

AI-enabled target discovery in traditional Chinese medicine: from computational prediction to experimental validation

Chengyang Guo^{1#}, Qingmeng Li^{2#}, Weishan Liang³, Jinyuan Lu², Wenjing Yue¹, Saisai Tian^{1*} and Weidong Zhang^{1,2,4*}

¹ Department of Phytochemistry, School of Pharmacy, Second Military Medical University, Shanghai 200433, China

² The Research Center for Traditional Chinese Medicine, Shanghai Institute of Infectious Diseases and Biosafety, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

³ Pharmacy College, Guangdong Pharmaceutical University, Guangdong 350122, China

⁴ State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China

[#] Authors contributed equally: Chengyang Guo, Qingmeng Li

* Correspondence: saisai_tian@foxmail.com (Tian S); wdzhangy@hotmail.com (Zhang W)

Abstract

Traditional Chinese Medicine (TCM), with a history of thousands of years, has been increasingly supported by modern clinical studies for its effectiveness in disease prevention, treatment, and rehabilitation. However, the complex, multi-component, and multi-target characteristics of TCM make it challenging to identify bioactive compounds and understand the molecular mechanisms underlying their effects. These features also complicate the standardization of clinical applications. However, investigation of TCM targets remains limited. Systems biology and network pharmacology have become valuable approaches for analyzing the complex multi-component and multi-target interactions of TCM. Omics-based approaches, including transcriptomics and single-cell sequencing, have further improved the ability to characterize gene expression changes and cell heterogeneity. Recently, artificial intelligence (AI) has emerged as a powerful tool for identifying potential targets within TCM. The continuous improvement of professional TCM databases supports the integration of information, including prescriptions, bioactive constituents, target interactions, disease associations, and pharmacological annotations, gradually building a more comprehensive data system for TCM research. These databases serve as key resources, allowing AI-based approaches to identify targets in TCM. This review summarizes recent progress in AI-assisted TCM target identification, discusses experimental strategies for validating computational predictions, and outlines the main methodological challenges and future directions.

Citation: Guo C, Li Q, Liang W, Lu J, Yue W, et al. 2026. AI-enabled target discovery in traditional Chinese medicine: from computational prediction to experimental validation. *Targetome* 2(3): e029 <https://doi.org/10.48130/targetome-0026-0027>

Introduction

Traditional Chinese Medicine (TCM) has a history of several thousand years and has evolved into a distinct medical system through centuries of clinical practice and progressive theoretical refinement. Its formulations are mainly derived from natural sources, including plants, animals, and minerals^[1]. The active ingredients in these natural materials are the basis of the clinical efficacy of TCM. From 1981 to 2019, over 60% of small-molecule drugs approved by the US Food and Drug Administration (FDA) were either directly or indirectly derived from natural products^[2]. Accordingly, TCM represents an important natural compound source for innovative drug discovery and development^[3].

In practice, TCM is commonly administered either as single herbs or as multi-herb formulas. *Astragalus membranaceus* is one representative medicinal herb commonly used for fatigue and related symptoms associated with qi deficiency^[4]. Multi-herb formulas are more frequently used than single herbs in clinical practice^[5]. For instance, *Si-Jun-Zi-Tang* is a representative TCM formula, which includes *Panax ginseng*, *Atractylodes macrocephala*, *Poria cocos*, as well as *Glycyrrhiza uralensis*^[6]. It is used to treat spleen-stomach qi deficiency and related symptoms, including loss of appetite, abdominal distension, and indigestion^[7]. Another representative formula is *Niu Huang Jie Du Wan*, which contains *Bubali cornu*, *Coptis chinensis*, *Scutellaria baicalensis*, *Gardenia jasminoides*, *borneol*, and several other ingredients, and is commonly used to clear heat, remove toxins, or treat conditions like acute upper respiratory tract infection^[8]. Additional formulas and TCM preparations, including

Yupingfeng Granule, *Qingfei Paidu Decoction*, *Xuebijing Injection*, and *Qiliqiangxin Capsule*, have also shown preventive or therapeutic potential across respiratory, infectious, inflammatory, and cardiovascular conditions^[9–14].

Although TCM is widely used in clinical practice, its pharmacological mechanisms remain unclear, especially at the molecular level^[1]. To achieve mechanistic understanding, ensure pharmacological standardization, and support clinical translation, systematic identification of TCM molecular targets has become essential^[15]. However, this remains challenging^[16]. Unlike many conventional drugs that act through relatively defined targets, TCM often works through multiple components, targets, and pathways. The isolation and characterization of bioactive constituents from TCM are often time-consuming, while their chemical composition and stability may differ from source materials and environmental conditions. Resource limitations may further constrain related TCM research in some settings^[17].

Recent advances in analytical and experimental technologies have improved the precision and scalability of TCM target research. Cellular Thermal Shift Assay (CETSA), Drug Affinity Responsive Target Stability (DARTS), and other methods can directly verify the interactions between bioactive constituents and their targets. High-throughput screening, based on an automated platform, can rapidly evaluate a large number of compounds and accelerate the identification of promising compounds and potential targets^[18]. Despite the continuous progress of modern analytical and experimental technology, TCM target-discovery research remains limited by chemical complexity, pharmacological diversity, and interindividual variability^[16].

AI is an important branch of computer science that aims to develop intelligent systems with cognitive functions^[19]. In 1956, John McCarthy and others first proposed the concept of 'artificial intelligence' at the Dartmouth Conference, marking the official birth of AI as an independent discipline^[20]. In the past few decades, the field of AI has developed rapidly^[21]. Recent advances in machine learning, deep learning, and generative AI have increased the ability to analyze complex biomedical and pharmacological data. These advances have created new opportunities for large-scale data mining, pattern recognition, and predictive modeling in drug discovery^[22]. In data integration and analysis, deep learning (DL) can model complex feature representations, while machine learning (ML) remains effective for identifying informative patterns across multiple data sources^[23]. With these technologies, researchers can extract valuable information from large and diverse datasets and improve decision-making in drug discovery and development^[24].

In pharmacology, AI can integrate heterogeneous data through multimodal learning and knowledge mapping, enabling the analysis and interpretation of complex target discovery in TCM research^[25]. AI algorithms can support virtual screening and docking-based prioritization of TCM-derived compounds against candidate protein targets, thereby facilitating lead identification and reducing time and cost in early-stage drug discovery^[26]. The integration of graph neural networks with network pharmacology provides a framework to infer potential targets and system-level mechanisms of TCM. By integrating multi-omics data, including genomics, transcriptomics, proteomics, and metabolomics, with bioinformatics analysis and multimodal machine learning, researchers can identify potential targets of active TCM ingredients^[27]. Current generative AI research is mainly centered on large language models (LLMs) based on Transformer architectures. Specialized generative AI models can efficiently predict potential targets related to TCM by utilizing multi-omics data and algorithms^[28]. Although AI-based models still face challenges such as uneven data quality and insufficient algorithm interpretability, AI shows potential in TCM target discovery^[29].

This review aims to summarize the application of AI in TCM target discovery and to help researchers in the field better understand its challenges and potential. It focuses on the roles of AI technologies in identifying targets and discusses experimental strategies for validating AI-based predictions. The importance of experimental confirmation for obtaining reliable and biologically relevant results is emphasized. Based on a systematic overview of current methods and research progress, the review highlights limitations of existing studies and proposes directions for future work to support the integration of AI and TCM and further exploration of TCM mechanisms. Figure 1 shows the schematic diagram of AI-assisted target discovery in traditional TCM.

TCM dataset

TCM databases are crucial data sources for target discovery research and are of great importance for AI-assisted target discovery^[5]. With the rise of data-driven approaches in TCM research, a range of databases have been developed to collect data on TCM compounds, TCM formulas, chemical compounds, targets, and disease relationships^[30]. These databases include ITCM^[31], HIT 2.0^[32], TCM Bank^[33], TCMIO^[34], ETCM v2.0^[35], HERB^[36], and SuperTCM^[37], which contain data on TCM-derived compounds, predicted or verified potential targets, and pharmacological effects (Table 1).

Besides TCM-related sources, AI-assisted target discovery also benefits from a wide range of public biomedical resources that provide complementary chemical, pharmacological, genomic, and disease-related information^[23]. These include chemical and bioactivity-related resources such as PubChem^[38], ChEMBL^[39], and BindingDB^[40], which provide extensive information on chemical structures, bioactive constituents, and experimentally supported compound–target interactions. Similarly, drug-related resources such as DrugBank^[33], DrugCentral^[41], and SuperDrug2^[42] offer a large amount of information regarding drug properties and

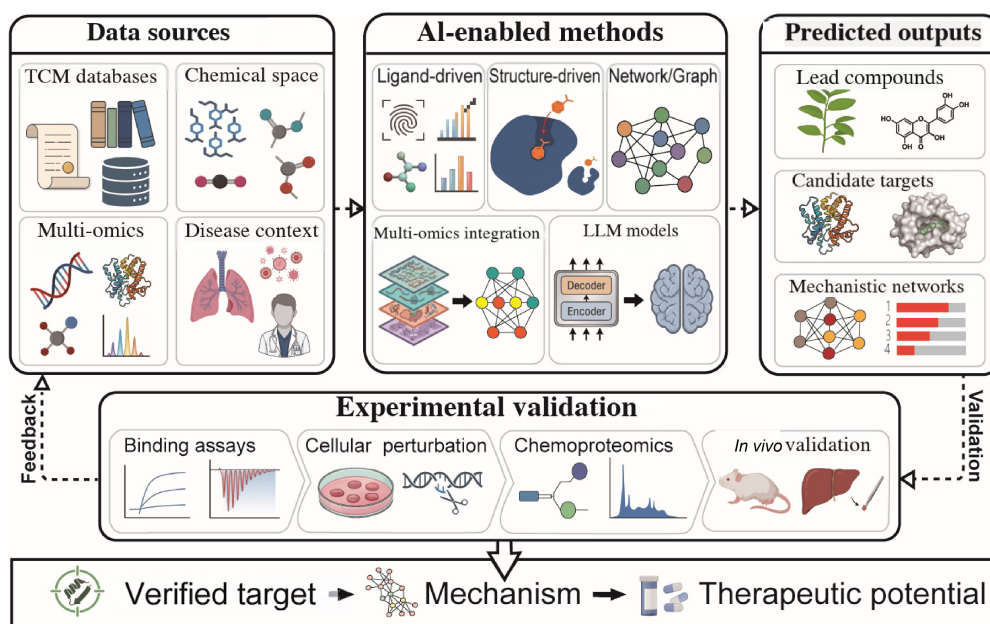


Fig. 1 Schematic diagram of AI-assisted target discovery in traditional Chinese medicine (TCM). The workflow integrates multiple data sources, including TCM databases, chemical space, multi-omics data, and clinical data, and infers potential TCM targets through ligand-driven, structure-driven, network-/graph-theory-based, multi-omics integration, large language model (LLM), and other AI-driven target prediction strategies. The predicted targets are then validated through biochemical binding experiments, cell-perturbation experiments, chemical proteomics, and *in vivo* pharmacological evaluation.

Table 1. Summary of major TCM-related databases and biological resources for AI-driven target discovery.

Database name	Last update	Data types	Description	Websites	Target types	Ref.
ETCM v2.0	2023	48,442 prescriptions, 38,298 ingredients, 1,040 targets, 9,872 patent drugs	Links TCM formulas and herbs to targets and diseases to reveal holistic mechanisms.	www.tcmip.cn/ETCM2/front	Predicted and experimentally validated	[35]
TCM bank	2023	9,192 herbs, 61,966 ingredients, 15,179 targets, 32,529 diseases	Provides chemical structures and ADMET properties for screening and target prediction.	https://tcmbank.cn/	Predicted and experimentally validated	[46]
HERB (v2.0)	2024	7,263 herbs, 49,258 ingredients, 12,933 targets	Integrates clinical trials with experimental data to validate herb–target interactions.	http://herb.ac.cn/v2	Predicted and experimentally validated	[47]
ITCM	2023	25,857 prescriptions, 8,454 herbs, 43,430 ingredients, 1,488 sequencing profiles	Focuses on the 'herb-ingredient-target' network to explain formula mechanisms.	http://itcm.biotcm.net/browse.html	Predicted and experimentally validated	[31]
Hit 2.0	2021	1,237 ingredients, 2,208 target sites, 10,031 active pairs	Offers experimentally verified ingredient-target pairs with binding affinity data.	www.badd-cao.net:2345	Experimentally validated	[32]
TM-MC 2.0	2024	635 herbs, 34,107 compounds, 13,992 targets, 27,997 diseases	Connects herbs and ingredients to molecular targets and metabolic enzymes.	https://tm-mc.kr/	Predicted	[48]
TCM-suite	2022	DNA marker sequences, Holmes/Watson module	A cloud platform for drug property prediction and network pharmacology analysis.	http://tcm-suite.aimicrobiome.cn/	Predicted	[49]
DCABM-TCM	2023	1,816 blood-entering components, 192 prescriptions, 194 herbs	Bridges TCM syndromes and medical cases with specific biological markers.	http://bionet.ncpsb.org.cn/dcabm-tcm/#/Home	Predicted and experimentally validated	[50]
TCMSID	2022	499 herbs, 20,015 ingredients, 3,270 targets	Provides structural and botanical data for identifying TCM-specific active ingredients.	https://bidd.group/TCMID/	Predicted	[51]
SuperTCM	2021	6,516 herbs, 55,000 ingredients, 254 pathways	Supports target discovery through functional enrichment and network analysis tools.	http://tcm.charite.de/supertcm	Predicted and experimentally validated	[37]
SymMap (v2)	2019	1,717 symptoms, 499 herbs, 4,302 targets	Maps TCM symptoms to modern diseases and targets via a large knowledge graph.	www.symmap.org	Predicted and experimentally validated	[52]
TCMID 2.0	2018	46,929 prescriptions, 43,413 ingredients, 17,521 targets, MS spectra	Maps TCM formulas and herbs to molecular targets and diseases to facilitate network pharmacology and mechanism study.	www.megabionet.org/tcmid	Predicted and experimentally validated	[53]
YaTCM	2018	1,813 prescriptions, 6,220 herbs, 18,697 targets, 47,696 compounds	Provides associations between TCM herbs, ingredients, and targets with a focus on ADMET properties and drug-likeness evaluation.	http://cadd.pharmacy.nankai.edu.cn/yatcm/home	Predicted and experimentally validated	[54]
TCMSP	2014	499 herbs, 29,384 ingredients, 3,311 targets, ADME properties	Links herbs and ingredients to targets and diseases based on ADME properties to evaluate drug-likeness and screening potentials.	https://old.tcmsp-e.com/tcmsp.php	Predicted	[55]
ccTCM	2023	273 medicinal materials, 1,449 quantitative components	Integrates chemical ingredients and molecular targets with cancer-related pathways to facilitate TCM-based anti-tumor drug discovery.	www.cctcm.org.cn	Predicted and experimentally validated	[56]
PubChem	2026	>100 million compounds, bioassay data	Provides extensive chemical structures, bioactivity data, and standardized identifiers to support virtual screening and target interaction studies.	https://pubchem.ncbi.nlm.nih.gov/	Experimentally validated	[57]
ChEMBL	2026	2.8 million substance, 1.8 million bioassays	A large-scale bioactivity database linking drug-like molecules to therapeutic targets through experimentally validated binding and functional assays.	www.ebi.ac.uk/chembl	Experimentally validated	[58]
BindingDB	2026	3.2 million data points, 1.4 million compounds, 114,000 targets	Focuses on the quantitative binding affinities of drug-like molecules with proteins to support molecular docking and QSAR model development.	www.bindingdb.org/rwd/bind/index.jsp	Experimentally validated	[40]
DrugBank	2026	13,575 substances, comprehensive target and pathway data	Provides detailed drug-target information and pharmacological data to facilitate target identification and drug repurposing for TCM ingredients.	https://go.drugbank.com/	Predicted and experimentally validated	[33]
DrugCentral	2026	Regulatory data, indications, side effects	Integrates comprehensive information on approved drugs, including molecular targets, bioactivity, and clinical indications to support target validation.	https://drugcentral.org/	Experimentally validated	[41]

