DCABM-TCM, a Database of Constituents Absorbed into Blood and Metabolites of Traditional Chinese Medicine

Xinyue Liu

Beijing Proteome Research Center   https://orcid.org/0000-0001-7905-3831

Jinying Liu

College of Traditional Chinese Medicine, Chengde Medical University   https://orcid.org/0000-0001-5350-0373

Feng Xu

School of Pharmaceutical Sciences, Peking University   https://orcid.org/0000-0002-9642-6406

Runa Li

School of Biomedicine, Beijing City University   https://orcid.org/0000-0003-4457-4375

Lin Xing

School of Biomedicine, Beijing City University   https://orcid.org/0000-0001-5263-4038

Liying Yuan

Hebei University College of Life Sciences   https://orcid.org/0000-0002-3546-429X

Shuzhen Guo

Beijing University of Chinese Medicine, School of Traditional Chinese Medicine   https://orcid.org/0000-0002-8980-9812

Feifei Guo

China Academy of Chinese Medical Sciences   https://orcid.org/0000-0003-3624-4704

Yuan Liu

Beijing Proteome Research Center   https://orcid.org/0000-0002-2866-4904

Jiale Liu

Beijing Proteome Research Center   https://orcid.org/0000-0001-7841-7728

Xun Wang

Beijing Proteome Research Center   https://orcid.org/0000-0003-4559-5450

Yaning Qi

Beijing Proteome Research Center   https://orcid.org/0000-0002-7303-0417

Bangze Fu

Beijing City University   https://orcid.org/0000-0003-0764-0951

Dong Li

Beijing Proteome Research Center   https://orcid.org/0000-0002-8680-0468

Zhongyang Liu (✉ liuzy1984@163.com)

State Key Laboratory of Proteomics, Beijing Proteome Research Center,National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics   https://orcid.org/0000-0002-3371-2392

Database

Abstract

Traditional Chinese Medicine (TCM) not only maintains the health of peoples in Asia but also provides a great resource of active natural products for modern drug development. However, owing to its diverse constituents and the complex interactions with the human body, it is still challenging to clarify its effective constituents and molecular mechanism. This greatly hinders its widespread application worldwide and TCM-based modern drug development. The vast majority of constituents in TCM can produce pharmacological effects only after they are absorbed into bloodstream. Compared with ordinary constituents generally detected *in vitro*, these that are verified to reach the bloodstream, including original constituents absorbed into blood and metabolites of original constituents produced by the gastrointestinal tract, intestinal microflora, and liver, are more likely to be really effective ones responsible for the pharmacological effects. Here we developed DCABM-TCM (a Database of Constituents Absorbed into Blood and Metabolites of TCM), the first database systematically collecting blood constituents of TCM prescriptions and herbs, including prototypes and metabolites experimentally-detected in blood, together with the corresponding detailed detection conditions by manual literature mining. Currently DCABM-TCM has collected 4206 blood constituents of 192 prescriptions and 194 herbs, and also integrated various related annotations including physicochemical properties, ADMET properties, and associated targets, functional terms, pathways, and diseases *etc*. Furthermore, DCABM-TCM supported two analysis functions, the network pharmacology analysis for TCM molecular mechanism elucidation and the target/pathway/disease-based screening of candidate blood constituents, herbs or prescriptions for TCM-based drug discovery. DCABM-TCM is freely assessable by http://bionet.ncpsb.org.cn/dcabm-tcm/ and is browsable and searchable for each of 6 kinds of entities, including prescriptions, herbs, blood constituents, targets, pathways, and diseases. DCABM-TCM will contribute to not only the elucidation of effective constituents and molecular mechanism of TCMs but also the discovery of TCM-derived drug-like compounds which are both bioactive and bioavailable, the feature investigation of absorbable natural compounds, and the quality control of herbs.

Introduction

Traditional Chinese Medicine (TCM) not only plays an important role in maintaining the health of peoples in Asia [1] but also provides a great resource of bioactive natural products for modern drug development. Many successful western drugs are derived from TCMs, including artemisinin extracted from *Artemisia annua* L. for malaria (of which the discoverer won the 2015 Nobel Prize in Physiology or Medicine), ephedrine from *Ephedra Sinica Stapf* for asthma treatment *etc*. Especially, because of its synergistic action of multiple components and multiple targets, TCM research can also inspire the treatment of complex diseases such as cancers and diabetes, caused by abnormal interactions among multiple molecules and pathways [2]. However, owing to TCM's diverse constituents and their complex interactions with the human body, it is still challenging to clarify its effective constituents and molecular mechanism. This greatly hinders not only TCM's more widespread application but also TCM-based modern drug development.

Actually the vast majority of constituents in TCM can produce pharmacological effects only after they are absorbed into the bloodstream [3]. These "constituents detected in blood" (also denoted as "blood constituents" in this paper) include not only original TCM constituents directly absorbed into blood (*i.e.* "constituents absorbed into blood"/"prototypes") but also "metabolites" of original constituents that are produced by the gastrointestinal tract, intestinal microflora, and liver. Compared with ordinary TCM constituents generally detected *in vitro*, these constituents that are verified to reach the bloodstream, including prototypes and metabolites, are more likely to be really active ones responsible for the pharmacological effects. The identification of the constituents detected in blood greatly avoids the false positives and false negatives in the identification of TCM effective constituents, where the false positives refer to those active *in vitro* but cannot be absorbed or become inactive after metabolism and the false negatives refer to those inactive *in vitro* but become active after metabolism.

Systematical collection of these constituents detected in blood of TCMs is important not only for 1) the elucidation of effective material basis and molecular mechanism of TCMs and 2) the selection of quality markers of herbs [4] but also for 3) the discovery of TCM-derived drug-like compounds which are both bioactive and absorbable and 4) the property study of natural compounds which can be absorbed, such as structural, physicochemical features. Compared with ordinary constituents of TCMs generally detected *in vitro*, the identification of blood constituents is a great step forward for all issues above.

Recently blood constituents for more and more TCM prescriptions and herbs were identified by various techniques such as liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). However these data are scattered in various publications. Meanwhile often a prescription/herb's blood constituents were detected by multiple researches. In these researches, detection conditions may be different to some extent, leading to that the detected blood constituents are different to some extent. It is necessary to systematically collect and compile these data.

However, none of current public databases have recorded blood constituent information of TCM prescriptions and herbs. Current TCM databases, including ETCM [5], SymMap [6], TCMID [7], TCM-ID [8], TCM Database@Taiwan [9], TCM-Mesh [10], YaTCM [11], TCMSP [12], and even ETM-DB [13], TM-MC [14], IMPPAT [15], PharmDB-K [16] for herbs used in other countries and CMAUP [17] for a wider range of useful plants only have collected ordinary constituents generally detected *in vitro*. HIM, the herbal ingredients *in-vivo* metabolism database, has recorded *in-vivo* metabolic information of herb-derived single compounds [18]. Different from HIM, here we focused on the prescription/herb as a whole. Compared with HIM, the collection of blood constituents of prescriptions or herbs has unique advantages on the elucidation of effective constituents and multi-component synergistic molecular mechanism of a prescription/herb, the herbal quality control, the property study of absorbable natural products and TCM symptom-inspired discovery of lead compounds derived from TCM (*i.e.* phenotype-based drug discovery) [6].

