Supplementary information
for
DCABM-TCM: A Database of Constituents Absorbed into the Blood and Metabolites of Traditional Chinese Medicine

Xinyue Liu 1,##, Jinying Liu 2,##, Bangze Fu 3,##, Ruzhen Chen 1,##, Jianzhou Jiang 1,4, He Chen 4, Runa Li 3, Lin Xing 3, Liying Yuan 4, Xuetai Chen 5, Jing Zhang 5, Honglei Li 6, Shuzhen Guo 7, Feifei Guo 8, Jiachen Guo 4, Yuan Liu 1, Yaning Qi 1, Biyue Yu 4, Feng Xu 5,*; Dong Li 1,*; Zhongyang Liu 1,4,*

1 State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China
2 College of Traditional Chinese Medicine, Chengde Medical University, Chengde 067000, China
3 School of Biomedicine, Beijing City University, Beijing 100094, China
4 School of Life Sciences, Hebei University, Baoding 071002, China
5 School of Pharmaceutical Sciences, Peking University, Beijing 100191, China
6 Beijing Cloudna Technology Company, Limited, Beijing 100029, China
7 School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China
8 Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
## These authors contributed equally to this work.
* Corresponding authors: Zhongyang Liu, liuzy1984@163.com; Dong Li, lidong.bprc@foxmail.com; Feng Xu, xufeng76@hsc.pku.edu.cn
Tutorials
for
the analysis functions of
DCABM-TCM

Updated on 2023-06-16
By Zhongyang Liu and Jinying Liu

Tutorial for the network pharmacology analysis function 2

Tutorial for the analysis function of the “the priorization of blood constituents, herbs and prescriptions targeting the target gene” 13

Tutorial for the analysis function of the “the priorization of blood constituents, herbs and prescriptions targeting the pathway” 15

Tutorial for the analysis function of the “the priorization of blood constituents, herbs and prescriptions targeting the disease” 17

Tutorial for the network pharmacology analysis function
Taking the herb “ZHI MU” as an example, we show the tutorial for the network
pharmacology analysis function.

**Step 1**: By database browse (A) or search page (B), users can enter the detailed annotation page of “ZHI MU”, on which the results of the network pharmacology analysis of “ZHI MU” will be given in the bottom column (shown in the next step).

![Browse DCABM-TCM](A)

![Search DCABM-TCM](B)

**Step 2**: The network pharmacology analysis results are given here. The results include three sections: Result1, target prediction result; Result2, bioinformatics analyses of potential targets (including KEGG pathway, GO functional term, CTR/OMIM disease enrichment analyses); Result3: blood constituent-target-pathway-disease association network visualization.

**Step 3**: Parameter adjustment

**Score_cutoff**: For each blood constituent, DCABM-TCM ranks its predicted
candidate targets according to the order of decreasing scores given by the
target prediction algorithm previously constructed by us [Sci Rep. 2016,
6:21146] for the drug-target interaction prediction. The predicted candidate
targets with scores>="Score_cutoff" (including known direct targets backed by
DrugBank, KEGG and TTD) will be hypothesized as the potential targets of the
blood constituent. The following network pharmacology analyses are based on
these potential targets. The default value is 20.

**Adjusted P-value cutoff:** The significantly enriched GO functional terms,
KEGG biological pathways and TTD/OMIM diseases among the potential
targets are analyzed. The cutoff of the P-value after Benjamini-Hochberg
multiple testing correction (i.e. adjusted P-value cutoff) for the significant level
can be set. The default value is 0.05.

Users can change the two parameters here. If the parameters are changed, all
results will be updated.
Result 1: Target prediction result

For each blood constituent of “ZHI MU”, the predicted candidate targets (denoted by Gene Symbol) with scores>=Score_cutoff ranked according to the order of decreasing scores given by the target prediction method are listed in the result table (including known targets).
A: According to user-defined parameter, only the predicted candidate targets with scores>=Score_cutoff (also including known targets) will be presented in the result table and be considered as potential targets.

B: Only for blood constituents with chemical structures, targets can be predicted.

C: The target prediction result table. For each blood constituent, known targets reported by DrugBank (version: 20150726), KEGG (version: July 31, 2014) and TTD (version: 4.3.02) database (marked by "known target in DrugBank, KEGG or TTD") will be given first if there are, followed by predicted targets.

D: Users can search the interested gene among these targets.

E: The complete target prediction result can be downloaded.