(to be continued)

Table 1. (continued)

Database name	Last update	Data types	Description	Websites	Target types	Ref.
DRESIS (v2.0)	2025	> 20,000 drugs, metabolic reprogramming data	Links TCM ingredients to gene expression profiles and disease-related targets to facilitate mechanism-based drug discovery and repositioning.	http://dresis.idrblab.net/	Predicted and experimentally validated	[59]
DrugMap	2025	33,000 drugs, 50,180 interactions, tissue expression profiles	Maps the relationships between drugs, targets, and diseases through integrated network pharmacology to support target identification and mechanism study.	https://drugmap.idrblab.net/	Predicted and experimentally validated	[60]
SuperDrug2	2017	4,587 compositions, 2D/3D structure, PBPK simulation	Provides extensive structural and pharmacological data for approved and experimental drugs to facilitate target prediction and repositioning for TCM ingredients.	http://bioinf.charite.de/superdrug	Predicted and experimentally validated	[42]
GeneCards	2025	Human genome disease target annotation	A comprehensive human gene database providing detailed information on functions, pathways, and associated diseases to support target identification.	www.genecards.org	Predicted and experimentally validated	[61]
TTD	2025	932 Validation targets, QSAR model library	Provides comprehensive information on clinical and experimental targets, including their associated drugs and clinical trial statuses.	https://ttd.idrblab.cn/	Predicted and experimentally validated	[44]
DisGeNET	2024	GWAS data, clinical variant association	Integrates human gene-disease associations from curated databases and literature to support disease-target discovery and mechanism analysis.	https://disgenet.com/	Predicted and experimentally validated	[45]
BATMAN-TCM (v2.0)	2023	Predictive interactions, GO/KEGG analysis	An online bioinformatics tool for predicting TCM molecular targets and functional mechanisms through network-based association analysis.	http://bionet.ncpsb.org.cn/batman-tcm/index.php	Predicted and experimentally validated	[62]
NPASS	2017	35,032 NPs, 25,041 species, quantitative activity values	Provides curated bioactivity and structural data for natural products and their targets to support drug discovery and mechanism studies.	http://bidd2.nus.edu.sg/NPASS	Experimentally validated	[63]

mechanisms. On the other hand, target- and disease-related sources such as GeneCards^[43], Therapeutic Target Database (TTD)^[44], and DisGeNET^[45] provide information on genes, targets, and diseases that support the understanding of targets and mechanisms.

Despite differences among these databases in scope, curation strategies, update frequency, and data quality, they collectively provide a comprehensive data resource for AI-driven studies in TCM target discovery. The integration of heterogeneous data from these resources enables more comprehensive analyses of compounds, targets, and diseases, thereby facilitating the identification and validation of potential targets in TCM research^[64].

AI-enabled target discovery in TCM

Ligand-based target discovery in TCM

Ligand-driven models use known active ingredients as references to infer biological activities or potential target associations without requiring specific structural information about the target. These models are primarily based on the similarity principle, which states that compounds with similar structures tend to exhibit similar biological activities^[65]. This consideration is especially relevant in TCM research, as many compounds remain incompletely characterized and may lack corresponding structural information for their targets. Therefore, ligand-driven methods are particularly valuable for hypothesis generation, multi-target strategies, and virtual screening of TCM compounds.

Chemical similarity remains one of the most commonly used approaches in ligand-driven target discovery. It is typically

quantified using molecular fingerprints, such as extended connectivity fingerprints (ECFP), combined with similarity metrics like the Tanimoto coefficient^[66]. Graph neural networks (GNNs) have demonstrated strong potential for learning molecular representations and can identify compounds with similar biological activities, thereby enhancing chemical similarity-based approaches^[67,68].

Ligand-based virtual screening (LBVS) represents another major branch of ligand-driven target discovery (Fig. 2a). This approach ranks candidate compounds according to their similarity to known bioactive constituents, generally using similarity search methods or machine learning classifiers. LBVS is particularly useful when target structural information is unavailable, which is a common situation in TCM-related research. Compared with conventional screening methods, AI-assisted LBVS improves the enrichment of bioactive constituents by capturing nonlinear relationships between molecular features and target-binding activity^[69]. Active learning strategies have also been integrated with LBVS to systematically explore TCM compound libraries. This combination facilitates the identification of bioactive constituents while maintaining cost-efficiency^[70].

For instance, quantitative structure–activity relationship (QSAR) models describe quantitative relationships between molecular properties and bioactivities and have been a fundamental component in ligand-driven discovery for a long time. Traditional QSAR methods rely on statistical models built from predefined molecular properties, whereas recent AI-based QSAR models heavily utilize deep learning neural networks that learn directly from chemical structures^[71,72]. GNN-based models have been shown to more accurately describe the nonlinear relationship between TCM molecule structures and target-binding affinities by treating TCM molecules

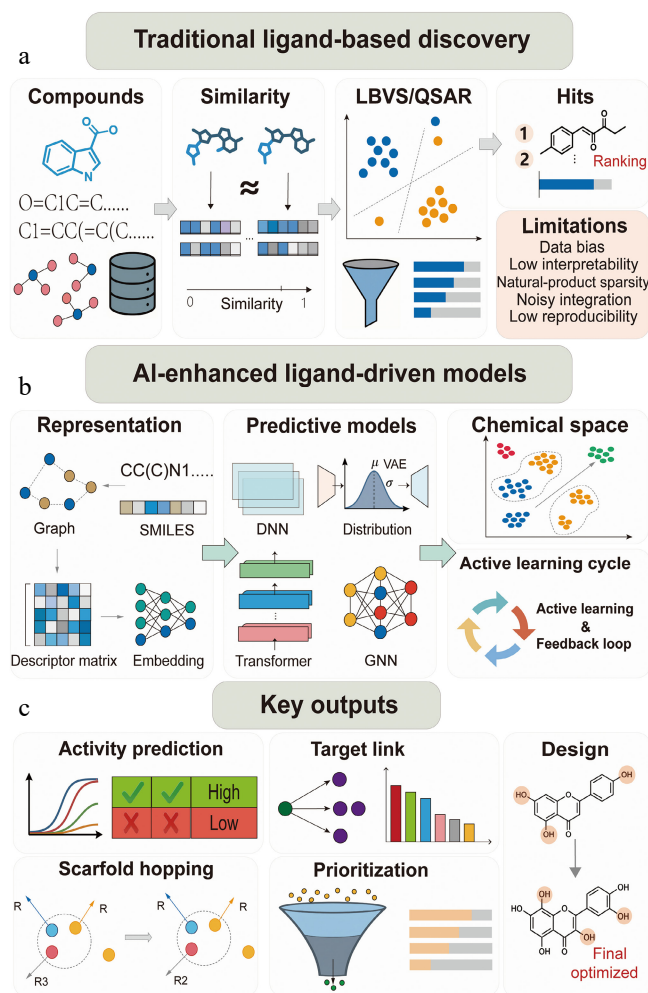


Fig. 2 Integrated ligand-driven target discovery workflow for TCM compounds. (a) Traditional ligand-based discovery: compounds from TCM libraries are screened via chemical similarity and ligand-based virtual screening (LBVS)/quantitative structure-activity relationship (QSAR) models to rank potential bioactive molecules. (b) AI-enhanced ligand-driven models: graph- or simplified molecular input line entry system (SMILES)-based embeddings feed deep neural network (DNN), graph neural network (GNN), or transformer models to predict activity, explore chemical space, and enable scaffold hopping. Active learning cycles iteratively refine predictions with experimental feedback. (c) Key outputs: generates activity predictions, target links, scaffold hopping opportunities, prioritized leads, and informs rational compound design for experimental validation.

as molecular graphs^[73]. Therefore, it is reasonable to believe that AI-based QSAR models have enhanced the robustness and scope of application for ligand-driven target discovery from various types of compounds in TCM (Fig. 2a, c). The Ligand-Transformer model proposed by Zhang et al. does not require compound structures and instead uses protein sequences and small molecule topologies for virtual screening of compounds. It has shown a hit rate greater than 50% and inhibitors with low nanomolar potency^[74].

Chemical-space analysis allows for more effective visualization and understanding of the diversity and bioactivities present in a large set of compounds. Dimensionality reduction techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) are often used for this purpose, and AI-based embeddings have also been proposed as a tool for better organization and understanding of chemical space and identification of

biologically active regions within it^[75]. More recent developments in generative models have extended this concept to allow for the projection of TCM bioactive constituents into continuous spaces. Variational autoencoders and other such models have been proposed as a tool for better visualization and understanding of activity cliffs and structurally informative regions and for target-based exploration of compounds (Fig. 2b, c). Thus, it has extended the scope of chemical-space analysis from purely descriptive visualization to more predictive and target-based understanding.

Significantly, AI has greatly enhanced ligand-driven target discovery by improving learning and predictability in molecular representations. Latest reports have shown that DL techniques, especially GNN and transformer models, have the potential to extract various molecular representations from chemical structures and SMILES strings (Fig. 2b), thereby facilitating better ligand similarity assessment and activity prediction compared to other models that use descriptor-based techniques^[76,77]. Moreover, AI-driven analysis in chemical spaces has enhanced scaffold-hopping and diversity discovery, especially in the case of natural products and TCM-related compound libraries^[78].

Despite these advancements, there are a few challenges. The public bioactivity datasets may be biased toward synthetic drug-like compounds, while the diversity of TCM is not well represented, making it difficult to generalize the results to TCM-derived compounds^[79]. In addition, the interpretability of DL algorithms, particularly in QSAR and similarity-based analysis, is not well understood, which may affect the understanding and translation of the results^[80]. The most challenging issue is the data integration of different data types, including ligand properties, phenotypic response, and pathway data, into a unified predictive framework, particularly in the presence of inconsistent annotations and noisy data^[81].

Future development of ligand-driven AI methods is likely to move toward multimodal or hybrid methods, which integrate chemistry, biology, omics technology, and network data, so as to build richer background information for target inference. This may be particularly useful for understanding the complex formulation characteristics commonly found in TCM-derived compounds. Deep learning methods presented in the form of conditional chemical language models may also contribute to target-directed molecular design because they are able to generate candidate compounds based on specific protein families or disease states. In addition, the integration of ligand similarity analysis, QSAR prediction, structure-based methods, and closed-loop methods for experimental target validation also helps to improve the reliability of computational target prediction. Another important aspect is adaptive or active learning methods, which help reduce false positives and improve the translational relevance of AI methods.

Structure-based target discovery in TCM

Structure-driven target discovery employs structural data related to biomolecular targets to predict ligand binding, as well as to clarify the binding process using atomic detail. Unlike ligand-driven methods, where activity is predicted using similarity between compounds, structure-driven methods offer more direct mechanistic insights using physical models of compound interaction within binding sites, as well as changes in conformation that affect ligand recognition^[82]. The data integration of AI with conventional physics-based methods has enhanced predictions of protein flexibility, interaction energetics, and even large-scale compound virtual screening^[76]. Thus, structure-driven methods have become critical tools for the discovery of binding sites, refining target hypotheses,

and even exploring allosteric or cryptic sites, especially in TCM pharmacology^[82].

Molecular docking aims to predict how ligands are embedded in protein binding pockets and score potential binding modes according to geometric and energy criteria^[83]. Traditional docking algorithms sample the ligand conformational space and rank the resulting poses using scoring functions (Fig. 2a)^[84]. Although classical docking tools (e.g., AutoDock, Glide) are still widely used, recent frameworks based on deep learning have improved the accuracy of conformation prediction and scoring. AI models trained based on protein–ligand complexes are able to infer binding interactions that are more closely related to experimental affinity, thereby reducing reliance on approximate scoring functions^[82]. The LABind model, which explicitly integrates ligand coding into binding site prediction, was developed by Zhang et al. By combining a graph converter with a cross-attention mechanism, the model can learn local protein–ligand interaction modes and maintain good generalization ability to unseen ligands^[85].

New AI-native docking approaches, like transformer-based approaches and unified scoring approaches, show promising efficiency in ligand virtual screening and binding pose prediction for a wide range of targets^[86]. With the help of Geometric deep learning and Equivariant Neural Networks, AI-aided docking approaches have unraveled the multi-target regulatory mechanism of Ginsenosides targeting the ACE2 receptor and Mpro protease (Fig. 2b, c)^[87].

In structure-based virtual screening (SBVS), ligands are prioritized based on the predicted interactions with a specified target structure. A standard process for SBVS is composed of target preparation, pocket characterization, docking, and ranking (Fig. 3a)^[84]. AI improves SBVS by using scoring functions and pose filters that eliminate the problems of heuristic docking score predictions. DL approaches that combine protein–ligand interaction patterns are able to differentiate between binders and false positives (Fig. 3b, c). SBVS is useful for assessing the interactions of various TCM compounds with complex targets and identifying novel associations in complex TCM formulas^[83].

Molecular dynamics (MD) simulation studies are based on the dynamics of the protein–ligand systems, including the time dependence of the systems, induced fit effects, and the stability of the binding process, in a dynamic manner^[88]. AI helps MD simulation studies achieve quantum accuracy in TCM natural products while maintaining the speed necessary for protein–ligand systems (Fig. 3b, c). ML helps in the acceleration of MD studies, the identification of important interaction patterns, and the development of enhanced sampling methods, which reduce the computational cost while maintaining the dynamics^[89]. In a recent structure-based drug discovery (SBDD) study, MD and SBVS were used on phytocompounds targeting a receptor involved in cancer, showing stable binding with dynamic properties, as observed *in vitro* studies, highlighting the role of MD in validating target prediction^[90].

