

# Ontology characterization, enrichment analysis, and similarity calculation-based evaluation of disease–syndrome–formula associations by applying SoFDA

Yudong, Jia Xu, Zecong Yu, Tong Chen, Ning Wang,  
Xia Du, Ping Wang, Xuezhong Zhou, Haiyu Xu, Yanqiong Zhang



Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China 1  
Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China 2  
Institute of Medical Intelligence and Beijing Key Lab of Traffic Data Analysis and Mining, School of  
Computer and Information Technology, Beijing Jiaotong University, Beijing, China 3  
National Resource Center for Chinese Materia Medica, Chinese Academy of Chinese Medical Sciences,  
Beijing, China 4

Yudong Liu, Jia Xu, Zecong Yu, Tong Chen, Ning Wang, Xia Du, *et al.* 2022. Ontology characterization, enrichment analysis, and similarity calculation-based evaluation of disease–syndrome–formula associations by applying SoFDA. *iMeta* 1: e1. <https://doi.org/10.1002/imt2.80>

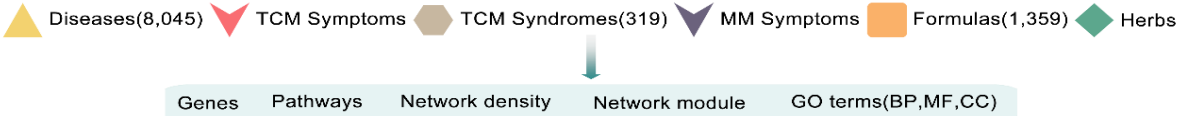
# Introduction / Results

- **SoFDA is the first manually curated public traditional Chinese medicine (TCM) syndrome annotation database.**
- **SoFDA explores the multiway associations among diseases, symptoms, genes, herbal formulas, drug targets, and pathways in heterogeneous biomedical networks with lots of customizations.**
- **A similarity-based association measurement was used to evaluate the association levels of disease–TCM syndrome, TCM syndrome–TCM syndrome and formula–TCM syndrome, and disease–TCM syndrome–formula by calculating the Jaccard/Cosine similarities.**

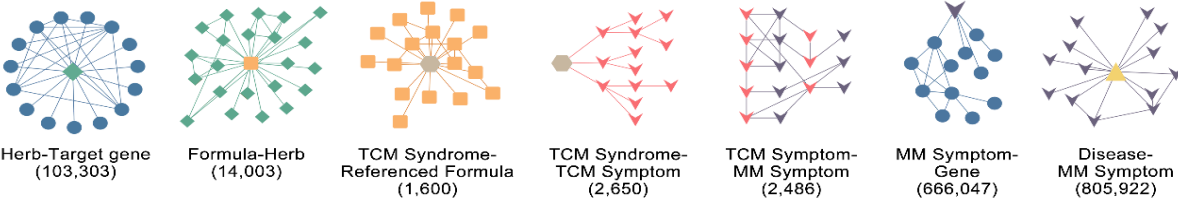


# Data support and sources of the SoFDA platform

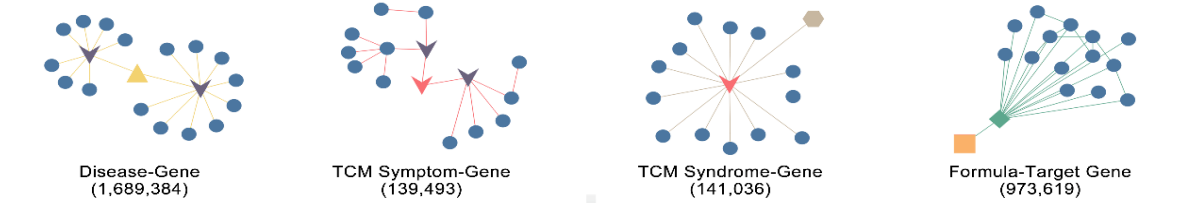
(A) Basic Data Sources



(B) Direct Associations



(C) Indirect Associations



(D) Evaluation

Association	Identical	High	Moderate	Low	Completely different
TCM Syndrome-TCM Syndrome(339,641)	16	49,703	154,698	113,074	22,150
Disease-TCM Syndrome(19,653,491)	0	4,091,345	7,423,106	5,820,898	2,318,142
TCM Syndrome-Formula(3,375,402)	10	819,151	1,221,983	970,170	364,008