Therefore we developed DCABM-TCM (a Database of Constituents Absorbed into Blood and Metabolites of Traditional Chinese Medicine) (http://bionet.ncpsb.org.cn/dcabm-tcm/), the first database

systematically collecting blood constituents of TCM prescriptions and herbs together with the corresponding detailed detection conditions, by manual literature mining. In addition, DCABM-TCM was also designed to support two data analysis functions for TCM molecular mechanism elucidation and TCM-based drug discovery.

Materials And Methods

Core data collection and curation

The core data of DCABM-TCM are constituents detected in blood (including prototype components and metabolites) of TCM herbs and prescriptions detected by various experimental techniques such as LC-MS/MS. In addition, considering that experimental conditions have important impact on the detected blood constituents of a prescription or herb, we also recorded the corresponding detailed detection conditions including the extraction method, experimental animal and animal model, administration method and dose, blood collection time and location *etc*. as well as the corresponding source publication.

All these data were manually mined from papers from PubMed and CNKI (China National Knowledge Infrastructure). Like PubMed, CNKI hosts articles published in Chinese. As shown in Figure 1, first, potentially related papers were identified by searching PubMed and CNKI abstracts (by Dec., 2019) with two keyword lists by use of self-developed bio-entity recognizer (which has been successfully used to establish the AAgAtlas database [19]). For English papers from PubMed, the first keyword list included "Traditional Chinese Medicine", "TCM", "Chinese herbal medicine", "CHM", "herb", "formula", "prescription" and their various variations. The other one included 61 phrases that potentially suggested the "constituent detected in blood", such as "constituent absorbed into blood", "absorbed component", "detected in plasma", "serum fingerprint", "serum pharmacochemistry". For Chinese papers from CNKI, corresponding keywords in Chinese were used. The potentially related papers were those that simultaneously contained these two classes of keywords in their abstracts. Next we carried out a preliminary screening by manually browsing all the titles and abstracts. Then for the qualified papers, we manually checked the full texts and even supplementary materials to extract the related information, where Chinese information was simultaneously translated into English. Every piece of manually-curated information was checked by at least three different experts to ensure the data quality.

In DCABM-TCM, blood constituents were uniformly mapped to PubChem CIDs [20] with the help of PubChemPy [21] or manually based on compound names and molecular formulas. For the blood constituents without mapped PubChem CIDs, we directly recorded them as the source publication gave.

Annotation data collection

To help the TCM molecular mechanism elucidation and TCM-based drug discovery, around the core data, we also collected various annotation data. The compositive herbs of a prescription were directly from the source paper which studied its blood constituents. The ordinary constituents of herbs were derived from the TCMID database [7]. For a blood constituent, its basic information such as structure and Chemical

Abstracts Service (CAS) number as well as the chemical and physical properties was from the PubChem database [20] with the help of PubChemPy [21]. Here only six classic physicochemical properties were considered, including the molecular weight, the logarithm of the partition coefficient between n-octanol and water (logP) [22], Topological Polar Surface Area (TPSA), number of hydrogen-bond donors in the structure, number of hydrogen-bond acceptors, and number of rotatable bonds, which are important for computationally evaluating the "drug likeness" [23]. Like the DrugBank database [24], the 22 Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties of the blood constituents were predicted by use of admetSAR [23], including the human intestinal absorption, blood-brain barrier penetration, Caco-2 permeability, P-glycoprotein substrate and inhibitor, CYP450 substrate and inhibitor (CYP1A2, 2C9, 2D6, 2C19, and 3A4), hERG inhibitors, AMES mutagenicity, carcinogens, rat acute toxicity *etc*. The hierarchical classification of blood constituents was given by ClassyFire [25], which provides a rule-based structural classification for compounds.

Furthermore, blood constituent-target gene associations were from BATMAN-TCM [26], including known and predicted ones. For a target gene, its basic annotation information such as synonyms, full name, cross-references was from NCBI Entrez Gene database (downloaded on 05/14/2020) [27]. Human gene-Gene Ontology (GO) term association data were also from Entrez Gene database (downloaded on 05/14/2020); biological pathway data were from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (downloaded on 02/26/2020) [28]. Human gene-disease associations were from the Online Mendelian Inheritance in Man (OMIM) [29] and the Comparative Toxicogenomics Database (CTD) [30] databases. For OMIM, we directly extracted gene-OMIM disease associations from Swiss-Prot (downloaded on 05/15/2020) [31]. For CTD, we only used its gene-disease associations with direct evidence (version: Jul.6, 2017).

The implementation of the two analysis functions

The network pharmacology analysis function of a prescription/herb/blood constituent was implemented by BATMAN-TCM, a bioinformatics analysis tool for molecular mechanism of TCM previously developed by us [26]. The analyses supported by BATMAN-TCM mainly include target prediction of constituents, functional analyses of targets and association network visualization. With respect to BATMAN-TCM released in 2016, in DCABM-TCM, the background data supporting the related analyses, mainly including human gene-GO term/disease associations and pathway data, were updated into the later version as described above.

In the function of "target/pathway/disease-based screening of candidate blood constituents, herbs and prescriptions", for a target gene, the candidate blood constituents targeting it were given with the help of BATMAN-TCM. Further, we thought that the significantly enriched prescriptions/herbs among these candidate blood constituents were potential candidate prescriptions/herbs targeting this target gene. Similarly, for a pathway/disease, we thought that significantly enriched blood constituents among the member genes of the pathway/the known related genes of the disease were potential candidate blood constituents targeting this pathway/disease. And further, significantly enriched herbs/prescriptions

among these candidate blood constituents were thought to be potential candidate herbs/prescriptions targeting this pathway/disease. The enrichment analyses were implemented based on the hypergeometric cumulative distribution test, and the multiple testing correction was based on Benjamini-Hochberg correction method [32].

Database implementation

The bottom of DCABM-TCM was a MongoDB database. Above this database, the analysis application was implemented in Perl and Node.js, and the web presentation application written in JavaScript and CSS. The structural similarity search was implemented based on Open Babel [33], where the FP2 fingerprint and the Tanimoto coefficient were used to compute the structural similarity of two compounds. DCABM-TCM web server is compatible with Chrome, Firefox, and Opera browsers for Windows; and Chrome, Firefox, Opera, and Safari for Mac operating system.

Results And Discussion

Overview of DCABM-TCM

DCABM-TCM can be freely assessed by http://bionet.ncpsb.org.cn/dcabm-tcm/. Its core data are constituents absorbed into blood and metabolites of TCM prescriptions and herbs together with the corresponding detailed detection conditions. Around the core data, annotation data include physicochemical properties, ADMET properties of blood constituents and the associated targets, GO functional terms, pathways, and diseases *etc.* Therefore, in total, in DCABM-TCM there are six kinds of entities: prescriptions, herbs, blood constituents, targets, pathways, and diseases (Figure 2). For every entity of these six types, DCABM-TCM provides a detailed annotation page presenting the related core and annotation information. For each of the six kinds of entities, DCABM-TCM supports data browse and search. DCABM-TCM also supports two data analysis functions: 1) the network pharmacology analysis for a prescription/herb/blood constituent to reveal the potential molecular mechanism, 2) the screening of candidate drugs (including blood constituents, herbs and prescriptions) potentially targeting a target/pathway/disease to help TCM-derived drug discovery. The two analysis functions and corresponding analysis results are presented on the detailed annotation page of the corresponding type of entity. In addition, data download and submission are also supported by DCABM-TCM.

