Cuihua Chen**, **Feiyan Chen**, **Ling Gu**, **Yucui Jiang**, and **Lin Chen:** Data curation. **Dongqing Yang:** Methodology, Software. **Jingyuan Zhang**, **Juan Wu**, and **Xingsu Gao:** Validation. **Junqin Wang** and **Guochun Li:** Methodology. **Yunan Zhao** and **Dongqing Yang**: Writing – review & editing. All authors have seen and approved the final version of the manuscript being submitted. We warrant that the article is the authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Data availability

All data provided by ccTCM is accessible for free at http://www.cctcm.org.cn/.

Acknowledgements

The authors thank Weichao Xu for authorizing ccTCM to use his TCM images. The study was financially supported by National Natural Science Foundation of China (No.82204647, 82003970), and supporting project of National Natural Youth Foundation of Nanjing University of Chinese Medicine (XPT82204647), and supporting project of Jiangsu Province "The 14th Five-year Plan" Key Discipline-Public Health and Preventive Medicine (035091005007).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.11.030.

References

[1] Oravecz M, Mészáros J. Traditional Chinese medicine: theoretical background and its use in China. Orv Hetil 2012;153:723–31. https://doi.org/10.1556/OH.2012.29365.
[2] Guo X, Chen X, Chen J, Tan Z, Yang Y, Zhang H. Current status and evaluation of randomized clinical trials of traditional Chinese medicine in the treatment of cardiovascular diseases. Evid Based Complement Alter Med ECAM 2022;2022:6181862. https://doi.org/10.1155/2022/6181862.
[3] Fan Y-L, Liu R-Z, Tan Q, Zhao H-L, Song M, Wang R, et al. A database-guided integrated strategy for comprehensive chemical profiling of traditional Chinese medicine. J Chromatogr A 2022;1674:463145. https://doi.org/10.1016/j.chroma.2022.463145.
[4] Wang M, Chen L, Liu D, Chen H, Tang D-D, Zhao Y-Y. Metabolomics highlights pharmacological bioactivity and biochemical mechanism of traditional Chinese medicine. Chem Biol Interact 2017;273:133–41. https://doi.org/10.1016/j.cbi.2017.06.011.
[5] Chen X, Zhou H, Liu YB, Wang JF, Li H, Ung CY, et al. Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation. Br J Pharmacol 2006;149:1092–103. https://doi.org/10.1038/sj.bjp.0706945.
[6] Chen CY-C. 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.
[7] Wu Y, Zhang F, Yang K, Fang S, Bu D, Li H, et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. Nucleic Acids Res 2019;47:D1110–7. https://doi.org/10.1093/nar/gky1021.
[8] Xue R, Fang Z, Zhang M, Yi Z, Wen C, Shi T. TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res 2012;41:D1089–95. https://doi.org/10.1093/nar/gks1100.
[9] Ru J, Li P, Wang J, Zhou W, Li B, Huang C, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Chemin 2014;6:13. https://doi.org/10.1186/1758-2946-6-13.
[10] Liu Z, Cai C, Du J, Liu B, Cui L, Fan X, et al. TCMIO: a comprehensive database of traditional Chinese medicine on immuno-oncology. Front Pharmacol 2020;11:439. https://doi.org/10.3389/fphar.2020.00439.
[11] Fang S, Dong L, Liu L, Guo J, Zhao L, Zhang J, et al. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. Nucleic Acids Res 2021;49:D1197–206. https://doi.org/10.1093/nar/gkaa1063.
[12] Li X, Tang Q, Meng F, Du P, Chen W. INPUT: an intelligent network pharmacology platform unique for traditional Chinese medicine. Comput Struct Biotechnol J 2022;20:1345–51. https://doi.org/10.1016/j.csbj.2022.03.006.
[13] Li X, Ren J, Zhang W, Zhang Z, Yu J, Wu J, et al. LTM-TCM: a comprehensive database for the linking of Traditional Chinese Medicine with modern medicine at molecular and phenotypic levels. Pharmacol Res 2022;178:106185. https://doi.org/10.1016/j.phrs.2022.106185.
[14] Tian S, Zhang J, Yuan S, Wang Q, Lv C, Wang J, et al. Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. Brief Bioinform 2023;24:bbad027. https://doi.org/10.1093/bib/bbad027.
[15] Zhang R, Zhu X, Bai H, Ning K. Network pharmacology databases for traditional Chinese medicine: review and assessment. Front Pharmacol 2019;10:123. https://doi.org/10.3389/fphar.2019.00123.