Association evaluation



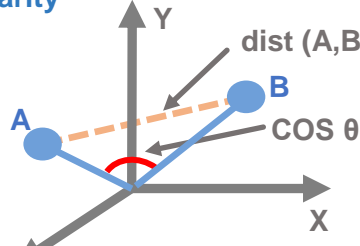
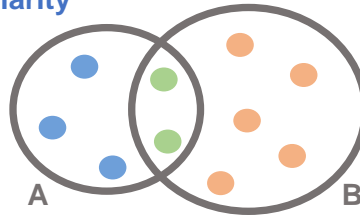
# Protocols for the data collection, the association evaluation, and the overview of the SoFDA platform

## Data Collection

- Syndrome data**
  - Syndrome name
  - Symptom
  - Alias of syndrome
  - Pathogenic syndrome element
  - Syndrome element of disease location
  - Formula against syndrome
- Disease data**
  - MalaCards (HUMAN DISEASE DATABASE)
  - OMIM (Human Genetics Knowledge for the World)
  - human phenotype ontology
  - orphanet
  - DisGeNET
  - Disease name
  - Global category
  - Symptom
  - Anatomical category
  - Disease-related gene
- Formula data**
  - ETCM ("Chinese Medicine Diagnostics")
  - Formula name
  - Symptom
  - Herbs containing the formula
  - Syndromes treated by formula
  - Formula-related genes

## Association evaluation

- Association**
  - Syndrome
  - Disease
  - Formula
- Arithmetic**
  - Cosine similarity
    - Functions
    - Pathways
    - Network features
  - Jaccard similarity
    - Genes
    - Symptoms
- Quantitative**
  - Identical
  - High
  - Moderate
  - Low
  - Completely different

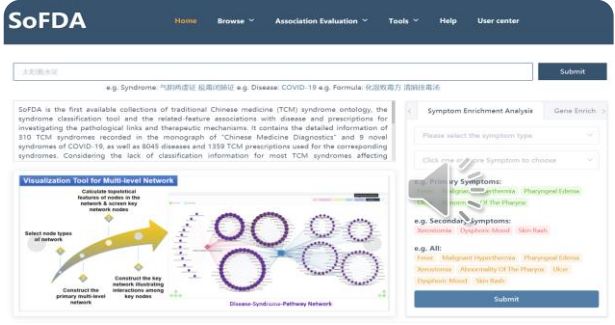



**Association similarity**

## SoFDA

- Function**
  - Syndrome-Syndrome association
  - Disease-Syndrome association
  - Formula-Syndrome association
  - Disease-Syndrome-Formula
  - Symptom/Gene Enrichment
  - Multi-level Network Visualization
  - Syndrome Classification Code Prediction
- Visualization**
  - Network visualization
  - Heatmap visualization

[www.tcmip.cn/Syndrome/front/#/](http://www.tcmip.cn/Syndrome/front/#/)



# The home page of the SoFDA platform

太阳蓄水证 (B)

Submit

e.g. Syndrome: 气阴两虚证 疫毒闭肺证 e.g. Disease: COVID-19 e.g. Formula: 化湿败毒方 清肺排毒汤

SoFDA is the first available collections of traditional Chinese medicine (TCM) syndrome ontology, the syndrome classification tool and the related-feature associations with disease and prescriptions for investigating the pathological links and therapeutic mechanisms. It contains the detailed information of 310 TCM syndromes recorded in the monograph of "Chinese Medicine Diagnostics" and 9 novel syndromes of COVID-19, as well as 8045 diseases and 1359 TCM prescriptions used for the corresponding syndromes. Considering the lack of classification information for most TCM syndromes affecting