In structure-based modeling, the accuracy of pose prediction is enhanced by the use of AI for the prediction of conformational patterns, scoring and ranking with deep interaction models that include physicochemical and data-driven features, and the efficiency of library virtual screening with the use of transformer and GNN-based scoring approaches^[82].

The major challenge for structure-based AI is target flexibility. Many existing AI models are based on a rather static description of proteins, but the interaction of TCM-derived compounds with their molecular targets is often highly dynamic. Furthermore, the 'Solvation Effect' of the complex physiological environment of TCM decoctions is hard for existing AI approaches to simulate^[91]. End-to-end generative models, such as AlphaFold 3, might provide new opportunities for a joint description of protein structure and ligand binding^[92]. Further progress might depend on closer data integration with machine learning force fields (MLFFs) and simulation approaches^[93].

Network pharmacology-based target discovery in TCM

Network pharmacology (NP), a term first proposed by Hopkins in 2007, is grounded in systems biology and network science

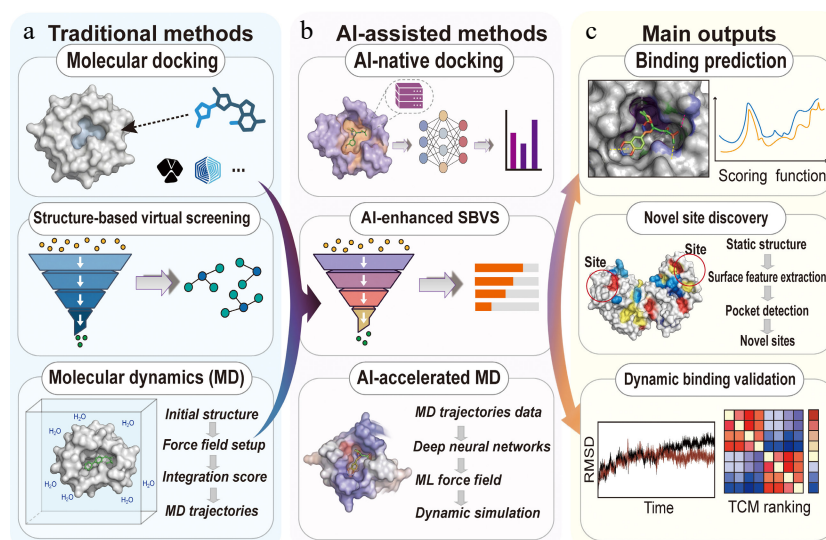


Fig. 3 Integrated structure-driven target discovery workflow. (a) Traditional methods: combines molecular docking, structure-based virtual screening (SBVS), and molecular dynamics (MD) to evaluate ligand–protein binding, rank candidate compounds, and assess interaction stability over time. (b) AI-assisted methods: integrates AI-native docking, AI-enhanced SBVS, and AI-accelerated MD to predict binding affinities, efficiently prioritize compounds, and model dynamic interactions with deep learning refinement. (c) Output applications: highlights core outcomes of target discovery, including binding prediction, binding-site discovery, and dynamic validation, supporting experimental prioritization of TCM compounds while reducing visual complexity for journal readability.

concepts^[94]. It treats drugs, targets, pathways, and diseases as interconnected networks, aligning with the multi-component and multi-target characteristics of TCM^[24]. A key feature of NP is the integration of biomolecular interaction data, which enables the combined analysis of multiple compounds and their potential targets, helping to identify major bioactive constituents and their mechanisms of action^[15]. In addition, NP provides a new perspective for visualizing complex interactions between TCM-derived compounds and molecular targets^[95]. By integrating the concepts of systems biology, multi-omics approaches, and high-throughput analysis, NP has become an important tool for investigating the holistic mechanisms of TCM^[96]. However, conventional NP has limitations, including noise accumulation, high dimensionality, and difficulty in integrating dynamic processes^[97]. AI provides new opportunities to enhance target discovery (Fig. 4c). The advance is achieved through the integration of ML, DL, and GNNs for the analysis of complex interactions among compounds, targets, pathways, and diseases^[27].

ML has become increasingly important in NP for target prioritization, association inference, and network-based prediction (Fig. 4c, d). These methods have shown promise in providing deeper insights into the mechanisms of TCM by identifying potential interactions between compounds, targets, and pathways based on patterns extracted from high-dimensional omics datasets and complex biomolecular networks^[98]. For example, Ying et al. analyzed TSGJ-related bioactive components and breast cancer-related targets and identified 160 common targets. Subsequently, they identified five key genes using four ML algorithms, including support vector machines (SVM) and random forests (RF)^[99]. A study by Ding et al. identified feature genes and key therapeutic pathways for idiopathic pulmonary fibrosis. They used a combination of LASSO logistic regression and SVM-recursive feature elimination (SVM-RFE) for this analysis^[100]. In another study by Zhao et al., potential targets for treating colorectal cancer were identified by combining database-derived targets with differential gene expression analysis and weighted gene co-expression network analysis^[101].

DL has the potential to extract and learn complex nonlinear relationships (Fig. 4c)^[102]. For instance, Chen et al. proposed a functional representation method for gene features based on DL. The authors showed that applying this method to predict compound-target interactions using the Broad Institute's L1000 dataset was more efficient than using models based solely on gene identity. This DL-based approach led to a substantial increase in the number of predicted compound-target interactions. DL models have also been applied to evaluate the effectiveness of TCM formulas in treating liver cancer by integrating multidimensional data from sources such as The Cancer Genome Atlas (TCGA) and the Traditional Chinese Medicine Systems Pharmacology Database (TCMSP), which may facilitate the identification of potential targets for compound combinations^[103]. Qu et al. proposed a DL model based on convolutional neural networks for the data integration of heterogeneous network data and the accurate prediction of drug-target interactions^[104]. Zhang et al. proposed a DL-based hybrid model using graph convolutional networks (GCN) and RNN for the accurate target prediction of the meridian associations of bioactive constituent metabolites^[105].

Among various DL techniques, GNNs have shown promise in network pharmacology due to their ability to handle graph-structured biological data and identify key hub nodes within complex biological networks (Fig. 4c, d). Graph theory-based GNN models can learn feature representations by propagating information across graph-structured data according to inter-node connections^[106]. In this regard, Guo et al. have shown that a GNN-based DL model could be applied for target prediction and virtual screening of compounds with *anti-Helicobacter pylori* activity using molecular graphs and extended connectivity fingerprints (ECFP)^[107]. Similarly, Duan et al. proposed a GNN-based target prediction framework consisting of three modules^[108].

However, some challenges still exist in AI-assisted NP analysis. Most AI models are highly dependent on the quality and quantity of input data^[109]. Issues such as data imbalance, inconsistent annotation, high collinearity among input features, and substantial noise

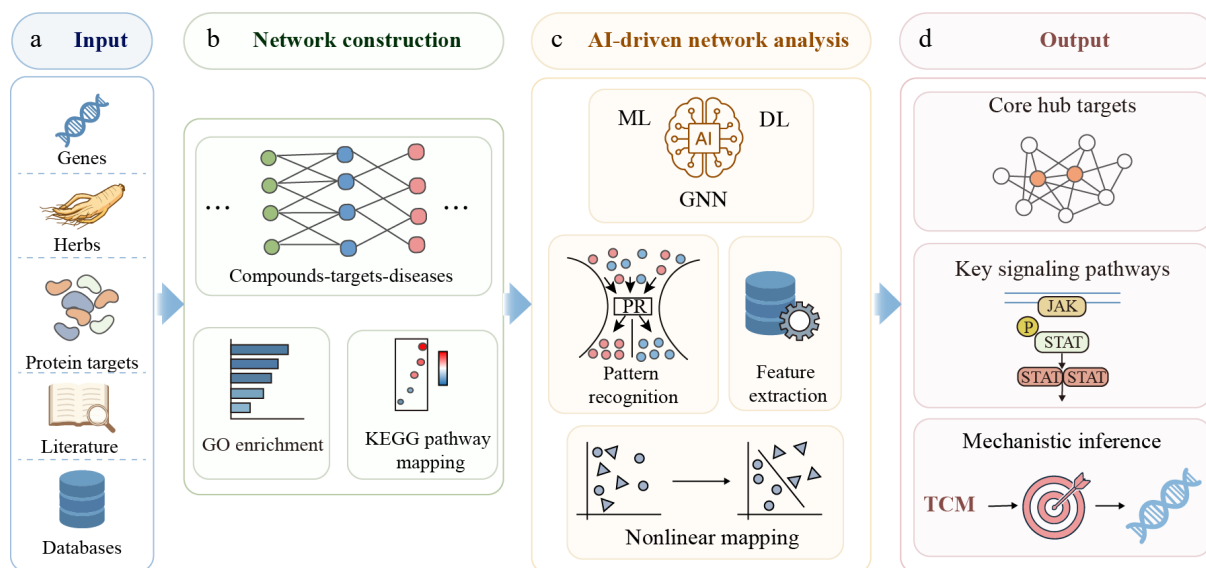


Fig. 4 Schematic workflow of AI-assisted network pharmacology for target discovery in traditional Chinese medicine (TCM). (a) Multi-source input data include genes, herbs, protein targets, literature, and databases. (b) Network construction integrates compound-target-disease associations with Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping. (c) AI-driven network analysis incorporates machine learning (ML), deep learning (DL), and graph neural networks (GNNs) for pattern recognition, feature extraction, and nonlinear mapping. (d) The resulting outputs include core hub targets, key signaling pathways, and mechanistic inference, thereby supporting target identification and biological interpretation in TCM.

can compromise the robustness and introduce bias in ML model predictions^[23]. Other challenges involve the strong reliance of models such as LASSO regression, SVM-RFE, decision trees, and ensemble methods on structured and high-quality input data^[110]. Similarly, clustering approaches may perform poorly in high-dimensional or noisy data environments^[24]. In deep learning models, additional limitations include restricted interpretability, dependence on structured experimental datasets, and a high risk of overfitting due to limited training samples. RNN models may be less effective in capturing long-range dependencies in data, and GNN models may be computationally expensive in high-dimensional heterogeneous networks^[22].

Recent studies have increasingly emphasized the integration of multiple database resources, cross-modality data fusion, and model-based target validation across various computational approaches^[5]. For example, Huang et al.^[111] explored the molecular targets and mechanisms of isoliquiritigenin (ISO) in chronic obstructive pulmonary disease (COPD) by using a combination of weighted gene co-expression network analysis, differential expression analysis, and target discovery from TCMSP, Comparative Toxicogenomics Database (CTD), PubChem, GeneCards, and Molecular Signatures Database (MSigDB). In addition, multi-modal fusion frameworks have been proposed. These frameworks integrate genomics, protein structure data, chemical information, pathway networks, and biomarkers into comprehensive input representations. For instance, Lu et al.^[112] proposed a multimodal fusion deep learning model that combined SMILES encoding vectors, ECFP fingerprints, and molecular graph information using a transformer encoder together with a bidirectional gated recurrent unit (BiGRU) and graph convolutional networks (GCNs). Their results suggest that integrating information from multiple modalities can improve predictive performance^[113].

Multi-omics-based target discovery in TCM

Although AI-driven network pharmacology has built a systematic framework for analyzing the multi-component and multi-target properties of TCM, most existing methods still rely mainly on relevance-based inference, which makes it difficult to distinguish between true treatment drivers and spurious associations caused by confounding factors (Fig. 5). This methodological bottleneck makes multi-omics integration and causal inference the next key step to

promote the discovery of TCM targets from correlation prediction to causal validation.

However, applying multi-omics technology and causal inference to TCM research presents significant challenges. TCM interventions are inherently multi-component, time-dependent, and dynamically adjustable, and are often tailored to patient responses^[114], while the vast majority of omics assays can only reflect static molecular snapshots. In addition, Mendelian randomization and other methods are based on lifelong genetic exposure models, which are difficult to directly match with the mode of short-term therapeutic intervention of TCM. If not handled properly, issues including high-dimensional noise, confounding factors, and time-scale mismatch will not only fail to reveal the true causal relationship but also generate spurious associations^[114].

Despite these inherent methodological challenges, the integration of multi-omics and AI technology provides an indispensable high-dimensional database for the mechanism analysis of TCM. In recent years, a comprehensive multidimensional analysis framework has been systematically established, which integrates genomics, transcriptomics, proteomics, and metabolomics, providing solid methodological support for decoding the multi-target characteristics of TCM^[115]. Under this multidimensional approach, the combination of multi-omics and network pharmacology successfully mapped the intervention network of the TCM system. For example, the comprehensive omics approach revealed that Fuyuan Xingnao Decoction and Huangqi Longdan granules play a role in the treatment of cerebral ischemia by regulating the gut microbial-metabolite axis and systemic signaling pathways, including the Janus kinase 2/signal transducer and activator of transcription 3 (JAK2/STAT3) pathway and T helper 17/interleukin 17 (Th17/IL-17) axis^[116]. In addition, AI algorithms have pushed this convergence upstream to decipher the complex material basis of TCM. Machine learning models have been combined with multi-omics data to decode the highly complex metabolic network of bioactive compounds in *Ganoderma lucidum*, providing an intelligent prediction model for identifying the effective material basis of TCM^[117].

However, although these AI-driven multi-omics studies have greatly enriched the pharmacological network of TCM, they mainly construct high-resolution correlation maps. Therefore, causal inference is needed as a key bridge from complex molecular associations to biologically meaningful target prioritization.

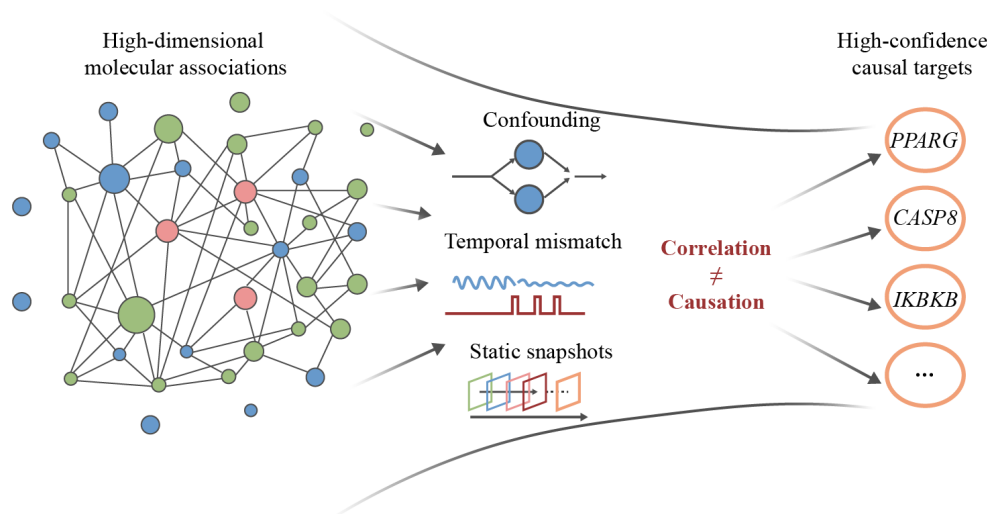


Fig. 5 Schematic overview of methodological challenges in multi-omics-based target discovery in traditional Chinese medicine (TCM). High-dimensional molecular associations are subject to confounding, temporal mismatch, and static omics snapshots, which complicate the distinction between correlation and causation and hinder the identification of high-confidence causal targets.

