Targetome

INVITED REVIEW

Open Access

https://doi.org/10.48130/targetome-0026-0036

Targetome **2026**, 2(4): e037

From target fishing to AI and single-cell multi-omics: an integrated framework for natural product target research

Haojie Du[#], Nana Chen[#] and Yang Sun[*]

State Key Laboratory of Pharmaceutical Biotechnology and Nanjing Drum Tower Hospital, School of Life Sciences, Chemistry and Biomedicine Innovation Center, Nanjing University, Nanjing, Jiangsu 210023, China

[#] Authors contributed equally: Haojie Du, Nana Chen

[*] Correspondence: yangsun@nju.edu.cn (Sun Y)

Abstract

Natural products are important sources of therapeutic agents, yet their structural diversity, polypharmacology, and context-dependent effects complicate target discovery. This review presents an integrated, decision-oriented framework linking experimental target fishing, orthogonal target-engagement and functional validation, single-cell and spatial multi-omics, and artificial intelligence (AI)-assisted prioritization. We compare representative label-based and label-free target-fishing strategies according to their biological applicability, evidential strength, limitations, and validation requirements. We also outline practical approaches for resolving drug-responsive cell states and tissue niches using single-cell and spatial technologies, and summarize AI methods for compound-target prediction, graph- and knowledge-based reasoning, perturbation modeling, and multimodal data integration. Particular attention is given to data quality, applicability domains, interpretability, and prospective validation. Finally, we propose an evidence-informed closed-loop workflow in which computational and omics-derived hypotheses are tested through direct-binding, cellular-engagement, genetic, pharmacological, and spatially resolved experiments. This framework aims to improve the rigor and efficiency of natural product target discovery and mechanism elucidation.

Keywords: Natural products, Target fishing, Target engagement, Target validation, Single-cell multi-omics, Artificial intelligence

Citation: Du H, Chen N, Sun Y. 2026. From target fishing to AI and single-cell multi-omics: an integrated framework for natural product target research. *Targetome* 2(4): e037 https://doi.org/10.48130/targetome-0026-0036

Introduction

Natural products have long served as a major source of medicines, lead compounds, and chemical probes[1]. Their pharmacological effects, however, frequently arise from structurally diverse metabolites acting across multiple targets, cell types, and tissue compartments. This complexity makes it difficult to distinguish a direct molecular target from an associated pathway component or a secondary response factor. Rigorous target discovery therefore requires an evidential sequence that connects phenotype, candidate-target generation, direct target engagement, functional causality, and disease-context validation[2,3]. In this review, target fishing refers primarily to experimental capture or proteome-wide nomination of compound-interacting proteins. Target identification or target deconvolution denotes the broader process of prioritizing the molecular entities responsible for a compound-induced phenotype. Target engagement indicates that a compound interacts with a candidate protein in a defined biochemical, cellular, or tissue context, whereas direct binding requires evidence of a physical ligand-protein interaction. Target validation further requires demonstration that engagement of the proposed target is causally linked to the pharmacological phenotype. These terms describe different evidential levels and should not be used interchangeably.

AI has become useful in natural product research primarily as a tool for prioritizing compounds, predicting candidate targets, and integrating heterogeneous data. Its value depends on whether predictions can be traced to chemical or biological evidence and tested experimentally[4,5]. In parallel, single-cell and spatial multi-omics can resolve treatment responses across cell types and tissue compartments, providing context that is not available from bulk measurements[6–8].

Combining these approaches creates a practical division of labor: single-cell and spatial data define the responsive cellular context, whereas AI organizes chemical, structural, and omics evidence into ranked hypotheses. Neither approach establishes a direct target on its own. Their contribution is strongest when candidate targets are subsequently evaluated by target-fishing, target-engagement, and causal perturbation experiments.

The novelty of this review lies not in cataloguing these technologies separately, but in organizing them into an end-to-end, decision-oriented workflow for natural product research. We compare experimental target-fishing methods according to their readouts and evidential boundaries, distinguish biophysical binding from cellular engagement and genetic dependence, provide practical guidance for selecting and integrating single-cell and spatial technologies, and describe how AI models accept, fuse, and transform chemical and multi-omics inputs into experimentally testable outputs.

Accordingly, the review proceeds from experimental target fishing and validation to cell-type-resolved mechanism analysis and AI-assisted target prioritization. General drug repurposing and pharmacokinetic applications are discussed only as supporting contexts. The final section presents an evidence-informed closed-loop framework and representative cases, while explicitly identifying the components that still require prospective validation in natural product studies.

Target fishing

Target fishing comprises experimental strategies used to enrich, detect, or prioritize macromolecules that interact with a bioactive small molecule. Depending on whether the ligand must be chemically derivatized, these strategies can be broadly classified as label-based and label-free approaches (Fig. 1). Importantly, the

https://doi.org/10.48130/targetome-0026-0036

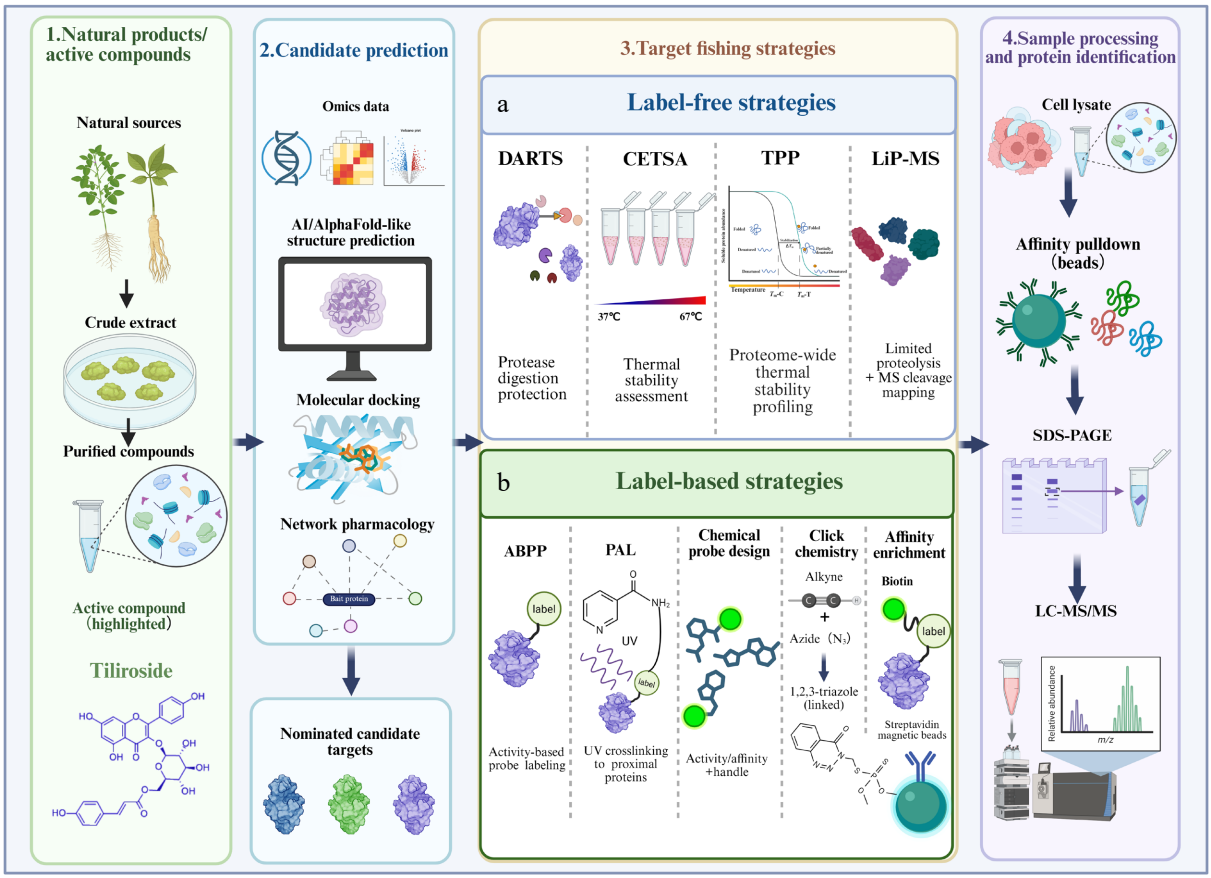

Fig. 1 Framework for the identification of direct targets of natural products. Natural products or bioactive compounds are isolated from natural sources and subjected to activity-guided purification, with tiliroside shown as an illustrative example. Candidate targets can first be nominated using omics-based analyses, artificial intelligence-assisted structure prediction, molecular docking, and network pharmacology. Experimental target-fishing strategies are subsequently divided into label-free and label-based approaches. (a) Label-free strategies detect ligand-induced changes in protein stability, protease susceptibility, or local conformation. DARTS identifies proteins protected from proteolysis after ligand binding; CETSA evaluates ligand-induced alterations in thermal stability; TPP extends thermal stability analysis to the proteome scale; and LiP-MS detects changes in local protease accessibility and maps ligand-responsive protein regions. (b) Label-based strategies use activity- or affinity-based chemical probes, photoaffinity labeling, click chemistry, and biotin-streptavidin enrichment to capture compound-interacting proteins. Depending on the target-fishing strategy, candidate proteins are recovered by affinity enrichment or collected from soluble or proteolytic fractions and subsequently identified by immunoblotting or LC-MS/MS. Candidate targets should be further evaluated using orthogonal binding and functional validation assays.

output is usually a candidate-target set rather than definitive proof of direct binding or functional relevance.