(C) Symptom Enrichment Analysis

Gene Enrich >

Please select the symptom type

Click one or more Symptom to choose

e.g. Primary Symptoms:

Fever Malignant Hyperthermia Pharyngeal Edema  
Ulcer Abnormality Of The Pharynx

e.g. Secondary Symptoms:

Xerostomia Dysphoric Mood Skin Rash

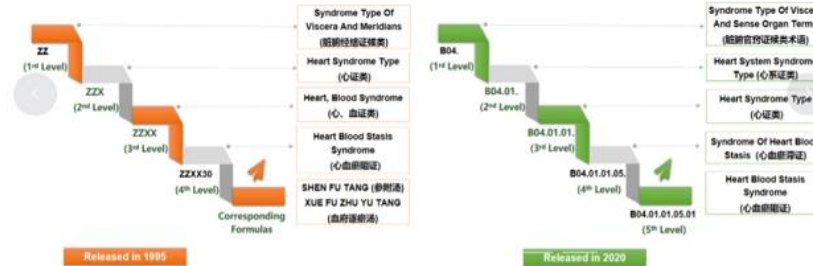
e.g. All:

Fever Malignant Hyperthermia Pharyngeal Edema  
Xerostomia Abnormality Of The Pharynx Ulcer  
Dysphoric Mood Skin Rash

Submit

### Syndrome Browse Page

Classification and codes of diseases and ZHENG of traditional Chinese medicine (GB/T 15657-1995)



Syndrome-Syndrome Association



Disease-Syndrome Association



Formula-Syndrome Association



Disease-Formula-Syndrome Association



User center



# A practical guide to browse the SoFDA platform

SoFDA

[Home](#)
[Browse](#)
[Association Evaluation](#)
[Tools](#)
[Help](#)
[User center](#)
中文 | English

### Syndrome Browse Page (A)

Classification and codes of diseases and ZHENG of traditional Chinese medicine (GB/T 15657-1995, Released in 1995)

- ▣ Cause Of Disease Syndrome Type
- ▣ Syndrome Type Of Six Meridian
- ▣ Syndrome Type Of Defense, Qi, Nutrient And Blood
- ▣ Syndrome Type Of Yin, Yang, Qi, Blood, Fluid And Humor, Phlegm
- ▣ Syndrome Type Of Viscera And Meridians
- ▣ \* Syndrome Type Of Six Meridian
- ▣ \* Syndrome Type Of Yin, Yang, Qi, Blood, Fluid And Humor, Phlegm
- ▣ \* Syndrome Type Of Viscera And Meridians
- ▣ \* Cause Of Disease Syndrome Type
- ▣ \* Syndrome Type Of Defense, Qi, Nutrient And Blood

Note: The predicted classification and code for the corresponding syndromes.

Classification and Codes of Diseases and Patterns of Traditional Chinese Medicine (Revised edition of GB/T15657-1995, Released in 2020)

- ▣ Syndrome Type Of Cause Of Disease Terms
- ▣ Syndrome Type Of Viscera And Sense Organ Terms
- ▣ Syndrome Type Of Six Meridian Terms
- ▣ Syndrome Type Of Defense, Qi, Nutrient And Blood Aspect Terms
- ▣ Syndrome Type Of Qi, Blood, Yin, Yang, Essence And Fluid Terms
- ▣ Syndrome Type Of Eight Principles Terms
- ▣ Syndrome Type Of Meridian And Collatera Terms
- ▣ Syndrome Type Of Triple Energizer Terms
- ▣ Unknown

