


# NP-TCMtarget: a network pharmacology platform for exploring mechanisms of action of traditional Chinese medicine

Aoyi Wang<sup>1</sup>, Haoyang Peng<sup>1</sup>, Yingdong Wang<sup>1</sup>, Haoran Zhang<sup>1</sup>, Caiping Cheng<sup>1</sup>, Jinzhong Zhao<sup>1</sup>, Wuxia Zhang<sup>1</sup>, Jianxin Chen<sup>2,\*</sup>, Peng Li <sup>1,\*</sup>

<sup>1</sup>Shanxi Key Lab for Modernization of TCM, College of Basic Sciences, Shanxi Agricultural University, 1 Mingxian South Road, Taigu District, Jinzhong, 030801, China

<sup>2</sup>School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, 11 North Third Ring Road East, Chaoyang District, Beijing 100029, China

\*Corresponding authors. Jianxin Chen, School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, 11 North Third Ring Road East, Chaoyang District, Beijing 100029, China. E-mail: [cjx@bucm.edu.cn](mailto:cjx@bucm.edu.cn); Peng Li, Shanxi Key Lab for Modernization of TCM, College of Basic Sciences, Shanxi Agricultural University, 11 North Third Ring Road East, Chaoyang District, 030801, China. E-mail: [lip@sxau.edu.cn](mailto:lip@sxau.edu.cn)

## Abstract

The biological targets of traditional Chinese medicine (TCM) are the core effectors mediating the interaction between TCM and the human body. Identification of TCM targets is essential to elucidate the chemical basis and mechanisms of TCM for treating diseases. Given the chemical complexity of TCM, both *in silico* high-throughput compound-target interaction predicting models and biological profile-based methods have been commonly applied for identifying TCM targets based on the structural information of TCM chemical components and biological information, respectively. However, the existing methods lack the integration of TCM chemical and biological information, resulting in difficulty in the systematic discovery of TCM action pathways. To solve this problem, we propose a novel target identification model NP-TCMtarget to explore the TCM target path by combining the overall chemical and biological profiles. First, NP-TCMtarget infers TCM effect targets by calculating associations between herb/disease inducible gene expression profiles and specific gene signatures for 8233 targets. Then, NP-TCMtarget utilizes a constructed binary classification model to predict binding targets of herbal ingredients. Finally, we can distinguish TCM direct and indirect targets by comparing the effect targets and binding targets to establish the action pathways of herbal component—direct target—indirect target by mapping TCM targets in the biological molecular network. We apply NP-TCMtarget to the formula XiaoKeAn to demonstrate the power of revealing the action pathways of herbal formula. We expect that this novel model could provide a systematic framework for exploring the molecular mechanisms of TCM at the target level. NP-TCMtarget is available at <http://www.bcxnfz.top/np-tcmtarget>.

**Keywords:** traditional Chinese medicine; herbal target; target discovery; transcriptomics; herbal ingredients

## Introduction

Traditional Chinese medicine (TCM), a treasured human heritage passed down through thousands of years, plays a pivotal role in disease prevention, treatment, and diagnosis. The therapeutic essence of TCM lies in their multifaceted action on biological molecular networks to exert pharmacological effects through a multi-component, multi-target, and multi-pathway approach. This polypharmacological nature is a unique strength of TCM in clinical practice, but it also underpins the complexity of interactions between TCM and the human body. This complexity hinders the elucidation of the chemical basis and underlying mechanisms of action (MoA) of TCM, thereby impeding its in-depth exploration and advancement [1, 2]. At the molecular level, TCM can be regarded as a composite of chemical ingredients. After entering the body, these compounds interact with various molecular targets to impact biological functions. The targets are the core effectors mediating the interactions between TCM and the human body. Therefore, the systematic identification of TCM targets is an indispensable step towards elucidating the TCM MoA.

Network pharmacology (NP), as an interdisciplinary of systems biology and pharmacology, intends to systematically interpret the occurrence and development of diseases as well as the MoA of drugs from the perspective of biological networks [3]. This paradigm naturally fits the systematic characteristics of TCM, giving rise to the methodology of TCM-NP [4, 5]. TCM-NP has been widely applied to explore the chemical basis and molecular mechanisms of TCM, resulting in the establishment of some specialized databases [6, 7], including SymMap [8], BATMAN-TCM [9], HIT [10], TCMID [11], TCMSP [12], ETCM [13], and YaTCM [14]. The main idea of TCM-NP lies in mapping the TCM targets and disease-related genes onto biological networks to build a multilayer TCM component—target—disease network, and systematically analyze the TCM MoA through network analysis. Given the numerous components in TCM, the efficient and rapid identification of the largely unknown biological targets of TCM components is the key problem of TCM-NP. Generally, TCM-NP adopts the efficient and high-throughput virtual prediction technology to capture potential compound-target interactions based on the structural information of TCM chemical components [15]. However, these

Received: August 23, 2024. Revised: January 17, 2025. Accepted: February 14, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

chemo-centric approaches usually tend to predict more binding interactions for existing proteins and ligands and fail to generalize to novel structures [16]. This poor generalizability should be more prominent in the TCM target prediction, as herbs with multiple components will predict more redundant targets. In addition, the aim of these methods is to predict structurally binding interactions between chemicals and targets, and neglect whether these targets are responsible for the biological effects of herbs, making researchers difficult to concentrate on pharmacologically related targets.

Alternatively, biological profile-based inference methods have the potential to address the limitations of chemo-centric approaches. These methods presume that drugs and their targets should induce similar biological activities, which can serve as the bridge to connect drugs and corresponding targets [17]. Among them, gene expression data that represent the overall molecular activities of perturbation, are the most suitable adaptors to connect activities and mechanisms of drugs and targets. A prototypical example of this type of study is the Connectivity Map (CMap) which is developed as a public database of >1.3 million gene expression profiles derived from chemical and genetic perturbations of various cell lines, providing opportunities to build functional connections between chemicals and genes at gene expression level [18, 19]. The CMap-based target prediction methods have been widely applied in various TCM fields, such as the mechanism study, synergistic effects within herbal formula, and drug repurposing [20]. Moreover, some pharmacotranscriptomic databases have been developed to facilitate the application of gene expression profile-based methods, including HERB, ITCM, and TMNP [21–23]. Biological profile-based methods have the advantage of exploring targets related to the biological effects of TCM. However, because these methods lack attention to the interactions between chemical components and targets, it is difficult to determine which targets directly bind with the TCM components to mediate the bioactivity of TCM.

Based on the above analysis, we find that the two strategies discover TCM therapeutic targets from different perspectives. Although they have made much significant contribution to the exploration of the molecular mechanisms of TCM, these methods still have some hierarchical and systematic deficiencies. It is necessary to strictly define the modes of action between drugs and biological targets. To elucidate the complexity of the interactions between drugs and biological targets, we divide the relevant targets for the effects of drugs into two categories: direct and indirect targets. Direct targets are those that directly bind with the drug to produce biological effects. Indirect targets are the subsequent targets influenced by the direct targets in the biological molecular network. Consequently, direct targets should be the intersection of binding targets (targets that structurally bind with the drug) and effect targets (targets that mediate the biological effects of the drug). Indirect targets are the remaining effect targets after excluding the direct targets. As shown in Fig. 1, after entering the body, the drug produces biological effects by binding to direct targets or indirectly impacting subsequent signaling pathways through indirect targets. Obviously, exploration of the TCM MoA involves the elucidation of the action pathway of component—direct target—indirect target.