Data statistics and analyses

Currently DCABM-TCM has recorded 4206 constituents detected in blood (among which 1306 can be mapped to PubChem CIDs) of 192 prescriptions and 194 herbs, including 1487 prototypes (703 mapped to PubChem CIDs) and 1783 metabolites (184 mapped to PubChem CIDs) (Table 1) (Additional files 1~3, and they can also be downloaded on "Download" page of DCABM-TCM). In our data, only 1/4 of source publications have stated that among the detected blood constituents of a prescription or herb which are prototypes and which are metabolites. DCABM-TCM involved 7585 prescription/herb-blood constituent

associations. In addition, there were 3838 target genes, further 333 KEGG pathways and 3987 CTD diseases associated with the constituents detected in blood (Table 1).

Table 1 Statistics of DCABM-TCM data

Data type	Number
Prescriptions	192
Herbs	194
Constituents detected in blood (All) (including prototypes and metabolites)	4206
Constituents detected in blood with PubChem CIDs (including prototypes and metabolites)	1306
Prototypes (All)	1487
Prototypes with PubChem CIDs	703
Metabolites (All)	1783
Metabolites with PubChem CIDs	184
Target genes (including the known and predicted ones with scores >= 10)	3838
Constituent detected in blood - target gene associations	47374
KEGG pathways involving the target genes	333
Target gene - KEGG pathway associations	17136
CTD diseases associated with the target genes	3987
Target gene - CTD disease associations	72794

In our data, the median (/average) number of constituents detected in blood is 13 (/22.5) for a prescription and 8 (/16.8) for a herb (Figure 3A and 3B). The median (/average) number of blood constituents with mapped PubChem CIDs is 9 (/11.7) for a prescription and 5 (/7.7) for a herb. By observation, the number of the detected blood constituents of a prescription was often smaller than the sum of the detected blood constituents of its compositive herbs. Most blood constituents detected in a prescription were also the blood constituents of its compositive herbs, while sometimes a few new blood constituents appeared which might be produced by the interactions between its compositive herbs such as those produced in the decocting process of the prescription. These data of constituents detected in blood were mined from 443 papers. In our data, for the majority of prescriptions and herbs, its blood constituents were studied by only one paper, while some of them were studied by multiple papers. On average, the blood constituents of a prescription were studied by 1.5 papers, and 2.8 papers for a herb (Figure 3C and 3D).

The recorded detection conditions mainly included the extraction method, experimental animal and animal model, administration method and dose, blood collection time and location. In our data, the extraction methods included water extraction, ethanol extraction, methanol extraction *etc*. Experimental animals included rats, mice, and rabbits. For the animal model, the majority of the researches used normal models and the minority used disease models such as the models of rheumatoid arthritis, acute heart failure, and cerebral ischemia-reperfusion injury. For the administration method, the vast majority of studies adopted the intragastrical administration (*i.e.*, ig), and the adopted other methods included intravenous injection (*i.e.*, iv), intraperitoneal injection (*i.e.*, ip), intravenous drip, intraduodental administration, intestinal circulatory perfusion *etc*. The administration dose were recorded typically in three forms: 1) "XX g/kg", representing a single administration; 2) "XX g/kg, XX times", representing multiple dosing in a short period of time; 3) "XX g/kg, XX times/day, XX days", representing a consecutive administration for XX days. The blood collection time was generally in 0.5~3 hours after the last administration, divided into, the single time point and multiple time points, two cases. Finally, blood collection locations mainly included the postorbital venous plexus, the abdominal aorta, the hepatic portal vein, the eyeball blood, and the fosse orbital vein *etc*.

At last, we described the physicochemical property distributions of 701 absorbed prototype constituents with structures of the prescriptions and herbs by oral administration (*i.e.*, intragastrical administration) in our data (Figure 4). We observed that most absorbed constituents by oral administration satisfied the traditional rule for the drug-like molecule screening (molecular weight <= 500, logP <= 5, hydrogen bond donor count <= 5, hydrogen bond acceptor count <= 10, rotatable bond count <= 10, TPSA <= 140 [34]), but indeed there were still many that did not satisfy the rule, suggesting the imperfection of the traditional rule in estimating the absorption and permeation of molecules.

Usage of DCABM-TCM

DCABM-TCM supports data browse, search, download, and submission. In addition, it also supports two data analysis functions as described in the next section.

For each of the six kinds of annotation pages, including prescriptions, herbs, constituents detected in blood, targets, pathways, and diseases, DCABM-TCM supports browse and search. Specially, for blood constituents, only those with mapped PubChem CIDs have their own annotation pages and are supported to be browsed and searched. All blood constituents can be viewed on the corresponding detailed annotation pages of prescriptions/herbs as described in the next section.

Among all the six kinds of entities that are searchable (Figure 5A), for the prescription or herb, users can search by English name or Pinyin name; for the target, by Entrez Gene ID, Gene symbol or Gene full name; for the disease, by Disease name or CTD disease ID; for the pathway, by KEGG pathway name or ID. Finally, for the constituent detected in blood, DCABM-TCM supports search by 1) Compound name or PubChem CID; 2) Structural similarity; 3) Physicochemical property range; or 4) Compound classification. When searching by the structural similarity, users input a compound of InChI format or draw the structure of a compound with the help of JSDraw [35], and meanwhile set the structural similarity cutoff (Figure

5B), then structurally similar blood constituents in the order of decreasing similarity scores (>= cutoff) will be returned. When searching by the property range, users can specify the range of the physicochemical properties of the returned blood constituents, including the molecular weight, logP, TPSA, hydrogen-bond donor count, hydrogen-bond acceptor count, and rotatable bond count (Figure 5C). All blood constituents with structures in DCABM-TCM were divided into 12 superclasses, given by a histogram graph on the "Compound classification" Search page of the blood constituent. Selecting a column will return blood constituents belonging to that superclass (Figure 5D).

On the submission page, users can submit the constituents detected in blood of a prescription or herb derived from their own research or other publications to DCABM-TCM, by simply filling information in some required and optional fields. Periodically, after manual verification, we will integrate them into DCABM-TCM.

Annotation page

For every entity of the six types, including prescriptions, herbs, blood constituents, targets, pathways, and diseases, DCABM-TCM provides a detailed annotation page presenting the related core and annotation information as well as the corresponding interactive analysis functions and analysis results.

For a prescription/herb, on its detailed annotation page, first, we give its basic information, including Chinese/English/Pinyin/Latin name, cross-references to TCMID [7] and CMAUP [17] (Figure 6A). Then, as a reference, its ordinary constituents generally detected *in vitro*, integrated from TCMID [7], are given, where for a prescription, the ordinary constituents of its every compositive herb and their sum are given (Figure 6B). Next, the constituents detected in blood are provided (Figure 6C). Here we provide the constituents detected in blood (as well as the prototypes and metabolites in them if the corresponding source publication gave) and their corresponding detailed detection conditions and source publications of the prescription/herb. Considering the potential difference of the blood constituents detected by different researches as a result of the difference of the detection conditions, here, we give the blood constituent-related information from different source publications, respectively, which can be browsed by the label switch of "Source1", "Source2", …. The content of the label "Sum" is the sum of blood constituents from different sources. In the next column, other prescriptions and herbs in DCABM-TCM which share blood constituents with the interested prescription/herb together with the shared blood constituents are listed (Figure 6D). Finally, at the end of the annotation page, the network pharmacology analysis function and analysis results are presented (Figure 6E). This analysis is implemented based on BATMAN-TCM, a bioinformatics analysis tool for the molecular mechanism of TCM previously developed by us [26]. In this function, known and predicted targets of the blood constituents of the prescription/herb, further the enriched GO functional terms, KEGG biological pathways, and CTD/OMIM diseases among the targets are analyzed. And the blood constituent-target-pathway/disease association network is visualized. Here, the target prediction score cutoff and the P-value of the enrichment analysis after multiple testing correction can be interactively changed and then the results will be re-analyzed. All analysis results as well as the network graph can be downloaded. This function aims to reveal the

potential molecular mechanism of the prescription/herb based on its blood constituents. We believe that compared with all constituents of a prescription/herb, the results of the network pharmacology analyses based on its blood constituents are more reliable for revealing its potential molecular mechanism.

