



FangNet: Mining herb hidden knowledge from TCM clinical effective formulas using structure network algorithm



Dechao Bu^{b,1}, Yan Xia^{a,1}, Jiayuan Zhang^{a,1}, Wanchen Cao^{a,1}, Peipei Huo^c, Zhihao Wang^c, Zihao He^a, Linyi Ding^a, Yang Wu^b, Shan Zhang^c, Kai Gao^a, He Yu^a, Tiegang Liu^a, Xia Ding^{a,*}, Xiaohong Gu^{a,*}, Yi Zhao^{a,b,*}

^a Beijing University of Chinese Medicine, ChaoYang District, Beijing 100029, China

^b Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

^c Chinese Academy of Sciences, Luoyang Branch of Institute of Computing Technology, Luoyang, China

ARTICLE INFO

Article history:

Received 31 July 2020

Received in revised form 22 November 2020

Accepted 23 November 2020

Available online 4 December 2020

Keywords:

Formulas

Herb

Symptom

TCM

ABSTRACT

The use of herbs to treat various human diseases has been recorded for thousands of years. In Asia's current medical system, numerous herbal formulas have been repeatedly verified to confirm their effectiveness in different periods, which is a great resource for drug innovation and discovery. Through the mining of these clinical effective formulas by network pharmacology and bioinformatics analysis, important biologically active ingredients derived from these natural products might be discovered. As modern medicine requires a combination of multiple drugs for the treatment of complex diseases, previously clinical formulas are also combinations of various herbs according to the main causes and accompanying symptoms. However, the herbs that play a major role in the treatment of diseases are always unclear. Therefore, how to rank each herb's relative importance and determine the core herbs, is the first step to assisting herb selection for active ingredients discovery. To solve this problem, we built the platform FangNet, which ranks all herbs on their relative topological importance using the PageRank algorithm, based on the constructed symptom-herb network from a collection of clinical empirical prescriptions. Three types of herb hidden knowledge, including herb importance rank, herb-herb co-occurrence, and associations to symptoms, were provided in an interactive visualization. Moreover, FangNet has designed role-based permission for teams to store, analyze, and jointly interpret their clinical formulas, in an easy and secure collaboration environment, aiming at creating a central hub for massive symptom-herb connections. FangNet can be accessed at <http://fangnet.org> or <http://fangnet.herb.ac.cn>.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Herbal medicine is an effective solution for primary health care and a great resource for drug innovation and discovery [1,2]. It have been received increasing attention by pharmaceutical companies in the past ten years [3,4]. Herbs are the starting materials for

isolation and further derivatization of natural biologically active ingredients [4]. Also, their larger number of medical use have recorded valuable effect on particular disease and phenotypes [2,5]. According to the statistics, more than 70% of 177 drugs approved for cancer treatment are based on herbs or natural products and other mimetics [6]. In June 2004, the U.S. FDA approved the use of herbs with unclear active ingredients but have definite efficacy in clinical practice [7]. In October 2010, the FDA approved a green tea extract called Veregen for the treatment of genital warts based on this rule [7].

In clinic practice, herbs exist in formula form, that is, multiple herbs are combined into one prescription, which is called "Fang" in traditional Chinese medicine (TCM) [2,8]. Ancient literature and the current medical system of TCM have accumulated numerous clinical formulas with definite efficacy, repeatedly verified to

Abbreviations: TCM, Traditional Chinese Medicine; THScore, Topological-Hub Score; PDD, Phenotype-based Drug Discovery; EBM, Evidence-Based Medicine; FOBT, Fecal Occult Blood Test; CNKI, China National Knowledge Infrastructure.

* Corresponding authors at: Beijing University of Chinese Medicine, ChaoYang District, Beijing 100029, China.

E-mail addresses: dingx@bucm.edu.cn (X. Ding), Guxh1003@126.com (X. Gu), biozy@ict.ac.cn (Y. Zhao), biozy@ict.ac.cn (Y. Zhao).

¹ Dechao Bu, Yan Xia, Jiayuan Zhang, Wanchen Cao have contributed equally to this work.

<https://doi.org/10.1016/j.csbj.2020.11.036>

2001-0370/© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

confirm their effectiveness in different periods [9]. In Asia, the use of TCM formulas can be traced back to 2,000 years ago and continuously accumulated and compiled by successive dynasties [10]. A total of 96,592 known prescriptions for 2,000 years have been included in the current *Chinese Medicine Prescriptions Dictionary*. Also, a large number of clinical formulas for the treatment of complex diseases such as immune diseases, cardiovascular diseases, and pain have been accumulated in the current medical system [11–14].

Clinically important TCM multi-herbal formulas usually contain a complex mixture of various biologically active ingredients [8,15]. Yet there are a number of bioactive ingredients have been explored from TCM formulas. Artemisinin, the first-line drug for malaria, was obtained by Tu Youyou, a 2015 Nobel Prize winner, from the approved prescription against malaria *Zhouhou Beiji Prescriptions* recorded in ancient books [16]. The anti-asthma drug ephedrine was discovered by Chen Kehui (1898–1988), inspired by the functional effect of ephedra in TCM clinical formulas [17]. Hypuconitine (HC) is the main chemical component of Fuzi in Sini Decoction (SND), and is considered to be the main chemical component for the treatment of cardiovascular diseases [8]. Paeonol from Mudanpi [15], Danshensu (DS) and Tanshinone I (TI) from Danshen [18] are the main active ingredients for the treatment of various cardiovascular [8]. These two herbs Mudanpi and Danshen are components of TCM classic prescription Shuang-Dan (SD), which has function of promoting blood circulation. Andrographolide from Chuanxinlian is the main active ingredient of clinically widely used injection, which has functions of liver-protecting, analgesic, anti-inflammatory, and anti-tumor activities [19].

As modern medicine requires a combination of multiple drugs for the treatment of complex diseases, previously clinical formulas are also combinations of various herbs. In clinical practice, TCM formulas exist in the form of herb combination, usually containing 2–20 herbs or even more, mainly due to the complexity of chemical constituents [2,8]. Its dosage size, dosage form (i.e. powder, liquid, suspension, tablet or capsule) exhibit quite complicated in application scenarios [18]. Different herbs or herb combinations can be added or subtracted into one prescription according to the characteristics of the disease with varied symptoms [20]. As a result, the formulas for treating the same disease vary in quite different forms. The relationship between herbs and diseases is non-dominant, and which herb plays a core role in specific diseases is unclear [21]. Therefore, how to rank and evaluate herb's importance, determine the most critical herbs from these complex combinations, and discover the associations with particular symptoms, is the primary key to maximizing the use of existing clinical formulas and promote herbal drug development [22].

At present, there are three main methods for mining core herbs from herbal formulas: 1) using the frequency of the herbs [23], 2) mining by association rules [24–25], 3) clustering [12,26]. Frequency is a commonly used indicator for herbal importance evaluation, however the importance it gives is misleading. For example, it will not rank drugs together that are used in combination. In the real scene, specific combinations of herbs are bundled together during the treatment of certain symptoms. Regardless of extracting rules guiding clinic or herbs selection for downstream active ingredient research, these combinations need to be ranked and studied together. Association rules and clustering methods take into account the herb combination but failed to provide an overview weight for each herb. Moreover, all three methods mentioned above only focus on herbs themselves to conduct herb evaluation but fail to make effective utilization of the symptom to herb associations in clinical prescriptions.