Here, we propose a network pharmacology model NP-TCMtarget that integrates the chemical and biological data of TCM to explore the target path. We first build specific functional gene signatures for 8233 gene targets. NP-TCMtarget can infer TCM effect targets by calculating associations between

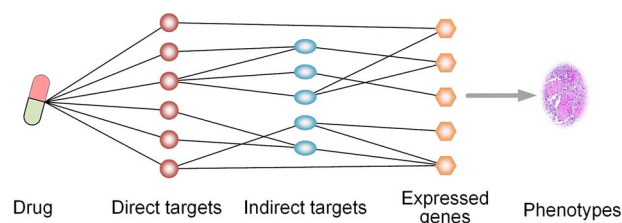


Figure 1. Drug target classification. Direct targets are those that bind to the drug and are responsible for the biological effects of the drug. Indirect targets are those that mediate the biological effects of the drug but do not bind to the drug. By interacting with its direct targets or sequentially modulating its indirect targets through cellular signal transduction, the drug plays its pharmacological action, such as changing the expression of some genes or changing some phenotypes.

herb/disease inducible gene expression profiles and specific target gene signatures. Then, we construct a binary classification model to predict compound-target interactions. NP-TCMtarget can utilize the model to predict the binding targets of herbal ingredients. After obtaining the potential effect targets and binding targets of TCM, we can distinguish direct and indirect targets by comparing the effect targets and binding targets. Finally, we can establish the action pathway of effective component—direct target—indirect target by mapping targets in the biological molecular network (Fig. 2). We apply NP-TCMtarget to the classic formula XiaoKeAn for treating type 2 diabetes to demonstrate the power of NP-TCMtarget to reveal the action pathways of TCM. In addition, we have embedded NP-TCMtarget in the online platform convenient for users. We believe NP-TCMtarget can provide a holistic framework for studying the molecular mechanisms of TCM.

## Materials and methods

### Implementation of NP-TCMtarget

NP-TCMtarget mines effect targets of herbal medicine by calculating associations between herb/disease inducible gene expression profiles and specific target gene signatures embedded in the drug effect target prediction (DETP) module, predicts binding targets of herbal ingredients by the binary classification model in the drug binding target prediction (DBTP) module, and maps drug targets onto biological pathways by the drug target mapping on signaling pathways (DTMP) module. NP-TCMtarget provides a convenient web interface for users to submit high-throughput transcriptional and chemical data to carry out target path analysis and download data. NP-TCMtarget is freely available at: <http://www.bcxnfz.top/NP-TCMtarget>. The website was built using the R shiny package and the analysis code was written in R. In addition, we also submitted the R code to GitHub (<https://github.com/lipi12q/NP-TCMtarget>).

### Gene expression data source and construction of target-specific gene signatures

Expanded CMap LINCS Resource 2020 (version: beta) is available from LINCS data releases app (<https://clue.io/releases/data-dashboard>). The L1000 platform carries out a rigorous five-step data processing pipeline to transform raw data from Luminex scanners to replicate-consensus signatures. The L1000 assay directly measures 978 landmark genes and infers additional 11 350 genes. Of the inferred genes, 9196 are well inferred. Our work only uses the high-fidelity 10 174 genes including 978 measured landmarks and 9196 well inferred genes. The perturbagens

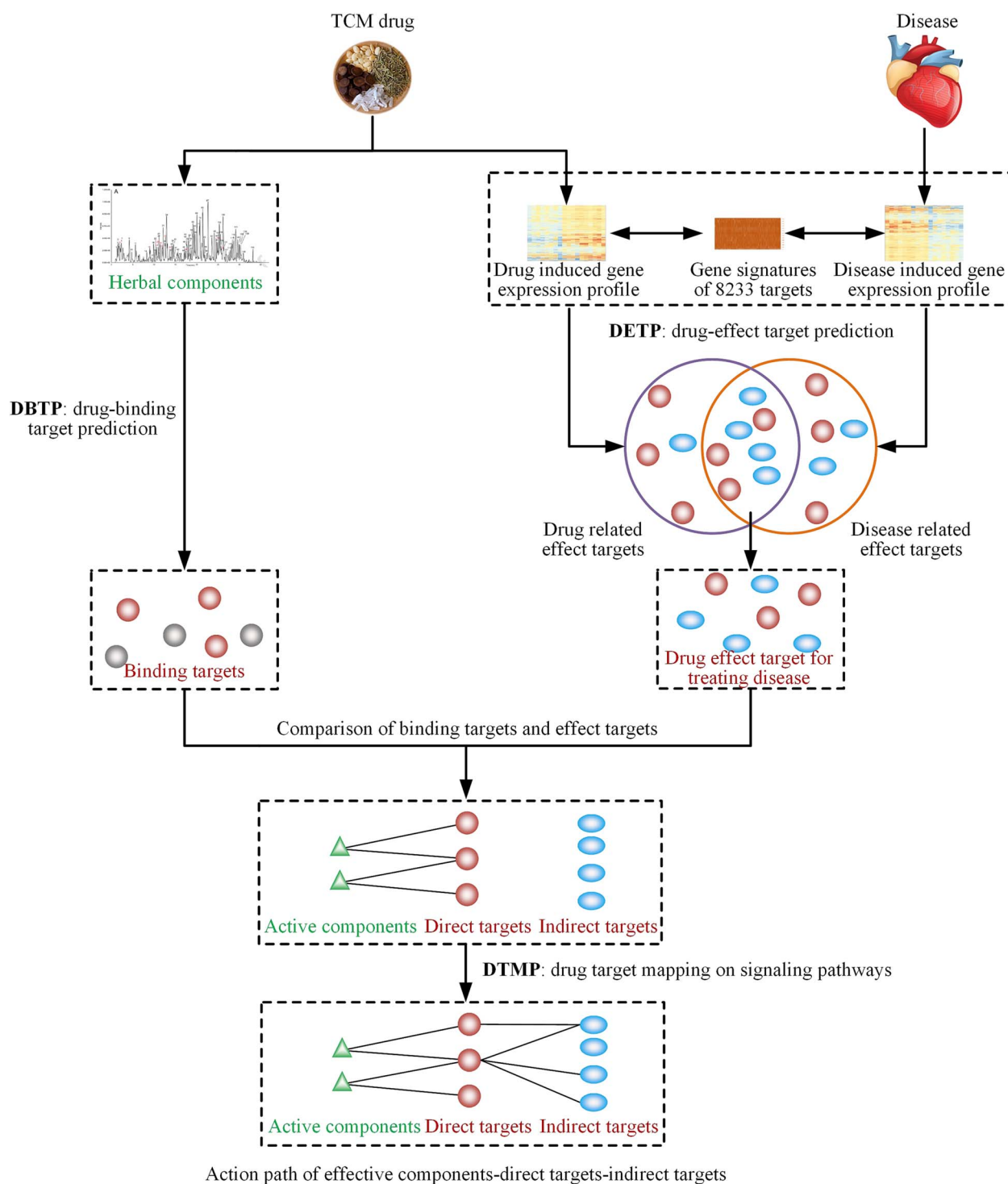


Figure 2. The framework of NP-TCMtarget.

include five types, compounds, shRNA treatments, CRISPR, over-expression (OE) treatments and other treatments. We use LEVEL 5 data for compounds, shRNA, CRISPR, and OE, which contains 921 123 signatures, corresponding to 12 487 compounds, 4345 shRNA, 5157 CRISPR and 4040 OE.

We collate gene targets for all chemical perturbagens from the metadata for compounds. Combining target information of compound and genetic perturbagens, all signatures are mapped to 8233 gene targets. For each target, to get a unique gene signature that accurately measures the on-target activity, we combine gene

expression signatures of each target to produce a consensus gene signature by the weighted average algorithm, which calculates a weighted average of the gene expression signatures of each target, with coefficients given by a pairwise Spearman correlation matrix between the expression profiles of all signatures [24].

For the expression signature of each target, we sort genes according to their values and select the top  $n$  ones ( $t_{up}$ ) of each gene list and the bottom  $n$  ones ( $t_{down}$ ) as the target-specific gene signature ( $t_{up}, t_{down}$ ) to represent the transcriptional activity induced by the target.

## Effect target prediction

Based on the drug or disease-induced gene expression data, we can predict effect targets for drugs or diseases by using the DETP module. The query is a gene list decreasingly ranked by gene metrics, e.g. differential gene expression profiles of the drug/disease. To evaluate the associations between the query gene list and targets, the connections between the query gene list  $L$  and each target-specific gene signature were calculated by using a modified gene set enrichment approach [25]. We defined the raw similarity score (also named Effect Target Score (ETS) in the present work) as follows:

$$ETS_L^t = \frac{ES_L^{up} - ES_L^{down}}{2}$$

Where  $ES_L^{up}$  is the enrichment score of  $t_{up}$  for  $L$  and  $ES_L^{down}$  is the enrichment score of  $t_{down}$  for  $L$ .  $ETS_L^t$  denotes the effect target score of the gene module pair  $(t_{up}, t_{down})$  of one target to  $L$ . EST values range between  $-1$  and  $1$ . It measures the degree of similarity between query  $L$  and target-induced gene expression profiles. It will be positive for targets that are positively related to  $L$  and negative for those that are inversely similar, and near zero for signatures that are unrelated. A zero value is assigned when both  $ES_L^{up}$  and  $ES_L^{down}$  are the same sign.

