News & Views

SoFDA: an integrated web platform from syndrome ontology to network-based evaluation of disease–syndrome–formula associations for precision medicine

Yanqiong Zhang, Ning Wang, Xia Du, Tong Chen, Zecong Yu, Yuwen Qin, Wenjia Chen, Meng Yu, Ping Wang, Huamin Zhang, Xuezhong Zhou, Luqi Huang, Haiyu Xu

Clinical manifestations of symptoms play a crucial role in the diagnosis and appropriate treatment of diseases and are considered one of the main clinical features for contemporary disease taxonomy (i.e., international classification of diseases, ICD) [1]. Deep investigation on molecular connections among symptoms is one of the key tasks for developing a disease-specific knowledge network and thus promoting the refinement of disease taxonomy toward precision medicine [2]. Given the wide accessibility of symptoms and significant values for the management of various diseases, the detection of symptom groups (i.e., two or more co-occurring symptoms) from large-scale clinical data for symptom science and precision health is a vital and novel promising approach [3]. Therefore, the perception of symptom groups in the clinical settings and underlying molecular networks could be the next frontier for personalized medicine.

Traditional Chinese medicine (TCM) syndromes are distinctive symptom groups summarized based on TCM theories and clinical empirical knowledge for thousands of years. TCM holds syndromes as the core diagnostic criteria and therapeutic guidance [4]. Moreover, ICD-11 has incorporated a subcategory with a total of 196 TCM syndromes, implying that modern healthcare systems recognize syndrome-based diagnosis in TCM. Compared with individual symptoms, TCM syndromes propose a more systematic approach for differential diagnosis and disease classification because co-occurring symptoms within a syndrome can accurately represent the clinicopathological characteristics of a disease subtype [5]. Accordingly, the treatment for one symptom may “cross-over” and reduce the severity of other symptoms included in a syndrome, implying that the complex relationships between and among symptoms may provide new targets for interventions to reduce the negative effect of multiple co-occurring symptoms on patients’ outcome.

Given the lack of clinical diagnosis and treatment standards, a modern molecular basis underlying TCM syndrome may better classify diseases and guide precise prescriptions of medications. In addition, most data on TCM syndromes and TCM formulas are written in ancient Chinese, and only a tiny fraction of Chinese intellectuals alive today can understand exactly the definitions of TCM syndrome, leading to the difficulty for modern biomedical and pharmaceutical scientists to understand knowledge about TCM syndromes. Therefore, we built an integrated web platform, SoFDA (http://www.tcmip.cn/Syndrome/front/#/) with a manually curated database of syndrome ontology, a network-based evaluation tool of multi-way associations among diseases, TCM syndromes, and herbal formulas, and a visualization tool for constructing and editing multidimensional networks.

SoFDA combines data from multiple sources and provides comprehensive information on the commonly used TCM syndromes, related diseases, and corresponding TCM formulas (Fig. 1a, Supplementary material online, Sections 1 and 2). The TCM syndrome section contains 319 TCM syndromes and related 1610 TCM symptoms, which are indirectly associated with 3955 genes using the intermediate relationships between TCM symptoms and modern medicine (MM) symptoms manually curated by experts. Three types of direct associations, namely, 2650 TCM syndrome–TCM symptom associations, 2486 TCM symptom–MM symptom associations, and 139,493 TCM symptom–gene associations, as well as 141,036 TCM syndrome–gene indirect associations, are identified. The disease section records also provide detailed information of 8045 diseases, 8937 MM symptoms, and 10,273 related genes and their associations, including 1,689,384 disease–gene associations, 805,922 disease–MM symptom associations, and 666,074 MM symptom–gene associations. Regarding the TCM formula section, 1359 TCM formulas are used for the treatment of the corresponding TCM syndrome according to the records in the monograph of TCM Syndrome Differentiation and Treatment Guidance [6]. The 1214 herbs and 1796 putative target genes of these TCM formulas are also provided based on the data collected from the database of the Encyclopedia of Traditional Chinese Medicine.

Related content

[1] National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[2] Institute of Medical Intelligence, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[3] Shaanxi Academy of Traditional Chinese Medicine, Xi’an 710003, China
[4] Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[5] E-mail addresses: xzzhou@bjtu.edu.cn (X. Zhou), huangluqi01@126.com (L. Huang), byxu@ccmcm.ac.cn (H. Xu).
[6] These authors contributed equally to this work.
Fig. 1. Schema and application of SoFDA. (a) Overview and data statistics of SoFDA. (b) Framework to evaluate and calculate the indirect associations among TCM syndromes, diseases, and formulas, including data sources, similarity calculation, and association quantification. (c) Roles of associations among diseases, TCM syndromes, and TCM herbal formulas in disease classification and personalized medicine. (d) Validation of the feasibility of SoFDA in the individualized treatment of COVID-19.
Moreover, there are 14,003 TCM formula–herb and 103,303 herb–target gene direct associations, as well as 973,619 indirect associations between TCM formulas and target genes.