To solve these problems and give a better importance rank of herbs for a disease, guiding the downstream ingredient research, we built the FangNet platform, which first constructed the

symptom-herb network for topological ranking of herbs using the PageRank algorithm [27]. Three types of herb hidden knowledge, including herb importance rank, herb-herb co-occurrence and mutual exclusivity, associations to specific symptoms, were provided in an interactive visualization. As a cloud-service platform, FangNet supports the teams to jointly analyze and interpret their clinical formulas in a collaborative way, towards constructing a central hub for massive symptoms and herbs, which would continuously provide essential clues for drug innovation and discovery.

2. Methods

2.1. Herb-herb interaction and symptom-herb association calculation

The Jaccard index [28] was used to calculate herb-herb interactions. It is a measure of similarity for the two sets, ranging from 0 to 1, to compare members for two sets to see how much proportion is shared. Confidence score in association rule mining [29] was calculated for symptom-herb associations. The calculation formula is as follows.

$$W(h_i, h_j) = \frac{|Rx(h_i) \cap Rx(h_j)|}{|Rx(h_i) \cup Rx(h_j)|} \quad (1)$$

$$W(s_t, h_i) = \frac{|Rx(s_t) \cap Rx(h_i)|}{|Rx(s_t)|} \quad (2)$$

where h_i stands for herb i , h_j stands for herb j , $Rx(h_i)$ stands for prescriptions containing the herb i , $Rx(h_j)$ stands for prescriptions containing the herb j . $Rx(s_t)$ stands for prescriptions containing symptom t .

2.2. Weighted interaction network construction

A weighted network was constructed, with symptom and herb as two types of nodes. The weight of the edges between herb-herb and symptom-herb is calculated as above. The initial value of the nodes was defined as follows.

$$V(s_t) = \frac{|Rx(s_t)|}{N} \quad (3)$$

$$V(h_i) = \frac{|Rx(h_i)|}{N} \quad (4)$$

where s_t stands for symptom t , h_j stands for herb i , $Rx(h_i)$ stands for prescriptions containing the herb i , $Rx(s_t)$ stands for prescriptions containing the symptom t , N stands for the number of total prescriptions.

2.3. Topological-Hub Score (THScore) using the PageRank algorithm

In order to reposition all the herbs in the clinical formulas, the Topological-Hub score was calculated using the PageRank algorithm for all symptom to herb associations. PageRank (PR) is an algorithm used by Google company to rank the retrieved web pages in their search engine results [27]. This algorithm has been widely used to discover community leaders in social networks [30,31] and identify important nodes in the networks [32–34]. The PageRank score measures the leadership role of a node based on all of its links, rather than simply calculating the degree of each herb node. Herbs with more interaction links, are given higher PageRank scores.

2.4. Segmented regression to determine driver herbs and passenger herbs

We have defined two types of driver herbs and passenger herbs, corresponding to those playing major roles or supporting roles for particular diseases. To determine a threshold distinguishing herb and passenger herbs, a segmented linear regression analysis was performed using the THScore calculated above. Briefly, this algorithm applied an iterative process to determine a segment breakpoint, at which a statistically significant change in the slope of adjacent regression lines occurred [35]. The herbs large than the thresholds are defined as the driver herbs, while the others defined as the passenger herbs.

2.5. Herb-herb co-occurrence and mutual exclusivity level calculation

Co-occurrence and mutual exclusivity level (Co_level) was defined by herb-herb interaction edge weight and the related p value calculated from Fisher's Test. The former is named Co_ratio in FangNet, which is the co-occurrence ratio for two herbs occur in the same prescription, ranging from 0 ~ 1, with 1 represents two herbs always occurring together, and 0 means that they never occur together. Totally nine levels of co-occurrence and mutual exclusivity were defined, including -4, -3, -2, -1, 0, 1, 2, 3, 4, while 0 represents no significance, -1 ~ -4 represents mutual exclusion, the smaller the value, the more significant, 1 ~ 4 represents co-occurrence, the larger the value, the more significant. The detailed definition of nine level by Co_ratio and p value is presented in [Supplementary Table S1](#).

2.6. Symptom-herb association mining

The high confidence relationship between the symptom and herb is filtered by the edge interaction weight and the co-occurrence event, which is defined as the number of prescriptions using the herb when having a specific symptom. The parameters of symptoms-herb association greater than 0.6 and co-occurrence event ≥ 20 , could be used as a feasible group of filtering parameters, according to our testing under the condition of prescriptions number about 100. However, this parameter is not universal for all kinds of inputs. For example, it is quite different when taking the same doctor's prescriptions as input, or taking literature collected prescriptions as input. Users are required to customize the parameters based on the scale and characteristics of their own data.

3. Results

3.1. Framework of FangNet

3.1.1. Input and output

FangNet requires a collection of empirical prescriptions for a particular disease as the input (Fig. 1A). In addition to uploading the multi-herb formula in the prescriptions, the corresponding typical symptoms also need to be uploaded. For the herb formula, herb name, dosage and processing method are required (Fig. 1C). For the symptoms, TCM symptoms, such as fatigue, insomnia, fever, cough, etc., as well as abnormal indicators found in clinic testing, such as Fecal Occult Blood Test (FOBT) positive, are required (Fig. 1B).

When a collection of prescriptions is constructed on the FangNet, users could start an analysis by clicking the analysis button within this collection. FangNet would automatically enter the process of network construction and topologic mining. Then FangNet will interactively visualize three forms of herb hidden knowledge, including herb driver/passage attributes, herb-herb co-occurrence

and mutually exclusivity, and symptom-herb associations (Fig. 1A). Users can dynamically change the thresholds of nodes and edges, and interactively obtain data tables and visual figures, which helps the users decipher the hidden rules behind the data from different perspectives easily and quickly.

3.1.2. Symptom/herb semantic repository and auto or manual tagging

As the core blocking of mining, the standard expressed of symptoms and herbs plays a vital role in the quality of mining. To standardize the input words, FangNet has built a database of 1717 symptoms and 618 herbs, by integrating SymMap [5], an integrative symptom database of TCM we developed earlier, and *Chinese Pharmacopoeia 2020 Edition* (Fig. 1A). The content entered by the user will be converted to terms in the database if they can match. However, given the complexity of symptom and herb terms used in TCM, FangNet's existing database may fail to convert all words and synonyms. To solve this, FangNet has come up with a solution by supporting users to specify tags for their content that cannot be automatically mapped (Fig. 1B). Using such a manual way to standardize the input terms, structured and normalized data would be generated for the subsequent mining. These tag terms added by users will serve as an essential data for expansion of the symptom/herb semantic repository, with a continuous pooling, counting and curation workflow.

3.2. Interactive visualization of herb hidden knowledge

3.2.1. Herb importance rank

Evaluating the weights of herb and extracting the core herbs is the key to understanding the herb's therapeutic effects on specific diseases. It would help summarize a more short and classical formula from a more extensive set of herbs, which would guide the clinical applications and provide productive candidates for drug research [11,36]. By establishing an symptom-herb network, FangNet calculates the THScore of herb in the network by PageRank algorithm, which is always used to discover community leader in social network [30,31], and identify essential nodes in the networks [32–34]. The herbs are then classified into two categories by segmented linear regression model according to the THScore, thus defined as driver herb and passenger herb respectively. Among them, driver herb is a set of herbs that play a major role in the treatment of diseases, and passenger herb tends to play a supporting role and can be added in or deleted in the presence of specific symptoms. These two types of herbs with different attributes are visualized in different colors. The weight of herb-herb interaction is converted into the gravity of the two vertices in the network (Fig. 2A),

3.2.2. Herb-herb co-occurrence and mutual exclusivity

Herbs appearing in the same empirical prescription have been proved to have some potential interactions, including mutual promotion, synergy, mutual restraint and mutual confrontation [37,38]. For example, herbal combinations were found to be more effective than using a single herb in the experiments of cancer cell lines [39]. By analyzing the co-occurrence and mutual exclusivity of two herbs, we can simplify the complex interactions among herbs, and better understand the combination rules of different herbs. Based on the herb-herb interaction network, Fisher's test was used to analyze the co-occurrence or mutual exclusivity between herb-herb pairs, which are divided into nine different levels, visualized in the form of triangle heatmap (Fig. 2B). Also, the highest level of co-occurrence or mutual exclusivity indicates that the two herbs have stronger potential interaction, which can be seen as essential candidates for experiment-based drug combination study.