For a constituent detected in blood, its annotation page gives basic information (including name, molecular formula, PubChem CID, CAS number, structure, cross-references), physicochemical properties, compound structural classification, ADMET properties, the lists of prescriptions and herbs the blood constituent belongs to, and the network pharmacology analysis results for the blood constituent.

For a target/pathway/disease, the content of its annotation page is only divided into two sections (Figure 7). One gives the basic information of the entry, including name, ID, cross-references, GO function terms a target entry belongs to, pathways a target entry participates in, diseases a target entry is associated with, member genes a pathway entry contains, a disease entry-related genes *etc.* (Figure 7A). The other is the analysis function of the candidate prescriptions, herbs, blood constituents targeting the target/pathway/disease (Figure 7B). The candidate blood constituents of a target are given based on known and predicted drug-target associations provided by BATMAN-TCM. The candidate blood constituents potentially targeting a pathway/disease are significantly enriched blood constituents among the pathway member genes/the disease-related genes. Further the candidate prescriptions/herbs potentially targeting a target/pathway/disease are significantly enriched ones among its candidate blood constituents (see Materials and Methods). Here users can specify the target prediction score cutoff and P-value of the enrichment analyses after multiple testing correction. This function aims to help the target/pathway/disease-based screening of candidate blood constituents, herbs and prescriptions.

In addition, for the user-friendliness, on each of the annotation pages, we provide various hyperlinks to external databases such as TCMID [7], CMAUP [17], DrugBank [24], ChEBI [36], KEGG [28], DGIdb [37], OMIM [29], Ensembl [38], and CTD [30], to internal pages and to the Help document in proper positions.

Conclusions

The vast majority of TCM constituents can produce pharmacological effects only after they reach the bloodstream. Compared with the ordinary constituents detected *in vitro*, which can be obtained from many existing TCM databases, the constituents detected in blood of a TCM, which have not been recorded by any previous database, are more likely to be the really effective ones responsible for its pharmacological effects. Here we manually collected so far the most comprehensive experimentally-detected blood constituents of TCM prescriptions and herbs together with the corresponding detailed detection conditions, and integrated associated annotations such as targets, pathways, diseases to establish the DCABM-TCM database. In addition to data browse, search, download, and submission, DCABM-TCM was also designed to support two typical applications of these blood constituents: 1) potential molecular mechanism elucidation of a prescription or herb by the network pharmacology analysis function; 2) TCM-based drug discovery based on the function of the priorization of candidate drugs (including blood constituents, herbs and prescriptions) potentially targeting a

target/pathway/disease. In summary, the unique data and the designed analysis functions of DCABM-TCM can benefit not only researchers interested in TCM modernization, but also a wider range of drug researchers. Not only for the elucidation of TCM molecular mechanism and quality control of herbs but also for TCM-inspired modern drug discovery, compared with previous ordinary TCM constituents detected *in vitro*, the blood constituents are a great step forward, which are more likely to be really effective/bioactive and meanwhile bioavailable.

Abbreviations

TCM: Traditional Chinese Medicine; ADMET: Absorption, Distribution, Metabolism, Excretion, and Toxicity; CAS: Chemical Abstracts Service; logP: the logarithm of the partition coefficient between n-octanol and water; TPSA: Topological Polar Surface Area; GO: Gene Ontology; LC-MS/MS: liquid chromatography coupled with tandem mass spectrometry; CNKI: China National Knowledge Infrastructure; KEGG: Kyoto Encyclopedia of Genes and Genomes; OMIM: Online Mendelian Inheritance in Man; CTD: The Comparative Toxicogenomics Database.

Declarations

Competing interests

All the authors declare that they have no competing interests.

Funding

Authors' contributions

ZL designed the database and wrote the manuscript. DL and FX provided the guidance and revised the manuscript. BF, JL, RL, LX, LY, YL performed the data curation. XL, SG, FG, JL, XW, and YQ implemented the database.

Acknowledgments

Availability of data and materials

The database can be accessed at http://bionet.ncpsb.org.cn/dcabm-tcm/, and the web server is compatible with Chrome, Firefox, and Opera browsers for Windows; and Safari, Chrome, Firefox, and Opera browsers for Mac operating system.

References

1. Cheung F. TCM: made in China. 2011;480:S82-3.

2. Kim HU, Ryu JY, Lee JO, Lee SY. A systems approach to traditional oriental medicine. Nat Biotechnol. 2015;33:264-8.

3. Wang X. In: Wang X, Zhang A, Sun H, editors. Serum Pharmacochemistry of Traditional Chinese Medicine. London: Academic Press; 2017. p. xix.

4. Li Y, Xie Y, He Y, Hou W, Liao M, Liu CX. Quality markers of traditional Chinese medicine: concept, progress, and perspective. Engineering. 2019;5:888-94.

5. Xu HY, Zhang YQ, Liu ZM, Chen T, Lv CY, Tang SH,et al. ETCM: an encyclopaedia of traditional Chinese medicine. Nucleic Acids Res. 2019;47:D976-82.

6. 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.

7. 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. 2013;41:D1089-95.

8. 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.

9. Chen CY. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One. 2011;6:e15939.

10. Zhang RZ, Yu SJ, Bai H, Ning K. TCM-Mesh: the database and analytical system for network pharmacology analysis for TCM preparations. Sci Rep. 2017;7:2821.

11. Li B, Ma C, Zhao X, Hu Z, Du T, Xu X, et al. YaTCM: yet another traditional Chinese medicine database for drug discovery. Comput Struct Biotechnol J.2018;16:600-10.

12. 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 Cheminform. 2014;6:13.

13. Bultum LE, Woyessa AM, Lee D. ETM-DB: integrated Ethiopian traditional herbal medicine and phytochemicals database. BMC Complement Altern Med. 2019;19:212.

14. Kim SK, Nam S, Jang H, Kim A, Lee JJ. TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. BMC Complement Altern Med. 2015;15:218.

15. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, et al. IMPPAT: a curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep. 2018;8:4329.

16. Lee JH, Park KM, Han DJ, Bang NY, Kim DH, Na H, et al. PharmDB-K: integrated bio-pharmacological network database for traditional Korean medicine. PLoS One. 2015;10:e0142624.

17. 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.

18. Kang H, Tang K, Liu Q, Sun Y, Huang Q, Zhu R, et al. HIM-herbal ingredients *in-vivo* metabolism database. J Cheminform. 2013;5:28.

19. Wang D, Yang L, Zhang P, LaBaer J, Hermjakob H, Li D, et al. AAgAtlas 1.0: a human autoantigen database. Nucleic Acids Res. 2017;45:D769-76.

20. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019;47:D1102-9.

21. Matt S. PubChemPy documentation. https://pubchempy.readthedocs.io/en/latest/#. Accessed 3 Mar 2020.

22. Cheng T, Zhao Y, Li X, Lin F, Xu Y, Zhang X, et al. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. J Chem Inf Model. 2007;47:2140-8.

23. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, et al. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. J Chem Inf Model. 2012;52:3099-105.

24. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46:D1074-82.

25. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. Classyfire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform. 2016;8:61.

26. Liu Z, Guo F, Wang Y, Li C, Zhang X, Li H, et al. BATMAN-TCM: a Bioinformatics Analysis Tool for Molecular mechANism of Traditional Chinese Medicine. Sci Rep. 2016;6:21146.

27. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2011;39:D52-7.

28. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353-61.

29. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43:D789-98.

30. Mattingly CJ, Rosenstein MC, Colby GT, Forrest JN Jr, Boyer JL. The Comparative Toxicogenomics Database (CTD): a resource for comparative toxicological studies. J Exp Zool A Comp Exp Biol. 2006;305:689-92.

31. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. Methods Mol Biol. 2007;406:89-112.

32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc. 1995;57:289-300.

33. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. J Cheminform. 2011;3:33.

34. Whitty A, Zhong M, Viarengo L, Beglov D, Hall DR, Vajda S. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. Drug Discov T 2016;21:712-7.

35. JSDraw - A Javascript Framework for Cheminformatics and Bioinformatics. http://elncloud.com/jsdrawapp/jsdraw/. Accessed 25 Apr 2020.

36. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008;36:D344-50.

37. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res. 2018;46:D1068-73.

38. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754-61.

Figures
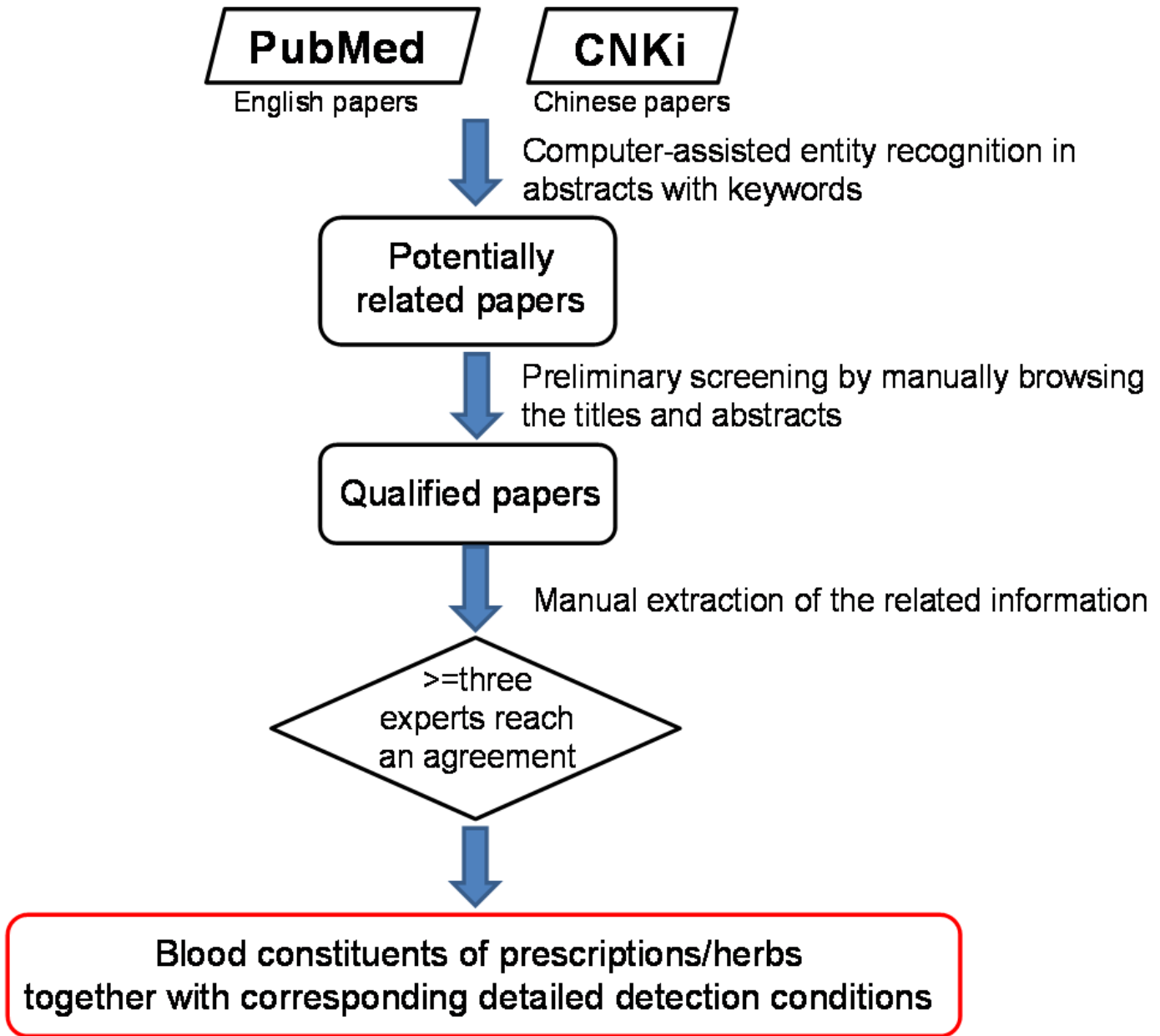
Figure 1

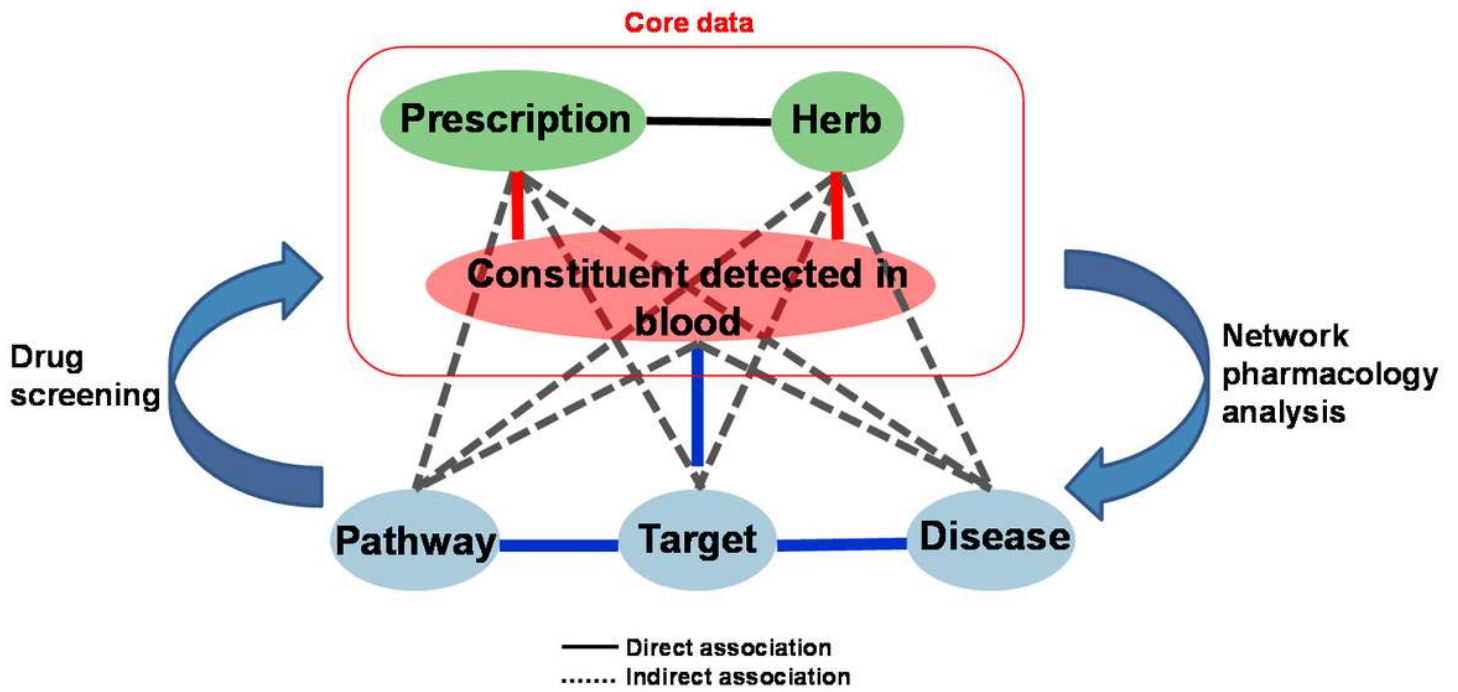The data curation workflow of the DCABM-TCM core data.