## Estimating significance of effect target score

We assess the significance of an actual ETS by comparing it with a set of random scores  $ETS_{NULL}$ . We generated the random space of query gene lists by randomly permuting genes in  $L$  for 1000 times, creating a null distribution of  $ETS_{NULL}$  for the signature. The nominal  $P$ -value for the actual ETS is estimated by using the positive or negative portion of the null distribution corresponding to the sign of the actual ETS, as follows:

$$P = \frac{N(ETS_{NULL} \geq ETS)}{N(ETS_{NULL} \geq 0)}, ETS \geq 0.$$

$$P = \frac{N(ETS_{NULL} < ETS)}{N(ETS_{NULL} < 0)}, ETS < 0$$

## Normalization of effect target score and adjustment for multiple hypothesis testing

When a query gene list is analyzed against multiple target signatures, we need to adjust the estimated significance level to account for multiple hypothesis testing. Firstly, we normalize the ETS for each target signature to account for the size of the signature, yielding a normalized effect target score (NETS):

$$NETS_L^i = \frac{ETS_L^i}{\text{mean}(ETS_{NULL}^i)}, \text{sign}(ETS_L^i) = \text{sign}(ETS_{NULL}^i)$$

where  $ETS_L^i$  is the actual effect target score of the signature  $i$ , and the  $ETS_{NULL}^i$  is the set of the effect target scores of the signature  $i$  against random permutations of the query gene list  $L$  with the same signs as  $ETS_L^i$ . By normalizing ETS, NP-TCMtarget accounts for different sizes in gene signatures and correlations between gene signatures and  $L$ . Therefore, NETS can be used to compare analysis results across gene signatures and multiple queries. We then control the proportion of false positives by calculating the false discovery rate (FDR) corresponding to each NETS. The FDR is the estimated probability that a signature with a given NETS

represents a false positive finding; it is computed by a ratio of two distributions: (i) the actual effect target score versus the scores for all gene signatures against all permutations of the query gene lists and (ii) the actual score versus the scores of all gene signatures against the actual gene lists:

$$A = \frac{N(NETS_{NULL} \geq NETS_L^i)}{N(NETS_{NULL} \geq 0)}, NETS_L^i > 0$$

$$A = \frac{N(NETS_{NULL} \leq NETS_L^i)}{N(NETS_{NULL} \leq 0)}, NETS_L^i < 0$$

$$B = \frac{N(NETS_{actual} \geq NETS_L^i)}{N(NETS_{actual} \geq 0)}, NETS_L^i > 0$$

$$B = \frac{N(NETS_{actual} \leq NETS_L^i)}{N(NETS_{actual} \leq 0)}, NETS_L^i < 0$$

$$FDR_L^i = \frac{A}{B}$$

Where  $NETS_L^i$  is the actual normalized correlation score of the signature  $i$  against  $L$ .  $NETS_{NULL}$  are the effect target scores for all target signatures against all permutations of the query gene lists.  $NETS_{actual}$  are the effect target scores of all target signatures against the actual gene lists.

## Determination of size of target signature

The size  $n$  of  $t_{up}$  and  $t_{down}$  is heuristically determined as follows. We vary the size  $n$  from 100 to 1000 with an interval of 50 and calculated ETS between all target gene expression profiles and gene signatures. It is observed that the overall ETS values (measured by the average of all ETS absolute values at each size) gradually decrease with the increase of the size of signatures. Moreover, we observe that the extent of decrease of ETS values is gradually slowing with the size increases, suggesting that with the size of gene module pair increase to some extent, its influence on the overall ETS values can be minimal. We quantify the decrease extent of ETS by the decrease rate (DR):

$$DR = \frac{\text{mean}(|ETS_{i+50}| - |ETS_i|)}{\text{mean}(|ETS_i|)}$$

Finally, we choose  $n=350$  as a good compromise, as the decrease rate of ETS at this size is small enough ( $DR \sim 0.01$ , [Supplementary Fig. 1](#)).

## Compound-target interaction dataset

Compound-target interaction datasets are collected from DrugBank, which is the most widely used drug information resource. We download the latest dataset from the DrugBank website (<https://go.drugbank.com/releases/latest>). Finally, 19 891 interactions with 3894 drugs and their 6898 protein targets are retained. All these data are used to build the compound-binding target prediction model. These data can be available at <https://github.com/lipi12q/NP-TCMtarget>.

## Drug binding target prediction by DeepPurpose

DeepPurpose is a deep learning library for encoding and downstream prediction of compound-protein interactions, in which models are usually considered as state-of-the-art methods for the prediction of drug-target interactions [26]. In the present work, the model is implemented in the DBTP module. DBTP takes the



simplified molecular-input line entry system (SMILES) string of drugs and amino acid sequences of proteins as input, which are both encoded with the convolutional neural network. Then the drug and protein embeddings are fed into a multilayer perceptron decoder to generate binary outputs indicating the probability that a compound binds to a protein (also named drug binding score). During training, the learning rate was set to 0.0008, the batch size was 256, and the overall training was conducted for 100 epochs. DBTP uses Binary Cross Entropy as the loss function and area under the receiver operating characteristics and area under precision-recall as performance metrics.

## Functional enrichment analysis

DTMP enriches the effect targets into their corresponding functional modules and distinguishes between direct and indirect targets within the modules. The default option includes widely used Kyoto Encyclopedia of Genes and Genomes (KEGG), gene ontology (GO) and disease analyses in the DTMP section. NP-TCMtarget employs the R package WebGestaltR (version: 0.4.6) to conduct these functional enrichment analyses. The basic principle is the over representation analysis (ORA), which is based on the hypergeometric distribution or binomial distribution to calculate the probability of observing a certain number of genes in common between the gene set of interest and a predefined gene set. The *P*-value was calculated by the hypergeometric test for ORA. FDR was calculated to correct the *P*-value for multiple testing.

## Results

### Scheme of drug target path prediction

NP-TCMtarget includes three main modules: drug effect target prediction based on the drug-induced gene expression data (DETP), drug binding target prediction using the deep learning model (DBTP), and drug target mapping on signaling pathways (DTMP). DETP has built a library of target-specific signatures derived from four types of gene expression data induced by four types of perturbations: compound, shRNA, CRISPR and OE. Then using gene expression data induced by drugs or diseases as a query, DETP can evaluate the effects of drugs on targets by measuring the correlation between the drug/disease inducible gene expression profiles and target signatures using the association algorithm (See Materials and methods). DBTP has constructed a deep learning drug-target interaction prediction model, which can be utilized to predict the binding probability between chemical ingredients of herbs and targets (See Materials and methods). DTMP enriches drug targets into biological pathways and distinguishes direct and indirect targets within the pathways, allowing for the exploration of MoA of herbal medicine.

### Application of NP-TCMtarget

The core objective of NP-TCMtarget is to explore the molecular mechanisms of herbal medicine for treating diseases based on high-throughput transcriptional and chemical data of herbal medicine. The primary input of NP-TCMtarget has two parts: herbal medicine or disease inducible gene expression profiles and chemical ingredients of the corresponding herbs (Figs 2 and 3). In the calculation, the query gene expression data should be ranked lists in descending order of some differential metrics (e.g. the Signal2Noise or fold change values between drug-treated versus control samples). DETP estimates the impact of the drug or disease on each target by the resultant normalized effect target score (NETS) and the corresponding FDR, which are calculated based on the correlation between the gene expression profiles

and the catalog of target signatures. Generally, those targets with  $FDR \leq 0.05$  are regarded as potential effect targets of the query drug or disease. More importantly, the sign (positive or negative) of NETS values reflects different action modes on the target. If the drug and disease have opposite NETS values for the same targets, the drug has the potential to treat the disease by these targets. Therefore, by comparing the effect targets of the drug and the disease, we can obtain the therapeutic targets that may play beneficial effects in the drug treatment for the disease.