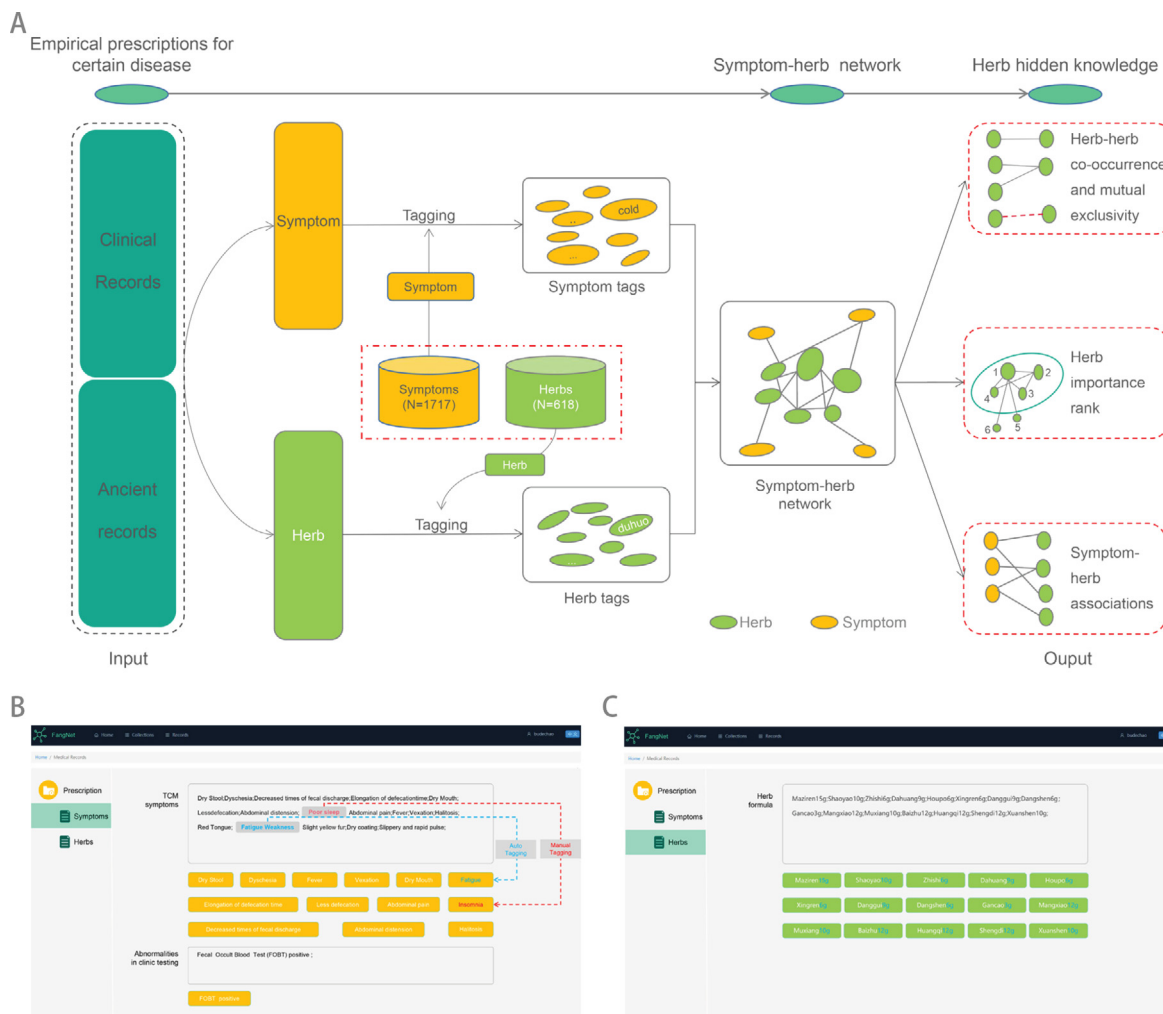


Fig. 1. A framework of FangNet. A. The workflow of mining hidden knowledge from empirical prescriptions for FangNet. Orange stands for symptom, green stands for herb. B. Automatic and manual tagging of input symptoms, take a constipation prescriptions as an example. C. The input herb formula. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.3. Symptom-herb association

Tailoring of medical treatment to the individual characteristics of each patient is the basic concept of personalized medicine. TCM exhibits a typical personalized mode in clinical diagnosis and treatment [40]. To discover how symptoms affect the personalized diagnosis and treatment, FangNet tries to explore the intimate associations between symptoms and herbs in TCM. With the symptom-herb association module of the platform, the correlation between herb and specific symptoms can be obtained by screening according to different support and confidence thresholds as conditions. Symptoms and herbs can be shown as many-to-many patterns with a relational circular layout (Fig. 2C). This part of herb knowledge is a further supplement to mining results of driver/passenger herbs, which indicates the conditions of using specific herbs according to particular symptoms.

3.3. Validation of knowledge extracted

3.3.1. Three benchmark prescription collections

To verify the extracted knowledge, FangNet has constructed three benchmark prescription collections for diseases of headache, abdominal pain and hiccups. In detail, totally 106 prescriptions of headache, 97 prescriptions of abdominal pain, 125 prescriptions of hiccups, were collected from *Chinese Medicine Prescriptions*

Dictionary and TCM Knowledge Database (Zhong Yi Zhi Ku) (Fig. 3). Their symptoms and herb formula info were uploaded into FangNet and then standardized using our online auto/manual tagging module.

3.3.2. Cross-validation with literates and expert knowledge base

Numerous literatures and herbal expert knowledge systems and have been published in concern of herbs [2,41–43]. For the extracted results of these three prescription collections, we have conducted the cross-validation both with the literates and the expert knowledge base. For headache prescriptions, the top ten THScore-ranked herbs are Chuanxiong, Baishao, Gancao, Danshen, Fuling, Tianma, Gouteng, Dangu, Shengdi, Juhua (Fig. 3A, Table S2). Chuanxiong has the highest THScore, also the most related literatures (No. = 1412). Expert knowledge base TCMKB [41] (TCM Knowledge Service Platform, <http://www.tcmkb.cn>) records 1470 entries of Chuanxiong's critical role in treating various headaches types. The ingredients of Ligustrazine and Ferulic acid in Chuanxiong have significant sedative and analgesic effects on the central nervous system [44,45], while anti-platelet aggregation is the main treatments for migraine Mechanism [46–48]. Moreover, a correlation analysis showed that THScore of the top 100 herbs is highly consistent with the number of literatures on