Primary coding of TCM Syndromes

1/2

Differentiation System for Syndrome Element of Disease Location

1/4

Differentiation System for Pathogenic Syndrome Element

1/4

Syndrome Name	Alias of Syndrome	Formula Against Syndrome
Syndrome Of Toxin Congesting The Upper Energizer	Syndrome Of Toxin Congesting Qifen	QING XIN LIANG GE SAN QING YAN ZHI CHI TANG XI LU SAN YU WU SHI
Syndrome Of Wind-Heat In The Qi Wheel	Syndrome Of Wind-Heat In The White Of The Eye	SANG BAI PI SAN XIE FEI TANG
Syndrome Of Nutritional Damage Of Greater Yang And Nutr	Syndrome Of Nutritional Damage And Generalized Pain Affe	GU ZHI JIA SHAO YAO SHENG JIANG GE YI LIANG REN SHI
Syndrome Of Pathogen Invasion In Greater Yang And Splee	Syndrome Of Barmu Crumensmi And Rhubarb Decoction	GU ZHI JIA DA HUANG TANG
Syndrome Of Pathogen Stagnation In The Exterior Skin	Syndrome Of Similar To Malaria Of Greater Yang, Syndrome	GU ZHI MA HUANG GE BAN TANG
Syndrome Of Heat Pathogenic Distressing The Lung In Great	Syndrome Of Pathogenic Heat Distressing The Lung As Ash	MA HUANG XING REN GAN CAO SHI GAO TANG
Syndrome Of Spleen Deficiency And Qi Stagnation In Great	Syndrome Of Magnoliae Officinalis, Ginger Plus Phellia Tem	HUI PO SHENG JIANG BAN XIA GAN CAO REN SHEN TAN
Syndrome Of Stomach Qi Ascending Counterflow In Greater	Syndrome Of Kudalflower And Homatite Decoction	XUAN FU DAI ZHE TANG
Syndrome Of Superfices Cold And Interior Heat In Greater Y	Syndrome Of Exterior Excess And Retained Fluid, Syndrome C	XIAO QING LONG TANG
Syndrome Of Superfices Cold And Interior Heat In Greater Y	Syndrome Of Decoction Da Qing Long	DA QING LONG TANG

### Disease Browse Page (B)

Global category

1/1

Anatomical category

1/4

Disease MCID	Disease Name	Symptom	Category
11P002	11p154 Microduplication Syndrome	Appendicular Hypotonia Macrotia Anteverted Nares Synophrys Aggressive Behavior Delayed Speech And Language Development Tongue Abuse	Global category Anatomical category
11Q001	11q22.2q22.3 Microdeletion Syndrome	Thick Eyebrow Obsessive-Compulsive Behavior Anxiety Delayed Speech And Language Development Autism With High Cognitive Abilities Poor Eye Contact Hemiparesis/Paralysis Of The Limb	Global category Anatomical category

### Formula Browse Page (C)

Formula Distribution according to the Corresponding Syndrome Coding

1/5

English Name of Formula	Herbs containing Formula	Syndrome Treated by Formula
BU ZHONG YI QI TANG	HUANG QI GAN CAO REN SHEN DANG GUI JU PI SHENG MA (2543 341)	Syndrome Of Qi Deficiency In The Flesh Wheel Syndrome Of Qi Deficiency In Water Wheel Sunken Middle Qi Syndrome Syndrome Of Qing Yang Non-Promotion Sunken Spleen Qi Syndrome
DA CHENG QI TANG	DA HUANG HOU PO ZHI SHI MANG XIAO	Lesser Yin Combined Yang Brightness Syndrome Syndrome Of Liver Wind And Phlegm-Fire In Water Wheel Syndrome Of Deficiency Fire Invading Throat In Lesser Yin Large Intestinal Heat Bind Syndrome Syndrome Of Accumulated Heat In Stomach And Intestines Syndrome Of Yang Brightness And Bowel Excess



# The syndrome–syndrome association of the SoFDA platform

## Syndrome-Syndrome Association

**Syndrome-Syndrome Association**

\* Syndrome Type 1:  Syndrome 1:  \* Syndrome Type 2:  Syndrome 2:

\* Symptom:  \* Association Item:

Primary Symptom  
Secondary Symptom  
Primary & Secondary Symptom

Symptom  
Symptom-related Genes  
GO-BP  
GO-CC  
GO-MF  
Network Density  
Network Module  
Reactome Pathway