Label-based target identification

Affinity-based chemical probes are widely used to identify the protein targets of small molecules. A typical probe comprises a bioactive moiety that retains target-binding activity, a linker, and a reporter or enrichment handle. After target engagement, the probe enables visualization or affinity enrichment of the interacting proteins, followed by identification using immunoblotting or mass spectrometry. Common strategies include immobilization on affinity matrices, biotinylation, click-chemistry-enabled labeling, and photoaffinity labeling.

Biotin-based affinity labeling

Biotinylation introduces a biotin tag into a bioactive molecule, enabling enrichment of interacting proteins through the high-affinity biotin-streptavidin interaction. Owing to its robustness and compatibility with mass spectrometry, the biotin-streptavidin system remains one of the most widely used platforms for affinity-based target fishing.

Using biotinylated artesunate (bio-ATS), Liu et al. demonstrated that probe derivatization did not abolish the biological activity of artesunate. Affinity capture and thermal-stability assays identified the mitochondrial protease LONP1, rather than CYP11A1, as the direct binding protein of artemisinins. Engagement of LONP1 promoted its interaction with CYP11A1 and subsequently accelerated CYP11A1 degradation[9].

Gambogic acid (GA), a natural product isolated from *Garcinia hanburyi*, promotes proteasomal degradation of KRAS. To identify the protein responsible for this effect, Wang et al. synthesized a biotinylated GA probe that retained the ability to induce KRAS degradation. Affinity pull-down coupled with mass spectrometry identified USP2 as a direct target of GA and also revealed the involvement of HSP90-associated protein quality-control machinery[10].

Protein microarrays provide another high-throughput platform for affinity-based target identification. Recombinant proteins carrying His or GST tags are arrayed on a solid support, and their interactions with labeled small molecules are detected using fluorescence or other reporter systems. Such arrays enable parallel interrogation of thousands of potential binding proteins[11,12].

α-Mangostin (α-MG), a natural xanthone derived from mangosteen pericarp, induces pyroptosis in osteosarcoma cells. A

biotinylated α-MG probe combined with a HuProt human proteome microarray identified reticulon 4 (RTN4) as its target. Mechanistically, α-MG functions as a molecular glue that recruits the E3 ubiquitin ligase UBR5, promotes K48-linked ubiquitination and proteasomal degradation of RTN4, remodels the endoplasmic reticulum membrane, and thereby facilitates tumor-cell pyroptosis[13].

The HuProt proteome microarray was also used to identify the deubiquitinase USP7 as a target of Eupalinolide B (EB). EB binds the noncatalytic HUBL domain of USP7 rather than its catalytic domain, promotes ubiquitin-dependent degradation of Keap1, and alleviates behavioral abnormalities in mouse models of dementia and Parkinson's disease[14].

Click-chemistry-enabled labeling

Click chemistry provides rapid and selective reactions for conjugating chemical probes to reporter or enrichment groups. Frequently used reactions include copper-catalyzed azide-alkyne cycloaddition (CuAAC) and strain-promoted azide-alkyne cycloaddition (SPAAC), both of which generate stable triazole products. In chemical proteomics, a minimally modified alkyne- or azide-containing probe is first allowed to interact with proteins in cells or tissues and is subsequently conjugated to biotin, a fluorophore, or another analytical handle.

Click chemistry is frequently combined with photoaffinity labeling, activity-based probes, and quantitative mass spectrometry, allowing target capture and identification in complex biological systems. An alkyne-containing probe and CuAAC-based quantitative proteomics were used to characterize the interactions between 7-O-cinnamoyl paclitaxel and mitochondrial proteins, illustrating the utility of this strategy for identifying the targets of natural products[15].

Bioorthogonal reactions are particularly suitable for biological applications because they proceed selectively under physiological conditions without substantially perturbing endogenous biochemical processes. Bioorthogonally activated reactive species (BARS) have recently been developed as an alternative to conventional photoaffinity labeling. In this platform, a chemically caged precursor is activated through a bioorthogonal reaction to generate a reactive intermediate *in situ*, thereby improving target-labeling efficiency and reducing nonspecific background[16].

Bioorthogonal chemistry has also been extended to molecular imaging. Wang et al. developed a dual-locked enzyme-activatable bioorthogonal fluorescence (DEBOF) turn-on imaging system comprising a dual-locked bioorthogonal targeting agent (DBTA) and a bioorthogonally activatable fluorescent imaging probe (BAP). The bioorthogonal reaction and fluorescence signal are activated only in cells displaying both cancer- and senescence-associated enzymatic activities, enabling selective detection of senescent cancer cells[17].

Photoaffinity labeling

Photoaffinity labeling uses probes containing a photoreactive group that can be activated by light to form a covalent bond with proteins in close proximity. This strategy can stabilize otherwise transient or weak ligand–protein interactions and is therefore widely used in target identification. Common photoreactive groups include aryl azides, diazirines, and benzophenones. A bifunctional photoaffinity probe derived from chlorogenic acid, designated PAL-CGA, was developed to identify the direct targets of chlorogenic acid. Chemical proteomic analysis identified mitochondrial acetyl-CoA acetyltransferase 1 (ACAT1) as a major chlorogenic acid-binding protein and linked ACAT1 engagement to the anticancer activity of chlorogenic acid[18].

Gao et al. developed an artemisinin photoaffinity probe (APP) and combined it with activity-based protein profiling to characterize the protein targets of artemisinin during the intraerythrocytic developmental cycle of *Plasmodium falciparum*[19]. APP captured both covalent and noncovalent interacting proteins in the ring, trophozoite, and schizont stages, indicating that artemisinin can be activated by heme throughout parasite intraerythrocytic development.

Martín-Acosta et al. developed LBL1-P, a clickable photoaffinity probe derived from the pentacyclic triterpenoid betulinic acid[20]. Chemical proteomic analysis using this probe identified tropomyosin as a previously unrecognized binding protein of betulinic acid. Related affinity-based workflows have also contributed to the identification of natural-product targets. Rg3-polyethylene glycol acrylamide pull-down followed by mass spectrometry identified the E2F transcriptional complex as a target of 20(S)-ginsenoside Rg3. CETSA, co-immunoprecipitation, and reporter assays further demonstrated that Rg3 disrupts E2F-DP dimerization, thereby suppressing E2F-dependent transcription and gastric cancer cell proliferation[21].

Bruceine A was identified as a direct inhibitor of HSP90AB1. Its interaction with HSP90AB1 was validated by surface plasmon resonance (SPR) and CETSA, and target engagement destabilized several HSP90 client proteins, including EGFR, PIK3CG, and KDM5C[22]. Chemoproteomic profiling further demonstrated that ailanthone directly binds the glycolytic enzyme PKM2, thereby suppressing metabolic reprogramming and hepatocellular carcinoma progression[23]. In breast cancer, nobiletin directly targets AKR1C1 and promotes AKR1C1-dependent ubiquitination and degradation of GPX4, ultimately inducing ferroptosis[24].

PROTAC-based target identification

Proteolysis-targeting chimeras (PROTACs) are heterobifunctional molecules consisting of a ligand for a protein of interest, an E3 ubiquitin ligase ligand, and a chemical linker. By recruiting the target protein to an E3 ligase, PROTACs induce target ubiquitination and proteasomal degradation. Unlike occupancy-driven inhibitors, PROTACs act through an event-driven and potentially catalytic mechanism and do not necessarily require binding to an enzymatic active site. These properties make PROTAC-based strategies particularly useful for identifying low-affinity targets or proteins that are difficult to interrogate using conventional inhibitors.

Wu et al. developed a PROTAC-based approach to identify the target of lathyrane diterpenoids isolated from *Euphorbia lathyris*[25]. A PROTAC derivative of the active compound ZCY-001, termed ZCY-PROTAC, was synthesized and evaluated using tandem mass tag-based quantitative proteomics. MAFF was selectively depleted after ZCY-PROTAC treatment, supporting MAFF as a functional target of ZCY-001.

Ni et al. subsequently established degradation-based protein profiling using celastrol as a model natural product[26]. This approach recovered previously reported celastrol targets, including IKKβ, PI3Kα and CIP2A, and identified additional candidate targets, such as CHK1, O-GlcNAcase and ERCC6L. Induced protein degradation therefore provides an orthogonal readout for natural-product target identification, particularly when conventional affinity enrichment is limited by weak or transient binding.

Label-free target identification

Chemical derivatization may alter the permeability, subcellular localization, binding affinity, or pharmacological activity of a natural product. Moreover, some small molecules lack suitable sites for probe installation. Label-free methods circumvent these limitations by detecting ligand-induced changes in protein stability, protease susceptibility, residue accessibility, oxidation, or solubility. Representative approaches include DARTS, CETSA/TPP, LiP-MS/PELSA, SPROX, TRAP, and solubility-based profiling.