Result 2: Bioinformatics analyses of potential targets

Further for the potential targets of blood constituents, DCABM-TCM provides three enrichment analyses, including KEGG biological pathway, GO functional term and OMIM/TTD disease enrichment analyses.
The significantly enriched KEGG biological pathways, GO functional terms and OMIM disease phenotypes/TTD diseases among the potential targets of blood constituents together with corresponding adjusted P-value and targets mapped to this term will be presented in the result table.

Attention: The enrichment analyses are based on the predicted candidate targets with scores>=Score_cutoff (also including known targets).

**KEGG pathway enrichment analysis result**
As you set, these enrichment analyses are based on predicted candidate targets with scores $> 20$. The significantly enriched functional terms (Gene ontology term, KEGG pathway and OMIM/CTD disease) are highlighted in red, whose adjusted $p$-values are smaller than 0.05.

**Target Prediction Result**

**Download all the enrichment analysis results**

**KEGG Pathway**

**Disease**

**Gene Ontology**

Select table view

**KEGG pathway**

**Enriched KEGG pathways**

**Table View**

**KEGG pathway ID**

**KEGG pathway name**

**herb_ZHI_MU**

**Adjusted p-value**

**Targets**

**Targets mapped to this term**

**GAA, GANC, MGAM**

**KEGG Pathway**

**Disease**

**Gene Ontology**

Select table view

**KEGG Hierarchy**

**Enriched KEGG pathways**

**Table View**

**KEGG pathway ID**

**KEGG pathway name**

**herb_ZHI_MU**

**Adjusted p-value**

**Targets**

**Targets mapped to this term**

**GAA, GANC, MGAM**
A: Here we provide two kinds of views to present the KEGG pathway enrichment analysis results. Tree view (“KEGG hierarchy”) shows the hierarchy of pathways.

B: The “KEGG pathway IDs” are crosslinked to the KEGG database.

C: The significantly enriched pathways with adjusted P-value smaller than the cutoff set by users are highlighted in red.

D: “Targets” are referred to as the targets mapped to the pathway and clicking on the number will present the detailed target list.

### Disease enrichment analysis result

| KEGG Pathway | Disease | Gene Ontology |
|--------------|---------|---------------|
| Select disease data | CTD | |

#### CTD disease enrichment analysis result

| Enriched CTD diseases | Pathway | Adjusted p-value |
|------------------------|---------|------------------|
| Nutritional and Metabolic Diseases | 2.99e-001 | 3 |
| Nervous System Diseases | 4.91e-001 | 3 |
| Neoplasms | 7.81e-001 | 3 |
| Neoplasms by Site | 7.96e-001 | 2 |
| Breast Diseases | 1.90e-001 | 2 |
| Vascular Diseases | 3.40e-001 | 2 |
| Glucose Metabolism Disorders | 1.20e-001 | 2 |
| Acetyltransferases | 1.23e-001 | 2 |
| Immune System Diseases | 4.49e-001 | 2 |
| Digestive System Diseases | 7.50e-001 | 2 |
| Catechol-O-Methyltransferase | 1.41e-001 | 2 |
| Enzymes and Coenzymes | 7.50e-002 | 2 |

#### OMIM disease enrichment analysis result

| OMIM ID | Disease name | Adjusted p-value | Targets |
|---------|--------------|------------------|---------|
| OMIM:209110 | Catechol-O-Methyltransferase II Deficiency, Myopathic, | 1.86e-002 | 1 |
| OMIM:610036 | Charcot-Marie-Tooth Disease, Axonal, Type 2A | 1.08e-001 | 1 |
| OMIM:614212 | Erythropoietic Progenitor Cell Deficiency, Erythroblast Induced, Succinyl-CoA Transferase Deficiency, Type 1 | 1.36e-004 | 1 |
| OMIM:6008036 | Erythropoietic Progenitor Cell Deficiency, Lethal Neonatal | 1.36e-004 | 1 |

Linked to OMIM database
The GO enrichment analysis result is presented as a tree structure which shows the hierarchical relationship between GO terms. GO terms of three categories are painted by different colors.

**Result 3: Blood constituent-target-pathway/disease association network visualization**

The network graph is drew based on the predicted candidate target proteins with scores $\geq$ Score cutoff (user defined) of the constituents.
Blood constituent-target-pathway/disease network

As you set, the network graph is drawn based on the known and predicted candidate target proteins with scores not smaller than 20. And in the "Simplified network view", only significantly enriched KEGG pathways and OMIM/CTD disease phenotypes with adjusted P-value smaller than 0.05 are shown.