In order to bridge the key gap from statistical correlation to biological causality, researchers began to adapt the causal reasoning framework, especially Mendelian randomization (MR)^[118,119], to the unique characteristics of TCM pharmacology. It is worth noting that the high-dimensional multi-omics foundation may support important causal inference analysis: multi-layer omics data may help prioritize candidate genes, improve biological interpretation, and support complementary analyses, such as colocalization and triangulation^[120,121], thereby improving the robustness of causal target identification (Fig. 5). However, the core limitation of traditional MR is that it relies on a lifelong genetic exposure model, which cannot directly simulate the short-term therapeutic intervention mode of TCM^[122]. In order to address this time-scale mismatch, a practical research paradigm has emerged. MR is first used to identify causal targets associated with disease at the population genetic level and is then used to evaluate the potential of the multi-component regulatory network of TCM formulas to regulate these genetically supported nodes. Recently, this two-step strategy has been applied to multi-omics driven target research in TCM^[123].

For example, one study identified seven putative causal targets of rheumatoid arthritis through MR analysis, including *CASP8*, *PPARG*, and *IKKBK*. Subsequent network-based analysis showed that the regulatory network of Guizhi Shaoyao Zhimu Decoction was significantly enriched in these genetically supported targets and may regulate some of these key nodes. Molecular docking further supported the potential interaction between the main active compounds in the formula (such as quercetin and kaempferol) and the identified targets^[124]. Although the evidence itself cannot establish direct target involvement or therapeutic causality, it provides a feasible framework for linking genetically supported disease targets with the multi-component action characteristics of TCM formulas. This reverse matching strategy helps to solve the key methodological bottleneck of applying MR to TCM research and provides a

feasible framework for linking omics associations with genetically supported therapeutic hypotheses.

However, an ongoing limitation of current TCM omics research is the excessive reliance on cross-sectional sampling^[125], which fails to capture the dynamic, time-dependent nature of TCM treatment responses^[126]. In order to fully tap the potential of AI-driven causal inference in TCM target discovery, future research efforts must turn to longitudinal and perturbation research designs, combining time-series molecular monitoring and controlled therapeutic intervention^[127]. Only by inputting time-resolved multi-omics data into an AI causal reasoning framework^[128] can dynamic biological responses be described, rather than static correlations, thus providing a solid and operable foundation for causal target identification and mechanism exploration of TCM.

Generative AI-based target discovery in TCM

Whereas multi-omics and causal inference strengthen biologically grounded target prioritization, generative AI offers a complementary route by transforming large-scale unstructured TCM knowledge into computable representations for hypothesis generation and mechanistic reasoning. In recent years, great progress has been made in the application of large language models (LLMs) in TCM research^[129]. In addition to their powerful natural language processing capabilities, LLMs can also transform unstructured TCM literature and clinical knowledge into computable structured data representations, thereby providing a basic framework for TCM target inference and pharmacological mechanism analysis. In addition, when combined with the curated structured database and domain knowledge base, LLMs can promote the construction and expansion of TCM knowledge graphs (Fig. 6a, b).

BATMAN-TCM 2.0 is a TCM knowledge resource that focuses on compound–target interactions. By integrating manually curated and

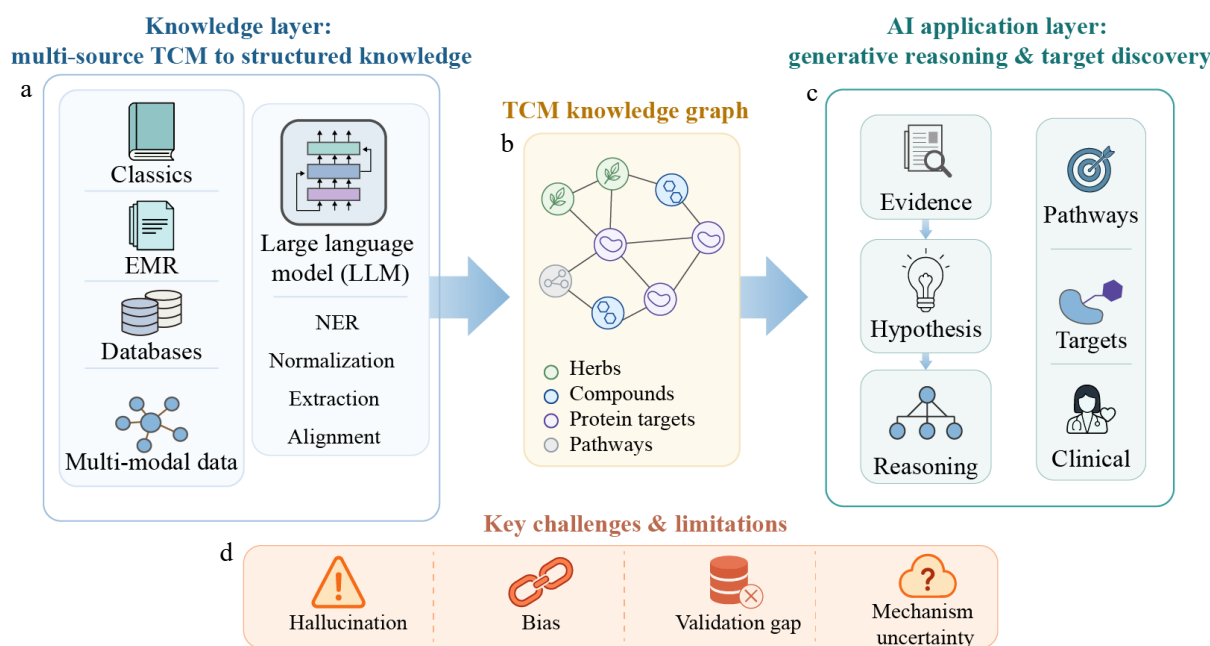


Fig. 6 Schematic framework of generative AI and knowledge graph-based modeling for traditional Chinese medicine (TCM) target discovery. (a) Knowledge layer for transforming multi-source TCM information, including classical texts, electronic medical records (EMRs), databases, and multimodal data, into structured representations with the assistance of large language models (LLMs). (b) Construction of a TCM knowledge graph linking herbs, compounds, protein targets, and pathways. (c) AI application layer for evidence integration, hypothesis generation, reasoning, and downstream applications in pathway analysis, target discovery, and clinical decision support. (d) Key challenges and limitations, including hallucination, bias, validation gaps, and mechanistic uncertainty.

computationally predicted data, the platform systematically organized more than 2.3 million high-confidence records of TCM ingredient–target protein interactions^[62]. This structured resource provides important support for knowledge graph construction, high-precision knowledge retrieval, and relational reasoning in TCM target discovery and pharmacological mechanism research. The knowledge graph supported by LLM and structured databases may help to represent the complex associations among formulas, herbs, ingredients, targets, and pathways, thus supporting multi-target discovery and mechanism interpretation in TCM research^[130].

Although large-scale language models and knowledge graph technologies have made rapid progress, TCM is still one of the most challenging knowledge engineering environments faced by AI systems. The characteristics of TCM knowledge include semantic ambiguity, inconsistent terminology, historical language changes, and a lack of standardized ontology. These characteristics make entity resolution, relationship extraction, and knowledge construction much more difficult than in the modern biomedical field^[131]. Therefore, building a reliable TCM knowledge graph is not a simple digitization process but a complex interpretive task that requires careful normalization, expert involvement, and validation. In order to systematically address these challenges and bridge the gap between historical literature and computable insights, a specialized modeling framework is needed (Fig. 6b).

However, the complexity of TCM theory, the profundity of syndrome differentiation, and ancient and abstruse medical classics have brought great challenges to the general LLMs applied in clinical diagnosis and treatment scenarios. In some cases, these problems may even lead to misinformation and model hallucinations, resulting in unsafe clinical recommendations^[129, 132, 133]. Therefore, in order to solve these key problems, researchers have designed and developed a series of generative AI models specifically used in the field of TCM (Fig. 6c). For example, TCMChat has been developed, which is a large generative model specially tailored for the field of TCM^[134].

In order to further improve the organization and reasoning ability of the AI models for complex TCM knowledge, researchers have developed a full-dimensional large TCM model named Tianyi, based on a massive amount of ancient Chinese medicine literature and real-world electronic medical record data. Benchmark evaluations suggest that the model performs well in complex tasks such as syndrome prediction, clinical decision support, and TCM formula generation^[135]. The above research results show that the model can help organize and clarify the complex correlation network among syndrome types, formulas, TCM, ingredients, and targets, which may be valuable for mechanistic interpretation and candidate target discovery in TCM research.

In addition, to improve AI-driven medical consultation, researchers have proposed BianQue, a health-oriented model centered on the logic of continuous inquiry. By simulating the iterative questioning process between clinicians and patients, BianQue may help overcome the limitation of single-step recommendations commonly seen in conventional large models. Preliminary studies suggest that, in health consultation scenarios, this strategy may improve the safety of computer-aided diagnosis and the relevance of personalized recommendations^[136].

From a research perspective, such interactive models may also facilitate more structured and traceable acquisition of clinically relevant TCM information. However, hallucinations in LLMs remain a major bottleneck for both TCM target discovery and clinical application (Fig. 6d). False compound–target interactions, unsupported mechanistic claims, and plausible but incorrect medical recommendations may all arise if model outputs are generated without

retrieval augmentation, graph-based reasoning, traceability, and expert review^[130, 132]. Therefore, although LLMs and TCM knowledge graphs provide unprecedented opportunities for organizing large-scale unstructured knowledge and exploring complex multi-component relationships, they should currently be primarily positioned as hypothesis generation and knowledge aids, rather than autonomous decision-making systems. Any new targets or mechanistic hypotheses generated by these models must still be validated through a rigorous experimental framework.

In summary, AI-based analytical methods can generate a large number of candidate targets and hypotheses regarding their potential mechanisms of action within a relatively short period^[24, 137]. However, prediction results often differ in evidence strength, model confidence, cross-method consistency, and biological interpretability^[25]. In this context, the transition from computational prediction to experimental investigation should not aim to validate all candidate targets in parallel^[137, 138]. Rather, it should rely on a stepwise prioritization framework that integrates evidence strength, model confidence, cross-method consistency, biological relevance, druggability, and experimental feasibility^[24, 25, 138]. Within this framework, experimental research serves not merely as a passive confirmation of computational outputs (Supplementary Table S1) but as an active process for target filtering, prioritization, mechanistic refinement, and causal validation^[18, 138].

Experimental validation of AI-predicted targets in TCM

Following computational prediction and target prioritization, experimental validation is essential to determine whether predicted targets are genuinely engaged and functionally relevant in TCM research. However, computational predictions are often prone to false-positive results because of algorithmic limitations, data bias, and the multicomponent, multitarget nature of TCM systems^[25, 139]. Therefore, a critical task in AI-driven TCM research is to convert *in silico* predicted targets into experimentally supported targets through a hierarchical evidence framework. This process requires a progressive, multilevel validation strategy that advances logically from *in vitro* biochemical binding assays to intracellular target engagement profiling, cellular functional validation, and ultimately, system-level *in vivo* pharmacological evaluation.

A range of experimental strategies are available to support the hierarchical validation of AI-predicted targets in TCM research, including *in vitro* biochemical and biophysical assays, intracellular target engagement approaches, cellular perturbation methods, and *in vivo* pharmacological evaluation (Table 2)^[140].

Biochemical and biophysical validation of binding activity

It is essential to confirm the *in vitro* interaction of compounds with their purified target proteins during the early stages of validation, as this provides direct molecular support for predicted targets (Fig. 7a). SPR is a real-time, label-free technique for the quantitative analysis of biomolecular interactions and binding kinetics^[141]. SPR has been widely adopted for the validation of target molecules for bioactive components in TCM. While the binding kinetics are primarily characterized by SPR, the thermodynamic analysis is carried out by ITC. ITC is used to measure the changes in heat during binding to determine simultaneously important parameters such as affinity (K_D), binding stoichiometry (n), and enthalpy change (ΔH)^[142].

Table 2. Experimental validation strategies for AI-predicted targets in TCM research.