Drug affinity-responsive target stability

Lomenick et al. introduced drug affinity-responsive target stability (DARTS) as a label-free method for identifying small-molecule-binding proteins[27]. DARTS is based on the observation that ligand binding can alter the susceptibility of a target protein to proteolysis. A protein mixture is incubated with the test compound and then subjected to limited protease digestion. Ligand-protected proteins are subsequently detected by gel electrophoresis, immunoblotting, or mass spectrometry. The method was applied to resveratrol-treated yeast lysates, in which resveratrol protected eukaryotic translation initiation factor 4A from proteolytic degradation, supporting eIF4A as a resveratrol-binding protein[28].

Aloperine is a quinolizidine alkaloid isolated from *Sophora alopecuroides* with antitumor activity in several cancer models. DARTS analysis revealed an aloperine-protected protein band of approximately 50 kDa. Mass spectrometry and immunoblotting identified the protein as vacuolar protein sorting-associated protein 4A (VPS4A), which was subsequently validated as a functional target involved in the inhibition of autophagosome–lysosome fusion[29].

Oleanolic acid is a pentacyclic triterpenoid with anti-inflammatory, antioxidant, and hepatoprotective activities. It directly binds and allosterically activates the open conformation of SHP2, resulting in sustained low-level SHP2 activation, reduced STAT3 phosphorylation and Th17 differentiation, and amelioration of experimental colitis[30].

Cellular thermal shift assay and thermal proteome profiling

Martinez et al. developed the cellular thermal shift assay (CETSA) to monitor drug–target engagement in cells and tissues[31]. CETSA is based on ligand-induced changes in protein thermal stability. During heating, proteins progressively unfold and aggregate; ligand binding may stabilize or destabilize a target protein, thereby shifting its apparent melting profile. The remaining soluble proteins are subsequently quantified by immunoblotting or other analytical methods.

CETSA can be performed using cell lysates, intact cells, or tissues. However, heat treatment may alter membrane permeability, protein complexes, and intracellular compound distribution, which can complicate data interpretation. Miettinen & Björklund used CETSA to identify NAD(P)H quinone dehydrogenase 2 (NQO2) as a reactive oxygen species-generating off-target of acetaminophen[32].

CETSA has evolved into several complementary formats. Conventional western blot-based CETSA is primarily used for hypothesis-driven target validation. Thermal proteome profiling (TPP), also known as MS-CETSA, combines thermal stability measurements with quantitative mass spectrometry for proteome-wide target identification. High-throughput CETSA is used for compound screening, hit characterization, and lead optimization. Ji et al. developed the matrix-augmented pooling strategy (MAPS) to increase the throughput of TPP-based target deconvolution[33]. Multiple compounds are arranged into optimized pools, and the target profile of each compound is reconstructed computationally. Application of MAPS to 15 compounds increased experimental throughput by approximately 60-fold while retaining high sensitivity and specificity.

Ginkgolic acid was shown by DARTS, CETSA, and microscale thermophoresis to directly bind HSPA8. Target engagement enhanced HSPA8-mediated chaperone-mediated autophagy, promoted GPX4 degradation, and induced ferroptosis in hepatocellular carcinoma cells[34].

Usenamine A was identified as a direct ligand of MYH9 using mass spectrometry, SPR, CETSA, and molecular modeling. By disrupting the MYH9–actin interaction, usenamine A impaired cytoskeletal remodeling and induced apoptosis and autophagic cell death in hepatoma cells[35]. In gastric cancer, CETSA and DARTS supported HSP90AA1 as a functional target of mulberrin. Inhibition of the HSP90AA1/PI3K/AKT/GSK3β/Snail pathway suppressed epithelial-mesenchymal transition and increased sensitivity to oxaliplatin[36].

Peptide-centric local stability assay

Li et al.[37] developed the peptide-centric local stability assay (PELSA) for proteome-scale identification of ligand-binding proteins and their binding regions. In PELSA, native protein mixtures are subjected to high-concentration trypsin digestion, directly generating peptides suitable for mass-spectrometric analysis. Ligand binding alters the local accessibility and stability of specific protein regions, resulting in reproducible changes in peptide abundance.

Unlike methods that measure the global solubility of intact proteins, PELSA amplifies local structural changes at the peptide level. It does not require chemical modification of the ligand and can be applied to complex samples such as cell lysates. The method enables the simultaneous identification of ligand-binding proteins, affected protein regions, and local binding characteristics.

Stability of proteins from rates of oxidation

Strickland et al.[38] established the stability of proteins from rates of oxidation (SPROX) method to evaluate protein–ligand interactions in complex biological mixtures. SPROX measures ligand-induced changes in protein thermodynamic stability by monitoring the oxidation of methionine residues.

Protein samples are exposed to a chemical denaturant gradient and an oxidizing reagent. Methionine residues that become accessible during protein unfolding are oxidized, and the corresponding peptides are quantified by mass spectrometry. Because ligand binding shifts the folding equilibrium of a target protein, SPROX can identify ligand-engaged proteins without prior purification.

Ogburn et al. combined large-scale protein folding and stability measurements with quantitative proteomics to investigate the targets of tamoxifen and *N*-desmethyl tamoxifen in MCF-7 cells[39]. Y-box-binding protein 1 (YBX1) was identified as a candidate target, and its relationship with estrogen receptor signaling was subsequently characterized.

Target-responsive accessibility profiling

Tian et al. developed target-responsive accessibility profiling (TRAP) to map ligand-induced changes in protein accessibility at the proteome scale[40]. Unlike DARTS and CETSA, which primarily detect changes in protease susceptibility or global thermal stability, TRAP quantifies changes in the accessibility of reactive lysine residues after ligand binding. Peptides showing significant abundance changes in the presence of a ligand are defined as target-responsive peptides. TRAP was initially used to map the targetome of glycolytic metabolites in cancer cells. Global labeling of reactive lysines identified accessibility changes induced by 10 major glycolytic metabolites, yielding 913 responsive candidate proteins and 2,487 metabolite–protein interactions.

Yan et al. applied living cell-target responsive accessibility profiling (LC-TRAP) to characterize the intracellular targetome of silibinin in HepG2 cells[41]. Covalent lysine labeling was combined with multiplexed quantitative proteomics to identify drug-induced accessibility changes in living cells. Subsequent validation identified ACSL4 as an important functional target through which silibinin counteracts ferroptosis.

SPR-MS and SPR-guided ligand or target fishing

Surface plasmon resonance is a real-time, label-free method for analyzing molecular interactions. It can provide association and

dissociation rate constants as well as equilibrium binding affinities. In natural-product research, SPR is commonly used to validate direct binding between a purified protein and a candidate ligand.

SPR complements cellular target-engagement methods such as CETSA and DARTS. Whereas CETSA and DARTS provide evidence of target engagement in cellular or proteomic environments, SPR provides direct biophysical evidence under defined experimental conditions. Together, these methods help distinguish cellular target engagement from direct protein–ligand binding.

Epigallocatechin gallate (EGCG) was shown to bind STAT3 and inhibit its nuclear translocation and transcriptional activity. SPR, CETSA, chromatin immunoprecipitation-qPCR, and dual-luciferase reporter assays demonstrated that EGCG represses STAT3-dependent PLXNC1 transcription, thereby inhibiting M2 macrophage polarization induced by gastric cancer cell-derived exosomal miR-92b-5p[42]. SPR and CETSA were also used to validate HSP90AB1 as a direct target of Bruceine A[22], whereas MYH9 binding by usenamine A was supported by SPR, CETSA, mass spectrometry, and molecular modeling[35].

The combination of SPR with mass spectrometry has expanded its application from binary interaction validation to ligand and target fishing in complex natural-product systems. Ni et al. established a trace-component fishing strategy based on offline two-dimensional liquid chromatography combined with PRDX3-SPR. Fractionation reduced interference from abundant constituents, after which recombinant peroxiredoxin 3 (PRDX3) was immobilized for SPR screening. Twenty-nine candidate PRDX3-binding alkaloids were detected in 13 two-dimensional fractions of *Uncaria*. Subsequent affinity and functional analyses showed that several trace alkaloids enhanced PRDX3-mediated hydrogen peroxide removal[43].

Tan et al. developed an SPR-guided workflow to identify quinolone alkaloids from the fruit of *Tetradium ruticarpum* as inhibitors of ferroptosis suppressor protein 1 (FSP1)[44]. Recombinant FSP1 was immobilized on a CM5 sensor chip, and icFSP1 was used as a positive control to confirm target activity. SPR screening followed by HPLC-MS-guided isolation yielded 12 quinolone alkaloids, including the previously undescribed compounds Ruticarponine A and Ruticarponine B.

18β-Glycyrrhetinic acid (18β-GA) is a licorice-derived triterpenoid with anti-inflammatory activity. In an SPR-MS workflow, immobilized 18β-GA was exposed to cell lysates, and β-glucuronidase (GUSB) was identified as a direct binding protein. Mass spectrometry, molecular docking, and enzymatic assays further supported this interaction. Engagement of GUSB modulated GUSB/ATF2 signaling, reduced CCL20 expression, and disrupted the inflammatory feedback loop between keratinocytes and CCR6-positive immune cells[45].