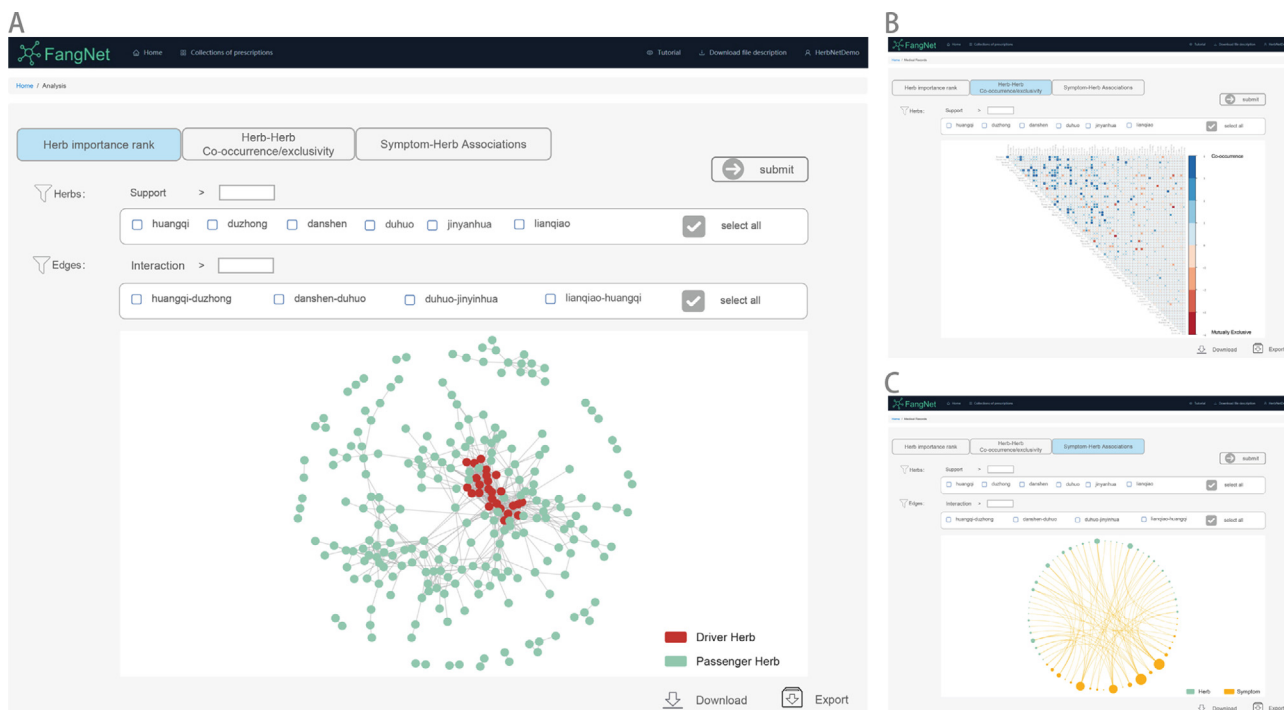


Fig. 2. Interactive visualization of herb hidden knowledge. A: Herb importance rank and driver/passenger classification. The figures can be redrawn by controlling the value of the node and the weight of the edge. Herbs with different importance rank are showed in different colors, while driver herbs are shown in red and passenger herbs are shown in green. B: Herb-herb co-occurrence and mutual exclusivity. Totally 9 levels are defined, Blue means a higher level of co-occurrence, while red means a higher level of mutual exclusivity. C: Symptom-herb associations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

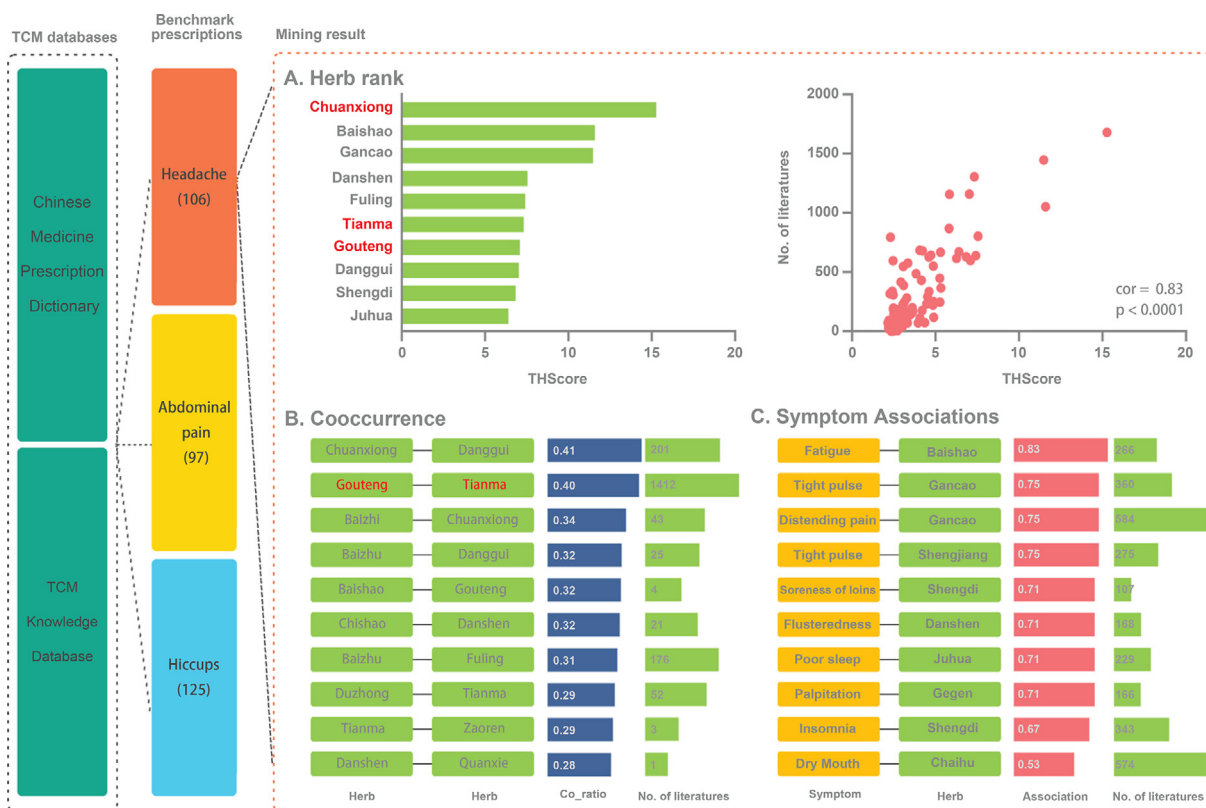


Fig. 3. Cross-validation for headache prescription mining results. 106 prescriptions of headache were collected from *Chinese Medicine Prescriptions Dictionary* and *TCM Knowledge Database (Zhong Yi Zhi Ku)*. A: Herb importance rank. The left is top 10 THScore-ranked herb. The right is the correlation analysis of THScore and number of literatures for top 100 ranked herb. B: Herb-herb co-occurrence (Co-occurrence Level = 4, Co-occurrence event ≥ 10 , top 10 ranked by Co_ratio). C: Symptom-herb associations (Co-occurrence event ≥ 5 , top 10 ranked by symptom-herb association).

CNKI using “headache” and each herb as keywords (cor = 0.83, p-value = 0.001) (Fig. 3A).