↓

Syndrome1	Syndrome2	Association Value	Association Level	Association Item
Syndrome Of Dampness Obstructing And Ar	Dual Deficiency Of The Lung-Spleen Syndro	0.533	High	Symptom
Cold-Dampness Obstructing The Lung Synd	Dual Deficiency Of The Lung-Spleen Syndro	0.5	High	Symptom
Dampness Toxin Depression Lung Syndrom	Cold-Dampness Obstructing The Lung Synd	0.467	High	Symptom
Syndrome Of Dampness Obstructing And Ar	Cold-Dampness Obstructing The Lung Synd	0.389	High	Symptom
Dampness Toxin Depression Lung Syndrom	Dual Deficiency Of The Lung-Spleen Syndro	0.25	High	Symptom

## For visualization

**Illustration Filter (Primary & Secondary Symptom)**

Syndrome1:  Syndrome2:  Association Level:  Association Item:



# The disease–syndrome and the formula–syndrome association of the SoFDA platform

## Disease-Syndrome Association

**Disease-Syndrome Association**

\* Syndrome Type:  Syndrome:  Disease:  \* Symptom:

\* Association Item:

Symptom  
Symptom-related Genes  
GO-BP  
GO-CC  
GO-MF  
Network Density  
Network Module  
Reactome Pathway

Primary Symptom  
Secondary Symptom  
Primary & Secondary Symptom

Syndrome	Disease	Association Value	Association Level	Association Item
Large Intestinal Heat Bind Syndrome	13q12.3 Microdeletion Syndrome	0.051	High	Symptom
Gallbladder Heat Syndrome	13q12.3 Microdeletion Syndrome	0.025	Moderate	Symptom
Gallbladder Heat Syndrome	12q14 Microdeletion Syndrome	0.193	High	Symptom-related Genes
Gallbladder Qi Deficiency Syndrome	16p12.1p12.3 Triplication Syndrome	0.184	High	Symptom-related Genes
Large Intestinal Heat Bind Syndrome	12q14 Microdeletion Syndrome	0.184	High	Symptom-related Genes

## For visualization

**Illustration Filter (Primary Symptom)**

Syndrome:  Disease:  Association Level:  Association Item:

## Formula-Syndrome Association

**Formula-Syndrome Association**

\* Syndrome Type:  Syndrome:  \* Formula:  \* Symptom:

\* Association Item:

Symptom  
Symptom-related Genes  
GO-BP  
GO-CC  
GO-MF  
Network Density  
Network Module  
Reactome Pathway

Primary Symptom  
Secondary Symptom  
Primary & Secondary Symptom

Syndrome1	Syndrome2	Association Value	Association Level	Association Item
Syndrome Of Dampness Obstructing And A	Dual Deficiency Of The Lung-Spleen Syndro	0.533	High	Symptom
Cold-Dampness Obstructing The Lung Synd	Dual Deficiency Of The Lung-Spleen Syndro	0.5	High	Symptom
Dampness Toxin Depression Lung Syndrome	Cold-Dampness Obstructing The Lung Synd	0.467	High	Symptom
Syndrome Of Dampness Obstructing And A	Cold-Dampness Obstructing The Lung Synd	0.389	High	Symptom
Dampness Toxin Depression Lung Syndrome	Dual Deficiency Of The Lung-Spleen Syndro	0.25	High	Symptom

**Illustration Filter (Primary & Secondary Symptom)**

Syndrome:  Formula:  Association Level:  Association Item:



# The disease-syndrome-formula association of the SoFDA platform

(A) Disease-Syndrome-Formula Association with the Same Disease

Disease	Syndrome	Formula	Correlation
12q14 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BA ZHEN TANG	0.042
12q14 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BA ZHENG SAN	0.053
12q14 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BA ZHEN YI MU GAO	0
12q14 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BA ZHENG HE JI	0
12q14 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAI BU WAN	0