SPR coupled with liquid chromatography-tandem mass spectrometry was used to screen xanthohumol-binding proteins in bone marrow-derived macrophage lysates. Heterogeneous nuclear ribonucleoprotein K (hnRNPK) was identified as a direct target. SPR affinity measurements, CETSA, and molecular docking supported this interaction. Mechanistically, xanthohumol reduced the nuclear accumulation of hnRNPK and its binding to the *Nlrp3* promoter, thereby suppressing NLRP3 inflammasome-associated macrophage pyroptosis and alleviating heatstroke-induced tissue injury[46].

Other label-free approaches

Energetics- and solubility-based proteomic methods provide additional strategies for studying small-molecule–protein interactions directly in complex cell lysates.

Zhang et al. developed solvent-induced protein precipitation (SIP) for proteome-scale drug-target discovery[47,48]. SIP exploits the greater resistance of ligand-bound proteins to organic solvent-induced denaturation and precipitation. In the original workflow, an acetone/ethanol/acetic acid mixture was used to perturb protein stability, and the remaining soluble proteins were quantified by mass spectrometry. Proteins showing ligand-dependent resistance to precipitation were prioritized as candidate targets. The same group subsequently developed pH-dependent protein precipitation (pHDPP), which detects the increased resistance of ligand-bound proteins to acid-induced denaturation. pHDPP is compatible with structurally diverse ligands, including folate derivatives, ATP analogues, and kinase inhibitors. Application of this approach to dihydroartemisinin identified 45 candidate binding proteins.

Differential precipitation of proteins (DiffPOP) profiles compound-induced changes in protein solubility across an organic solvent gradient. Using DiffPOP, Xu et al. identified serine hydroxymethyltransferase 2 (SHMT2) as a direct target of the histone demethylase inhibitor JIB-04[49]. SHMT2 was further linked to the BRCC36/BRISC deubiquitinase complex and the regulation of HIV-1 Tat K63-linked ubiquitination and autophagic degradation.

Functional genomic screening provides a conceptually distinct strategy for identifying proteins required for a drug-induced phenotype. Genome-wide CRISPR-Cas9 screening of RNF43-mutant pancreatic ductal adenocarcinoma cells identified a selective dependency on the WNT7B-FZD5 signaling circuit. Genetic and antibody-based validation further supported FZD5 as a therapeutically tractable cell-surface target[50]. Although functional genomic screening does not directly demonstrate physical ligand binding, it can prioritize functionally relevant targets and complement affinity- or stability-based target identification.

Collectively, label-based and label-free strategies generate candidate target lists through distinct physicochemical readouts, including affinity enrichment, covalent capture, induced degradation, protease protection, thermal stabilization, residue accessibility, oxidation, and solubility changes. These readouts differ in their compatibility with intact cells, requirement for ligand modification, proteome coverage, throughput, and susceptibility to indirect effects. Method selection should therefore be driven by the biological question and compound properties rather than by platform availability alone. Table 1 summarizes the major decision variables. Regardless of the discovery method, nonspecific binders, indirect interactors, and proteins stabilized within larger complexes must be excluded through competition, orthogonal binding assays, and functional experiments.

A practical decision rule is to combine methods with non-overlapping biases. For example, affinity enrichment provides physical capture but requires probe validation; CETSA or TPP preserves cellular context but may detect indirect thermal shifts; DARTS is modification-free but depends on protease accessibility; and peptide-level methods can localize responsive regions but do not alone establish a binding site. Prospective target confirmation should therefore combine at least one discovery-scale method with an orthogonal direct-binding or cellular-engagement assay and a causal perturbation experiment.

Resource requirements also differ substantially. DARTS and hypothesis-driven CETSA are comparatively accessible but generally low-throughput. Affinity and photoaffinity workflows require probe synthesis and extensive controls, whereas TPP, LiP-MS/PELSA, TRAP, and solubility-based proteomics provide broader coverage at the cost of quantitative mass spectrometry, biological replication, and more intensive data analysis. Protein microarrays and pooled TPP can increase throughput, but they shift the burden toward platform access, reagent preparation, and follow-up validation. Method selection should therefore consider not only discovery coverage,

Table 1. Decision-oriented comparison of representative natural product target-fishing strategies.

Method/readout	Label and biological setting	Main advantages	Main limitations	Best use and required validation
Affinity or biotin pull-down	Requires an immobilized or tagged ligand; lysates or intact-cell-compatible probes	Direct enrichment; compatible with competition and quantitative proteomics	Probe modification may alter permeability or affinity; matrix and abundant protein background	Unbiased capture when a validated probe is available; confirm with free-compound competition and an orthogonal binding assay
Photoaffinity/click chemistry	Minimal photo-crosslinker and clickable handle; usually intact cells or lysates	Captures weak or transient interactions; preserves spatial proximity	Photochemical background and crosslinking-radius effects; synthesis and controls are demanding	Transient or low-affinity interactions; require inactive-probe, no-UV, and competition controls
Degradation-based profiling	Ligand incorporated into a degrader or molecular-glue workflow; intact cells	Event-driven signal amplification; can reveal low-occupancy binders	Depends on ternary-complex geometry, E3 expression, and proteasome competence	Functional target nomination when degradation chemistry is feasible; validate direct binding and degradation dependence
DARTS	No ligand modification; native lysates and limited proteolysis	Simple, inexpensive, and compatible with chemically intractable ligands	Biased by protein abundance, protease accessibility, and indirect conformational changes	Focused or discovery-scale screening; validate by dose-dependent protection and a biophysical assay
CETSA/TPP	No ligand modification; lysates, intact cells, tissues; immunoblot or MS readout	Measures engagement in a biologically relevant environment; proteome-wide with TPP	Not all binders shift thermal stability; complexes and downstream effects can produce indirect shifts	Cellular target engagement and proteome-wide deconvolution; combine with purified-protein binding and genetics
PELSA/LiP-MS	No ligand modification; peptide-level proteolysis in native mixtures	Detects local structural responses and can suggest responsive protein regions	Peptide detectability and protease accessibility limit coverage; responsive regions are not necessarily binding sites	Mapping local conformational responses; confirm by mutagenesis or structural analysis
SPROX/TRAP	No ligand modification; oxidation or residue-accessibility readout in complex proteomes	Orthogonal physicochemical evidence; sensitive to local folding or accessibility changes	Requires appropriate reactive residues and specialized quantitative proteomics	Complementary discovery when thermal or proteolytic shifts are weak; validate direct engagement
SIP/pHDPP/ DiffPOP	No ligand modification; solvent-, pH-, or gradient-induced precipitation	Scalable and applicable to structurally diverse compounds	Solubility changes may be indirect and are influenced by protein physicochemical properties	Proteome-wide prioritization; require orthogonal engagement and functional testing
SPR-MS or target-immobilized fishing	Immobilized protein or ligand; fractions, extracts, or lysates	Links real-time binding detection with MS identification; useful for trace constituents or complex mixtures	Immobilization can alter conformation; mass transport and nonspecific surface binding require controls	Ligand fishing or target fishing in mixtures; confirm affinity, activity, and cellular relevance

but also the resources needed to exclude false-positive or indirect candidates.

Target validation

Target-identification strategies frequently generate multiple candidate proteins. These candidates should be prioritized according to the strength of the binding evidence, their known biological functions, subcellular localization, and relevance to the phenotype induced by the compound. Appropriate negative controls, competition assays, and inactive structural analogues are essential for distinguishing specific interactions from nonspecific binding. Moreover, a protein captured by affinity- or stability-based methods may represent an indirect interactor or a component of a larger protein complex rather than the direct molecular target. Target validation must therefore establish both direct target engagement and functional causality (Fig. 2).

Biophysical validation of direct binding

Demonstrating direct binding between a natural product and a candidate protein is a central step in target validation. Commonly used biophysical approaches include surface plasmon resonance (SPR)[51,52], isothermal titration calorimetry (ITC)[53,54], fluorescence polarization (FP)[55,56], homogeneous time-resolved fluorescence (HTRF)[57], and microscale thermophoresis (MST)[58–60]. Table 2 summarizes the main applications of these methods, which provide complementary information on small-molecule–protein interactions. SPR enables real-time measurement of association and dissociation kinetics and can provide the association rate constant, dissociation rate constant, and equilibrium dissociation constant. ITC directly measures the heat released or absorbed during binding and simultaneously determines binding affinity, stoichiometry, enthalpy, and entropy under label-free, solution-phase conditions. FP and HTRF are homogeneous assay formats suitable for high-throughput screening and competitive binding analyses. MST detects ligand-induced changes in molecular thermophoresis and requires only small amounts of sample. Wienken et al.[58] further demonstrated that MST can quantify protein–ligand interactions in complex biological fluids, including serum and cell lysates. A single affinity assay is generally insufficient to establish a functional direct target. The apparent affinity obtained using different techniques may vary because of differences in protein conformation, immobilization, labeling, buffer composition, and assay format. Key candidate targets should therefore be examined using at least two orthogonal approaches, such as SPR combined with MST or ITC, or a cellular target-engagement assay such as CETSA combined with SPR. Competition experiments, concentration-dependent binding, inactive analogues, and structurally related compounds should also be included to assess specificity.