It is interesting in our rank, although the support of Gouteng is smaller than Danggui and Shengdi, it ranks before Danggui and Shengdi, and be together with Tianma (Fig. 3A). Tianma-Gouteng is in the top-ten co-occurrence herb pairs, and 1412 literatures have studied on their combination (Fig. 3B, Table S3), in which Gouteng can increase the dissolution of the active ingredients in Tianma [49], and Tianma can enhance the function of Gouteng in dilating blood vessels and lowering blood pressure to a certain extent [50]. Tianma and Gouteng being ranked together could better reflect their relative importance. Finally, the top 10 significant symptoms-herb associations have more than 100 related literatures (Fig. 3C, Table S4). Among them, Gegen treatment of palpitation, Danshen treatment of flustered, Chaihu treatment of dry mouth is recorded in *TCM Knowledge Database* (<http://www.zk120.com>), and *Family of Traditional Chinese Medicine* (<http://www.zysj.com.cn>). Similar to the headaches, the results of abdominal pain and hiccups also have cross-validation support from literates and Expert knowledge Base. The details are described in the [supplementary materials](#) (Fig. S1, Fig. S2).

3.4. Operation mechanism towards a data central

3.4.1. Collaborative mining of empirical prescriptions across multiple institutions

FangNet has designed role-based permission for a collection of prescriptions, to provides an easy and secure collaboration environment for teams working globally across multiple institutions. Collections are the core building blocks of FangNet Platform. Each collection corresponds to a distinct scientific investigation and serves as a container for its prescriptions and analysis results. Access to a collection of prescriptions is restricted to the collaborators in the survey. There are four different permission roles, including the Creator, the Administrator, the Data Clerk and the Passenger (Fig. 4A). To be specific, 1) Each collection must have one Creator, who has the ownership of the collection and the highest authority of controlling the other members' permissions. IT is also the only authority executor to delete the collection. 2) The Administrator has the authority to control the Data Clerk and Passenger members' permissions, and invite other members to join in the collection. 3) The Data Clerk has the authority to upload and edit the empirical prescriptions. 4) The Passenger can only view the prescriptions, without other authorities such as data modification or download.

3.4.2. Expansion of the semantic repository and symptom-herb network

The FangNet is a central hub for teams to store, analyze, and jointly interpret their empirical prescription data. Two types of big data will be accumulated with the continuous tagging and analysis of symptoms and herbs (Fig. 4B). The first is a semantic repository of symptoms/herbs. The manual tag module in the platform is an entrance to collect a large amount of symptom/herb corpus, which could be used towards generating curation of new semantic terms as we processed in the SymMap work [5]. By pooling the corpus every six months, analyzing the word frequency and confirming them the by experts, a much larger semantic repository would be built. The second is the symptom-herb and herb-herb associations. The relationship of herb-herb and symptom-herb would be dynamically changed due to adding of new prescriptions data. A pan-disease network would be generated, taking massive prescriptions from various diseases as the input. This network could be utilized as a baseline knowledge of TCM symptoms and herbs. Guide by this network, FangNet could give more sensitive

and accurate annotation of herb-herb and symptom-herb associations for a particular disease.

3.5. Perspective view of FangNet's symptom-herb network

3.5.1. Network construction

Totally 6,271 prescriptions from 127 TCM teams have been uploaded on the FangNet platform, covering more than 50 TCM diseases. To generate a perspective view of current data on FangNet, we have built a symptom-herb network according to the method part. However, our primary purpose is not to analyze any empirical prescriptions for specific diseases, but to verify the way of transforming big data into big knowledge by applying the method above. The network constructed initially contains 433 herbs, 104 symptoms, and 35,469 herb-herb edges and 12,095 symptom-herb edges. Filter by herb-herb interaction weight greater than 0.1 and co-occurrence event ≥ 5 , symptom-herb interaction weight greater than 0.1 and co-occurrence event ≥ 10 , we got 813 herb-herb and 76 symptom-herb edges, which were visualized in Fig. 5B. The Top 10 most frequently used herbs in the network are Danggui, Fuling, Gancao, Baizhu, Dangshen, Chuanxiong, Chenpi, Chaihu, with support value range from 0.2 ~ 0.35 (Supplementary Table T11).

3.5.2. Herb-herb co-occurrence

115 significant co-occurrence herb-herb pairs and 51 mutual exclusive herb-herb pairs are discovered from the network with a co-occurrence level of 2/3/4 and -2/-3/-4 (See the method). We have adopted a search on China National Knowledge Infrastructure (CNKI) using herb pairs as keywords. Significantly, the herb pairs with high co-occurrence tend to have been paid more research attention (Fig. 5A). For 17 herb pairs with the highest co-occurrence (level = 4), average 36 (2 ~ 195) corresponding studies were retrieved, while for 10 mutually exclusive pairs (level = -4/-3), only 2 studies for two pairs were obtained, which means that the herb pairs are generally rarely used simultaneously in Chinese medicine (Supplementary Table T12). This result shows that despite possible bias caused by the small diversity and scale of the current data, the herb-herb associations calculated from the FangNet network is highly consistent with the observation from the real-world study.

3.5.3. Symptom-herb association

Totally 76 significant associations between symptom and herb were found using the filtering parameters of edge weight ≥ 0.6 and co-occurrence event ≥ 20 (Fig. 5C, Supplementary Table T13). Most of the associations can be explained in the existing research of TCM. Among them, the rules of using Quanxie, Jiangcan and Baifuzi to treat mouth and eye skew are clearly recorded in lots of TCM studies, in which they are commonly used to treat facial palsy, trigeminal neuralgia and migraine [51,52 53,54]. Gancao is an useful herb to treat depression, proved by the previous pharmacological network study. Glycyrrhizic Flavone in Gancao involves the regulation of monoamine transmitters and their receptors [55]. For Danggui to treat depression, it is reported that Ferulic Acid in Danggui may regulate the antioxidant system by affecting the nervous system of 5-HT [56] and play an antidepressant role. Compound triterpenoid saponin in Jiegeng is often to treat rhinitis, which has the symptoms of nasal obstruction, rhinocnesmus and thin nasal secretion [57]. Overall, these results show that the symptom-herb associations discovered by FangNet network are consistent with the known studies of using specific herbs to treat corresponding symptoms.