In the NP-TCMtarget platform, the structural information of herbal ingredients is represented by canonical SMILES and inputted into the DBTP module (Fig. 2 and 3). DBTP uses the embedded compound-target interaction prediction model to predict the binding targets of chemical ingredients. The predicted targets with binding scores above a threshold value are regarded as potential binding targets of herbal ingredients. We can compare effect targets to binding targets of herbs and obtain direct targets of herbs which are the intersection of effect targets and binding targets. The remaining effect targets are seen as indirect targets of herbs.

DTMP enriches the effect targets into their corresponding biological signaling pathways and distinguishes between direct and indirect targets within the pathways. Various biological networks such as protein-protein interaction networks, gene co-expression networks, gene function networks, and molecular signaling pathway networks can be utilized to map all targets to their respective networks, allowing for the exploration of signaling pathways of herbal targets and laying the foundation for the study of the molecular mechanism. As various known biological networks are applied in different fields without a clear distinction of superiority or inferiority, the choice of the appropriate molecular network can be flexible according to one's own research background and research objectives. The default option includes widely used KEGG, GO and disease analyses in the DTMP module. In the NP-TCMtarget website, the above procedures are performed automatically on the server and all results can be easily reviewed and downloaded by users on the 'Results' page (Fig. 3).

## Use case

### Exploring molecular mechanisms of XiaoKeAn against type 2 diabetes

Type 2 diabetes is a common chronic disease with persistently elevated blood glucose concentration and the pathophysiology is characterized by insulin resistance and initial hyperinsulinemia [27]. XiaoKeAn is a TCM preparation specifically designed to manage type 2 diabetes. Classified as a national Class III new drug in China, XiaoKeAn demonstrates significant efficacy in reducing blood sugar levels and alleviating associated symptoms. However, despite its widespread use, there remains a dearth of research exploring the molecular mechanisms underlying its therapeutic effects in diabetes treatment. This section aims to employ NP-TCMtarget, leveraging transcriptomic data and chemical information of XiaoKeAn, to elucidate the mechanism of XiaoKeAn in treating type 2 diabetes at the molecular level.

Following the instruction of NP-TCMtarget, we first collect the transcriptional data from the microarray experiment deposited in GEO (GSE accession: GSE62087) (Supplementary Table 1). The experiment includes three groups: the normal control group (C57BL/6 J mice), the disease model group (KKAY mice) and the corresponding XiaoKeAn treatment group. The differential gene profiles induced by the diabetes model and XiaoKeAn treatment are calculated and submitted to the DETP module

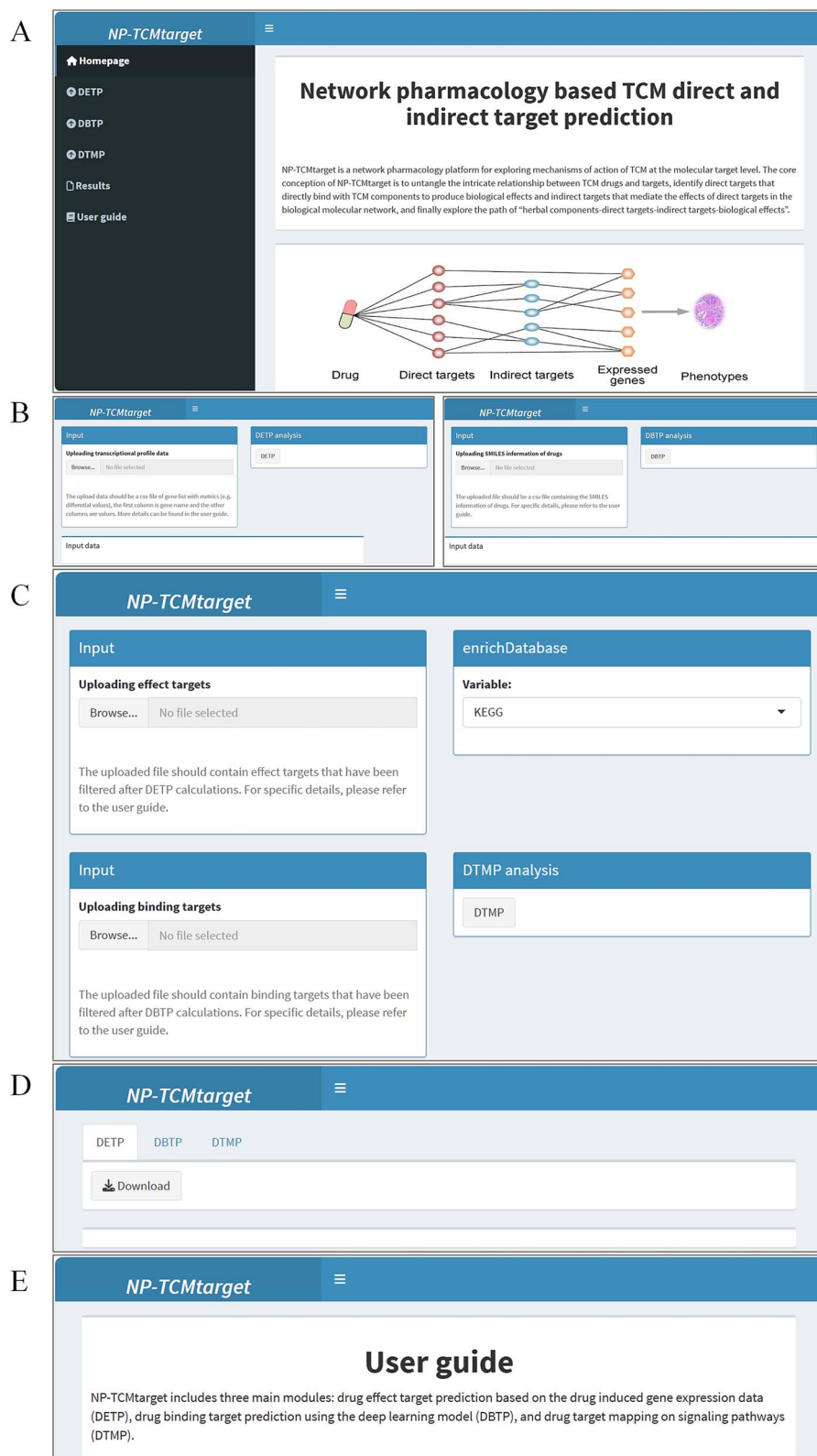


Figure 3. An illustration of the NP-TCMtarget website. (A) the 'homepage' of NP-TCMtarget shows the overview. (B) The DETP module for uploading the query drug/disease gene expression data and the DBTP module for uploading the chemical information represented by SMILES. (C) The DTMP module enriches the effect targets into their corresponding biological signaling pathways. (D) the result data corresponds to the analysis in the DETP, DBTP and DTMP modules. (E) the 'user guide' page details the procedure and the principle of NP-TCMtarget.

to predict effect targets. We find that the disease model and XiaoKeAn treatment are significantly associated with 1612 and 2157 effect targets among all 8233 targets, respectively (Fig. 4A; Supplementary Table 2;  $FDR \leq 0.05$ ). We observe a negative correlation between the disease model and XiaoKeAn treatment that is quantified by the Pearson correlation of NETS values for all targets ( $PCC = -0.50$ ; Fig. 4B), indicating that XiaoKeAn can reverse the impact of diabetes. The absolute value of NETS reflects the magnitude of the herb/disease's effects on a target, with a higher absolute value indicating a stronger impact. If the NETS values for the herb medicine and disease on a specific target are opposite, it suggests opposing effects of the herb medicine and disease on that target. By comparing the NETS values of the effect targets that are significantly associated with the diabetes model and XiaoKeAn treatment, we identify 734 targets that were significantly affected by the diabetes model but could be reversed by the XiaoKeAn treatment (indicated by opposing signs in NETS values;  $FDR \leq 0.05$ ; Supplementary Table 2). More importantly, we observe that there is a strong negative correlation between the disease model and XiaoKeAn treatment if we only focus on these 734 targets ( $PCC = -0.98$ ; Fig. 4C). These results suggest that the 734 targets can be potential therapeutic targets for XiaoKeAn in the treatment of diabetes mellitus.