Figure 2

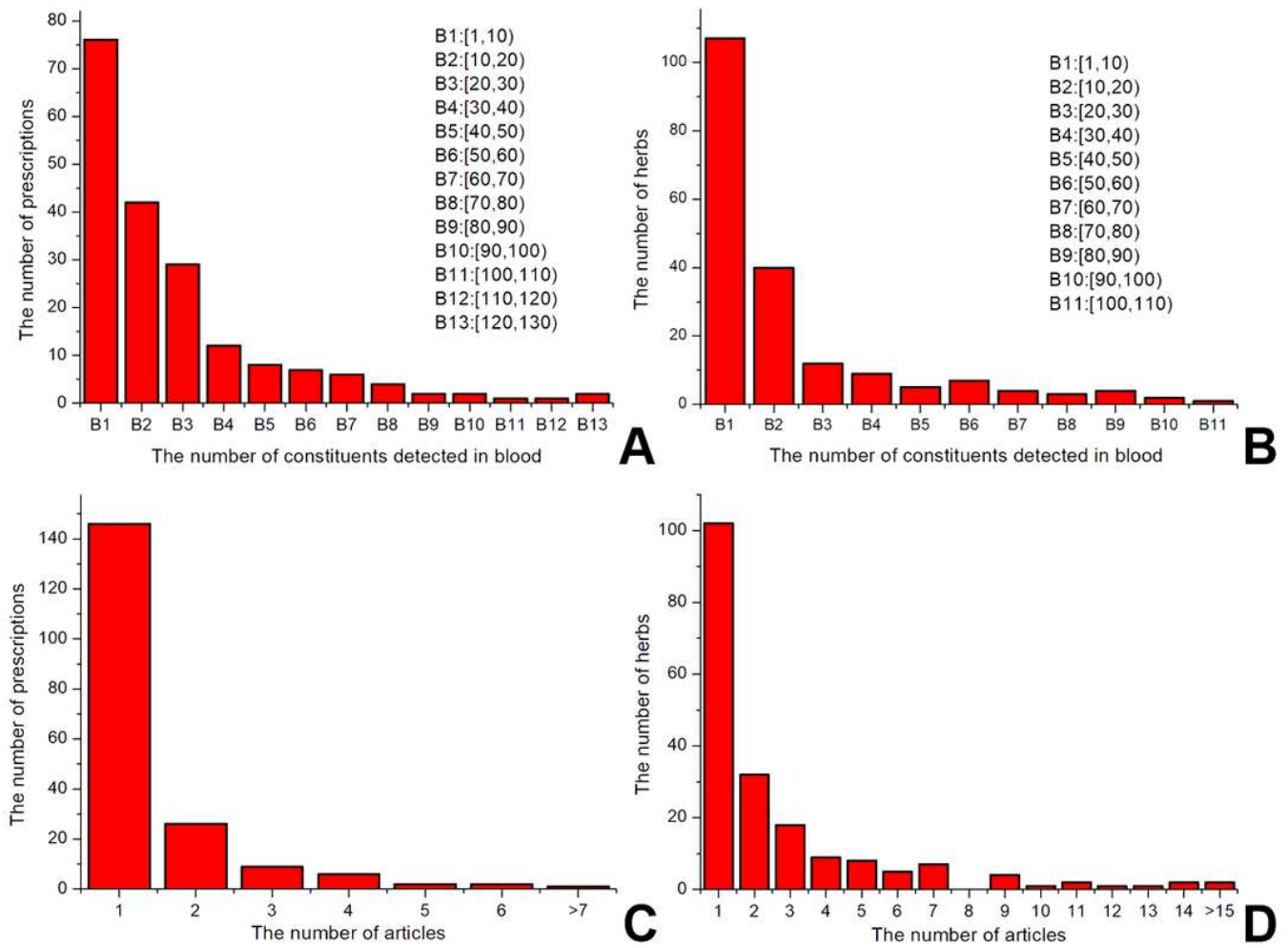Overview of DCABM-TCM database.

Figure 3

Distributions of DCABM-TCM data. Distributions of the number of blood constituents of prescriptions (A) and herbs (B). Distributions of the number of papers that studied the blood constituents of a prescription (C) and herb (D).
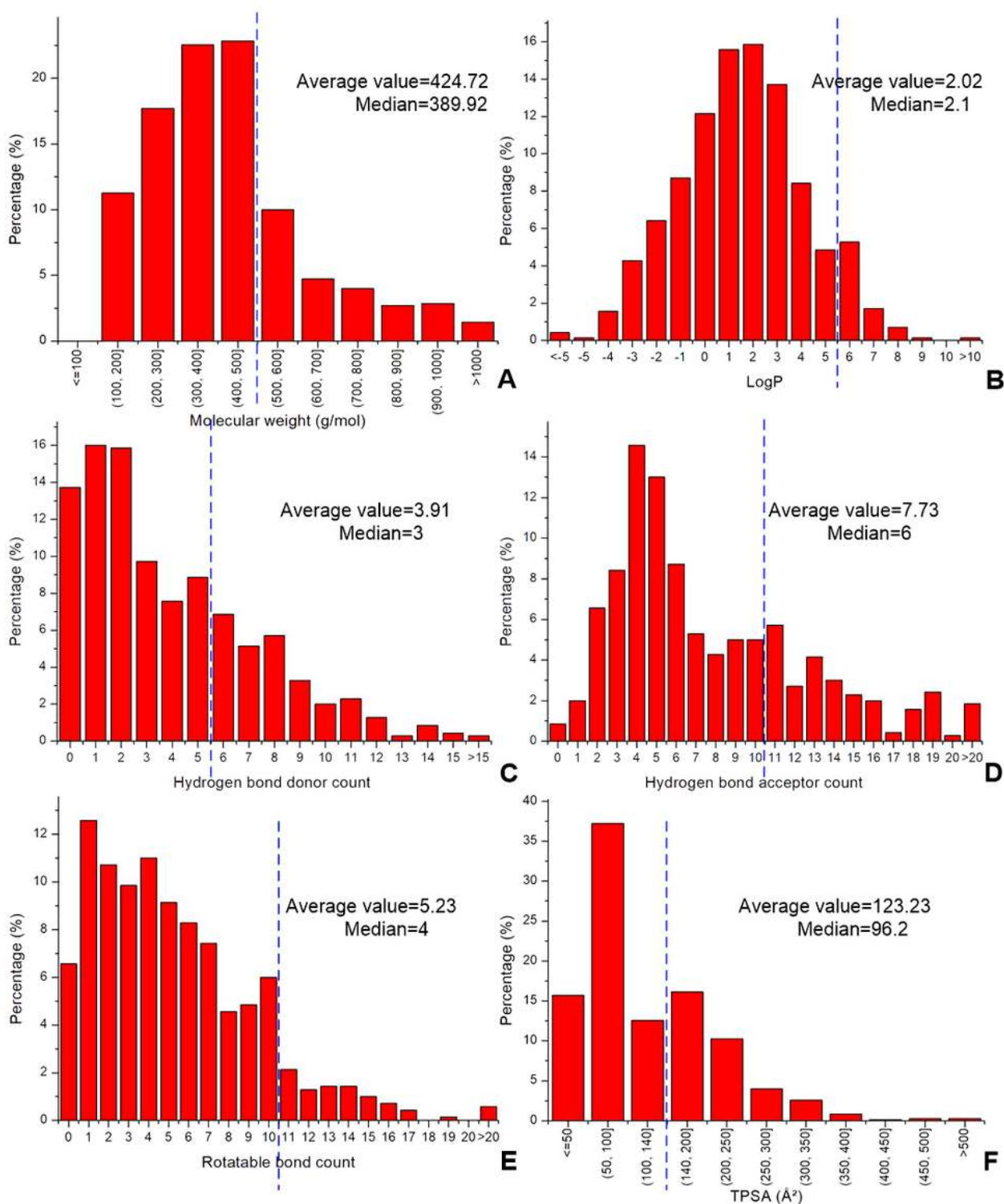
Figure 4