Where feasible, nuclear magnetic resonance spectroscopy, small-angle X-ray scattering, X-ray crystallography, or cryogenic electron microscopy can be used to resolve ligand-binding sites, conformational changes, and critical interacting residues. In particular, protein X-ray crystallography can provide atomic-resolution structures of protein–ligand complexes and remains an important tool for structure-guided drug development[61].

Biological function verification of target proteins

Direct binding between a small molecule and a candidate protein does not necessarily indicate that the protein mediates the pharmacological phenotype. After target engagement or direct binding has been demonstrated by DARTS[27], CETSA[31], thermal

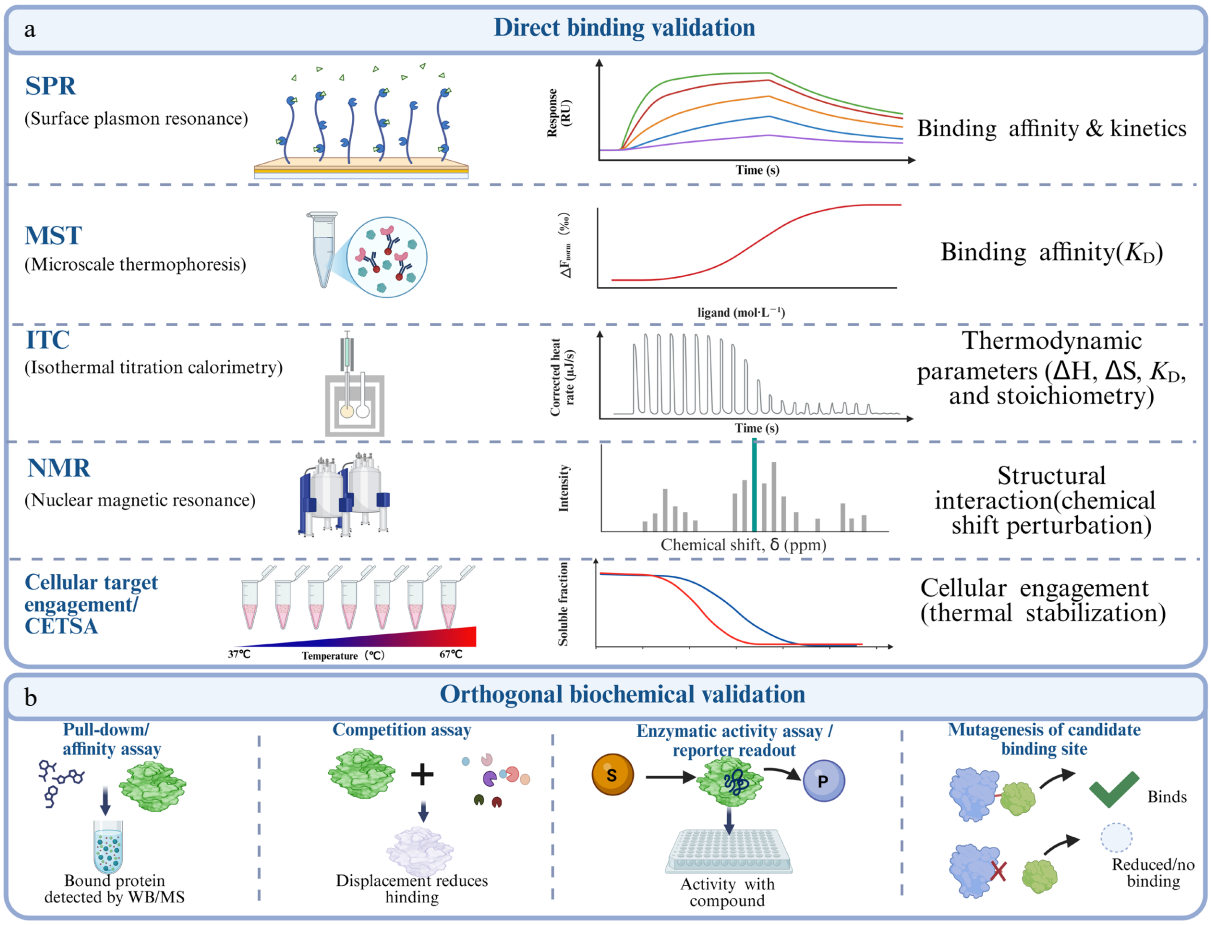

Fig. 2 Framework for validating direct binding and functional causality. (a) Biophysical and cellular target-engagement assays offer complementary evidence for compound–protein interactions. SPR, MST, ITC, and NMR assess binding kinetics, affinity, thermodynamics, and structural changes, respectively. CETSA measures target engagement in cells or lysates *via* thermal stability shifts, but should be combined with purified-protein biophysical assays to confirm direct binding. (b) Orthogonal biochemical validation uses affinity pull-down, competition assays, functional readouts, and site-directed mutagenesis to confirm binding specificity and identify critical residues. Competition by excess unmodified compound supports specific binding, while loss of binding or activity after mutation indicates a defined binding site.

Table 2. Complementary methods for validating small-molecule-protein interactions.

Method	Core readout	Main strengths	Main limitations	Primary evidential role
SPR	Surface refractive-index change during binding	Real-time K_a, K_d, and K_D; low sample use	Immobilization, mass transfer, and nonspecific binding	Direct binding and kinetics
ITC	Heat change during solution-phase titration	K_D, stoichiometry, enthalpy, and entropy	High sample demand; weak or low-heat interactions are difficult	Direct binding and thermodynamics
FP/HTRF	Binding-dependent rotation or time-resolved energy transfer	Homogeneous, scalable, and suitable for competition	Requires tracers or paired reagents; interference risk	Screening and displacement evidence
MST	Binding-dependent thermophoretic movement	Low sample use; broad affinity range; complex matrices possible	Fluorescence, adsorption, and aggregation artifacts	Orthogonal affinity measurement
NMR	Chemical-shift or relaxation changes	Weak-binding detection and interaction-surface mapping	Protein size, labeling, solubility, and instrument access	Direct binding and residue-level information
X-ray/cryo-EM	Atomic or near-atomic complex structure	Binding pose, pocket geometry, and critical contacts	Sample preparation and conformational-state limitations	Structural confirmation and mutation design

proteome profiling[62], SPR[51,52], MST[58,59], or ITC[54], functional experiments are required to establish a causal relationship among target engagement, target regulation, and the observed phenotype.

For enzymatic targets, biochemical activity assays should determine whether compound binding alters catalytic activity. Ganoderic acid T (GAT) was identified as an activator of pyruvate carboxylase (PC). DARTS, CETSA, biolayer interferometry, and molecular modeling supported the direct interaction between GAT and PC, whereas enzyme activity and metabolic assays demonstrated that PC activation contributed to the effects of GAT on hepatocellular carcinoma metabolism and proliferation[63].

Genetic perturbation provides a complementary strategy for establishing functional causality. Loss-of-function approaches, including siRNA, shRNA, CRISPR-mediated knockout, or CRISPR interference, should be combined where appropriate with cDNA overexpression and rescue experiments[64]. When structural information is available, mutation of predicted binding residues can further determine whether direct ligand binding is required for the

pharmacological response. Ideally, depletion of the target should attenuate or phenocopy the compound-induced effect, whereas re-expression of the wild-type target, but not a binding-deficient mutant, should restore drug responsiveness.

Chikusetsusaponin IVa (CS-IVa) was shown to bind directly to yes-associated protein (YAP), as supported by SPR, CETSA, and molecular docking. CS-IVa inhibited YAP/TAZ signaling and reduced hepatic stellate cell activation and liver fibrosis. Notably, genetic depletion or pharmacological inhibition of YAP did not further enhance the inhibitory effects of CS-IVa, supporting the functional involvement of YAP in its antifibrotic activity[65]. Zhao et al. developed a selection-based genetic screening platform that combines CRISPR loss-of-function screening with small-molecule phenotypic selection for intracellular target deconvolution[66]. This strategy identifies genetic perturbations that alter compound sensitivity and can therefore prioritize proteins required for a drug-induced phenotype. Nevertheless, genetic screening alone does not demonstrate a physical interaction and should be combined with biochemical or biophysical target-engagement assays.

Computational integration of drug-response and functional-genomic data provides another means of prioritizing candidate targets. Sinha et al. developed DeepTarget, which integrates large-scale drug-response profiles, CRISPR knockout viability screens, and matched omics data. Based on the premise that knockout of a functional drug target may phenocopy compound treatment, DeepTarget predicts primary targets, context-dependent secondary targets, and mutation-specific drug responses[67]. Because these predictions may include both direct binding targets and indirect pathway components, DeepTarget should be regarded as a target-prioritization tool rather than definitive evidence of physical binding. Celastrol is a pentacyclic triterpenoid derived from *Tripterygium wilfordii*. It directly binds adenylyl cyclase-associated protein 1 (CAP1) and disrupts the interaction between CAP1 and resistin. This interaction suppresses resistin-induced cAMP-protein kinase A-NF-κB signaling, reduces macrophage-mediated inflammation, and ameliorates high-fat-diet-induced metabolic syndrome in mice[68]. This study illustrates how direct binding, disruption of a protein-protein interaction, pathway modulation, and an *in vivo* phenotype can be integrated into a target-validation framework.