In addition, a total of 820 chemical components are gathered from the SymMap database for eight herbs in the XiaoKeAn formula (Supplementary Table 3). Utilizing the SMILE information of all these components, the DBTP module is employed to predict 5030 binding targets of XiaoKeAn ingredients (Supplementary Table 4; Binding score  $\geq 0.90$ ). By comparing these 5030 binding targets of XiaoKeAn and its 734 effect targets for treating diabetes, we obtain 421 direct targets and 313 indirect targets of XiaoKeAn for treating diabetes (Fig. 4A; Supplementary Fig. 2 and Supplementary Table 5). These targets can further be mapped to biological pathways to elucidate the molecular mechanisms of XiaoKeAn in diabetes treatment.

Through enrichment analysis in the DTMP module, the 734 potential therapeutic targets are significantly enriched in 46 functional pathways ( $FDR \leq 0.05$ ; Fig. 4D and Supplementary Table 6). Detailed examination of these pathways reveals three primary targeted mechanisms of XiaoKeAn in diabetes treatment: (i) regulation of glucose metabolism, such as 'Carbohydrate digestion and absorption', 'insulin signaling pathway' and 'insulin secretion', (ii) regulation of endocrine-related functions, such as 'Aldosterone-regulated sodium reabsorption', 'Thyroid hormone signaling pathway', 'Neurotrophin signaling pathway', 'Proximal tubule bicarbonate reclamation', 'Bile secretion', 'Mineral absorption', 'Thyroid hormone synthesis' and 'Endocrine and other factor-regulated calcium reabsorption', and (iii) regulation of inflammatory immune responses, such as 'AGE-RAGE signaling in diabetic complications', 'Human T-cell leukemia virus 1 infection', 'Influenza A', 'C-type lectin receptor signaling pathway', 'TNF signaling pathway', 'B cell receptor signaling pathway' and 'JAK-STAT signaling'. These findings indicate that XiaoKeAn regulates multiple functions to treat Type 2 diabetes, aligning with previous reports on its effectiveness against this condition [28, 29]. In addition, we find that functional enrichment analysis of GO has similar results (Supplementary Table 6).

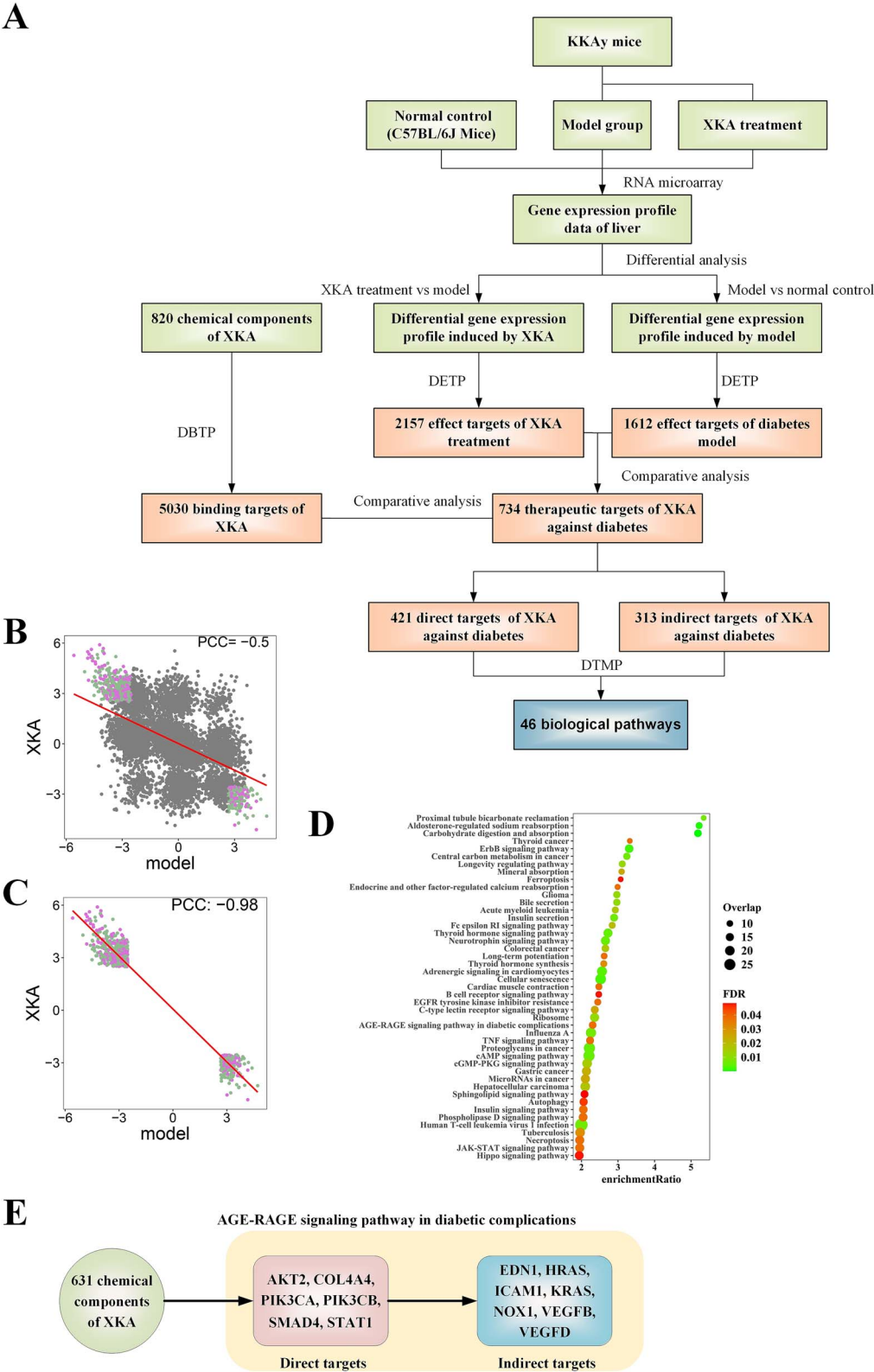
DTMP also indicates the direct and indirect targets in each pathway, in which we can inspect the detailed target route of XiaoKeAn. For example, the DTMP analysis indicates that 'AGE-RAGE signaling pathway in diabetic complications' is one of all 46 pathways significantly related to type 2 diabetes ( $FDR = 0.04$ ). Actually, it has been well known that the AGE-RAGE signaling

pathway plays a crucial role in the development of diabetic complications. AGEs (advanced glycation end products) are accumulated under hyperglycemic conditions and activate the receptor of AGEs (RAGE). This activation triggers a cascade of adverse responses, including inflammation, oxidative stress, endothelial cell dysfunction with apoptosis, and eventually diabetic complications, including retinopathy, neuropathy, cardiomyopathy, microvascular complications, and nephropathy [30, 31]. Therefore, drug development and therapeutic strategies targeting the AGE-RAGE signaling pathway may offer new directions for preventing and treating diabetic complications. The DTMP analysis displays that XiaoKeAn can regulate this pathway by 6 direct targets including AKT2, COL4A4, PIK3CA, PIK3CB, SMAD4 and STAT1, and 7 indirect targets including EDN1, HRAS, ICAM1, KRAS, NOX1, VEGFB and VEGFD (Fig. 4E and Supplementary Table 6). Moreover, we can find 631 chemical components that hit the 6 direct targets from the results of the DBTP analysis. Based on these results, we can explore the molecular mechanisms of XiaoKeAn for regulating the AGE-RAGE signaling pathway from the route of chemical component—direct target—indirect target (Fig. 4E). In addition, we can further focus key targets and their corresponding components from the overall herbal component—direct target—indirect target network based on effect target scores or network topological properties. For example, in the herbal component—direct target interaction network, we extract top 38 direct targets with NETS absolute values  $\geq 4$  and top 100 herbal compounds interact with these targets by the network degree (Supplementary Fig. 3). Interestingly, we find that these targets can be significantly related to pathways similar to all therapeutic targets, such as 'Carbohydrate digestion and absorption', 'Proximal tubule bicarbonate reclamation' and 'Aldosterone-regulated sodium reabsorption'. Especially, the pathway 'Carbohydrate digestion and absorption' is also most significantly enriched these 38 direct targets (Supplementary Table 7). Therefore, we further explore the herbal component—direct target—indirect target path in this pathway (Fig. 5). We find that 13 therapeutic targets inducing 10 direct targets and 3 indirect targets can be mapped onto this pathway. Among them, five direct targets including 'HK3', 'ATP1B3', 'ATP1A3', 'ATP1A4' and 'ATP1A2', have NETS absolute values  $\geq 4$ . These targets play a critical role in glucose metabolism and energy metabolism. For example, HK3 is an enzyme that phosphorylates glucose to produce glucose-6-phosphate, marking the first step in most glucose metabolic pathways. ATP1B3, ATP1A3, ATP1A4 and ATP1A2 belong to the subunits of Na<sup>+</sup>/K<sup>+</sup>-ATPases which is responsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane. These gradients are essential for osmoregulation and sodium-coupled transport of a variety of biological molecules. These targets can be targeted by many active herbal components such as chlorogenic acid (com434), quercetin (com436), and forsythoside A (com666).