[16] Wu H, Guo J, Chen S, Liu X, Zhou Y, Zhang X, et al. Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography-mass spectrometry. J Pharm Biomed Anal 2013;72:267–91. https://doi.org/10.1016/j.jpba.2012.09.004.
[17] Ren J-L, Zhang A-H, Kong L, Han Y, Yan G-L, Sun H, et al. Analytical strategies for the discovery and validation of quality-markers of traditional Chinese medicine. Phytomed Int J Phytother Phytopharm 2020;67:153165. https://doi.org/10.1016/j.phymed.2019.153165.
[18] Wu M, Ma S, Wu M, Cao H, Zhang Y, Ma Z. Simultaneous qualitative and quantitative analysis of 10 bioactive flavonoids in Aurantii Fructus Immaturus (Zhishi) by ultrahigh-performance liquid chromatography and high-resolution tandem mass spectrometry combined with chemometric methods. Phytochem Anal PCA 2022;33:710–21. https://doi.org/10.1002/pca.3122.
[19] Tan M, Chen J, Wang C, Zou L, Chen S, Shi J, et al. Quality evaluation of ophiopogonis radix from two different producing areas. Molecules 2019;24:3220. https://doi.org/10.3390/molecules24183220.
[20] Sun P, Tong J, Li X. Evaluation of the effects of paclobutrazol and cultivation years on saponins in Ophiopogon japonicus using UPLC-ELSD. Int J Anal Chem 2020;2020:5974130. https://doi.org/10.1155/2020/5974130.
[21] Luo S, Ren X, Shi X, Zhong K, Zhang Z, Wang Z. Study on enhanced extraction and seasonal variation of secondary metabolites in Eucommia ulmoides leaves using deep eutectic solvents. J Pharm Biomed Anal 2022;209:114514. https://doi.org/10.1016/j.jpba.2021.114514.
[22] Han L, Wang R, Zhang X, Yu X, Zhou L, Song T, et al. Advances in processing and quality control of traditional Chinese medicine Coptidis rhizoma (Huanglian): a review. J AOAC Int 2019;102:699–707. https://doi.org/10.5740/jaoacint.18-0303.
[23] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. Nucleic Acids Res 2023;51:D1373–80. https://doi.org/10.1093/nar/gkac956.
[24] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res 2017;45:D945–54. https://doi.org/10.1093/nar/gkw1074.
[25] Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. J Chem Inf Model 2015;55:2324–37. https://doi.org/10.1021/acs.jcim.5b00559.
[26] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 2018;46:D608–17. https://doi.org/10.1093/nar/gkx1089.
[27] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34:D668–72. https://doi.org/10.1093/nar/gkj067.
[28] Davis AP, Wiegers TC, Johnson RJ, Sciaky D, Wiegers J, Mattingly CJ. Comparative Toxicogenomics Database (CTD): update 2023. Nucleic Acids Res 2022:gkac833. https://doi.org/10.1093/nar/gkac833.
[29] Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, et al. NPASS: natural product activity and species source database for natural product research, discovery and tool development. Nucleic Acids Res 2018;46:D1217–22. https://doi.org/10.1093/nar/gkx1026.
[30] Zeng X, Zhang P, Wang Y, Qin C, Chen S, He W, et al. CMAUP: a database of collective molecular activities of useful plants. Nucleic Acids Res 2019;47:D1118–27. https://doi.org/10.1093/nar/gky965.
[31] Li R, Chen Y, Shi M, Xu X, Zhao Y, Wu X, et al. Gegen Qinlian decoction alleviates experimental colitis via suppressing TLR4/NF-κB signaling and enhancing antioxidant effect. Phytomed Int J Phytother Phytopharm 2016;23:1012–20. https://doi.org/10.1016/j.phymed.2016.06.010.
[32] Wang X, Dalkic E, Wu M, Chan C. Gene module level analysis: identification to networks and dynamics. Curr Opin Biotechnol 2008;19:482–91. https://doi.org/10.1016/j.copbio.2008.07.011.
[33] Huang L, Xie D, Yu Y, Liu H, Shi Y, Shi T, et al. TCMID 2.0: a comprehensive resource for TCM. Nucleic Acids Res 2018;46:D1117–20. https://doi.org/10.1093/nar/gkx1028.
[34] Xu H-Y, Zhang Y-Q, Liu Z-M, Chen T, Lv C-Y, Tang S-H, et al. ETCM: an encyclopaedia of traditional Chinese medicine. Nucleic Acids Res 2019;47:D976–82. https://doi.org/10.1093/nar/gky987.
[35] Yan D, Zheng G, Wang C, Chen Z, Mao T, Gao J, et al. HIT 2.0: an enhanced platform for Herbal Ingredients' Targets. Nucleic Acids Res 2022;50:D1238–43. https://doi.org/10.1093/nar/gkab1011.