To facilitate the understanding of TCM syndromes and TCM herbal formulas from the perspective of molecular biology, enriched gene ontology (GO) terms or pathways associated with TCM syndromes, or targeted by TCM herbal formulas, as well as the protein–protein interaction network visualization and network module information, are also included in the SoFDA.

To calculate the associations of clinical entities (i.e., TCM syndrome, disease, and formula) in SoFDA, we adopted two widely used association measures (i.e., Jaccard and Cosine similarity [71]) to compute the degree of indirect associations between the aforementioned three entities in terms of six shared features, namely, symptoms, genes, enriched GO terms, enriched pathways, network modules, and network density (Fig. 1b and Supplementary materials online, Sections 1 and 2). This means that more shared symptoms between TCM syndrome and disease would result in higher similarity (thus stronger association). It is similar to other types of shared features. To validate the reliability of these two measures, we further generated random samples with shuffled features for the three clinical entities to filter the significant indirect associations (see Supplementary materials online, Section 3, Fig. S1 and Fig. S2 online, for detailed information). Herein, we list the main results of the four indirect associations, as shown below.

Disease–TCM syndrome association: TCM practitioners often perform syndrome differentiation to determine the clinical characteristics of diseases for preparing individualized treatment. Together with the diagnosis of a modern disease, the patterns of disease–syndrome associations hold significant evidence for the principles of TCM individualized treatment of different diseases with similar therapies (when different diseases hold the same syndrome) and the same disease with different therapies (when a disease holds different syndromes). Furthermore, these patterns would propose a very promising novel disease subtyping approach for precision medicine. Therefore, using the Jaccard and Cosine similarity measures, SoFDA offers a total of 4,091,345 high and 7,423,106 moderate association pairs between 8045 diseases and 319 TCM syndromes, which are included in the “disease–TCM syndrome association” section. For example, all three types of coronary heart diseases (CHDs; MalaCards IDs. CRN270, CRN113, and CRN304) are strongly associated with blood stasis syndrome (SoFDA ID. syndrome159) based on the symptom-related genes, GO terms, Reactome pathways, and network modules (Table S1 online).

Accumulating clinical evidence shows that blood stasis syndrome is the core syndrome of CHDs, and approximately 85% of patients with CHD may have this syndrome [8]. Mechanically, both CHD- and blood stasis syndrome-related genes are involved in several common pathways associated with platelet aggregation, fibrin clot formation, etc. Therefore, the “disease–TCM syndrome association” evaluation in the SoFDA may not only allow a correct diagnosis of diseases but also contribute to therapeutic decision-making and prognosis.

TCM syndrome–TCM syndrome association: correct differentiation of various syndromes is significant to ensure the reasonable treatment of diseases. To facilitate syndrome differentiation and treatment, SoFDA evaluates the associations among syndromes according to the clinical symptoms and related genes, enriched GO items, pathways, network density, and network modules. As a result, there are 16 identical, 49,903 high, 154,698 moderate association pairs among 319 TCM syndromes in the SoFDA.

TCM formula–TCM syndrome association: the concept of “formulas corresponding to the syndromes” may be the main therapeutic principles of TCM, which refers to a correct match between the “syndrome” and the “formula” based on the correspondence between clinical symptoms of syndrome and indications of formula, syndrome genes and formula targets, biological functions, and pathways involved syndrome– and formula-related target genes. Thus, the evaluation of syndrome–formula association may facilitate the prescription of an appropriate formula based on syndrome differentiation, which may improve the clinical efficacy, support the scientific research, and reinforce the standardization of TCM. Regarding all SoFDA data, there are 10 identical, 819,151 high, and 1,221,983 moderate association pairs among 319 TCM syndromes and 1359 herbal formulas.

Disease–TCM syndrome–TCM formula association: to quantitatively evaluate the association of syndrome, disease, and formula, we first identify the syndrome features for further stratification of the conditions of patients with a certain disease, which could help improve the efficacy of the selected intervention. Thereafter, an objective evaluation on the correspondence of formulas to the combination of disease and syndrome was performed according to the targeting symptoms, targeting pathological changes-related genes, functions, and pathways, which may help reveal the complex scientific connotation of formulas based on “disease–TCM syndrome–TCM formula association,” guide clinical practice, present a new strategy for studies of “treating the same disease with different formulas” and “treating different diseases with the same formula” theories, and notably make a difference to the innovation development of basic theories of TCM, research, and production of modern Chinese medicines.