Stage	Method	Measured readout	Strengths and TCM-specific pitfalls	Recommended application scenario	Ref.
<i>In vitro</i> binding	SPR	Direct binding kinetics (K_D , k_{on} , k_{off})	real-time quantitative binding; prone to nonspecific adsorption or aggregation with some polyphenols, quinones, and flavonoids	Orthogonal validation of AI-predicted binders/Single compound ★★★/Formula χ	[141]
	ITC	Binding thermodynamics (K_D , ΔH , stoichiometry)	Gold-standard thermodynamic readout; limited sensitivity for weak-affinity ligands; high concentrations may cause precipitation	Secondary confirmation of AI-predicted binders/Single compound ★★★/Formula χ	[140]
Intracellular engagement (probe-based)	ABPP	Active-site occupancy	High enzyme specificity; limited applicability to non-covalent compounds and probe-incompatible TCM ingredients	Enzyme-target validation/Single compound ★/Formula χ	[146,147]
	PAL	Reversible noncovalent interactions	Captures reversible binding; probe derivatization may reduce bioactivity or permeability	Target fishing when chemical modification is feasible/Single compound ★/Formula χ	[146,147, 149]
Intracellular engagement (label-free)	CETSA	Cellular thermal stabilization (ΔT_m)	Label-free <i>in situ</i> engagement; low intracellular exposure may yield false negatives	Cellular target engagement validation/Single compound ★★★/Formula ★	[151,152]
	DARTS	Protease resistance shift	Simple and scalable; sensitive to high-concentration artifacts and nonspecific protein protection	Rapid preliminary target screening/Single compound ★★★/Formula ★	[18, 147, 153]
	TPP	Proteome-wide thermal shift profiling	Unbiased proteome-wide target discovery; indirect responders or stress-related proteins may appear as hits	Identification of indirect or missed targets/Single compound ★★★/Formula ★★★	[151,152]
Functional necessity	RNAi	Knockdown-associated phenotypic change	Operationally accessible; incomplete knockdown and off-target effects may compromise causal interpretation	Preliminary functional validation of AI-predicted targets/Single compound ★★/Formula ★★	[150, 154]
	CRISPR knockout	Knockout-dependent loss of drug response	Strong causal evidence; partial attenuation may still be informative in multi-target settings	Core causal validation of AI-predicted targets/Single compound ★★★/Formula ★	[150, 156, 161]
	CRISPR screening	Functional network hub identification	Supports systems-level interpretation; herbal synergy complicates hit annotation	Pathway and module validation/Single compound ★★★/Formula ★★★	[156,157, 169]
<i>In vivo</i> validation	Target-dependent <i>in vivo</i> efficacy	Target-linked pharmacological efficacy	High translational relevance; multi-component exposure obscures single-ingredient contributions	Final <i>in vivo</i> validation of AI-derived hypotheses/Single compound ★★★/Formula ★★★	[150, 161, 174]
	PK-PD modeling	Exposure-response relationship	Clarifies formula-level synergy; herb-herb PK interactions complicate modeling	Validation of formula compatibility mechanisms/Single compound ★/Formula ★★★	[158, 162]

Applicability was rated as follows: ★★★, highly suitable; ★★, moderately suitable; ★, limited suitability; χ , generally not recommended. Ratings reflect practical feasibility, interpretability, and compatibility with AI-guided target validation in TCM research. Methods are organized as a progressive workflow from direct binding to intracellular engagement, functional necessity, and *in vivo* confirmation. TCM, traditional Chinese medicine; AI, artificial intelligence; SPR, surface plasmon resonance; ITC, isothermal titration calorimetry; ABPP, activity-based protein profiling; PAL, photoaffinity labeling; CETSA, cellular thermal shift assay; DARTS, drug affinity responsive target stability; TPP, thermal proteome profiling; RNAi, RNA interference; PK-PD, pharmacokinetic-pharmacodynamic; PAINS, pan-assay interference compounds.

However, these methods have inherent limitations when applied to TCM compounds. In particular, polyphenols and quinones may exhibit Pan-Assay Interference Compounds (PAINS)-like behavior or assay-interference risks under certain experimental conditions^[143]. They may undergo non-specific colloidal aggregation or nonspecific binding to SPR sensor chips, resulting in deceptive false-positive binding signals^[144]. Therefore, it is strongly recommended to add detergents (e.g., Tween-20) to the running buffer or use dynamic light scattering (DLS) to exclude aggregation^[145]. In addition, binding to purified proteins does not guarantee binding to targets in living cells, so subsequent intracellular validation is crucial.

Chemoproteomics-based evidence for target engagement

Although *in vitro* biophysical assays, such as SPR and ITC, can confirm the direct interaction between compounds and purified proteins, they cannot guarantee that such binding events actually occur in a highly complex intracellular environment. Chemical proteomics can capture the interaction between drugs and proteins under near-physiological cellular conditions, thus converting *in silico* predictions into verifiable intracellular molecular evidence and providing direct experimental verification for bridging the gap between *in vitro* binding and cellular function (Fig. 7b)^[146].

Probe-based chemical proteomics approaches

In competitive activity-based protein profiling (ABPP), it is possible to determine whether the test compound interacts with the active site of a specific enzyme by analyzing the fluorescence intensity difference of the fluorescently labeled probe compound before and after drug treatment. The traditional ABPP method is most effective when used with ligands that can covalently bind to their target proteins. However, many bioactive compounds derived from TCM bind to proteins through reversible, noncovalent interactions^[147]. To extend the utility of ABPP to noncovalent ligands, photoaffinity labeling (PAL) was adapted. PAL involves the use of photoaffinity probes. These probes form covalent bonds with the target amino acids under UV irradiation, thereby enabling the capture of non-covalent interactions that would otherwise be lost^[148].

However, the application of probe-based methods to the research of TCM faces a major practical bottleneck. When TCM molecules with complex structures are modified with bulky photosensitive groups and reporter tags (e.g., biotin or alkyne), they often suffer from significant steric hindrance. This modification frequently alters their intrinsic biological activity, target binding profile, or membrane permeability^[149,150]. Therefore, for many bioactive ingredients of TCM, the successful synthesis of probes is still challenging, which requires an alternative strategy without modification.

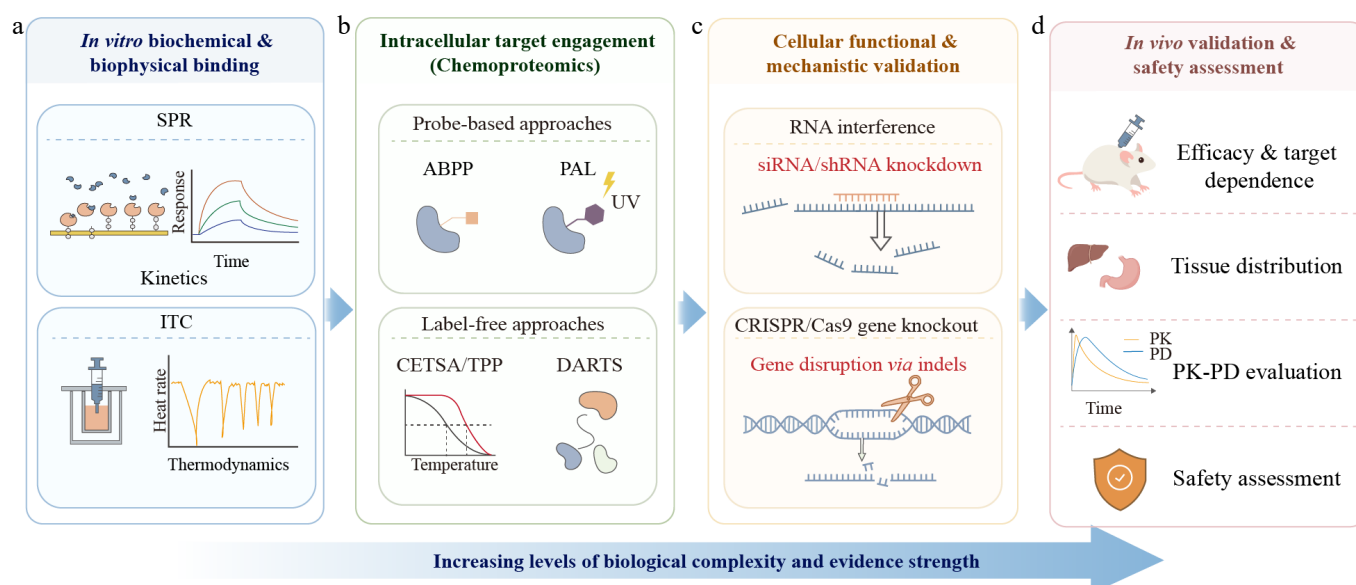


Fig. 7 Hierarchical experimental framework for validation of AI-predicted targets from traditional Chinese medicine (TCM)-derived compounds. (a) Biochemical and biophysical binding validation *via* surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) using purified target proteins. (b) Intracellular target engagement profiling *via* chemoproteomics, including probe-based (activity-based protein profiling, ABPP; photoaffinity labeling, PAL) and label-free (thermal proteome profiling, TPP/CETSA; drug affinity responsive target stability, DARTS) approaches, for intracellular target profiling under near-physiological conditions. (c) Cellular functional and mechanistic validation *via* gene perturbation (RNA interference, siRNA/shRNA; CRISPR/Cas9-mediated gene knockout) to confirm target necessity for the observed pharmacological effects. (d) *In vivo* validation and safety assessment in animal models, evaluating target-dependent efficacy, pharmacokinetic-pharmacodynamic (PK-PD) profiles, tissue distribution, and safety. From left to right, the framework indicates increasing levels of biological complexity and evidence strength.

Label-free target engagement approaches

Label-free target engagement approaches bypass the need for ligand modification, and can avoid potential interference with endogenous binding interactions. For example, in Cellular Thermal Shift Assay (CETSA), the thermal stability of proteins is usually improved when ligands are bound to proteins. This principle can be used to identify direct interactions between candidate compounds and their predicted targets in whole cells or cell extracts^[151]. Thermal proteomics, also known as thermal proteome analysis (TPP), further expands its application by combining this principle with quantitative mass spectrometry. This method can simultaneously measure the thermal stability changes of thousands of proteins in one experiment, thus comprehensively and unbiasedly revealing the interactions between drugs and targets in the proteome^[147,152]. Drug Affinity Responsive Target Stability (DARTS) is another major label-free method for target identification. When the ligand binds, the target protein will change its structure, making it resistant to proteolytic degradation. This technique is often used for target validation to demonstrate the direct interaction between small molecules and their target proteins without thermal denaturation^[153].

However, in TCM research, it is critical to interpret the results of label-free assays^[150,151]. In order to observe significant changes in thermal stability or resistance to proteolysis, methods such as CETSA and DARTS usually require incubation of cells with relatively high concentrations of TCM compounds (e.g., 10–100 $\mu\text{mol}\cdot\text{L}^{-1}$). These experimental doses may significantly exceed their true physiological steady-state concentrations *in vivo*. Therefore, confirming that these interactions are functionally essential for the pharmacological phenotype is the next step.

Cellular perturbation and mechanistic validation

Although chemical proteomics and label-free methods can confirm the physical binding of compounds to intracellular targets,

they cannot verify whether such binding events are functionally necessary to drive the pharmacological actions of compounds. To establish a clear causal relationship between target binding and phenotypic outcomes, cellular perturbation techniques at the transcriptional or genomic level must be utilized (Fig. 7c).

RNA interference (RNAi) technologies

RNA interference reagents, such as small interfering RNA (siRNA) and short hairpin RNA (shRNA), are widely used tools for post-transcriptional gene silencing, which enable researchers to rapidly test computational predictions at an early stage. For example, studies have shown that siRNA knockdown of AMPK can significantly inhibit berberine-induced apoptosis and autophagy in hepatocellular carcinoma cells. This suggests that the AMPK pathway may mediate an important part of the pharmacological response to berberine^[154]. However, RNAi reagents are not always effective enough to knock down target genes and may also exhibit sequence-dependent off-target effects. This problem is particularly prominent in TCM because TCM has the characteristics of being multi-component and multi-target, which may lead to compound off-target effects^[155]. Therefore, the data obtained from RNAi-based experiments should generally be regarded as preliminary evidence and need further rigorous verification.

CRISPR/Cas9-mediated gene knockout and screening

Unlike RNAi technology, CRISPR/Cas9-mediated gene editing can precisely, irreversibly, and stably disrupt gene function at the genomic level, providing a more reliable method for functional verification. Cell lines with complete gene knockout can be generated to evaluate the effect of gene deletion on the biological activity of compounds, so as to directly determine whether the target gene is necessary for drug activity^[155,156]. Genome-wide CRISPR screening has been used to systematically identify important regulatory nodes from a pool of targets identified by computational prediction or

omics analysis^[157]. Because most bioactive compounds in TCM have the characteristics of a synergistic multi-target network, knockout of a single predicted target gene may only partially attenuate the pharmacological phenotype, but cannot eliminate it^[94]. Therefore, researchers should comprehensively interpret CRISPR results and recognize that partial loss of efficacy still constitutes effective evidence for target validation of TCM, rather than rigidly adopting the typical 'all-or-nothing' validation criteria for single-target small-molecule drugs.

***In vivo* validation and safety assessment**

Although *in vitro* and cellular experiments provide strong evidence for target binding and mechanistic necessity, active compounds often fail to translate into *in vivo* efficacy due to off-target effects, complex tissue distribution, and poor pharmacokinetic (PK) properties. This makes *in vivo* research the final and indispensable checkpoint in the hierarchical validation framework. Animal models enable researchers to simultaneously assess pharmacokinetics, pharmacodynamics (PD), pharmacokinetic–pharmacodynamic (PK-PD) relationships, and disease progression, thus bridging the critical translational gap between computational predictions and true pharmacological value^[158,159].

For bioactive single compounds of TCM, the core goal of *in vivo* validation is to confirm the target dependence of drug efficacy (Fig. 7d). The validation of GLI1, the core target of berberine against colorectal cancer, is a typical case. Following AI-based target prediction and subsequent validation of target binding and cellular engagement, researchers used tumor-bearing mouse models to confirm that berberine exerts significant antitumor effects by inhibiting the GLI1/STAT3-ferroptosis negative regulator (FNR) pathway, inducing ferroptosis, and interfering with energy metabolism^[160]. The key is that the *in vivo* inhibition of this specific pathway verifies the target dependence of its efficacy^[161]. However, for complex TCM formulas, a single rigid target validation paradigm is not enough to meet the research needs. *In vivo* validation should focus on the compatibility and synergy between the overall target network predicted by AI and TCM formulas. For example, a multi-dimensional PK-PD model was used to verify the compatibility mechanism of Lianhua Qingwen (LHQW). The study provides *in vivo* system-level evidence for the AI-predicted multi-target network and the principle of TCM compatibility^[162].

In addition to validating the efficacy, *in vivo* studies are equally crucial for evaluating the safety and druggability of AI-predicted targets. Because most AI-predicted targets participate in a variety of physiological processes, long-term pharmacological regulation may produce unexpected biological effects. AI toxicology models can predict potential organ toxicity and off-target risk in advance, but these predictions still need to be systematically evaluated in animal models to comprehensively evaluate their tissue specificity and metabolic stability^[115]. This final *in vivo* verification link closes the translational loop from computational hypothesis to preclinical confirmation and serves as the critical pathway to translate AI-driven TCM target discovery into clinical application.

Challenges and strategies

Although the hierarchical experimental validation framework discussed provides a robust way to validate computational predictions, the fundamental reliability of AI-driven target discovery in TCM depends not only on downstream experimental validation but also on upstream parameters, including data quality, computational robustness, modeling bias, and generalization ability, as well as

interpretability, traceability, and reproducibility^[163,164]. Therefore, these considerations are not only technical barriers but also the basic requirements for the scientific credibility of AI-driven target research in TCM.

Data quality, heterogeneity, scarcity, and concept drift

The reliability and upper bound of prediction performance of any AI model is inherently constrained by the quality of its underlying training data^[139]. Inconsistent terminology, missing structural annotations, and serious data noise are common problems in TCM datasets^[138]. A major downstream consequence of the lack of a universal gold standard for target validation of TCM is that false-positive component–target interaction records continue to accumulate, thereby systematically biasing model training. In addition, there is a serious lack of validated negative interaction data in the research of TCM. Therefore, most AI models have to treat the unobserved interactions as negative samples, thus introducing significant systematic bias. The third persistent challenge is concept drift: the dynamic distribution of newly acquired real-world clinical and experimental data often deviates from the static distribution of the original training dataset, resulting in a gradual decline in the performance of the model in practical applications^[165].