This section displays that NP-TCMtarget can effectively reveal the chemical basis and relevant mechanisms of XiaoKeAn in the treatment of type 2 diabetes from the target level, and provide an example for the exploration of the MoA of herbal formula for treating complex diseases.

## Comparative analysis of Alzheimer's disease and Huntington's disease

This section displays the application of NP-TCMtarget on multiple queries of gene-expression profiles of different diseases. As an example, we compare the effect targets of Alzheimer's disease





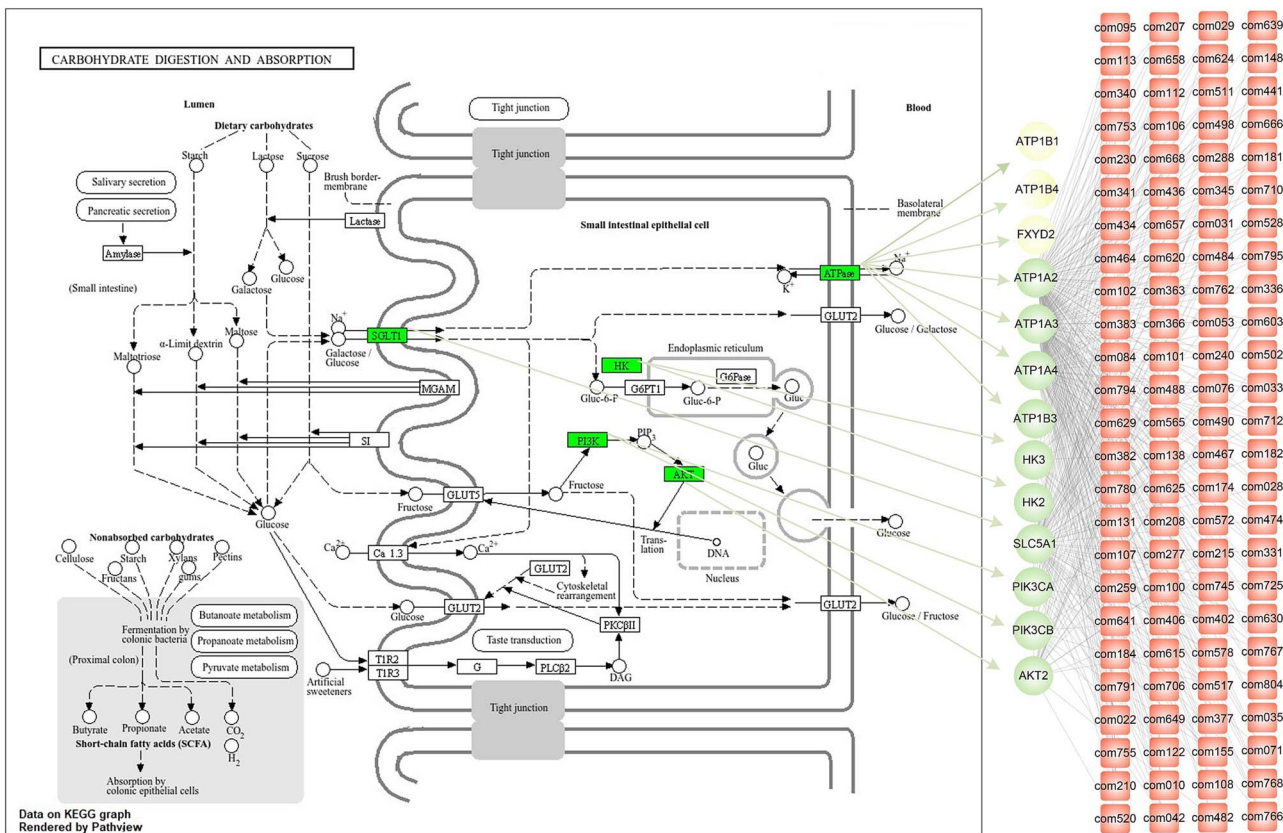


Figure 5. The herbal component—Direct target—Indirect target path in the ‘Carbohydrate digestion and absorption’. Red squares represent the top 100 herbal compounds ranked by the network degree. Round green nodes represent the direct targets and round yellow nodes represent the indirect targets.

(AD) and Huntington's disease (HD) in NP-TCMtarget based on the transcriptional data induced by the two neurological diseases. AD and HD are two most common neurodegenerative disorders, causing neurological illness among older adults [32]. Their transcriptional data are obtained from the microarray experiment deposited in GEO (GSE accession: GSE33000) (Supplementary Table 8) and are submitted to the NP-TCMtarget website.

In the DETP module, we find that the AD and HD are linked to 5909 and 6529 target genes, respectively ( $\text{FDR} \leq 0.05$ ; Fig. 6A; Supplementary Table 9). Interestingly, on the whole, the two neurological diseases have similar regulations for all target genes. This can be observed more clearly in the heatmap of NETS values (Fig. 6B). To quantify the similarity between the two diseases, we calculate the Pearson correlation of all NETS values for all targets and find that the two diseases have correlations of 0.86 (Fig. 6C). If we only focus on the targets that are significantly influenced by AD and HD, there is an extremely high correlation between them ( $\text{PCC} = 0.94$ ; Fig. 6D). We compare their detailed targets and find that the two diseases have 5187 overlapping targets with same signs, accounting for 87.8% and 79.4% of all significant targets for AD and HD, respectively (Fig. 6E). These results indicate that AD and HD can induce many similar effects in the brain.

To further detect the mechanistic similarity of AD and HD, we do the enrichment analysis for the 5187 overlapping targets in the DTMP module, and find these targets are significantly enriched in 167 functional pathways ( $FDR \leq 0.05$ , [Fig. 6F](#) and [Supplementary Table 10](#)). Detailed examination of these pathways reveals these overlapping targets are mainly related to the

regulation of inflammatory immune responses, such as 'Human cytomegalovirus infection', 'TNF signaling pathway', 'Influenza A', 'Chemokine signaling pathway', and 'Human immunodeficiency virus 1 infection'.

This section uses two queries of gene-expression profiles from AD and HD to explore their similar mechanisms. We find the two neurodegenerative diseases have many same effect targets. And further analysis of these indicates that the regulation of inflammatory immune function is pivotal for the two diseases, and provides plenty of clues for further mechanistic and therapeutic studies.

## Discussion

Elucidating the intricate interactions between herbs and the human body remains a pivotal issue in the TCM field. Herbal targets, acting as crucial mediators in these interactions, are essential for elucidating the underlying mechanisms of these interactions. We here systematically delineate the sequential pipeline of 'herbal components-direct target-indirect target-biological effects' through which TCM drugs interact with the human body. Based on this framework, we introduce the NP-TCMtarget strategy to facilitate the identification of TCM targets and further enable the traversal of the path of 'herbal component-direct target-indirect target-biological effect' by integrating the chemical and biological information of TCM cohesively.