Taken together, the above all-versus-all pairwise associations among diseases, TCM syndromes, and herbal formulas enable biomedical and pharmaceutical scientists to rank and filter the most promising associations for disease diagnosis and tailored interventions (Fig. 1c).

SoFDA provides a user-friendly web interface that enables users to search, browse, analyze, and download data of diseases, TCM syndromes, and TCM herbal formulas (Supplementary materials online, Section 4). On the “home” page (Fig. S3 online), users can perform a global search for SoFDA data using any words found in the disease name, TCM syndrome, TCM herbal formula, symptom, or gene. Notably, users can obtain the significantly associated TCM syndromes of a certain symptom list and the TCM/MM symptoms, TCM syndromes, TCM formulas, functional GO items, and pathways involved by a certain gene list by using the enrichment tool of SoFDA. By clicking on the “browse” button, users can select the types of information to browse, including TCM syndromes, diseases, or TCM herbal formulas, and click the SoFDA ID for each term to jump onto the details page. Especially on the “syndrome browse” page, the browsing results include the classification and codes of TCM syndromes and corresponding TCM herbal formulas (Fig. S4 online). For detailed information regarding TCM syndromes, diseases, and TCM herbal formulas, basic information, related functional annotation, and network visualization are available (Fig. S5 online).

The “association evaluation” panel calculates and visualizes the all-versus-all relationships among TCM syndromes, diseases, and TCM herbal formulas, and a list panel displays tables and images of the evaluation results. For example, on the “disease–syndrome–formula association” page, users can choose a certain disease, multiple syndromes, and corresponding formulas and subsequently set the clinical symptom type, association items, and levels. By clicking the “submit” button, detailed results on the selected association are obtained with various presentation forms, such as tables, heatmaps, networks, and upset plots (Fig. S6a online).

More worth mentioning, SoFDA provides a tool for a multilevel network of disease–syndrome–formula visualization, which can be used to illustrate the associations among diseases, TCM syndromes, symptom genes, herbal formulas, drug targets, and pathways according to users’ designation and modification. By selecting
one or more terms, users can build a disease–TCM syndrome–pathway, TCM herbal formula–drug target gene–pathway–TCM syndrome–disease, and many other networks (Fig. S6b online). They can also mark or modify nodes and edges of a network to facilitate further research.

Furthermore, SoFDA provides a syndrome–disease–formula–centric database with associated interconnections and rich annotations. All SoFDA data are available for both research purposes via an academic collaboration agreement and commercial licenses. We provide the details in the help section (No.12) of SoFDA (http://www.tcmip.cn/Syndrome/front/#/Help). A clear user manual is also provided.

To validate the efficacy of SoFDA in improving the accuracy of diagnoses and therapy response assessments, we herein took coronavirus disease–2019 (COVID–19) as an example (Supplementary materials online, Section 5). SoFDA provides detailed information of nine TCM syndromes and the corresponding 18 herbal formulas against COVID–19. Among these formulas, the Huashan Baidu formula (HBF) is the first TCM prescription with clinical approval in China and abroad and drug registration approval by the Ministry of Health and Prevention of the United Arab Emirates. According to the National Medical Products Administration, HBF is recommended for the treatment of patients with COVID–19 having the syndrome of epidemic toxin obstructing the lung. However, the molecular basis of patients with COVID–19 having the syndrome of epidemic toxin obstructing the lung and the reason for their favorable response to the HBF therapy remain unclear. In SoFDA, the main clinical symptoms of the syndrome of epidemic toxin obstructing the lung include fever, cough, headache, fatigue and lack of strength, dry mouth and sore pharynx, anorexia, nausea and vomiting, unconsciousness and delirious speech, chest oppression, choking sensation in the chest, myalgias, hasty panting, dyspnea, and abnormal stool and urine, which are similar to that of severe cases with COVID–19 based on World Health Organization (WHO) (2021) COVID–19 COVID–19 Clinical Management Living Guidance (https://www.who.int/zh/emergencies/diseases/novel-coronavirus–2019/technical-guidance, updated 25 January 2021).

According to the association evaluation of “COVID–19–syndrome of epidemic toxin obstructing the lung–HBF”, the symptom-related genes of COVID–19 with a syndrome of epidemic toxin obstructing the lung were significantly enriched into various immune-modulatory and inflammatory pathways, which have been indicated to play important roles in the occurrence and progression of COVID–19 [9]. Interestingly, several putative targets of HBF, such as TLR4, tumor necrosis factor, interleukin (IL)–6, IL1B, and nuclear factor–kappa B, also participate in the above pathways, which may be the molecular basis of the therapeutic efficacy of HBF in alleviating the clinical symptoms of COVID–19, such as fever, cough, chest discomfort, dry mouth, and sore throat [10–12] (Fig. 1d and Fig. S7a).