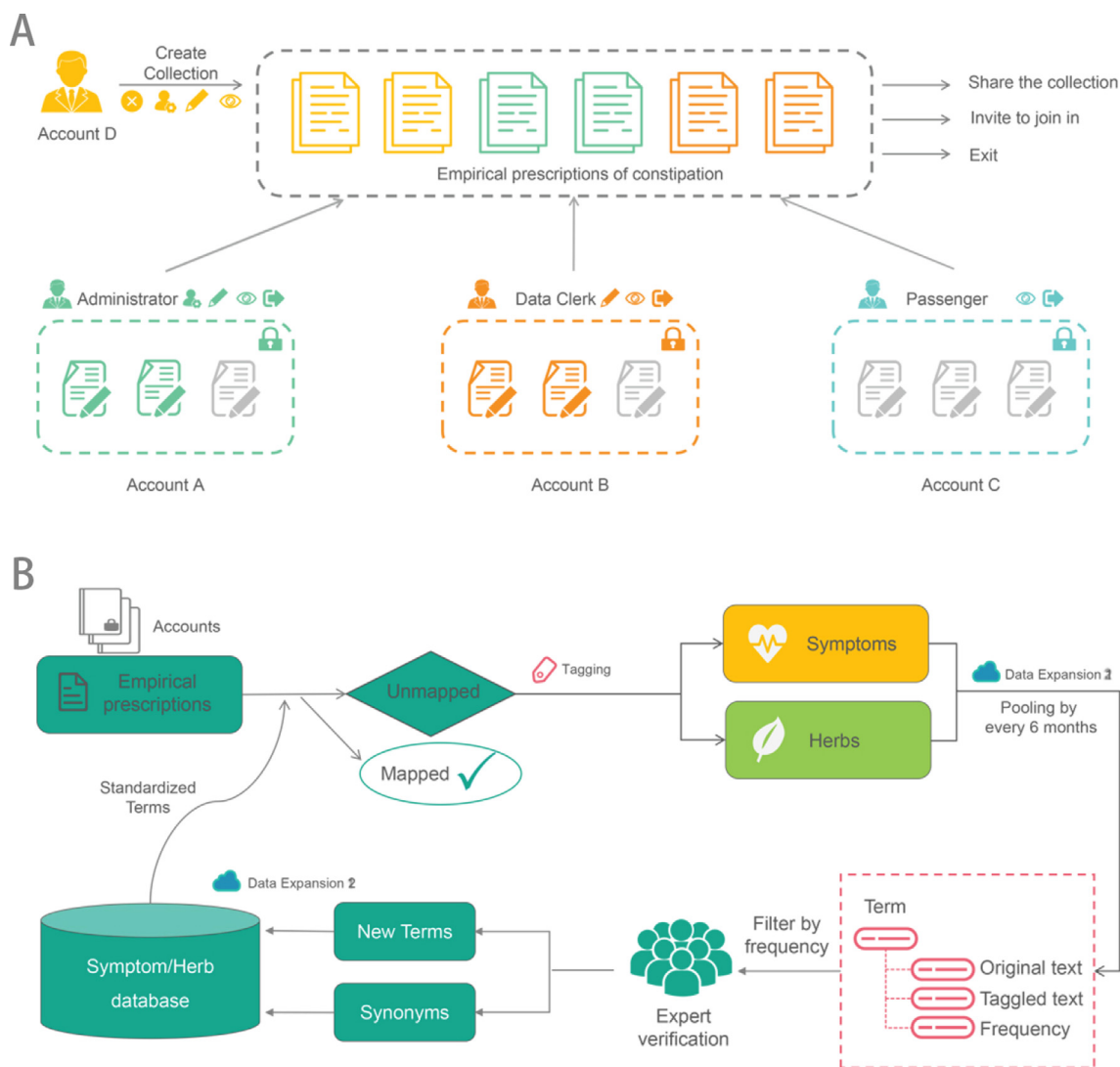


Fig. 4. Operation Mechanism of FangNet. A. Role-based permission for a collection of prescriptions. take a Take a collection of constipation prescriptions as an example. Prescriptions with different colors are created by different accounts. Different accounts have separate permission to create, authorize, edit, view, exit, invite to a collection. B. Expansion of the semantic repository and symptom-herb big data.

4. Discussion

Herbal medicines are the oldest and most universally used manner of treatment for human health and welfare [8]. They have already been proven for their remarkable potential in the treatment of a wide range of complex diseases. Herbs are essential for pharmacological research and drug development and have received increasing attention from pharmaceutical companies in the past ten years [3,4]. Its role as the starting materials for derivatization of natural biologically active ingredients, and its numerous valuable effect on particular disease phenotypes, make it have great value in drug discovery.

In this study, we built FangNet, a tool for retrospective analysis to a collection of empirical prescriptions for particular diseases, which can rank each herb's relative importance, determine the core herbs and find the associations to specific symptoms. Although it is illustrated by a collection of empirical prescriptions as the input in this study. However, more broadly, diverse prescriptions collections share a common characteristic could be used for analysis, such as a prescription collection from the same doctor, or a prescription collection with the same side effects.

FangNet only provides a limited semantic repository of symptoms and herbs, far from enough to convert all the input prescriptions. To solve this problem, the manual tag module was introduced into FangNet. In detail, users is allowed to specify tags for their content, and in turn, these tags provide an abundant corpus for the platform to expand its semantic repository in a semi-automatic increment way. This design can bring improved curation of semantic repository and more accurate of entities extraction.

However, our study still needs more improvement. In a real-world study, it is critical to evaluate the effectiveness of a prescription. In the FangNet platform, a prescription uploaded by the users is recommended as user self-conformed effective prescriptions. This may bring biases due to the different subjective evaluation criteria. Although it has less impact on the mining result of a single user, it may disturb quality of increment symptom-herb associations in the entire platform. Therefore, more research methods for evidence-based medicine (EBM) efficacy assessment would be introduced into the platform in the future, to distinguish the levels of evidence and applied mining with different confidence weights.

The current FangNet focuses on how to obtain the candidate herbs in a collection of empirical prescriptions. However, there is

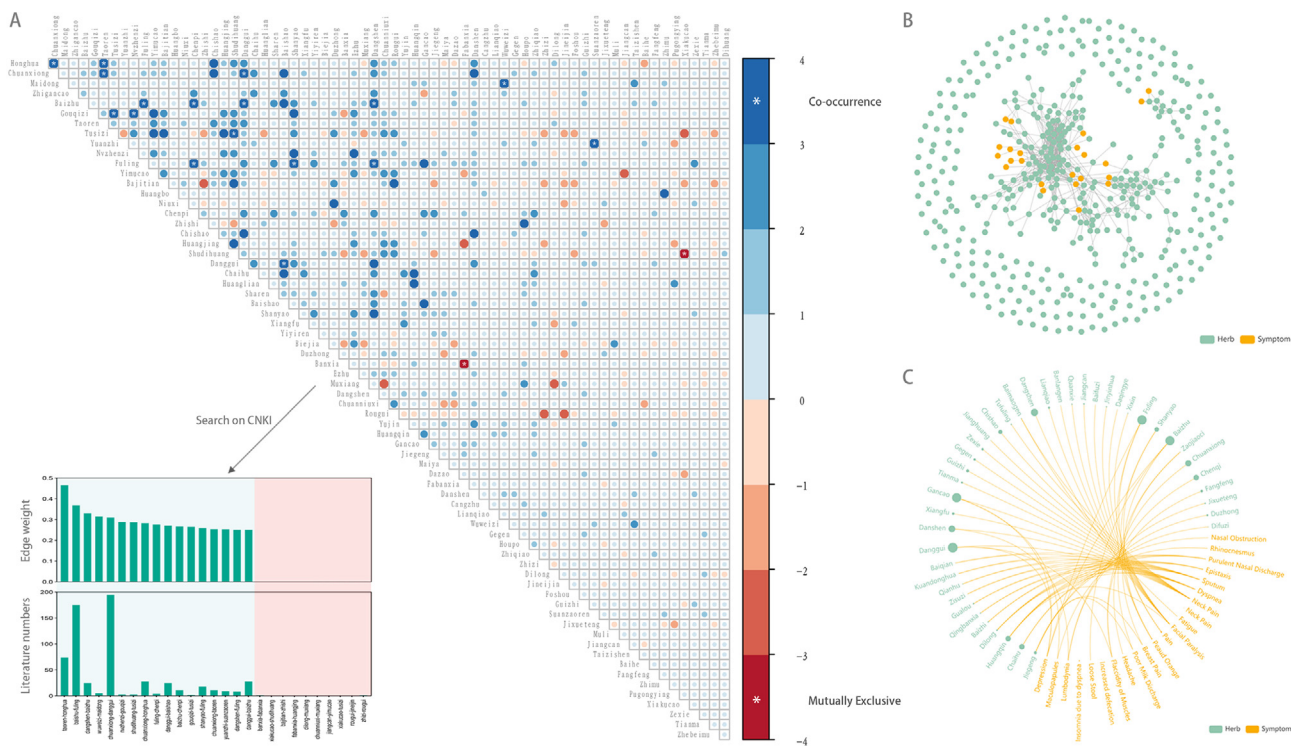


Fig. 5. Perspective view of FangNet's symptom-herb network. A. Herb-herb co-occurrence and Mutual Exclusivity. The triangular heat map is an illustration of co-occurrence and mutual exclusivity for 69 herbs with a frequency more than 0.05. The inner figure on the left is the result of the literature search on CNKI for herbs with significant co-occurrence and mutual exclusivity. 17 Herb pairs on the left with light green background are those herbs with high co-occurrence. 10 herb pairs on the right with light red background are those herbs with high mutual exclusivity. B. Symptom-herb network with 813 herb-herb and 76 symptom-herb edges. C. 76 high confidence symptom-herb associations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