Distributions of physicochemical properties of orally absorbed prototype constituents with structures. (A) Molecular weight. (B) LogP (the logarithm of the partition coefficient between n-octanol and water). (C) Hydrogen bond donor count. (D) Hydrogen bond acceptor count. (E) Rotatable bond count. (F) TPSA (Topological Polar Surface Area). The average values and medians of the properties are given on the

graphs. The property cutoff of the traditional rule of the drug-like molecule screening is marked on each of the graphs by a dotted line.
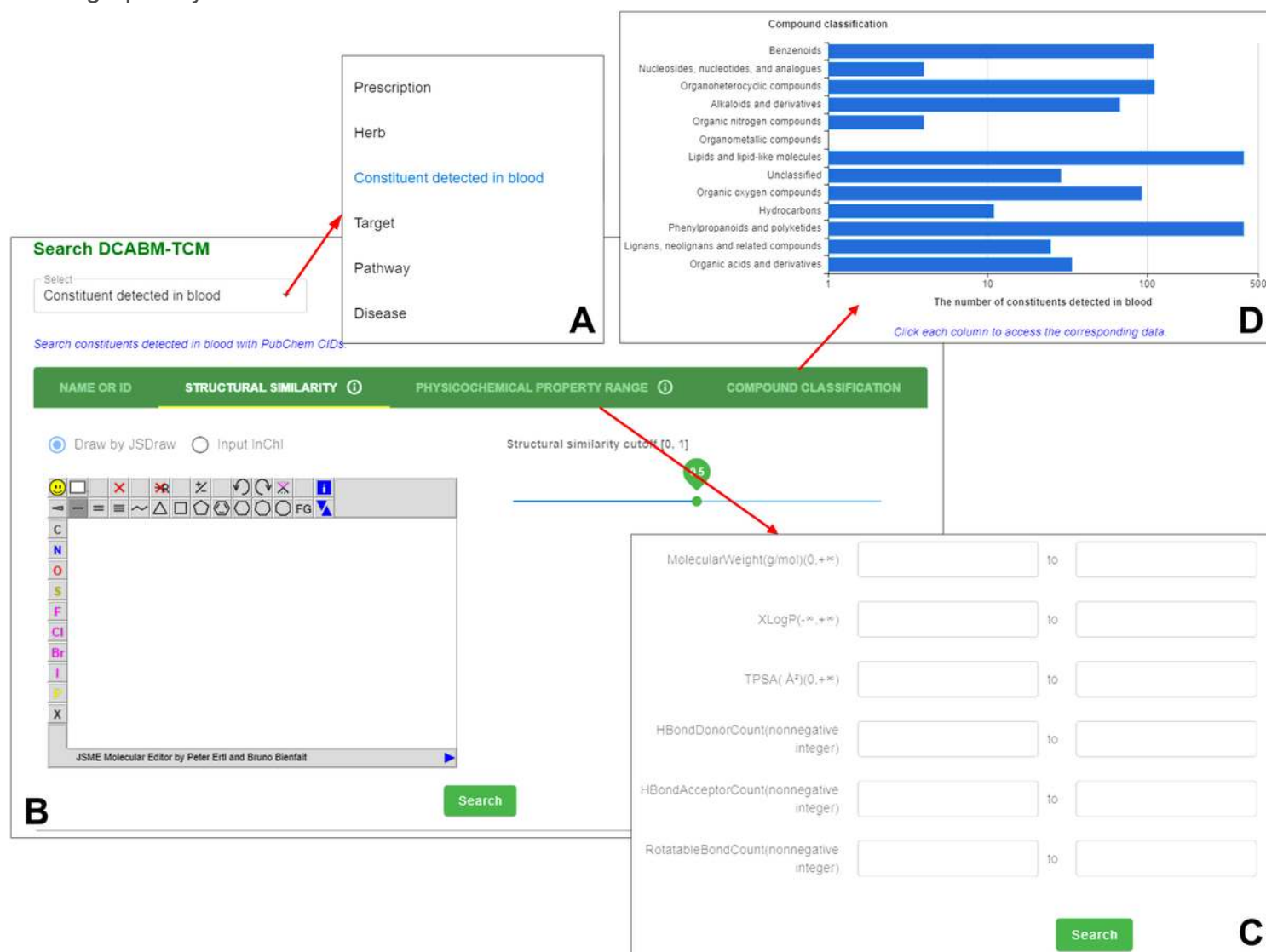