(B) Disease-Syndrome-Formula Association with the Same Syndrome

Disease	Syndrome	Formula	Correlation
11q22.2q22.3 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAI HU TANG	0.121
11q22.2q22.3 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAO HE WAN	0.078
11q22.2q22.3 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAI TONG TANG	0.032
11q22.2q22.3 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAO AN WAN LING DAN	0.082
11q22.2q22.3 Microdeletion Syndrome	Syndrome Of Cold Congealing In The Uterus	BAI ZI ZI XIN WAN	0.037



# Multilevel network visualization tool of the SoFDA platform

Custom Network

Syndrome: Syndrome Of Cold ... +2

Symptom: Primary & Secondary Symptom

Disease: Core Pain, Painful... +2

Formula: MI GONGJIANG Y... +2

Show: Disease +3

Submit

Disease  
Syndrome  
Formula  
Symptom  
Gene  
Reactome Pathway

Primary Symptom  
Secondary Symptom  
Primary & Secondary Symptom

The list of core network targets and their topological eigenvalues

Gene #	network target (True/False) #	Degree #	Closeness #	Betweenness #
RPS29	true	21	0.015	0.292
RPS17	true	21	0.015	0.292
RPS26	true	21	0.015	0.292
RPS20	true	21	0.015	0.292
RPS14	true	22	0.015	24.292

For visualization

Network visualization showing clusters of nodes and edges, with a central hub node. The nodes are represented by blue circles, and the edges are represented by red lines. The network is divided into several distinct clusters, with a central hub node connecting to multiple clusters. The visualization is presented in a window titled 'For visualization' with an 'Edit' button and navigation controls.

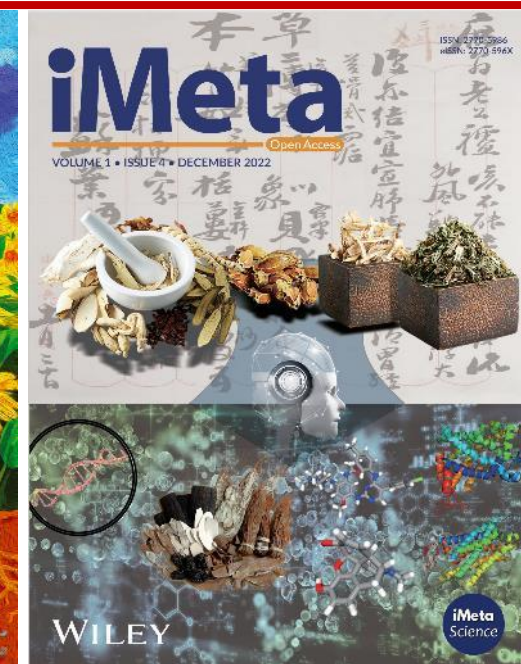


# Summary

SoFDA may be a promising platform that bridges information from TCM syndromes, diseases, and TCM formulas to molecular mechanisms, which will deepen our understanding of ancient systematic medicine, TCM, and the corresponding medical intervention.

Yudong Liu, Jia Xu, Zecong Yu, Tong Chen, Ning Wang, Xia Du, et al. 2022. Ontology characterization, enrichment analysis, and similarity calculation-based evaluation of disease–syndrome–formula associations by applying SoFDA. *iMeta* 1: e1. <https://doi.org/10.1002/imt2.80>





“***iMeta***” is an open-access Wiley partner journal launched by iMeta Science Society consist of scientists in bioinformatics and metagenomics world-wide. iMeta aims to promote microbiome, and bioinformatics research by publishing research, methods/protocols, and reviews. The goal is to publish high-quality papers (top 10%, IF > 15) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. Four issues were released in [March](#), [June](#), [September](#), and [December](#) 2022. Index by [Google Scholar](#), [Crossref](#), [Dimensions](#), [PubMed\(partial\)](#), [DOAJ](#) and [Scopus](#).



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

Submission: <https://mc.manuscriptcentral.com/imeta>



[office@imeta.science](mailto:office@imeta.science)



[Promotion Video](#)



[iMetaScience](#)



[iMetaScience](#)