In order to solve these data-related bottlenecks, a number of targeted strategies have been proposed. First, the establishment of a standardized TCM terminology ontology and a multi-center data-sharing alliance may help alleviate data heterogeneity, inconsistent labeling, and the lack of universal validation standards. Second, zero-shot learning and few-shot learning frameworks based on transfer learning (e.g., Therapy-related Graph Neural Network [TxGNN]) have successfully improved model generalization in new tasks with limited data^[166]. Third, matched negative sampling strategies based on TCM compound structural similarity and target functional classification have been developed to mitigate systematic bias caused by the lack of verified negative samples. Finally, to address concept drift, a continuous quality control and performance evaluation system should be established, with regular model retraining or updating and Bayesian uncertainty quantification to reduce the risks of dynamic data distribution shifts^[167].

Computational demands and the need for shared resources

Deep learning models for TCM target discovery often require substantial computational resources, relying heavily on graphics processing units (GPUs) and tensor processing units (TPUs), with considerable hardware and operating costs. The interdisciplinary nature of TCM requires substantial computational power: the multi-component and multi-target characteristics of TCM lead to an exponential growth of the potential component–target interaction space^[168], while target prediction requires the integration of cross-domain heterogeneous data, including chemical structures, protein targets, gene expression profiles, and TCM syndrome (Zheng) information. At present, there is a lack of open-source pre-trained molecular base models for TCM. This means that most research teams must either spend additional computational resources to build the model architecture from scratch or perform extensive fine-tuning on general chemical models to adapt to the unique multidimensional characteristics of TCM.

There are several practical strategies to address these computational bottlenecks. First, the establishment of an open, shared computing platform and cloud-based TCM-specific pre-trained models can significantly lower the hardware barrier for researchers

and avoid redundant model training and waste of computing resources. Second, the development of lightweight AI models for TCM through model pruning and knowledge distillation can reduce the computational requirements while maintaining the prediction performance. Third, the open-source collaborative framework for TCM target prediction can standardize the model training workflow, promote the sharing of pre-trained weights and validation datasets, and further reduce the redundant computational overhead.

Modeling bias, overfitting, and generalization ability

Robustness and generalization ability are the core methodological criteria of AI-driven target prediction. Due to data-related factors and model-related factors, there are often significant deviations between *in silico* predictions and experimental verification results. Overfitting and poor generalization ability are closely related to data bias and data leakage: if the molecular distribution of the training dataset overlaps with the test set too much, the performance of the model will be spuriously inflated^[25].

Beyond these general challenges, the unique challenge of TCM lies in its syndrome differentiation system: the same disease may correspond to different TCM syndrome types, and different syndrome types may reflect partially different molecular network states^[169]. If the model cannot correctly stratify and identify syndrome types, the efficacy of TCM formulas may be incorrectly attributed to unrelated target pathways, and the model bias may also be manifested in the so-called 'Clever Hans' phenomenon; that is, the model relies on spurious associations to obtain high prediction performance, but does not capture the real biological mechanism^[170]. To address these challenges, several targeted strategies have been proposed. First, entropy-based methods and Monte Carlo (MC) dropout have been widely used to estimate prediction uncertainty and improve the reliability of AI models^[171]. Second, in order to reduce data leakage and overfitting problems, structure-based dataset splitting strategies have been developed, which exclude training samples with high structural similarity to the test set to ensure the unbiased evaluation of the model^[25]. Finally, syndrome-stratified model training and multi-center external validation should be included in the model development workflow to ensure that the prediction results are generalizable across different TCM syndrome types and clinical scenarios.

Interpretability, traceability, and reproducibility

The development of interpretable, traceable, and reproducible AI models is the basis for bridging the gap between *in silico* hypotheses and scientific validity. Interpretability mainly aims at the black-box nature of deep learning: the opaque decision-making process of AI models hinders the biological interpretation of predicted targets and mechanisms^[172]. At present, there are two main methods to improve model interpretability: one is to design an inherently interpretable model architecture under the guidance of knowledge in the fields of TCM and biomedical knowledge; the other uses post hoc methods to generate visual and mechanistic interpretations. However, accurately linking abstract AI-derived mathematical features to specific physiological processes in TCM remains a major bottleneck^[137].

Traceability refers to the ability to trace the prediction results back to the original training data, processing steps, and model versions. For complex natural products in TCM, a traceable computational workflow can ensure that every step of target prediction is auditable, which is a prerequisite for subsequent experimental verification^[173]. To this end, Findable, Accessible, Interoperable, and

Reusable (FAIR) principles must be strictly applied to all training data, model codes, and processing protocols to ensure reliable data sharing and result verification^[137].

Reproducibility, the cornerstone of scientific inquiry, is facing a severe crisis in the AI era. To address this, AI-driven standardized reporting guidelines for TCM target prediction should be developed, requiring researchers to disclose complete model codes, training datasets, and detailed experimental protocols. At the same time, publishing large-scale, standardized and annotated TCM datasets can effectively reduce batch effects, ensure the consistency of model input, and support robust external validation of AI models^[137].

Ultimately, solving the challenges of interpretability, traceability and reproducibility in TCM is not only a technical necessity; it is a prerequisite for legitimately integrating AI-identified targets into both modern pharmacological research and TCM theoretical frameworks. This represents a key way for AI to bridge the gap between high-dimensional modern omics data and the holistic and multi-target therapeutic principles of TCM.

Validation gaps, experimental inconsistencies, and translational challenges

Although the above hierarchical experimental evidence framework provides a systematic method to reduce the high false-positive rate in AI-driven target prediction, it still has significant limitations in practical application^[174]. First, the current validation technology is largely derived from the research and development of the paradigm of conventional single-target drugs, and its applicability is still limited in fully analyzing the inherent multi-target synergy effect of complex TCM formulas^[94]. Second, for the whole chain validation analysis from *in vitro* biophysical binding to *in vivo* target-dependent confirmation, there is still a lack of standardized and reproducible experimental protocols between laboratories. This lack of standardization directly leads to poor reproducibility and inconsistent conclusions when AI-driven computational predictions are translated into reliable preclinical and clinical outcomes^[94].

To overcome these bottlenecks, future optimization of this framework should integrate cutting-edge biological models. For example, patient-derived organoids (PDOs) can bridge the translational gap between 2D cell cultures and whole-animal models^[175], provide a physiologically relevant human-based platform to simultaneously verify the multi-target synergistic effect of TCM formulas and standardize validation workflows^[176]. This will not only systematically reduce false positives in AI-driven TCM target prediction but also accelerate the modernization and clinical transformation of TCM.

Conclusions and future prospects

From the perspective of methodological integration and transformation of experimental validation, this review systematically summarizes the application pathways, advantages, and limitations of AI in TCM target discovery, as well as the research framework of synergistic mechanism analysis and target confirmation with experimental validation. Meanwhile, biochemistry, cell biology, chemical proteomics, and *in vivo* validation strategies are still indispensable for translating computational predictions into targets with biological significance and therapeutic value.

Despite these advances, there are still some major challenges in AI-assisted target discovery of TCM. There are still problems of heterogeneity and uneven quality in the underlying data ecosystem. There are differences in annotation standards, update frequency,

experimental evidence level, and the representation of natural products in different databases. Furthermore, extracting true causal targets from high-dimensional multi-omics associations is frequently hindered by confounding variables, temporal mismatches between static omics snapshots and dynamic physiological responses, and the fundamental difficulty of distinguishing correlation from causation. Many existing models still have problems such as overfitting, bias, limited generalization ability, and insufficient representation of chemical diversity and pharmacological complexity of CHM-derived compounds and prescriptions. Black box prediction is still a key bottleneck because its poor interpretability, lack of traceability, and reproducibility make it difficult to translate the computational predictions into experimentally verifiable and clinically meaningful hypotheses. In addition, the dynamics of protein conformation, the complexity of biological background, and the multi-target characteristics of TCM also continue to challenge the existing structure-based and network-based AI models.

Rigorous experimental validation therefore remains indispensable for confirming AI-derived predictions. Current experimental strategies can assess the binding reliability, biological activity, and physiological relevance of candidate targets at multiple levels, thereby providing critical support for computational results and reducing false-positive discoveries. Nevertheless, existing validation technologies still possess methodological limitations, highlighting the need for more robust, scalable, and standardized experimental frameworks. AI prediction and experimental validation can be more closely integrated through iterative feedback mechanisms that continuously improve model training and optimization. Such collaborative workflows may improve the reliability and translational potential of AI-assisted target discovery.

Future progress may depend on several converging directions. A high-quality TCM database is one of the most fundamental requirements, and it is important to build a higher-quality, standardized, and interoperable special dataset for TCM. At the same time, generative AI and foundation models can further broaden the scope of target discovery by improving knowledge extraction, hypothesis generation, candidate drug prioritization, and even targeted molecular design. However, before these models can be reliably applied to research or clinical practice, they need to control their possible hallucinations more strictly, enhance their domain adaptability, and conduct more rigorous benchmarking.

In addition, experimental validation may become more tightly integrated with AI by incorporating iterative feedback to improve model training and establishing a collaborative optimization strategy between AI and experiments. Ultimately, combining AI predictions with rigorous experimental validation and a robust, standardized data infrastructure will accelerate the translational impact of TCM. This integration will bridge the gap between historical empirical practice and modern clinical applications, providing clear, actionable insights for personalized and precision medicine.

Ethical statements

Not applicable.

Author contributions

The authors confirm contributions to the work as follows: conception and design: Zhang W, Tian S; draft manuscript preparation: Guo C, Li Q, analysis and interpretation: Liang W, Lu J, Yue W. All authors reviewed and approved the final version of the manuscript.

Data availability

Data sharing is not applicable to this review as no datasets were generated or analyzed.

Acknowledgments

This research was funded by the Innovative Drug Research and Development-National Science and Technology Major Project (2025ZD1800300), the National Natural Science Foundation of China (82574713, 82430119), the Computational Biology Program (25JS2830400) of Science and Technology Commission of Shanghai Municipality (STCSM), the Shanghai Municipal Science and Technology Major Project (ZD2021CY001), the Chenguang Program of Shanghai Education Development Foundation, and Shanghai Municipal Education Commission (23CGA45, Saisai Tian), the Natural Science Foundation of Shanghai (25ZR1402574), the ability establishment of sustainable use for valuable Chinese medicine resources (2060302), and the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2023-I2M-3-009).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: <https://doi.org/10.48130/targetome-0026-0027>.

Dates

Received 26 April 2026; Revised 29 May 2026; Accepted 3 June 2026; Published online 30 June 2026

References

- [1] Guo DA, Lu A, Liu L. 2012. Modernization of traditional Chinese medicine. *Journal of Ethnopharmacology* 141:547–548
- [2] Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *The Journal of Natural Products* 83:770–803
- [3] Saldivar-González FI, Aldas-Bulos VD, Medina-Franco JL, Plisson F. 2022. Natural product drug discovery in the artificial intelligence era. *Chemical Science* 13:1526–1546
- [4] Banihashemi ZS, Azizi-Fini I, Rajabi M, Maghami M, Yadollahi S. 2025. Chronic fatigue syndrome post-COVID-19: triple-blind randomised clinical trial of Astragalus root extract. *BMJ Support Palliat Care* 15:359–366
- [5] Zhang R, Zhu X, Bai H, Ning K. 2019. Network pharmacology databases for traditional Chinese medicine: review and assessment. *Frontiers in Pharmacology* 10:123
- [6] Wang Y, Liu B, Fu X, Tong T, Yu Z. 2021. Efficacy and safety of Si-Jun-Zi-Tang-based therapies for functional (non-ulcer) dyspepsia: a meta-analysis of randomized controlled trials. *BMC Complementary Medicine and Therapies* 21:11
- [7] Wang X, Shi X, Xi Z, Zhang Z, Luo Z, et al. 2026. The scientific basis of synergy in traditional Chinese medicine: physicochemical, pharmacokinetic, and pharmacodynamic perspectives. *Chinese Medicine* 21:15
- [8] Liu J, Wei LX, Wang Q, Lu YF, Zhang F, et al. 2018. A review of cinnabar (HgS) and/or realgar (As₄S₄)-containing traditional medicines. *Journal of Ethnopharmacology* 210:340–350
- [9] Huang C, Zhu Z, Cao X, Chen X, Fu Y, et al. 2017. A pectic polysaccharide from Sijunzi decoction promotes the antioxidant defenses of SW480 cells. *Molecules* 22:1341
- [10] Xu B, Li X, Hu S, Bao Y, Chen F, et al. 2022. Safety and efficacy of Yupingfeng granules in children with recurrent respiratory tract infection: a randomized clinical trial. *Pediatric Investigation* 6:75–84