Figure 5

Search modes supported by DCABM-TCM. (A) Users can search the database by prescriptions, herbs, blood constituents, targets, pathways, or diseases. Specially, for blood constituents, in addition to routine search by the name and PubChem CID, DCABM-TCM also supports search by structural similarity (B), physicochemical property range (C) and compound classification (D).
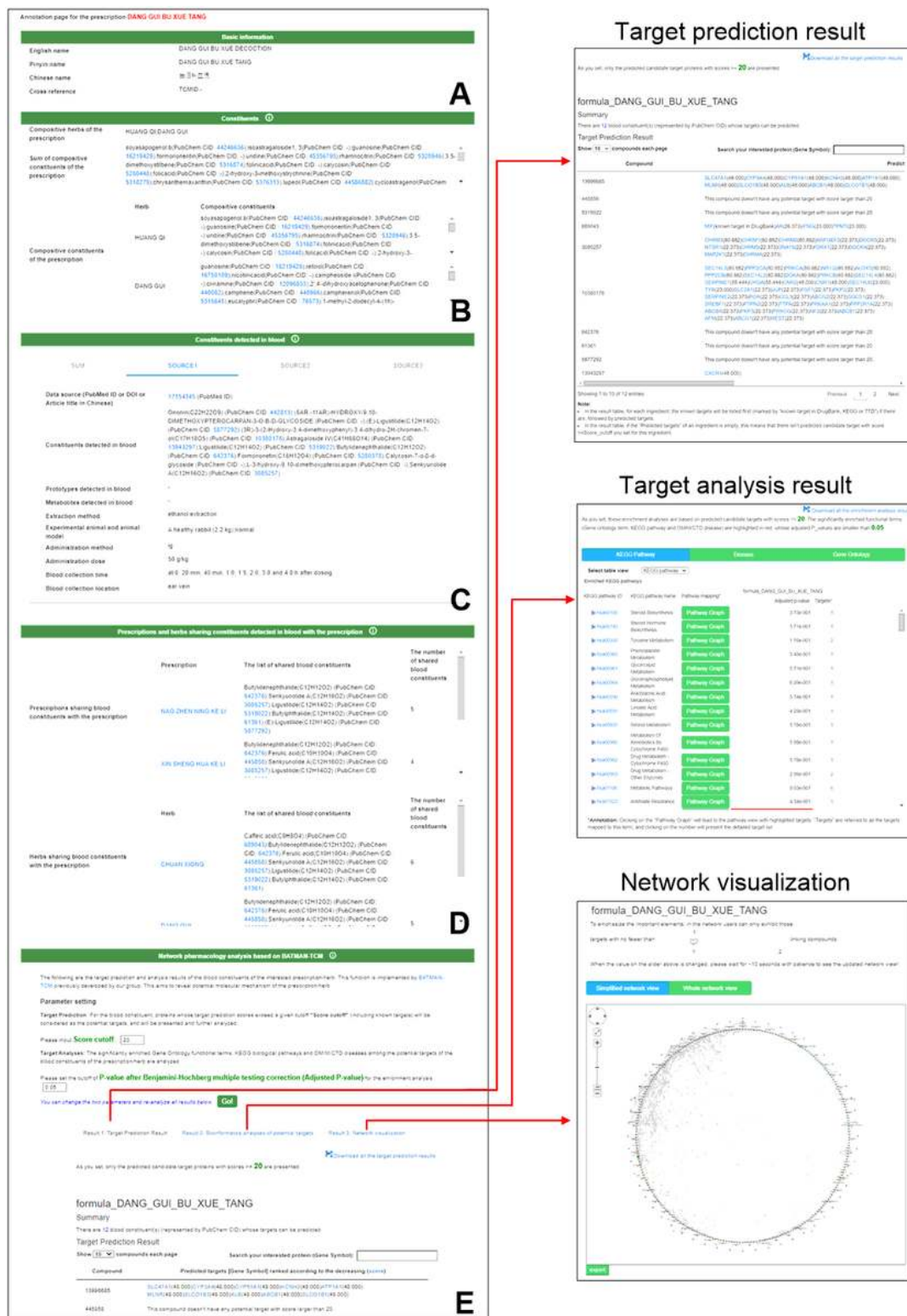
Figure 6

Prescription annotation page. (A) Basic information. (B) Ordinary constituents, including three data fields: the compositive herbs of the prescription, ordinary constituents of every compositive herb and their sum. (C) Constituents detected in blood. This column gives the blood constituents of the prescription (including prototypes and metabolites) together with the detailed detection conditions. (D) Other prescriptions and herbs in DCABM-TCM that share blood constituents with the prescription. (E) Network

pharmacology analysis. Here, users can specify the target prediction score cutoff and P-value of enrichment analyses after multiple testing correction. The analysis results include target prediction of blood constituents of the prescription and further GO/pathway/disease enrichment analyses as well as the blood constituent-target-pathway/disease association network visualization.
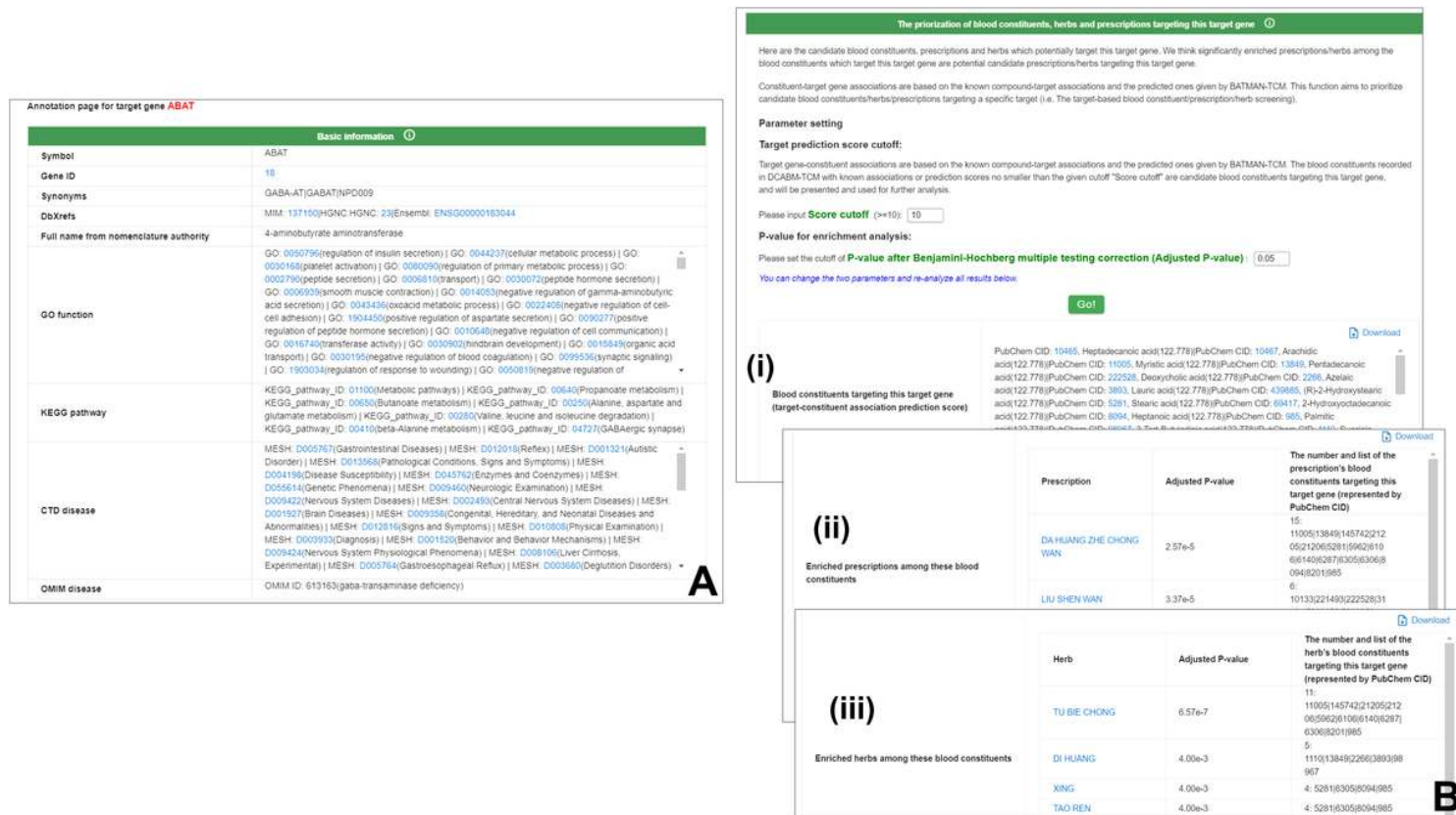

Figure 7

Target annotation page. (A) Basic information. (B) The analysis of candidate blood constituents (i), prescriptions (ii) and herbs (iii) targeting the target gene. In this analysis, users can specify the target prediction score cutoff and the P-value of enrichment analyses after multiple testing correction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.zip
- Additionalfile2.xlsx
- Additionalfile3.xlsx