Landmark studies of thalidomide and its analogues further demonstrate the importance of connecting direct binding with genetic and phenotypic evidence. Ito et al. identified cereblon (CRBN) as a primary thalidomide-binding protein by affinity purification and showed in zebrafish and chick embryos that CRBN was required for thalidomide-induced developmental abnormalities[69]. Subsequent studies established CRBN as a direct target required for the immunomodulatory and antiproliferative effects of lenalidomide and pomalidomide[70]. Krönke et al. and Lu et al. subsequently used quantitative proteomics and biochemical analyses to demonstrate that lenalidomide promotes CRBN-dependent recruitment, ubiquitination, and degradation of the lymphoid transcription factors IKZF1 and IKZF3[71,72]. A single amino-acid substitution in IKZF3 conferred resistance to lenalidomide-induced degradation and rescued lenalidomide-mediated growth inhibition, thereby providing strong evidence that neosubstrate recruitment and degradation are causally linked to the pharmacological phenotype[71].

Cell-based validation should ultimately be extended to physiologically relevant animal models[69]. Conditional floxed alleles combined with tissue- or cell-type-specific Cre drivers enable the role of a candidate target to be examined in defined biological compartments. Heffner et al. established a comprehensive Cre-characterization resource to support the construction and validation of conditional mouse models[73], whereas Kim et al. summarized the general principles and experimental considerations of Cre-loxP-based tissue-specific genetic manipulation[74]. Such models can determine whether a candidate target is required for both disease progression and the therapeutic activity of a natural product *in vivo*.

Overall, target validation should integrate biophysical, biochemical, genetic, cellular, structural, and *in vivo* evidence[75,76]. A direct molecular target should be clearly distinguished from downstream signaling proteins and secondary response factors. Establishing this hierarchy is essential for defining the mechanism of action, evaluating therapeutic relevance, guiding compound optimization, and identifying potential off-target effects.

scMulti-omics for elucidating pharmacological mechanisms

Single-cell and spatial multi-omics can be used at two points in the evidence chain. After target engagement and functional causality have been established, they define the cell states and tissue niches in which the target acts. Earlier in a study, the same data can nominate responsive cell populations and mechanistic nodes for subsequent target testing. The direction of inference should be stated explicitly, because an omics association is not evidence of direct binding.

Advances in scMulti-omics

Single-cell multi-omics combines transcriptomic, chromatin accessibility, protein, and spatial measurements to resolve pharmacological responses beyond population averages (Fig. 3). Its practical value depends as much on experimental design and analytical choices as on the sequencing platform. Biological replication, balanced processing, appropriate controls, and independent validation are essential because batch effects, dissociation bias, sparse counts, and cell-composition changes can otherwise be mistaken for drug-specific mechanisms.

Single-cell and spatial analyses should therefore be positioned within a validation hierarchy. They can identify responsive cell types, prioritize candidate regulatory axes, and test whether a validated target is expressed and active in the relevant compartment. Direct targets nominated from these analyses must still be examined using chemical proteomics or orthogonal engagement assays such as DARTS, CETSA/TPP, affinity capture, SPR, MST, or ITC, followed by genetic perturbation and rescue. A practical analytical workflow is outlined below.

Practical analytical workflow and technology selection

Experimental design should define the biological unit before sequencing. Drug and control samples should include independent biological replicates, matched tissue handling, balanced library preparation, and, where possible, dose or time-course information. Cell recovery, viability, dissociation-induced stress, doublets, ambient RNA, mitochondrial read fractions, and sample-specific cell loss should be evaluated before downstream analysis. Pseudoreplication should be avoided by testing treatment effects at the sample level rather than treating individual cells as independent replicates.

For scRNA-seq, Seurat or Scanpy can support quality control, normalization, dimensionality reduction, clustering, and annotation. Harmony, Seurat integration, or scVI can reduce technical batch effects, but overcorrection may remove genuine treatment biology and should be assessed using both biological markers and sample mixing. Differential expression should be complemented by differential-abundance analysis because an apparent bulk-like

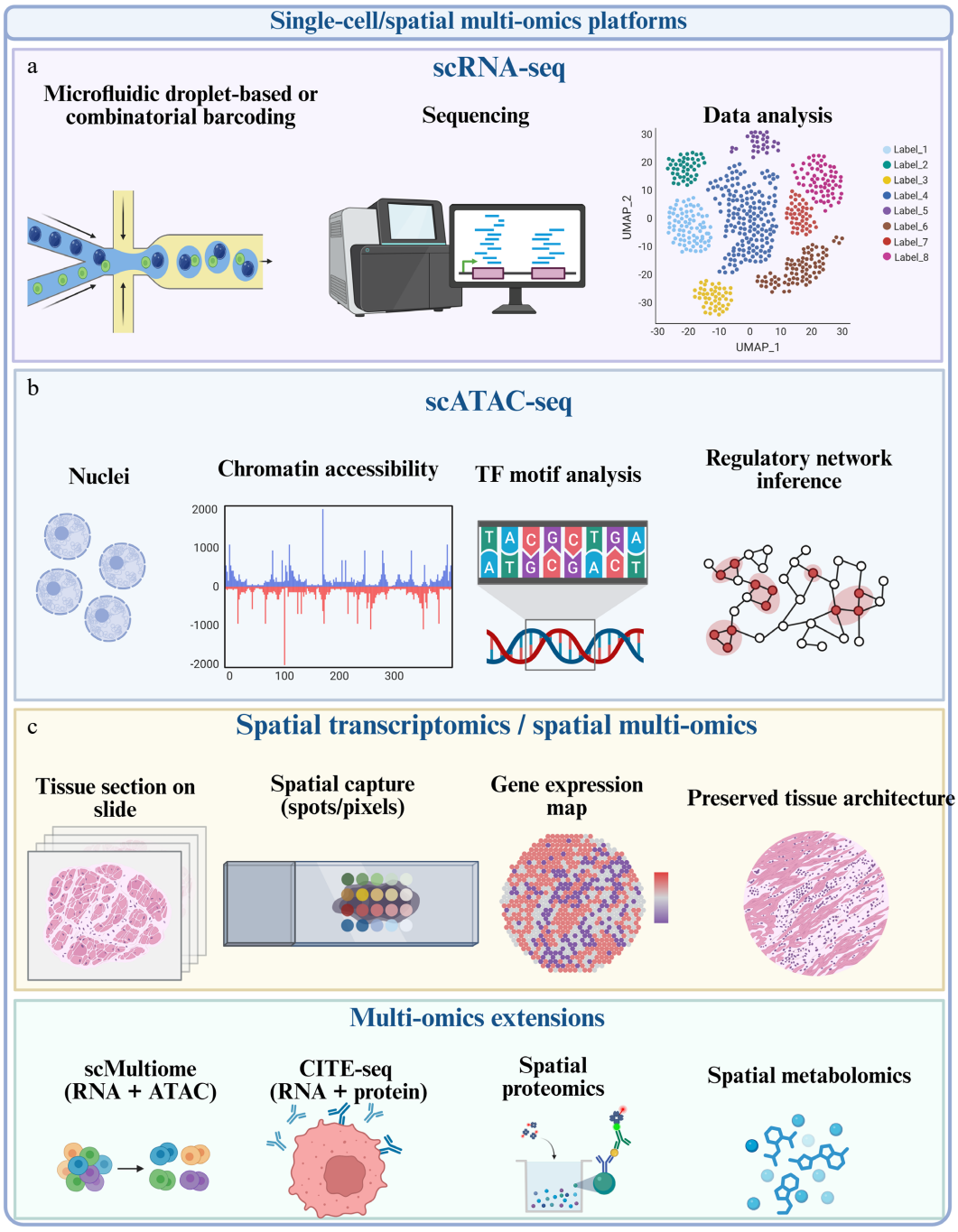

Fig. 3 Single-cell and spatial multi-omics platforms for resolving the pharmacological mechanisms of natural products. (a) scRNA-seq identifies drug-responsive cell populations and transcriptional states after quality control, normalization, integration, clustering, and annotation. (b) scATAC-seq and single-cell multiome approaches profile chromatin accessibility, transcription-factor motifs, and regulatory links between accessible elements and gene expression. (c) Spatial transcriptomics and spatial multi-omics retain tissue architecture and map treatment-responsive pathways or cellular interactions to defined pathological niches. These modalities provide contextual and mechanistic evidence but do not independently establish direct compound-target binding.

expression change may result from altered cell composition. Reference-based annotation should be reconciled with tissue-specific markers and manual biological review[77–79].