- [11] Song T, Hou X, Yu X, Wang Z, Wang R, et al. 2016. Adjuvant treatment with yupingfeng formula for recurrent respiratory tract infections in children: a meta-analysis of randomized controlled trials. *Phytotherapy Research* 30:1095–103
- [12] Xin S, Cheng X, Zhu B, Liao X, Yang F, et al. 2020. Clinical retrospective study on the efficacy of Qingfei Paidu decoction combined with Western medicine for COVID-19 treatment. *Biomedicine & Pharmacotherapy* 129:110500
- [13] Liu S, Yao C, Xie J, Liu H, Wang H, et al. 2023. Effect of an herbal-based injection on 28-day mortality in patients with sepsis: the EXIT-SEP randomized clinical trial. *JAMA Internal Medicine* 183:647–655
- [14] Li X, Zhang J, Huang J, Ma A, Yang J, et al. 2013. A multicenter, randomized, double-blind, parallel-group, placebo-controlled study of the effects of qili qiangxin capsules in patients with chronic heart failure. *Journal of the American College of Cardiology* 62:1065–1072
- [15] Li S, Zhang B. 2013. Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chinese Journal of Natural Medicines* 11:110–20
- [16] Qu J, Zhang T, Liu J, Su Y, Wang H. 2019. Considerations for the quality control of newly registered traditional Chinese medicine in China: a review. *Journal of AOAC International* 102:689–694
- [17] Newman DJ. 2019. The impact of decreasing biodiversity on novel drug discovery: is there a serious cause for concern? *Expert Opinion on Drug Discovery* 14:521–525
- [18] Ren YS, Li HL, Piao XH, Yang ZY, Wang SM, et al. 2021. Drug affinity responsive target stability (DARTS) accelerated small molecules target discovery: principles and application. *Biochemical Pharmacology* 194:114798
- [19] Xu Y, Liu X, Cao X, Huang C, Liu E, et al. 2021. Artificial intelligence: a powerful paradigm for scientific research. *Innovation* 2:100179
- [20] Kaul V, Enslin S, Gross SA. 2020. History of artificial intelligence in medicine. *Gastrointestinal Endoscopy* 92:807–812
- [21] Alakhdar A, Poczos B, Washburn N. 2024. Diffusion models in *de novo* drug design. *Journal of Chemical Information and Modeling* 64:7238–7256
- [22] Jiang J, Chen L, Ke L, Dou B, Zhang C, et al. 2025. A review of transformer models in drug discovery and beyond. *Journal of Pharmaceutical Analysis* 15:101081
- [23] He X, Sun M, Battulga T, Copp BR, Rustamovna DK, et al. 2025. Application of artificial intelligence in the development of traditional Chinese medicine. *Basic & Clinical Pharmacology and Toxicology* 137:e70066
- [24] Chen W, Liu X, Zhang S, Chen S. 2023. Artificial intelligence for drug discovery: resources, methods, and applications. *Molecular Therapy Nucleic Acids* 31:691–702
- [25] Graber D, Stockinger P, Meyer F, Mishra S, Horn C, et al. 2025. Resolving data bias improves generalization in binding affinity prediction. *Nature Machine Intelligence* 7:1713–1725
- [26] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, et al. 2021. Artificial intelligence in drug discovery and development. *Drug Discovery Today* 26:80–93
- [27] Noor F, Asif M, Ashfaq UA, Qasim M, Tahir ul Qamar M. 2023. Machine learning for synergistic network pharmacology: a comprehensive overview. *Briefings in Bioinformatics* 24:bbad120
- [28] Liu P, Huang F, Zheng X, Hao H. 2025. Targetome-guided combination drug discovery as next-generation therapeutics. *Targetome* 1:e002
- [29] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, et al. 2023. Artificial intelligence for natural product drug discovery. *Nature Reviews Drug Discovery* 22:895–916
- [30] Li X, Li X, Wang L, Hou Y, Liu Y, et al. 2025. Advancing traditional Chinese medicine research through network pharmacology: strategies for target identification, mechanism elucidation and innovative therapeutic applications. *The American Journal of Chinese Medicine* 53:2021–2042
- [31] Tian S, Zhang J, Yuan S, Wang Q, Lv C, et al. 2023. Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. *Briefings in Bioinformatics* 24:bbad027
- [32] Yan D, Zheng G, Wang C, Chen Z, Mao T, et al. 2022. HIT 2.0: an enhanced platform for Herbal Ingredients' Targets. *Nucleic Acids Research* 50:D1238–D1243
- [33] Knox C, Wilson M, Klinger CM, Franklin M, Oler E, et al. 2024. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research* 52:D1265–D1275
- [34] Liu Z, Cai C, Du J, Liu B, Cui L, et al. 2020. TCMIO: a comprehensive database of traditional Chinese medicine on immuno-oncology. *Frontiers in Pharmacology* 11:439
- [35] Zhang Y, Li X, Shi Y, Chen T, Xu Z, et al. 2023. ETCM v2.0: an update with comprehensive resource and rich annotations for traditional Chinese medicine. *Acta Pharmaceutica Sinica B* 13:2559–2571
- [36] Fang S, Dong L, Liu L, Guo J, Zhao L, et al. 2021. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Research* 49:D1197–D1206
- [37] Chen Q, Springer L, Gohlke BO, Goede A, Dunkel M, et al. 2021. SuperTCM: a biocultural database combining biological pathways and historical linguistic data of Chinese Materia Medica for drug development. *Biomedicine & Pharmacotherapy* 144:112315
- [38] Kim S, Chen J, Cheng T, Gindulyte A, He J, et al. 2019. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research* 47:D1102–D1109
- [39] Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, et al. 2024. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* 52:D1180–D1192
- [40] Liu T, Hwang L, Burley SK, Nitsche CI, Southan C, et al. 2025. BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data. *Nucleic Acids Research* 53:D1633–D1644
- [41] Avram S, Wilson TB, Curpan R, Halip L, Borota A, et al. 2023. DrugCentral 2023 extends human clinical data and integrates veterinary drugs. *Nucleic Acids Research* 51:D1276–D1287
- [42] Siramshetty VB, Eckert OA, Gohlke BO, Goede A, Chen Q, et al. 2018. SuperDRUG2: a one stop resource for approved/marketed drugs. *Nucleic Acids Research* 46:D1137–D1143
- [43] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, et al. 2016. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* 54:1.30.1–1.30.33
- [44] Zhou Y, Zhang Y, Zhao D, Yu X, Shen X, et al. 2024. TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Research* 52:D1465–D1477
- [45] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, et al. 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 48:D845–D855
- [46] Lv Q, Chen G, He H, Yang Z, Zhao L, et al. 2023. TCMBank: bridges between the largest herbal medicines, chemical ingredients, target proteins, and associated diseases with intelligence text mining. *Chemical Science* 14:10684–10701
- [47] Gao K, Liu L, Lei S, Li Z, Huo P, et al. 2025. HERB 2.0: an updated database integrating clinical and experimental evidence for traditional Chinese medicine. *Nucleic Acids Research* 53:D1404–D1414
- [48] Kim SK, Lee MK, Jang H, Lee JJ, Lee S, et al. 2024. TM-MC 2.0: an enhanced chemical database of medicinal materials in Northeast Asian traditional medicine. *BMC Complementary Medicine and Therapies* 24:40
- [49] Yang P, Lang J, Li H, Lu J, Lin H, et al. 2022. TCM-Suite: a comprehensive and holistic platform for Traditional Chinese Medicine component identification and network pharmacology analysis. *iMeta* 1:e47
- [50] Liu X, Liu J, Fu B, Chen R, Jiang J, et al. 2023. DCABM-TCM: a database of constituents absorbed into the blood and metabolites of traditional Chinese medicine. *Journal of Chemical Information and Modeling* 63:4948–4959
- [51] Zhang LX, Dong J, Wei H, Shi SH, Lu AP, et al. 2022. TCMsID: a simplified integrated database for drug discovery from traditional Chinese medicine. *Journal of Cheminformatics* 14:89
- [52] Wu Y, Zhang F, Yang K, Fang S, Bu D, et al. 2019. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Research* 47:D1110–D1117

- [53] Huang L, Xie D, Yu Y, Liu H, Shi Y, et al. 2018. TCMID 2. 0: a comprehensive resource for TCM. *Nucleic Acids Research* 46:D1117–D1120
- [54] Li B, Ma C, Zhao X, Hu Z, Du T, et al. 2018. YaTCM: yet another traditional Chinese medicine database for drug discovery. *Computational and Structural Biotechnology Journal* 16:600–610
- [55] Ru J, Li P, Wang J, Zhou W, Li B, et al. 2014. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *Journal of Cheminformatics* 6:13
- [56] Yang D, Zhu Z, Yao Q, Chen C, Chen F, et al. 2023. ccTCM: a quantitative component and compound platform for promoting the research of traditional Chinese medicine. *Computational and Structural Biotechnology Journal* 21:5807–5817
- [57] Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, et al. 2009. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research* 37:W623–W633
- [58] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, et al. 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40:D1100–D1107
- [59] Sun X, Wei Z, Yu X, Liu K, Yao S, et al. 2026. DRESIS 2.0: the comprehensive landscape of drug resistance information. *Nucleic Acids Research* 54:D1387–D1396
- [60] Li F, Mou M, Li X, Xu W, Yin J, et al. 2025. DrugMAP 2.0: molecular atlas and pharma-information of all drugs. *Nucleic Acids Research* 53:D1372–D1382
- [61] Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. 1997. GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics* 13:163
- [62] Kong X, Liu C, Zhang Z, Cheng M, Mei Z, et al. 2024. BATMAN-TCM 2.0: an enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. *Nucleic Acids Research* 52:D1110–D1120
- [63] Lin H, Hou D, Jiang X, Yin R, Yu S, et al. 2026. NPASS database update 2026: comprehensive quantitative composition, bioactivity, and ADME-Tox data of natural products for biomedical research. *Nucleic Acids Research* 54:D1519–d27
- [64] Lv Q, Chen G, He H, Yang Z, Zhao L, et al. 2023. TCMBank-the largest TCM database provides deep learning-based Chinese-Western medicine exclusion prediction. *Signal Transduction and Targeted Therapy* 8:127
- [65] Willett P. 2006. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* 11:1046–53
- [66] Rogers D, Hahn M. 2010. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling* 50:742–754
- [67] Duvenaud DK, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, et al. 2015. Convolutional networks on graphs for learning molecular fingerprints. *NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems, December 7–12, 2015, Montreal, Canada*. Vol. 2. Cambridge, United States: MIT Press. pp. 2224–2232 <https://dl.acm.org/doi/proceedings/10.5555/2969442>
- [68] Yang K, Swanson K, Jin W, Coley C, Eiden P, et al. 2019. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling* 59:3370–3388
- [69] Del Prete E, Facchiano A, Liò P. 2020. Bioinformatics methodologies for coeliac disease and its comorbidities. *Briefings in Bioinformatics* 21:355–367
- [70] Lin Y, Zhang Y, Wang D, Yang B, Shen YQ. 2022. Computer especially AI-assisted drug virtual screening and design in traditional Chinese medicine. *Phytomedicine* 107:154481
- [71] Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. 2015. Deep neural nets as a method for quantitative structure-activity relationships. *Journal of Chemical Information and Modeling* 55:263–74
- [72] Arora B, Coudrat T, Wootten D, Christopoulos A, Noronha SB, Sexton PM. 2016. Prediction of loops in G protein-coupled receptor homology models: effect of imprecise surroundings and constraints. *Journal of Chemical Information and Modeling* 56:671–686
- [73] Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, et al. 2018. MoleculeNet: a benchmark for molecular machine learning. *Chemical Science* 9:513–530
- [74] Zhang S, Huo D, Horne RI, Qi Y, Pujalte Ojeda S, et al. 2025. Sequence-based virtual screening using transformers. *Nature Communications* 16:6925
- [75] López-Pérez K, López-López E, Medina-Franco JL, Miranda-Quintana RA. 2023. Sampling and mapping chemical space with extended similarity indices. *Molecules* 28(17):28
- [76] Wu H, Liu J, Zhang R, Lu Y, Cui G, et al. 2024. A review of deep learning methods for ligand based drug virtual screening. *Fundamental Research* 4:715–737
- [77] Koirala M, Yan L, Mohamed Z, DiPaola M. 2025. AI-integrated QSAR modeling for enhanced drug discovery: from classical approaches to deep learning and structural insight. *International Journal of Molecular Sciences* 26:9384
- [78] Kattuparambil AA, Chaurasia DK, Shekhar S, Srinivasan A, Mondal S, et al. 2025. Exploring chemical space for "druglike" small molecules in the age of AI. *Frontiers in Molecular Biosciences* 12:1553667
- [79] Xie S, Zhu H, Huang N. 2025. AI-designed molecules in drug discovery, structural novelty evaluation, and implications. *Journal of Chemical Information and Modeling* 65:8924–8933
- [80] Pathan I, Raza A, Sahu A, Joshi M, Sahu Y, et al. 2025. Revolutionizing pharmacology: AI-powered approaches in molecular modeling and ADMET prediction. *Medicine in Drug Discovery* 28:100223
- [81] Niazi SK. 2025. Artificial Intelligence in Small-molecule drug discovery: a critical review of methods, applications, and real-world outcomes. *Pharmaceuticals* 18:1271
- [82] Sim J, Kim D, Kim B, Choi J, Lee J. 2025. Recent advances in AI-driven protein-ligand interaction predictions. *Current Opinion in Structural Biology* 92:103020
- [83] Zhou G, Rusnac DV, Park H, Canzani D, Nguyen HM, et al. 2024. An artificial intelligence accelerated virtual screening platform for drug discovery. *Nature Communications* 15:7761
- [84] Rashid MA, Radhika NM, Yogeeta OA, Sameer NG, Kartik TN, et al. 2025. Recent advances in molecular docking techniques: transforming perspectives in distinct drug targeting and drug discovery approaches. *Medicinal Chemistry* 00:in press
- [85] Zhang Z, Quan L, Wang J, Peng L, Chen Q, et al. 2025. LABind: identifying protein binding ligand-aware sites via learning interactions between ligand and protein. *Nature Communications* 16:7712
- [86] Yang Z, Ji J, He S, Li J, He T, et al. 2024. Dockformer: A transformer-based molecular docking paradigm for large-scale virtual screening. *arXiv* 00:2411.06740
- [87] Sharma G, Kumar N, Sharma CS, Alqahtani T, Tiruneh YK, et al. 2025. Identification of promising SARS-CoV-2 main protease inhibitor through molecular docking, dynamics simulation, and ADMET analysis. *Scientific Reports* 15:2830
- [88] Hasan GM, Mohammad T, Zaidi S, Shamsi A, Hassan MI. 2025. Molecular docking and dynamics in protein serine/threonine kinase drug discovery: advances, challenges, and future perspectives. *Frontiers in Pharmacology* 16:1696204
- [89] Wang Y, Li Y, Chen J, Lai L. 2025. Modeling protein-ligand interactions for drug discovery in the era of deep learning. *Chemical Society Reviews* 54:11141–11183
- [90] Wang S, Zhang L, Zhang W, Zeng X, Mei J, et al. 2025. Structure-based molecular screening and dynamic simulation of phytocompounds targeting VEGFR-2: a novel therapeutic approach for papillary thyroid carcinoma. *Frontiers in Pharmacology* 16:1583329
- [91] Li D, Hu J, Zhang L, Li L, Yin Q, et al. 2022. Deep learning and machine intelligence: new computational modeling techniques for discovery of the combination rules and pharmacodynamic characteristics of Traditional Chinese Medicine. *European Journal of Pharmacology* 933:175260
- [92] Abramson J, Adler J, Dunger J, Evans R, Green T, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500
- [93] Takaba K, Friedman AJ, Cavender CE, Behara PK, Pulido I, et al. 2024. Machine-learned molecular mechanics force fields from large-scale quantum chemical data. *Chemical Science* 15:12861–12878
- [94] Hopkins AL. 2008. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology* 4:682–690