To emphasize the important elements, the size of the target node, pathway node and disease node is proportional to their degree in the network, which is respectively defined as the number of compounds acting on the target, the number of targets involved in the pathway and the number of targets being known the disease-related genes.

There are two types of network view. Different from the "Whole network view", in the "Simplified network view" only those significantly enriched pathways/diseases (adjusted P-value < 0.05 cutoff set by users) are shown in the network.

The simplified network view

The search result of node "GANC"
The whole network view

The simplified network view only exhibiting those targets with no fewer than 2 linking blood constituents

herb_ZHI_MU

To emphasize the important elements, in the network users can only exhibit those targets with no fewer than 2 linking compounds.

When the value on the slider above is changed, please wait for ~10 seconds with patience to see the updated network view!
A: In the association network, there are four kinds of nodes distinguished by different shapes and colors including blood constituents, targets, biological pathways and OMIM/TTD diseases and three types of edges including constituents-target association (if the protein is a known or potential target of the constituents), target-pathway association (if the target protein is a member of the biological pathway) and target-disease association (if the target protein is a known related gene of the disease). In addition, to emphasize the important elements, the size of the target node, pathway node and disease node is proportional to their degree in the network, which is respectively defined as the number of compounds acting on the target, the number of targets involved in the pathway and the number of targets being known the disease-related genes.

B: There are two types of network view. Different from the “Whole network view”, in the “Simplified network view” only those significantly enriched pathways /diseases (adjusted P-value <= cutoff set by users) are shown in the network.

C: Users can only exhibit those targets with no fewer than M linking blood constituents (which can be adjusted by the slider) in the network.

D: The network graph navigation buttons. In addition, to further facilitate navigation, besides using these navigation buttons, users can also pan the network view by directly holding down the left mouse button in a blank area and moving the mouse. Users can also move the nodes on the graph by holding down the left mouse button on the node and moving it.

E: Node search function using the node name. Once the interested node is found in the network, the node will be highlighted in size and color on the network graph, and the area around it will be zoomed in.

F: The network graph and corresponding network file can also be downloaded.
Tutorial for the analysis function of the “the prioritization of blood constituents, herbs and prescriptions targeting the target gene”

For an interested target, this analysis function prioritizes the candidate blood constituents, prescriptions and herbs which potentially target this target gene (i.e. The target-based blood constituent/prescription/herb screening). Blood constituent-target gene associations are predicted by a prediction method (BATMAN-TCM) previously constructed by us [Sci Rep. 2016, 6:21146]. Further, we think significantly enriched prescriptions/herbs among the blood constituents which target this target gene are potential candidate prescriptions/herbs targeting this target gene.

Parameter adjustment

**Target prediction score cutoff:** Target gene-blood constituent associations are based on the known compound-target associations and the predicted ones given by BATMAN-TCM. The blood constituents recorded in DCABM-TCM with known associations or prediction scores no smaller than the given cutoff "Score cutoff" are candidate blood constituents targeting this target gene, and will be presented and used for further analyses. The default cutoff value is set to 10.

**P-Value cutoff:** We think significantly enriched prescriptions/herbs among the blood constituents which target this target gene are potential candidate prescriptions/herbs targeting this target gene. The enrichment analysis is performed by the hypergeometric cumulative distribution test. This parameter is the cutoff of P-value after Benjamini-Hochberg multiple testing correction for the enrichment analysis (Adjusted P-value). The default value is set to 0.05.

You can change the two parameters and re-analyze all results.
1) Change the parameter

2) Re-analyze all the results below
Tutorial for the analysis function of the “the priorization of blood constituents, herbs and prescriptions targeting the pathway”

Here are the candidate blood constituents, prescriptions and herbs which potentially target this pathway. This function aims to prioritize candidate blood constituents/herbs/prescriptions targeting a specific pathway (i.e. The pathway-based blood constituent/prescription/herb screening).

We think that significantly enriched blood constituents among the pathway’s member genes are potential candidate blood constituents targeting this pathway. Further significantly enriched herbs/prescriptions among these potential candidate blood constituents are thought to be potential candidate herbs/prescriptions targeting this pathway. Constituent-gene associations are based on the known compound-target associations and the predicted ones given by BATMAN-TCM.