Tool selection should follow the biological question. Monocle or Slingshot can reconstruct putative state transitions, but pseudotime is an inferred ordering rather than direct temporal evidence. CellChat or CellPhoneDB can compare global ligand-receptor networks, whereas NicheNet links candidate ligands to downstream target-gene programs in a defined receiver population. ArchR or Signac supports scATAC-seq analysis; Seurat weighted-nearest-neighbor analysis, MOFA+, and related latent-factor models integrate multiple modalities; and cell2location, Tangram, or SPOTlight map cell states into spatial data[80–86].

Every computationally inferred mechanism should be connected to a measurable validation endpoint. Cell-type-specific immunostaining, flow cytometry, sorted-cell assays, spatial colocalization, perturbation of the inferred ligand-receptor pair, and target-specific rescue can distinguish a reproducible mechanism from a

software-dependent association. Table 3 summarizes a question-driven approach to technology and tool selection.

Single-cell transcriptomics

Single-cell RNA sequencing (scRNA-seq) is currently the most widely applied single-cell technology in pharmacological research on natural products. Tang et al. reported whole-transcriptome mRNA sequencing of an individual mammalian cell in 2009, helping establish sequencing-based analysis at single-cell resolution[87]. Subsequent tag-based and full-length protocols considerably improved sensitivity, throughput, and transcript coverage. Liao et al. developed SCAN-seq2, a high-throughput and high-sensitivity full-length scRNA-seq method based on third-generation sequencing, enabling the detection of transcript isoforms and immune-receptor rearrangements in thousands of individual cells[88].

By resolving transcriptional heterogeneity, scRNA-seq can determine which cell populations respond to a natural product and which pathways are selectively altered within those populations. This information narrows the search space for candidate targets and distinguishes primary drug-responsive cells from secondary tissue-level responses.

In atopic dermatitis, scRNA-seq revealed aberrant keratinocyte differentiation, mitochondrial dysfunction, and oxidative stress. Trilinolein, a triacylglycerol derived from *Cannabis fructus*, activated AhR-Nrf2 signaling, attenuated NOX2-dependent mitochondrial dysfunction and oxidative injury, and restored epidermal barrier function[89].

In psoriasis, scMulti-omics identified a pathogenic fibroblast-macrophage communication circuit driven by fibroblast-derived CCL2. Subsequent target-validation experiments demonstrated that celastrol directly binds the β-chain of low-density lipoprotein receptor-related protein 1 (LRP1), disrupts the nuclear LRP1-c-Jun interaction, and suppresses CCL2 production, thereby inhibiting fibroblast-macrophage crosstalk[90].

In hepatocellular carcinoma, scRNA-seq showed that berberine remodeled intrahepatic T-cell heterogeneity. Treatment reduced dysfunctional or immunosuppressive T-cell states, including exhausted $CD8^+$ T cells and regulatory T cells, while restoring effector and cytotoxic T-cell activity[91]. These findings indicate that the antitumor activity of berberine involves immune-microenvironment remodeling in addition to its direct effects on malignant cells.

In diabetic nephropathy, scRNA-seq demonstrated that rosmarinic acid alleviated pathological transcriptional states associated with oxidative stress, inflammation, fibrosis, and metabolic dysfunction across renal cell populations[92]. Such analyses can nominate cell-specific protective pathways for subsequent target identification and validation.

Single-cell chromatin-accessibility profiling

Single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) maps accessible chromatin regions in individual cells and enables the inference of transcription-factor activity, enhancer usage, and upstream regulatory programs. Whereas scRNA-seq primarily measures transcriptional output, scATAC-seq provides information on the regulatory potential underlying drug-induced cell-state transitions.

In natural-product research, scATAC-seq can be used to identify transcription factors and cis-regulatory elements associated with treatment-sensitive or treatment-resistant states. Integration with scRNA-seq further links changes in chromatin accessibility to their downstream transcriptional consequences. In the celastrol study of psoriasis, integrated single-cell transcriptomic and chromatin-accessibility analyses helped identify fibroblasts as a disease-promoting population and characterize the regulatory programs underlying fibroblast-macrophage communication[91].

In colorectal cancer, combined scRNA-seq and scATAC-seq demonstrated that cannabidiol (CBD) remodeled both the abundance and regulatory state of tumor-associated macrophages. CBD suppressed M2-like macrophage programs, promoted M1-like macrophage states, inhibited PI3K-AKT-associated alternative activation, and shifted macrophage metabolism from oxidative phosphorylation and fatty-acid oxidation toward glycolysis. These changes enhanced antitumor immunity and improved the response to anti-PD-1 therapy[93].

Spatial transcriptomics and spatial multi-omics

Spatial transcriptomics and spatial multi-omics preserve tissue architecture while profiling gene expression, metabolic activity, protein abundance, or cell–cell communication. They therefore overcome the loss of positional information caused by tissue dissociation during conventional single-cell sequencing.

Spatial approaches are particularly relevant to natural-product pharmacology because drug responses may be confined to tumor margins, immune-infiltrated regions, perivascular niches, hypoxic areas, specific renal tubular segments, hepatic lobular zones, or the intestinal crypt–villus axis. Mapping candidate pathways to these regions can guide subsequent immunofluorescence colocalization, spatial protein analysis, and region-specific metabolic validation.

Spatially resolved multi-omics was used to investigate the renoprotective effects of paeoniflorin in IgA nephropathy. The analysis indicated that paeoniflorin modulated *Oat-*, *Aco1-*, and *Fh*-associated metabolic reprogramming and altered communication between renal tubular cells and immune populations[94]. This study illustrates how spatially resolved molecular changes can connect natural-product activity to specific pathological niches.

Table 3. Question-driven selection of single-cell and spatial multi-omics strategies.

Research question	Recommended modality and tools	Expected output	Key caution
Which cell populations respond to treatment?	scRNA-seq; Seurat/Scanpy; harmony or scVI when integration is required	Cell-type abundance, transcriptional states, and sample-level treatment effects	Dissociation and batch effects can mimic cell loss or induction; use biological replicates
Does treatment induce a cell-state transition?	scRNA-seq time course; monocle or Slingshot	Pseudotime ordering, branch points, and state-associated genes	Pseudotime is not direct lineage or chronological proof
Which cells communicate after treatment?	scRNA-seq or spatial data; CellChat/CellPhoneDB; NicheNet for ligand-to-target links	Altered ligand-receptor networks and predicted receiver-cell programs	Expression-based interactions require protein-level and perturbational validation
Is chromatin regulation altered?	scATAC-seq or RNA + ATAC multiome; ArchR/Signac	Accessible elements, motif activity, peak-to-gene links, and regulatory programs	Sparse peak counts and inferred links can reduce robustness
How should modalities be integrated?	Matched or unmatched multi-omics; Seurat WNN, MOFA+, totalVI or graph-based models	Shared latent states, modality-specific factors, and cross-modal regulatory links	Integration can obscure modality-specific biology; benchmark against unimodal results
Where does the response occur in tissue?	Spatial transcriptomics/proteomics; cell2location, Tangram or SPOTlight	Spatial niches, cell-state maps, and region-specific interactions	Spot resolution, deconvolution assumptions, and histological registration affect inference

Single-cell proteomics and post-translational-modification profiling

Single-cell transcriptomics does not necessarily predict protein abundance, stability, localization, or activity. Single-cell proteomics therefore provides a more direct view of the functional molecular state of individual cells. However, because proteins cannot be amplified in the same manner as nucleic acids, mass spectrometry-based single-cell proteomics remains constrained by limited sample input, protein dynamic range, peptide loss, missing values, and analytical throughput. Recent reviews and reporting guidelines have summarized the capabilities and quality-control requirements of this rapidly developing field[95–98]. Budnik et al. introduced single-cell proteomics by mass spectrometry (SCoPE-MS), demonstrating that the proteomes of individual mammalian cells could be quantified and used to distinguish cancer-cell types and differentiation states[99]. Schoof et al. subsequently applied quantitative single-cell proteomics to characterize cellular hierarchies in acute myeloid leukemia, revealing protein-level differences that were not fully captured by transcriptomic data[100]. Woo et al. developed a nested nanowell chip for high-throughput and high-efficiency preparation of single-cell proteomic samples[101]. Huffman et al. developed prioritized mass spectrometry, which increases proteome depth, sensitivity, and data completeness by preferentially analyzing peptides of interest[98]. Derks et al. further described strategies for increasing analytical depth and throughput and developed plexDIA, a multiplexed data-independent acquisition framework that improves throughput and data completeness in low-input and single-cell proteomics[102,103].

Mass spectrometry-based single-cell proteomics does not require a predefined antibody panel and is therefore suitable for detecting previously unanticipated protein changes. By contrast, antibody-oligonucleotide methods provide high-throughput measurements of selected proteins. Stoeckius et al. developed cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), which simultaneously measures transcriptomes and cell-surface proteins in individual cells[104]. CITE-seq is particularly useful for immune-cell classification, pharmacodynamic phenotyping, and localization of drug-responsive cell populations.

Post-translational modification profiling provides an additional functional layer. Phosphorylation, ubiquitination, acetylation, glycosylation, methylation, succinylation, lactylation, and lipidation can regulate protein activity, localization, stability, complex formation, and signal transduction. These events may therefore reflect pharmacological responses more directly than mRNA abundance or total protein expression.