- [95] Cui G, Li M, Guo W, Gao M, Zhu Q, et al. 2025. AI driven network pharmacology: multi-scale mechanisms of traditional Chinese medicine from molecular to patient analysis. *Computational and Structural Biotechnology Journal* 27:5087–5104
- [96] Jiang W, Ye W, Tan X, Bao YJ. 2025. Network-based multi-omics integrative analysis methods in drug discovery: a systematic review. *BioData Mining* 18:27
- [97] Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. 2021. Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal* 19:3735–3746
- [98] Wei S, Sasi C, Piepenbrock J, Huynen MA, t Hoen PAC. 2025. The use of knowledge graphs for drug repurposing: from classical machine learning algorithms to graph neural networks. *Computers in Biology and Medicine* 196:110873
- [99] Ying H, Kong W, Xu X. 2025. Integrated network pharmacology, machine learning and experimental validation to identify the key targets and compounds of TiaoShenGongJian for the treatment of breast cancer. *OncoTargets and Therapy* 18:49–71
- [100] Ding C, Liao Q, Zuo R, Zhang S, Guo Z, et al. 2024. Machine learning potential predictor of idiopathic pulmonary fibrosis. *Frontiers in Genetics* 15:1464471
- [101] Zhao H, He L, Yin D, Song B. 2019. Identification of β -catenin target genes in colorectal cancer by interrogating gene fitness screening data. *Oncology Letters* 18:3769–3777
- [102] LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436–444
- [103] Chen H, King FJ, Zhou B, Wang Y, Canedy CJ, et al. 2024. Drug target prediction through deep learning functional representation of gene signatures. *Nature Communications* 15:1853
- [104] Qu Z, Wang W, Hou C, Hou C. 2019. Radar signal intra-pulse modulation recognition based on convolutional denoising autoencoder and deep convolutional neural network. *IEEE Access* 7:112339–11234
- [105] Zhang S, Zhang X, Du J, Wang W, Pi X. 2024. Multi-target meridians classification based on the topological structure of anti-cancer phytochemicals using deep learning. *Journal of Ethnopharmacology* 319:117244
- [106] Zhang Z, Chen L, Zhong F, Wang D, Jiang J, et al. 2022. Graph neural network approaches for drug-target interactions. *Current Opinion in Structural Biology* 73:102327
- [107] Guo X, Zhao X, Lu X, Zhao L, Zeng Q, et al. 2024. A deep learning-driven discovery of berberine derivatives as novel antibacterial against multidrug-resistant *Helicobacter pylori*. *Signal Transduction and Targeted Therapy* 9:183
- [108] Duan YJ, Fu L, Zhang XC, Long TZ, He YH, et al. 2023. Improved GNNs for log $D(7.4)$ prediction by transferring knowledge from low-fidelity data. *Journal of Chemical Information and Modeling* 63:2345–2359
- [109] Kimber TB, Chen Y, Volkamer A. 2021. Deep learning in virtual screening: recent applications and developments. *International Journal of Molecular Sciences* 22:4435
- [110] Zakharov AV, Peach ML, Sitzmann M, Nicklaus MC. 2014. QSAR modeling of imbalanced high-throughput screening data in PubChem. *Journal of Chemical Information and Modeling* 54:705–12
- [111] Huang S, Zhang L, Liu X. 2025. Bioinformatics approach to identifying molecular targets of isoliquiritigenin affecting chronic obstructive pulmonary disease: a machine learning pharmacology study. *International Journal of Molecular Sciences* 26(8):3907
- [112] Lu X, Xie L, Xu L, Mao R, Xu X, et al. 2024. Multimodal fused deep learning for drug property prediction: integrating chemical language and molecular graph. *Computational and Structural Biotechnology Journal* 23:1666–1679
- [113] MacLean F. 2021. Knowledge graphs and their applications in drug discovery. *Expert Opinion on Drug Discovery* 16:1057–1069
- [114] Yu Z, Li T, Zheng Z, Yang X, Guo X, et al. 2025. Tailoring a traditional Chinese medicine prescription for complex diseases: a novel multi-targets-directed gradient weighting strategy. *Journal of Pharmaceutical Analysis* 15:101199
- [115] Zhang J, Li H, Zhang Y, Huang J, Ren L, et al. 2025. Computational toxicology in drug discovery: applications of artificial intelligence in ADMET and toxicity prediction. *Briefings in Bioinformatics* 26:bbaf533
- [116] Wu C, Wu C, Peng L, Wu M, Li Z, et al. 2024. Multi-omics approaches for the understanding of therapeutic mechanism for Huang-Qi-Long-Dan Granule against ischemic stroke. *Pharmacological Research* 205:107229
- [117] Khanal S, Kumar A, Kumar P, Thakur P, Chander AM, et al. 2025. Unraveling bioactive potential and production in *Ganoderma lucidum* through omics and machine learning modeling. *Chinese Herbal Medicines* 17:414–427
- [118] Mohammadi-Shemirani P, Sood T, Paré G. 2023. From 'omics to multi-omics technologies: the discovery of novel causal mediators. *Current Atherosclerosis Reports* 25:55–65
- [119] Chen LG, Tubbs JD, Liu Z, Thach TQ, Sham PC. 2024. Mendelian randomization: causal inference leveraging genetic data. *Psychological Medicine* 54:1461–74
- [120] Li Y, Miao Y, Feng Q, Zhu W, Chen Y, et al. 2024. Mitochondrial dysfunction and onset of type 2 diabetes along with its complications: a multi-omics Mendelian randomization and colocalization study. *Frontiers in Endocrinology* 15:1401531
- [121] Huang L, Liu J, Zheng X, Zhang K, Chen Y, et al. 2026. Integrated multi-omics mapping of the causal landscape of gout across the circulating-tissue axis. *Frontiers in Immunology* 17:1776456
- [122] Garcia-Argibay M, Wootton RE, Larsson H, Mattheisen M, Polimanti R, et al. 2026. Causal inference in psychiatric research: how to critically evaluate and interpret mendelian randomization studies. *Molecular Psychiatry* 31:3533–3543
- [123] Yazdani A, Yazdani A, Mendez-Giraldez R, Samiei A, Kosorok MR, et al. 2022. From classical mendelian randomization to causal networks for systematic integration of multi-omics. *Frontiers in Genetics* 13:990486
- [124] Yang J, Yang H, Wang F, Dai Y, Deng Y, et al. 2025. Bioinformatics identification based on causal association inference using multi-omics reveals the underlying mechanism of Gui-Zhi-Shao-Yao-Zhi-Mu decoction in modulating rheumatoid arthritis. *Phytomedicine* 136:156332
- [125] Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, et al. 2021. Toward causal representation learning. *Proceedings of the IEEE* 109:612–634
- [126] Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167:1853–1866.e17
- [127] Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020. Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights* 14:1177932219899051
- [128] Hasin Y, Seldin M, Lusis A. 2017. Multi-omics approaches to disease. *Genome Biology* 18:83
- [129] Shataer D, Cao S, Liu X, Aierken K, Bhattacharya P, et al. 2025. Application of large language models in traditional Chinese medicine: a state-of-the-art review. *American Journal of Chinese Medicine* 53:973–997
- [130] Guo P, Jiang M, Hu S, Jiang Q, Li L, et al. 2026. Advancing the modernization of traditional Chinese medicine through artificial intelligence and multimodal data integration. *Chinese Medicine* 21:54
- [131] Hogan A, Blomqvist E, Cochez M, D'amato C, De Melo G, et al. 2021. Knowledge graphs. *arXiv* 00:2003.02320
- [132] Han C, Yang G, Li H, Zhu L, Feng M. 2026. Tuning and clinical application of large language models in Traditional Chinese Medicine: scoping review. *Chinese Medicine* 21:71
- [133] Wang L, Tang K, Wang Y, Zhang P, Li S. 2025. Advancements in Artificial intelligence-driven diagnostic models for traditional Chinese medicine. *American Journal of Chinese Medicine* 53:647–673
- [134] Dai Y, Shao X, Zhang J, Chen Y, Chen Q, et al. 2024. TCMChat: A generative large language model for traditional Chinese medicine. *Pharmacological Research* 210:107530
- [135] Liu Z, Yang T, Wang J, Chen Y, Gao Z, et al. 2026. Tianyi: a traditional Chinese medicine all-rounder language model and its real-world clinical practice. *Information Fusion* 126:103663
- [136] Chen Y, Wang Z, Xing X, Zheng H, Xu Z, et al. 2023. BianQue: balancing the questioning and suggestion ability of health LLMs with multi-turn health conversations polished by ChatGPT. *arXiv* 00:2310.15896

- [137] Zhou W, Gao Y. 2026. Phenotype-target coupled drug screening: a high-efficiency framework for innovative drug discovery from CHMs. *Engineering* 57:10–13
- [138] Li Y, Liu X, Zhou J, Li F, Wang Y, et al. 2025. Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling. *Frontiers in Pharmacology* 16:1541509
- [139] Bittner MI, Farajnia S. 2022. AI in drug discovery: applications, opportunities, and challenges. *Patterns* 3:100529
- [140] Bastos M, Abian O, Johnson CM, Ferreira-da-Silva F, Vega S, et al. 2023. Isothermal titration calorimetry. *Nature Reviews Methods Primers* 3:17
- [141] Li L, Zhang J, Li Y, Huang C, Xu J, et al. 2024. Dielectric surface-based biosensors for enhanced detection of biomolecular interactions: advances and applications. *Biosensors* 14:524
- [142] Hong M, Du Y, Chen D, Shi Y, Hu M, et al. 2023. Martynoside rescues 5-fluorouracil-impaired ribosome biogenesis by stabilizing RPL27A. *Science Bulletin* 68:1662–1677
- [143] Baell JB, Holloway GA. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry* 53:2719–2740
- [144] McGovern SL, Caselli E, Grigorieff N, Shoichet BK. 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medicinal Chemistry* 45:1712–1722
- [145] Leelananda SP, Lindert S. 2016. Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry* 12:2694–2718
- [146] Chen X, Wang Y, Ma N, Tian J, Shao Y, et al. 2020. Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification. *Signal Transduction and Targeted Therapy* 5:72
- [147] Wang S, Zhang Y, Yu R, Chai Y, Liu R, et al. 2024. Labeled and label-free target identifications of natural products. *Journal of Medicinal Chemistry* 67:17980–17996
- [148] Wang Q, Du T, Zhang Z, Zhang Q, Zhang J, et al. 2024. Target fishing and mechanistic insights of the natural anticancer drug candidate chlorogenic acid. *Acta Pharmaceutica Sinica B* 14:4431–4442
- [149] Parker CG, Pratt MR. 2020. Click chemistry in proteomic investigations. *Cell* 180:605–632
- [150] Schenone M, Dančák V, Wagner BK, Clemons PA. 2013. Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* 9:232–240
- [151] Martinez Molina D, Jafari R, Ignatushchenko M, Seki T, Larsson EA, et al. 2013. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341:84–87
- [152] Savitski MM, Reinhard FB, Franken H, Werner T, Savitski MF, et al. 2014. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346:1255784
- [153] Lomenick B, Hao R, Jonai N, Chin RM, Aghajan M, et al. 2009. Target identification using drug affinity responsive target stability (DARTS). *Proceedings of the National Academy of Sciences of the United States of America* 106:21984–21989
- [154] Yu R, Zhang ZQ, Wang B, Jiang HX, Cheng L, et al. 2014. Berberine-induced apoptotic and autophagic death of HepG2 cells requires AMPK activation. *Cancer Cell International* 14:49
- [155] Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, et al. 2019. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science Translational Medicine* 11:eaaw8412
- [156] Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84–87
- [157] Tian S, Qin Y, Wu Y, Dong M. 2025. Design, performance, processing, and validation of a pooled CRISPR perturbation screen for bacterial toxins. *Nature Protocols* 20:1158–1195
- [158] Jansson-Löfmark R, Fridén M, Badolo L, Ahlström C, Gurrell I, et al. 2025. Translational PK/PD: a retrospective analysis of performance and impact from a drug portfolio. *Drug Discovery Today* 30:104417
- [159] Haid RTU, Reichel A. 2025. PK/PD modeling of targeted protein degraders: charting new waters and navigating the shallows. *Drug Discovery Today* 30:104311
- [160] Sun Q, Tu K, Xu Q, Yan L, Yang S, et al. 2025. Berberine suppresses colorectal cancer progression by inducing ferroptosis-mediated energy metabolism disorders. *Journal of Advanced Research* 00:In Press
- [161] Sweis RF. 2015. Target (In)Validation: a critical, sometimes unheralded, role of modern medicinal chemistry. *ACS Medicinal Chemistry Letters* 6:618–621
- [162] Qi H, Chen S, Chen X, Ma Q, Shi J, et al. 2025. Multi-dimensional PK-PD insights into Lianhua Qingwen's formula compatibility. *Journal of Ethnopharmacology* 353:120251
- [163] Topol EJ. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* 25:44–56
- [164] Waman VP, Sen N, Varadi M, Daina A, Wodak SJ, et al. 2021. The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies. *Briefings in Bioinformatics* 22:742–68
- [165] Patchipala S. 2023. Tackling data and model drift in AI: strategies for maintaining accuracy during ML model inference. *International Journal of Science and Research Archive* 10:1198–1209
- [166] Huang K, Chandak P, Wang Q, Havaladar S, Vaid A, et al. 2024. A foundation model for clinician-centered drug repurposing. *Nature Medicine* 30:3601–3613
- [167] Guo XX, An S, Bao F, Xu TR. 2023. Challenges and perspectives in target identification and mechanism illustration for Chinese medicine. *Chinese Journal of Integrative Medicine* 29:644–654
- [168] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589
- [169] Zhang P, Zhang D, Zhou W, Wang L, Wang B, et al. 2023. Network pharmacology: towards the artificial intelligence-based precision traditional Chinese medicine. *Briefings in Bioinformatics* 25:bbad518
- [170] Kauffmann J, Dippel J, Ruff L, Samek W, Müller KR, et al. 2025. Explainable AI reveals Clever Hans effects in unsupervised learning models. *Nature Machine Intelligence* 7:412–422
- [171] Tong X, Wang D, Ding X, Tan X, Ren Q, et al. 2022. Blood-brain barrier penetration prediction enhanced by uncertainty estimation. *Journal of Cheminformatics* 14:44
- [172] Vatansever S, Schlessinger A, Wacker D, Kaniskan H, Jin J, et al. 2021. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: state-of-the-arts and future directions. *Medicinal Research Reviews* 41:1427–1473
- [173] Xu H, Zhao H, Ding C, Jiang D, Zhao Z, et al. 2023. Celastrol suppresses colorectal cancer via covalent targeting peroxiredoxin 1. *Signal Transduction and Targeted Therapy* 8:51
- [174] Begley CG, Ellis LM. 2012. Drug development: raise standards for preclinical cancer research. *Nature* 483:531–533
- [175] Clevers H. 2016. Modeling development and disease with organoids. *Cell* 165:1586–1597
- [176] Vlachogiannis G, Hedayat S, Vatsiou A, Jamin Y, Fernández-Mateos J, et al. 2018. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 359:920–926



Copyright: © 2026 by the author(s). Published by Maximum Academic Press on behalf of China Pharmaceutical University. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.