Moreover, the symptom-related genes of COVID–19 with the syndrome of epidemic toxin obstructing the lung were demonstrated to participate in multiple metabolic pathways. Among them, the disturbance of energy metabolism leading to the reduced levels of tricarboxylic acid metabolites has been detected in the peripheral blood of patients with COVID–19 [13]. Significant changes in metabolites related to sphingolipid metabolism, glycolysis and gluconeogenesis, and the urea cycle were also observed in the sera of patients with COVID–19, especially in critical conditions [14]. By contrast, the putative targets of HBF may reverse the above metabolic disturbances by regulating similar pathways, subsequently alleviating various clinical symptoms of COVID–19 with the syndrome of epidemic toxin obstructing the lung, such as myalgias, fatigue and lack of strength [15], anorexia, nausea and vomiting, and abnormal stool and urine (Fig. 1d and Fig. S7b).

Importantly, both the symptom-related genes of COVID–19 with the syndrome of epidemic toxin obstructing the lung and the putative targets of HBF were significantly enriched into various pathways related to coagulation disorders, implying a novel indication of HBF in treating the impaired blood coagulation caused by severe acute respiratory syndrome coronavirus 2 infection by promoting blood circulation, removing blood stasis, and improving microcirculation (Fig. 1d and Fig. S7c).

This case study suggests that the computational in silico strategies in the SoFDA platform may aid in modeling clinical and biological relevance to find new disease/syndrome-relevant signatures and discovering novel drug–target and drug–symptom associations for symptom-triggered individualized treatment.

In summary, SoFDA may be the first integrated web platform with the curated ontology of TCM syndromes and their relationships with genes, diseases, and formulas. It bridges information from macroscopic TCM syndromes, diseases, and TCM formulas to microscopic molecular mechanisms, which will deepen our understanding of ancient systematic medicine, TCM, and corresponding medical intervention.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81830111, 82030122, 82174533, and 81774201), National Key Research and Development Program of China (2018YFC1705201), Innovation Project of China Academy of Chinese Medical Sciences (CI2021A04907), Youth Innovation Team of Shaanxi Universities and Shaanxi Provincial Science and Technology Department Project (2016SF–378), and Fundamental Research Funds for the Central Public Welfare Research Institutes (ZKXT17058 and ZZ13–YQ–095). We would also like to thank Mr. Pu Xue and Mr. Moyu Liu from EBHBIO Gene Technology (Beijing) Co., LTD, for their help in the construction of the SoFDA database.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2022.03.013.

References


Yanqiong Zhang is a professor at the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences. She received her Ph.D. degree from the Chinese Academy of Medical Sciences & Peking Union Medical College and completed her postdoctoral training in Massachusetts General Hospital/Harvard Medical School. Her research interest focuses on bioinformatics of traditional Chinese medicine, network-based elucidation on molecular mechanisms of diseases and traditional Chinese medicine-originated therapeutics.

Ning Wang is currently a Ph.D. candidate in the School of Computer and Information Technology at Beijing Jiaotong University. His research interests focus on machine learning, deep learning, artificial intelligence in medicine, medical data analysis, and mining. He has been engaged in data mining and biomedical computational analysis related research, dedicated to the use of computer technology for clinical data mining of traditional Chinese medicine.

Xia Du is an associate professor at Shaanxi Academy of Traditional Chinese Medicine. She received her Ph.D. degree from Liaoning Normal University and her postdoctoral training at the China Academy of Chinese Medical Sciences. Her current research focuses on the identification of the material basis and the investigation of pharmacological mechanisms of traditional Chinese medicine by combining bioinformatics calculation and experimental validation.

Xuezhong Zhou is a professor and director at the Institute of Medical Intelligence, School of Computer and Information Technology, Beijing Jiaotong University. He received his Ph.D. degree from the College of Computer Science and Technology, Zhejiang University, and completed his postdoctoral training at China Academy of Chinese Medical Sciences. His research interest focuses on network medicine for traditional Chinese medicine and biomedical informatics.

Luqi Huang is an academian of the Chinese Academy of Engineering and president of the China Academy of Chinese Medical Sciences. He received his Ph.D. degree from Peking University. He carried out the fourth national survey on Chinese Medicinal Material Resources and developed the first traditional Chinese medicine-originated new drug for the treatment of COVID-19 with intellectual property rights in China. His research interest focuses on molecular pharmacognosy and resource science of traditional Chinese medicine.

Haiyu Xu is a professor at the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences. He received his Ph.D. degree from Tianjin University of Traditional Chinese Medicine. His research interest focuses on elucidating the bioactive constituents and molecular mechanisms related to therapeutic effects of traditional Chinese medicine based on the research strategies of integrative pharmacology of traditional Chinese medicine.