The core conception of NP-TCMtarget is to untangle the intricate relationship between drugs and biological targets and identify direct targets that directly bind with the drug to produce

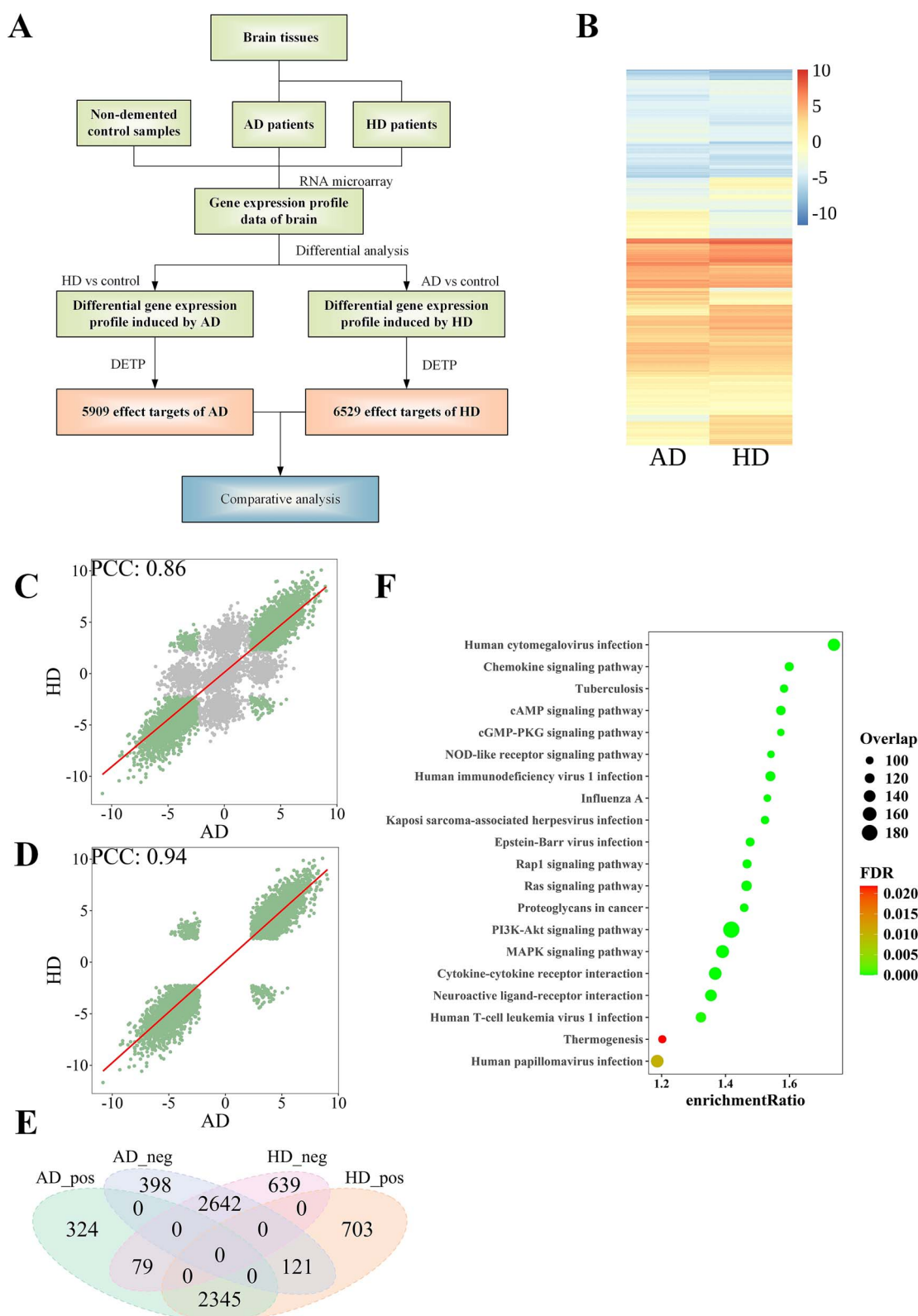


Figure 6. Comparative analysis of Alzheimer's disease and Huntington's disease. (A) the analysis process of AD and HD by NP-TCMtarget. (B) the heatmap of NETS values for AD and HD. (C) Pearson correlation of NETS values between AD and HD. (D) Pearson correlation of NETS values between AD and HD based on 5187 effect targets with  $FDR \leq 0.05$ . (E) the Venn diagram for the significantly associated gene signatures by AD and HD. AD\_neg indicates those signatures negatively related to the Alzheimer's disease. AD\_pos indicates those signatures positively related to the Alzheimer's disease. HD\_neg indicates those signatures negatively related to Huntington's disease. HD\_pos indicates those signatures positively related to Huntington's disease. (F) TOP 20 pathways enriched by overlapping targets between AD and HD (all enriched pathways can be found in [supplementary table 10](#)).

biological effects and indirect targets that are subsequently influenced by the direct targets in the biological molecular network (Fig. 1). To achieve this goal, NP-TCMtarget firstly predicts effect targets that mediate the biological effects of the drug and should be the union of direct and indirect targets based on the transcriptional profiles of drugs. Then NP-TCMtarget predicts binding targets that structurally bind with the drug. By comparing effect targets and binding targets, we can get direct targets and indirect targets of the drug, and further elucidate the action path of effective component—direct target—indirect target.

The idea of predicting drug effect targets in the DETP module of NP-TCMtarget is derived from the CMap-based methods, which consider two perturbagens with similar gene expression profiles that have similar molecular targets. However, although CMap dataset has curated numerous chemical and genetic perturbations with gene expression profiles (<https://clue.io/>) [19], the knowledge of the landmarked perturbations is still a severe constraint for the application of the CMap-based methods. NP-TCMtarget solves this problem by manually constructing gene signatures of each target from prior knowledge. The targets of a query drug can be inferred by directly comparing its inducible gene expression profiles with the target-specific gene signatures. It should be noted that although we propose that if the drug and disease have opposite NETS values for the same targets, the drug has the potential to treat the disease by these targets. The precise mechanisms by the drug treats the disease remains uncertain. For example, it is difficult to judge whether the drug treats the disease by alleviating symptoms or directly meliorates the cause of the disease from the predicted effect targets. Certainly, an initial assessment can be made by comparing the predicted targets with known symptom- or disease-causing genes. For more definitive evidence, however, would require further genetic manipulation studies.

The prediction of drug binding targets embedded in the DBTP module of NP-TCMtarget is based on the chemogenomic information to formalize binding interactions between chemicals and targets. Chemogenomics-based methods utilize both machine learning and artificial intelligence algorithms to learn molecular representations and capture the complex nonlinear relationships between chemicals and target proteins [33]. NP-TCMtarget uses the DeepPurpose framework which contains known state-of-the-art deep learning models that have been developed and leveraged for the prediction of drug-target interactions [26]. Based on the DeepPurpose, NP-TCMtarget can effectively predict binding targets for herbal ingredients. More importantly, by comparing effect targets and binding targets, we can get direct targets of herbal ingredients. The direct target space is responsible for the biological function of the drug and is greatly compressed compared with the predicted binding target set which is usually redundant. Moreover, the herbal compounds fished by the direct targets with high binding affinity can be regarded as potential effective compounds. Users can choose interested compounds based on the NETS and binding scores, and then explore specific effective compounds through in-depth experiments. However, it should be noted that components can be collected by different ways. For example, we can collect them from the existing literatures and databases, or identify them through mass spectrometry analysis. Different methods will lead to the collection of different components. Even with the same method, many factors such as herbal source, modes of administration, herbal prescriptions, as well as the differences in compound dosages as you mentioned, can also result in discrepancies in chemical components, thereby

influencing subsequent target prediction. Therefore, the chemical space of the herbal medicine could also be further improved, especially given that the main objective is to dig the underlying active compounds. In this respect, high-throughput component identification techniques such as liquid chromatography-mass spectrometry and metabolomics, which meets this challenge and can be coupled with the transcriptome analysis to characterizing the active compounds and evaluating their efficacy.

In summary, NP-TCMtarget combines TCM transcriptomic and chemical data to explore the underlying TCM MoA for treating diseases at the molecular target level. The ability of NP-TCMtarget can be impacted by some limitations. The result of predicting effect targets is affected by the quantity and quality of the target-specific signatures. We have created molecular signatures for 8233 gene targets by their inducible gene expression data in the CMap dataset. However, the gene expression data can be affected by many factors, including cell types, drug doses, genetic types, and time of duration. This data heterogeneity significantly impacts the quantity of resulting gene signatures for each target. In addition, the scale of 8233 targets is still inadequate relative to the overall genome and can be improved. To facilitate users to flexibly apply the algorithm for the customized transcriptomic data and gene signatures, except for the online NP-TCMtarget website, we have submitted the R code to GitHub (<https://github.com/lipi12q/NP-TCMtarget>).