still a gap existing between these candidates and a phenotype-based drug discovery (PDD) study. Discovering the biologically active ingredients is an important part of subsequent analysis from herb candidates. A PDD cellular phenotypic screening could only be carried out after these ingredients are identified or predicted. Thus, tools and sources for network pharmacology is necessary to be integrated or linked by FangNet. In our previous study, we have developed HERB database [58] (<http://herb.ac.cn/>), which collects experimental data of herb's high-throughput sequencing and provides herb's ingredients, gene targets and disease information. More database and other ingredients analysis tools would be linked or integrated with FangNet, to assist studies from prescriptions to drug ingredients discovery.

Ontology is widely used in sorting out the knowledge and building an easy-access expert knowledge system [42,43,59]. It is the key technology to reconstruct and present the knowledge net of TCM. In the FangNet platform's construction, the current extracted knowledge has not yet been organized in an ontology form. However, with more knowledge discovered, to sort out the knowledges in structured ontology hierarchy, would be an essential direction for optimizing the FangNet platform.

Recently, intelligent computational models such as deep learning and soft computing techniques have been widely applied in pattern understanding and knowledge-generating [60,61]. As a critical technology of big data analysis, it overcomes the restriction that traditional machine learning algorithms must rely on the selection of features. Intelligent computational technology yields multiple applications in the fields such as speech recognition, image recognition, object detection and drug target prediction. Limited by the data size, FangNet has just applied limited mining method in its knowledge mining. However, more intelligent computational models would be introduced to mine huge hidden

symptom to herb knowledge in the future, toward a more accurate and intelligent knowledge finding.

5. Conclusions

In this study, we propose a platform FangNet for mining herb hidden knowledge from a collection of empirical prescriptions, aiming at speed the path from raw clinic treatment observations to new drug discovery. A collaborative system is created to amplify the power of connection, by jointly building and interpreting their data from different teams, leading to a central hub for storage, mining and investigating TCM clinical important formulas on a secure cloud environment. Hidden symptoms-herb connections would be continuously accumulated and discovered, with the data expansion in FangNet, also with more advanced intelligent computational models brought in. It is expected that FangNet would be a great source and intelligent transformer for sorting herb hidden knowledge from clinic data, and provide import clues towards drug innovation and discovery.

CRedit authorship contribution statement

Dechao Bu: Methodology, Software, Writing - original draft, Writing - review & editing, Supervision. **Yan Xia:** Data curation, Writing - original draft, Methodology, Validation. **JiaYuan Zhang:** Methodology, Writing - original draft. **Wanchen Cao:** Validation, Data curation. **Peipei Huo:** Visualization, Software. **Zhihao Wang:** Software, Formal analysis. **Zihao He:** Software, Formal analysis. **Linyi Ding:** Validation, Data curation. **Yang Wu:** Investigation. **Shan Zhang:** Investigation. **Kai Gao:** Investigation. **He Yu:** Investigation. **Tiegang Liu:** Investigation. **Xia Ding:** Conceptualization.

Xiaohong Gu: Conceptualization. **Yi Zhao:** Conceptualization, Methodology, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Shihuan Tang for giving some suggestions on herb knowledge mining. We thank XiaoFei Ma for the web design and color schema advice.

Funding

National Key R&D Program of China [2018YFC1704100, 2019YFC1709801]; National Natural Science Foundation for Young Scholars of China [31701149,31701141]; National Natural Science Foundation of China [31672126,32070670]; National Natural Science Foundation of Zhejiang Province [LY20C060001, LY21C060003]. Funding for open access charge: National Key R&D Program of China [2018YFC1704100].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.11.036>.

References

- [1] Bent S. Herbal medicine in the United States: review of efficacy, safety, and regulation: grand rounds at University of California, San Francisco Medical Center. *J Gen Intern Med* 2008;23(6):854–9.
- [2] Xu HY et al. ETCM: an encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res* 2019;47(D1):D976–82.
- [3] Harvey AL, Edrada-Ebel RA, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 2015;14(2):111–29.
- [4] Cragg GM, Newman DJ. Natural products: A continuing source of novel drug leads. *Biochim Biophys Acta* 2013;1830(6):3670–95.
- [5] Wu Y et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Res* 2019;47(D1):D1110–7.
- [6] Vicki B. Back to Nature: Extinction of Medicinal Plants Threatens Drug Discovery. *J Natl Cancer Inst* 2008;100(12):838–9.
- [7] Qiu and Jane. Traditional medicine: a culture in the balance. *Nature* 2007;448(7150):126.
- [8] Mehta P, Lohidasan S, Mahadik KR. Pharmacokinetic behaviour of clinically important TCM prescriptions. *Orient Pharm. Exp* 2017;17(3):171–88.
- [9] Tang JL, Liu BY, Ma KW. Traditional Chinese medicine. *Lancet* 2008;372(9654):1938–40.
- [10] Che CT et al. Herb-Herb Combination for Therapeutic Enhancement and Advancement: Theory, Practice and Future Perspectives. *Molecules* 2013;18(5):5125–41.
- [11] Li Y et al. Herb Network Analysis for a Famous TCM Doctor's Prescriptions on Treatment of Rheumatoid Arthritis. *Evid Based Complement Alternat Med* 2015;2015.
- [12] Chu SM et al. Use of traditional Chinese medicine in patients with hyperlipidemia: A population-based study in Taiwan. *J Ethnopharmacol* 2015;168:129–35.
- [13] Xia P et al. Data Mining-Based Analysis of Chinese Medicinal Herb Formulae in Chronic Kidney Disease Treatment. *Evid Based Complement Alternat Med* 2020;2020:9719872.
- [14] Liu LQ et al. Data mining in Xu Runsan's Traditional Chinese Medicine practice: treatment of chronic pelvic pain caused by pelvic inflammatory disease. *J Tradit Chin Med* 2019;39(3):440–50.
- [15] Mehta P et al. Pharmacokinetic profile of phytoconstituent(s) isolated from medicinal plants—a comprehensive review. *J. Trad. Comp. Med* 2015;5(4):207–27.
- [16] Tu YY. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nature Medicine*, 2011. 17(10): p. 1217–20.
- [17] Chen KK. A pharmacognostic and chemical study of ma huang (*Ephedra vulgaris* var. *helvetica*). 1925. *J Am Pharm Assoc* 2012. 52(3): p. 406,409–407,412.