Parameter adjustment

Target prediction score cutoff: Constituent-gene associations are based on the known compound-target associations and the predicted ones given by BATMAN-TCM. For a pathway member gene, the blood constituents recorded in DCABM-TCM with known associations or prediction scores no smaller than the given cutoff "Score cutoff" are the candidate ones targeting it, and will be used in the analyses. The default value is 10.

P-value for enriched blood constituents: The enrichment analysis is performed by the hypergeometric cumulative distribution test. This parameter is the cutoff of P-value after Benjamini-Hochberg multiple testing correction for the enrichment analysis (Adjusted P-value). The default value is set to 0.05.

P-value for enriched herbs/prescriptions: The enrichment analysis is performed by the hypergeometric cumulative distribution test. This parameter is the cutoff of P-value after Benjamini-Hochberg multiple testing correction (Adjusted P-value). The default value is set to 0.05.

Users can change the three parameters and re-analyze all results.
1) Change the parameter
2) Change the parameter
3) Re-analyze all the results below
Tutorial for the analysis function of the “the priorization of blood constituents, herbs and prescriptions targeting the disease”

Here are the candidate blood constituents, prescriptions and herbs which potentially target this disease. This function aims to prioritize candidate blood constituents/herbs/prescriptions targeting a specific disease (i.e. The disease-based blood constituent/prescription/herb screening).

We think that significantly enriched blood constituents among the disease-related genes are potential candidate blood constituents targeting this disease. Further significantly enriched herbs/prescriptions among these potential candidate blood constituents are thought to be potential candidate herbs/prescriptions targeting this disease. Constituent-gene associations are based on the known compound-target associations and the predicted ones given by BATMAN-TCM.

Parameter adjustment

**Target prediction score cutoff:** For a disease-related gene, the blood constituents recorded in DCABM-TCM with known associations or prediction scores given by BATMAN-TCM no smaller than the given cutoff "Score cutoff" are the candidate ones targeting it, and will be used in the analyses. The default value is 10.

**P-value for enriched blood constituents:** The enrichment analysis is performed by the hypergeometric cumulative distribution test. This parameter is the cutoff of P-value after Benjamini-Hochberg multiple testing correction (Adjusted P-value). The default value is set to 0.05.

**P-value for enriched herbs/prescriptions:** The enrichment analysis is performed by the hypergeometric cumulative distribution test. This parameter is the cutoff of P-value after Benjamini-Hochberg multiple testing correction (Adjusted P-value). The default value is set to 0.05.

Users can change the parameters and re-analyze all results.
Parameter adjustment

1) Change the parameter

3) Re-analyze all the results below

Analysis results

Download

Linked to the detailed annotation page of the blood constituent

Blood constituent

| Blood constituent | Adjusted P-value | The number and list of disease-related genes targeted by the constituent (represented by geneID) |
|-------------------|------------------|--------------------------------------------------------------------------------------------------|
| PatChem CID: 637546 (DCABM ID: BC1094, S1-2) | 3.54e-2 | 1: 100509659 |
| PatChem CID: 1549106 (DCABM ID: BC1545, S1-3) | 3.54e-2 | 1: 100509659 |
| PatChem CID: 54708747 (DCABM ID: BC3327, S1-2) | 3.54e-2 | 1: 100509659 |

Enriched blood constituents among the disease-related genes

Linked to the detailed annotation page of the prescription

Prescription

| Prescription | Adjusted P-value | The number and list of the prescription for all constituents targeting the disease (represented by PubChem CID) |
|--------------|------------------|-----------------------------------------------------------------------------------------------------------|
| GUA LOU GUI ZHI TANG | 5.71e-4 | 3: 637546 (637541, 5327542) |
| NAND SHI JIAO NANG | 2.99e-3 | 2: 54708747 (5327542) |
| DANG GUI HOMI HUA XIA | 3.56e-2 | 1: 5327542 |

Enriched prescriptions among these blood constituents

Linked to the detailed annotation page of the herb

Herb

| Herb | Adjusted P-value | The number and list of the herb's bioactivities targeting the disease (represented by PubChem CID) |
|------|------------------|---------------------------------------------------------------------------------------------------|
| QING PI ZHU | 1.67e-4 | 2: 36510963 (5327542) |
| GU ZHI | 2.67e-4 | 1: 637542 |

Enriched herbs among these blood constituents

Download

Result download

Linked to the detailed annotation page of the herb

Result download

Linked to the detailed annotation page of the prescription

Result download

Linked to the detailed annotation page of the blood constituent

Result download

Linked to the detailed annotation page of the blood constituent

Result download

Linked to the detailed annotation page of the blood constituent