#### Key Points

- NP-TCMtarget is developed to explore the TCM molecular pathways of herbal component—direct target—indirect target for treating diseases.
- NP-TCMtarget infers effect targets by calculating associations between drug/disease inducible gene expression profiles and specific gene signatures for 8233 targets.
- NP-TCMtarget predicts binding targets of herbal ingredients by a deep learning binary classification model.

## Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflicts of interest: The authors declare no conflicts of interest.

## Funding

This research was supported by the National Natural Science Fund of China (No. 82274363, 82025036, 22338004), the Fundamental Research Program of Shanxi Province (No. 20210302124129), the Distinguished and Excellent Young Scholars Cultivation Project of Shanxi Agricultural University (No. 2022YQPYGC09), and the Postgraduate Research Innovation Project of Shanxi Province (No. 2023KY322).

## Data availability

The data used in the present work are available at <https://github.com/lipi12q/NP-TCMtarget>. All data used in this work are from public resources. The CMap source can be found at <https://clue.io/releases/data-dashboard>. The Drugbank source can be found at <https://go.drugbank.com/releases/latest>.



## Code availability

The codes of this work are freely available at GitHub (<https://github.com/lipi12q/NP-TCMtarget>).

## References

- Qiu J. Traditional medicine: A culture in the balance. *Nature* 2007;**448**:126–8. <https://doi.org/10.1038/448126a>
- Xu Z. Modernization: One step at a time. *Nature* 2011;**480**:S90–2. <https://doi.org/10.1038/480S90a>
- Hopkins AL. Network pharmacology: The next paradigm in drug discovery. *Nat Chem Biol* 2008;**4**:682–90. <https://doi.org/10.1038/nchembio.118>
- Li S, Zhang B. Traditional Chinese medicine network pharmacology: Theory, methodology and application. *Chin J Nat Med* 2013;**11**:110–20. [https://doi.org/10.1016/S1875-5364\(13\)60037-0](https://doi.org/10.1016/S1875-5364(13)60037-0)
- Luo TT, Lu Y, Yan SK. et al. Network pharmacology in research of Chinese medicine formula: Methodology, application and prospective. *Chin J Integr Med* 2020;**26**:72–80. <https://doi.org/10.1007/s11655-019-3064-0>
- Zhang R, Zhu X, Bai H. et al. Network pharmacology databases for traditional Chinese medicine: Review and assessment. *Front Pharmacol* 2019;**10**:123. <https://doi.org/10.3389/fphar.2019.00123>
- Li P, Fu Y, Wang Y. Network based approach to drug discovery: A mini review. *Mini Rev Med Chem* 2015;**15**:687–95. <https://doi.org/10.2174/1389557515666150219143933>
- Wu Y, Zhang F, Yang K. et al. SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Res* 2019;**47**:D1110–d1117. <https://doi.org/10.1093/nar/gky1021>
- Liu Z, Guo F, Wang Y. et al. BATMAN-TCM: A bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Sci Rep* 2016;**6**:21146. <https://doi.org/10.1038/srep21146>
- Ye H, Ye L, Kang H. et al. HIT: Linking herbal active ingredients to targets. *Nucleic Acids Res* 2011;**39**:D1055–9. <https://doi.org/10.1093/nar/gkq1165>
- Xue R, Fang Z, Zhang M. et al. TCMID: Traditional Chinese medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res* 2013;**41**:D1089–95. <https://doi.org/10.1093/nar/gks1100>
- Ru J, Li P, Wang J. et al. TCMSP: A database of systems pharmacology for drug discovery from herbal medicines. *J Chem* 2014;**6**:13. <https://doi.org/10.1186/1758-2946-6-13>
- Xu HY, Zhang YQ, Liu ZM. et al. ETCM: An encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res* 2019;**47**:D976–d982. <https://doi.org/10.1093/nar/gky987>
- Li B, Ma C, Zhao X. et al. YaTCM: Yet another traditional Chinese medicine database for drug discovery. *Comput Struct Biotechnol J* 2018;**16**:600–10. <https://doi.org/10.1016/j.csbj.2018.11.002>
- Zhang P, Zhang D, Zhou W. et al. Network pharmacology: Towards the artificial intelligence-based precision traditional Chinese medicine. *Brief Bioinform* 2023;**25**:bbad518. <https://doi.org/10.1093/bib/bbad518>
- Chatterjee A, Walters R, Shafi Z. et al. Improving the generalizability of protein-ligand binding predictions with AI-bind. *Nat Commun* 2023;**14**:1989. <https://doi.org/10.1038/s41467-023-37572-z>
- Schenone M, Dančik V, Wagner BK. et al. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 2013;**9**:232–40. <https://doi.org/10.1038/nchembio.1199>
- Lamb J, Crawford ED, Peck D. et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929–35. <https://doi.org/10.1126/science.1132939>
- Subramanian A, Narayan R, Corsello SM. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e1417. <https://doi.org/10.1016/j.cell.2017.10.049>
- Jiang H, Hu C, Chen M. The advantages of connectivity map applied in traditional Chinese medicine. *Front Pharmacol* 2021;**12**:474267. <https://doi.org/10.3389/fphar.2021.474267>
- Fang S, Dong L, Liu L. et al. HERB: A high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res* 2021;**49**:D1197–d1206. <https://doi.org/10.1093/nar/gkaa1063>
- Tian S, Zhang J, Yuan S. et al. Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. *Brief Bioinform* 2023;**24**:bbad027. <https://doi.org/10.1093/bib/bbad027>
- Li P, Zhang H, Zhang W. et al. TMNP: A transcriptome-based multi-scale network pharmacology platform for herbal medicine. *Brief Bioinform* 2022;**23**:bbab542. <https://doi.org/10.1093/bib/bbab542>
- Smith I, Greenside PG, Natoli T. et al. Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the connectivity map. *PLoS Biol* 2017;**15**:e2003213. <https://doi.org/10.1371/journal.pbio.2003213>
- Subramanian A, Tamayo P, Mootha VK. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50. <https://doi.org/10.1073/pnas.0506580102>
- Huang K, Fu T, Glass LM. et al. DeepPurpose: A deep learning library for drug-target interaction prediction. *Bioinformatics* 2021;**36**:5545–7. <https://doi.org/10.1093/bioinformatics/btaa1005>
- Ahmad E, Lim S, Lamprey R. et al. Type 2 diabetes. *Lancet* 2022;**400**:1803–20. [https://doi.org/10.1016/S0140-6736\(22\)01655-5](https://doi.org/10.1016/S0140-6736(22)01655-5)
- Yang ZZ, Liu W, Zhang F. et al. Deciphering the therapeutic mechanisms of Xiao-Ke-an in treatment of type 2 diabetes in mice by a Fangjiomics approach. *Acta Pharmacol Sin* 2015;**36**:699–707. <https://doi.org/10.1038/aps.2014.138>
- Yang Z, Wang L, Zhang F. et al. Evaluating the antidiabetic effects of Chinese herbal medicine: Xiao-Ke-an in 3T3-L1 cells and KKAY mice using both conventional and holistic omics approaches. *BMC Complement Altern Med* 2015;**15**:272. <https://doi.org/10.1186/s12906-015-0785-2>
- Rungratanawanich W, Qu Y, Wang X. et al. Advanced glycation end products (AGEs) and other adducts in aging-related diseases and alcohol-mediated tissue injury. *Exp Mol Med* 2021;**53**:168–88. <https://doi.org/10.1038/s12276-021-00561-7>
- Egaña-Gorroño L, López-Díez R, Yepuri G. et al. Receptor for advanced glycation end products (RAGE) and mechanisms and therapeutic opportunities in diabetes and cardiovascular disease: Insights from human subjects and animal models. *Front Cardiovasc Med* 2020;**7**:37. <https://doi.org/10.3389/fcvm.2020.00037>
- Panegyres PK, Chen HY. Alzheimer's disease, Huntington's disease and cancer. *J Clin Neurosci* 2021;**93**:103–5. <https://doi.org/10.1016/j.jocn.2021.09.012>
- Chen H, Engkvist O, Wang Y. et al. The rise of deep learning in drug discovery. *Drug Discov Today* 2018;**23**:1241–50. <https://doi.org/10.1016/j.drudis.2018.01.039>