- [18] Mehta P, Hapte V. A Comprehensive Review on Pharmacokinetic Profile of Some Traditional Chinese Medicines. *New J. Sci.*2016,(2016-10-27), 2016. 2016: p. 1-31.
- [19] Pawar A et al. Strategies for formulation development of andrographolide. *RSC Adv* 2016;6(73):69282–300.
- [20] Li J et al. Traditional Chinese Medicine-Based Network Pharmacology Could Lead to New Multicomponent Drug Discovery. *Evid Based Complement Alternat Med* 2012;2012(12).
- [21] Li S et al. Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinf* 2010;11(Suppl 11):1–12.
- [22] Yang M et al. Application of Genetic Algorithm for Discovery of Core Effective Formulae in TCM Clinical Data. *Comput Math Methods Med* 2013;2013.
- [23] Ren X et al. Identifying potential treatments of COVID-19 from Traditional Chinese Medicine (TCM) by using a data-driven approach. *J Ethnopharmacol* 2020;258.
- [24] Yu JS et al. Traditional Chinese medicine treatments for upper respiratory tract infections/common colds in Taiwan. *Eur J Integr Med* 2014;6(5):538–44.
- [25] Fu XJ et al. Study of the distribution patterns of the constituent herbs in classical Chinese medicine prescriptions treating respiratory disease by data mining methods. *Chin J Integr Med* 2013;19(8):621–8.
- [26] Tang SH, et al., Designing New TCM Prescriptions Based on Complex System Entropy Cluster. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*, 2009. 11(2): p. 225–228 [Chinese].
- [27] Page L et al., The PageRank Citation Ranking: Bringing Order to the Web. 1998.
- [28] Jost L. Entropy and diversity. *Oikos* 2010;113(2):363–75.
- [29] Liu F et al. Unobtrusive Mattress-Based Identification of Hypertension by Integrating Classification and Association Rule Mining. *Sensors* 2019;19(7).
- [30] Heidemann J, Klier M, Probst F. Identifying Key Users in Online Social Networks: A PageRank Based Approach. in *International Conference on Information Systems*. 2010.
- [31] Yan E, Ding Y. Discovering author impact: A PageRank perspective. *Inf Process Manage* 2011;47(1):125–34.
- [32] Gleich, David F. PageRank Beyond the Web. *SIAM Rev* 2014;57(3):321–63.
- [33] Thalhammer A, Rettinger A. PageRank on Wikipedia: Towards General Importance Scores for Entities. *International Semantic Web Conference*. 2016.
- [34] Zeng X et al. PageRank analysis reveals topologically expressed genes correspond to psoriasis and their functions are associated with apoptosis resistance. *Mol Med Rep* 2016;13(5):3969–76.
- [35] Muggeo VMR. Estimating regression models with unknown break-points. *Stat Med* 2010;22(19):3055–71.
- [36] Chen HY, et al., Identifying Core Herbal Treatments for Children with Asthma: Implication from a Chinese Herbal Medicine Database in Taiwan. *Evid Based Complement Alternat Med*,2013,(2013-8-28), 2013. 2013: p. 125943.
- [37] Williamson EM. Synergy and other interactions in phytomedicines. *Phytomedicine* 2001;8(5):401–9.
- [38] Patwardhan B, Gautam M. Botanical immunodrugs: scope and opportunities. *Drug Discovery Today* 2005;10(7):495–502.
- [39] Adams LS et al. Analysis of the Interactions of Botanical Extract Combinations Against the Viability of Prostate Cancer Cell Lines. *Evid Based Complement Alternat Med* 2006;3(1):117–24.
- [40] Zhou X et al. Clinical phenotype network: the underlying mechanism for personalized diagnosis and treatment of traditional Chinese medicine. *Front Med* 2014;8(3):337–46.
- [41] Yu T et al. Research on the Construction of Knowledge Services Platform s for Traditional Chinese Medicine. *Medical Innovation of China* 2014;11(15):120–3 [Chinese].
- [42] Chien HY, et al., An Ontology-based Herb Expert System for Digestion System Wellness. 2016.
- [43] Chien HY, et al., An Ontology-Based Herb Therapy Recommendation for Respiration System. in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 2018.
- [44] Yang XM. Research progress on the pharmacological effects of Ligustrazine. *Chinese Journal of Biochemical and Pharmaceuticals* 2010;31(3):215–8 [Chinese].
- [45] Xie W et al. Study on effects and mechanism of extract of *Rhizoma Chuanxiong* on radicular neuralgia in rats with cervical spondylofic radiculopathy. *Journal of Guangdong Pharmaceutical University* 2008;24(5):496–8 [Chinese].
- [46] Yuan J et al. Four traditional Chinese medicines and their preparations for the treatment of migraine. *Chinese Journal of Modern Drug Application* 2008;08(20):62–3 [Chinese].
- [47] Lin YH, Chen WW. The pharmacological effects of sodium ferulate and the prospect of molecular modification. *Acta Pharmaceutica Sinica* 1994;29(09):717–20 [Chinese].
- [48] Zhu QG et al. Pharmacological effects of ferulic acid and its derivatives. *Journal of Chinese Medicinal Materials* 2001;12(2):522–6 [Chinese].
- [49] Gu J. Study on the chemical characteristic map of the compatibility of *Gastrodia elata-Uncaria* and the content changes of active ingredients. *Electronic Journal of Clinical Medical Literature* 2019;6(97):188–90 [Chinese].
- [50] Li, X.Q., The study of the distribution of effect ingredients in macrovascular and the gene expression before and after compatibility of *unicaria* with *gastrodia*. 2012. Southwest Jiaotong University [Chinese].
- [51] Wu S. Clinical experience of centipede and scorpion. *Guangming Journal of Chinese Medicine* 2012;27(10). 2124-2124 [Chinese].

- [52] Jiang DH, Zong CC. Observation on Curative Effect of Modified Qianzheng Powder on Peripheral Facial Paralysis. *Journal of Practical Traditional Chinese Medicine* 2013;11. 901-901 [Chinese].
- [53] Jia RZ. Treatment of 85 Cases of Peripheral Facial Paralysis with Acupuncture and Chinese Medicine. *Clinical Journal of Traditional Chinese Medicine* 2009;021(003):231.
- [54] Yang B, Li YZ. Modified Qianzheng Powder in Treating 155 Cases of Facial Paralysis. *Inf Traditional Chin Med* 1995;012(003):32 [Chinese].
- [55] Zong Y et al. Antidepressant Mechanism and Application of Glycyrrhizma Radix et Rhizoma. *Chinese Journal of Experimental Traditional Medical Formulae* 2016;22(10):194–8 [Chinese].
- [56] Zeni ALB et al. Ferulic acid exerts antidepressant-like effect in the tail suspension test in mice: Evidence for the involvement of the serotonergic system. *Eur J Pharmacol* 2012;679(1–3):68–74 [Chinese].
- [57] Li Y et al. Progress on Chemical Constituents and Pharmacological Effects of Platycodon Grandiflora (Jacq.) A.DC. *Food and Drug* 2016;18(1):72–5 [Chinese].
- [58] Fang SS. et al., HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Research*. gkaa1063, <https://doi.org/10.1093/nar/gkaa1063>.
- [59] Yu-Liang et al. A chronic disease dietary consultation system using OWL-based ontologies and semantic rules. *J Biomed Inform* 2015.
- [60] Schmidhuber, Jürgen. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
- [61] Bhardwaj J, Gupta KK, Gupta R. Towards a cyber-physical era: soft computing framework based multi-sensor array for water quality monitoring. *Drinking Water Eng Sci* 2018;11(1):1–7.