Phosphoproteomics is particularly relevant to kinase and phosphatase signaling, receptor tyrosine kinase pathways, MAPK, PI3K-AKT, NF-κB, and JAK–STAT signaling[105]. Orsburn et al. used trapped ion mobility time-of-flight mass spectrometry to characterize proteins and multiple post-translational modifications in individual human cells, demonstrating the feasibility of resolving modification heterogeneity at single-cell resolution[95]. Blair et al. developed Phospho-seq, a scMulti-omics method that jointly profiles chromatin accessibility and intracellular proteins, including phosphorylated proteins[106].

In natural-product pharmacology, modification profiling can help determine whether target engagement produces a functional change in the candidate protein. Phosphoproteomics can assess systematic alterations in downstream substrates of a kinase pathway. Ubiquitinome analysis, protein-turnover measurements, and interactomics can determine whether a compound affects E3 ligases, deubiquitinases, or proteasomal degradation. Acetylation, succinylation, lactylation, and lipidation profiling may connect metabolic reprogramming to mitochondrial function, epigenetic regulation, or tumor immunity. Nevertheless, single-cell analysis of many modification classes remains technically immature and should be interpreted together with biochemical and genetic evidence.

Applications of single-cell multi-omics in natural product pharmacology

Single-cell multi-omics extends natural product pharmacology beyond average tissue responses by resolving cellular composition, cell state, regulatory activity, protein phenotypes, and spatial context. In traditional Chinese medicine and natural-product studies, scRNA-seq is particularly useful for identifying cell-type-specific actions and multicellular response networks, provided that these associations are linked to independent mechanistic validation[107].

Cell-type-specific drug effects

Following natural product treatment, drug responses often vary markedly across different cell types. For example, the same compound may simultaneously inhibit tumor cell proliferation, activate certain immune cell populations, and alter the states of stromal or endothelial cells. Single-cell transcriptomics enables comparison of cell composition and functional state differences between treated and model groups at subpopulation resolution, thereby identifying the major target cell populations affected by the drug. Compared with bulk omics approaches, this type of analysis is better suited to reveal the cell type-dependent pharmacological effects of natural products, particularly in disease models with complex cellular compositions, such as cancer, inflammation, kidney disease, metabolic disorders, and tissue repair. In a study of berberine in an HCC model, Hu et al. showed that scRNA-seq can be used to characterize changes in the composition and functional state of intrahepatic T-cell subsets after natural product administration, suggesting that its antitumor effects may not arise solely from direct inhibition of tumor cells, but may also be closely associated with remodeling of the tumor immune microenvironment[91].

Tumor microenvironment remodeling and immune reprogramming

The tumor microenvironment (TME) is a critical component in studies of the antitumor pharmacological effects of natural products. In addition to directly suppressing tumor cell proliferation, many natural products can reshape the tumor immune landscape by modulating macrophage polarization, T-cell exhaustion, dendritic cell antigen presentation, NK-cell cytotoxicity, cancer-associated fibroblast activation, and immune checkpoint signaling. scMulti-omics technologies enable simultaneous characterization of changes in cellular composition, functional state transitions, ligand-receptor communication, and transcription factor regulatory networks, and are therefore particularly well suited for elucidating natural product-mediated immune remodeling mechanisms. Recent research reported that extracellular vesicles derived from *Artemisia annua* could remodel the immune microenvironment of breast cancer by regulating macrophage polarization and promoting T-cell infiltration. Wang et al. further used scRNA-seq to delineate the resulting changes in immune cell states, providing single-cell-level evidence that plant-derived bioactive substances and their delivery systems are involved in tumor immune regulation[108].

Applications beyond oncology and current evidence gaps

The value of cell-resolved pharmacology extends to cardiovascular, neurodegenerative, and metabolic diseases. In coronary artery disease, a recent single-cell analysis combined disease-cell-state

profiling with network-based prediction and experimental evaluation of *Ginkgo biloba* extract, providing a cell-contextual view of endoplasmic-reticulum-stress pathways and candidate therapeutic mechanisms[109]. In metabolic and renal disorders, the rosmarinic acid and paeoniflorin studies discussed above illustrate how scRNA-seq or spatial multi-omics can separate tubular, immune, fibrotic, and metabolic responses that would be averaged in bulk tissue[92,94].

Natural-product-specific single-cell studies in neurodegenerative disorders remain comparatively limited. Single-cell-led network analysis in Alzheimer's disease has nevertheless shown how cell-type-specific disease signatures, genetic evidence, and drug-response information can be integrated to prioritize therapeutic candidates[110]. This work provides a methodological template rather than direct natural-product target evidence. The scarcity of studies that combine a natural product, cell-resolved disease biology, direct target engagement, and causal validation should therefore be regarded as a major opportunity for the field.

AI-assisted strategies relevant to natural product target discovery

AI can support several stages of natural product research, but its relevance to direct-target discovery depends on the question being modeled. Candidate prioritization requires chemical, structural, biological, or perturbational inputs that can be linked to experimentally testable proteins. By contrast, drug repurposing and pharmacokinetic prediction are useful supporting applications but do not independently identify direct targets. This section therefore prioritizes target-prediction tasks, model inputs and outputs, applicability domains, and validation requirements[111–118].

Supporting application: AI-assisted repurposing and candidate prioritization

AI-assisted repurposing uses chemical structure, known targets, disease signatures, knowledge graphs, and perturbational data to rank compounds for new indications. Classic examples such as sildenafil and tamoxifen illustrate the broader logic of repurposing, but they do not constitute natural-product target identification[119–123]. For natural products, a repurposing score should therefore be treated as an upstream hypothesis: chemical identity, exposure, target engagement, and disease-relevant activity still require experimental confirmation (Fig. 4).

Knowledge graphs can connect compounds, proteins, pathways, phenotypes, and diseases and rank mechanistically coherent hypotheses. CoV-KGE and TxGNN illustrate the scalability of graph-based prediction[124,125], but in natural-product research, the proposed path should be traceable to compound-specific structural or experimental evidence rather than generic herb-pathway associations.

Multimodal platforms such as AgeXtend, PTD-DEP, DeepDrugDiscovery, and GPS show how chemical structures, perturbational signatures, pathway information, and toxicity features can be combined[126–129]. Here, they are treated as methodological precedents rather than evidence of direct natural-product-protein binding.

Accordingly, repurposing is treated here as an upstream hypothesis-generation step. Natural product candidates emerging from these models should be advanced only after confirmation of chemical identity, bioactivity, exposure, target engagement, and disease-relevant efficacy.

Supporting application: AI-assisted pharmacokinetic prediction

Physiologically based pharmacokinetic (PBPK) modeling supports formulation, route, dose, and exposure decisions, including in populations for which direct clinical pharmacokinetic data are limited[130]. Machine-learning models can estimate physicochemical or pharmacokinetic parameters from molecular structure and then supply or complement mechanistic PBPK models[131]. This role is relevant to natural products because inadequate or variable exposure can confound a target-validation study, but PBPK prediction does not identify a molecular target.

Representative studies have coupled structure-based machine learning with PBPK to predict plasma exposure, used PBPK simulations to generate training data, and integrated quantitative structure–activity relationships with PBPK models for nanoparticle biodistribution[132–135]. Reported performance is dataset- and endpoint-dependent, and cross-species, cross-population, and real-world generalization remain key limitations.

At present, AI-PBPK models should therefore be used to support exposure and dose selection, with prospective pharmacokinetic validation in the intended biological setting.

AI-assisted target identification: task-oriented categories

AI-assisted target identification combines chemical representations, protein information, prior interaction data, omics responses, and biological networks to rank candidate targets for experimental testing[2] (Fig. 5). The appropriate model depends on whether the starting point is a compound, a protein target, a disease-associated cell state, or a perturbational signature. For natural products, structural novelty, stereochemistry, metabolites, and limited confirmed target annotations should be considered explicitly.

Drug-based target identification

Machine-learning models can combine chemical, protein, and pharmacological features to estimate drug-target interactions. Graph neural networks and convolutional architectures are among the approaches used for this purpose, although reported performance depends strongly on dataset construction and evaluation design.

Daina & Zoete used known compound-target pairs and structural similarity for reverse target prediction; the true target was ranked first for more than 51% of compounds in an external test set[136]. Rao et al. combined chemical-similarity prediction with cross-species transcriptomic evidence to identify non-canonical interactions among 2,766 approved drugs[137]. Natural-product examples provide a more relevant test of this strategy. For bufalin, multi-platform prediction and multitask QSAR prioritized ESR1, and SPR, biotin pull-down, CETSA, and cellular localization supported direct binding to ERα[138]. In a separate study, a graph-convolutional prediction nominated STING as a target of ginkgetin; biochemical and cellular experiments then supported ginkgetin-STING binding and pathway inhibition[139].

These examples show that AI is most informative when it narrows the candidate space and specifies a falsifiable compound-protein hypothesis. Prediction alone should not be equated with target identification.

Target-oriented screening and structure-enabled prioritization

A substantial fraction of disease-relevant proteins remains difficult to modulate with conventional small molecules because of conformational flexibility, shallow or transient pockets, conserved

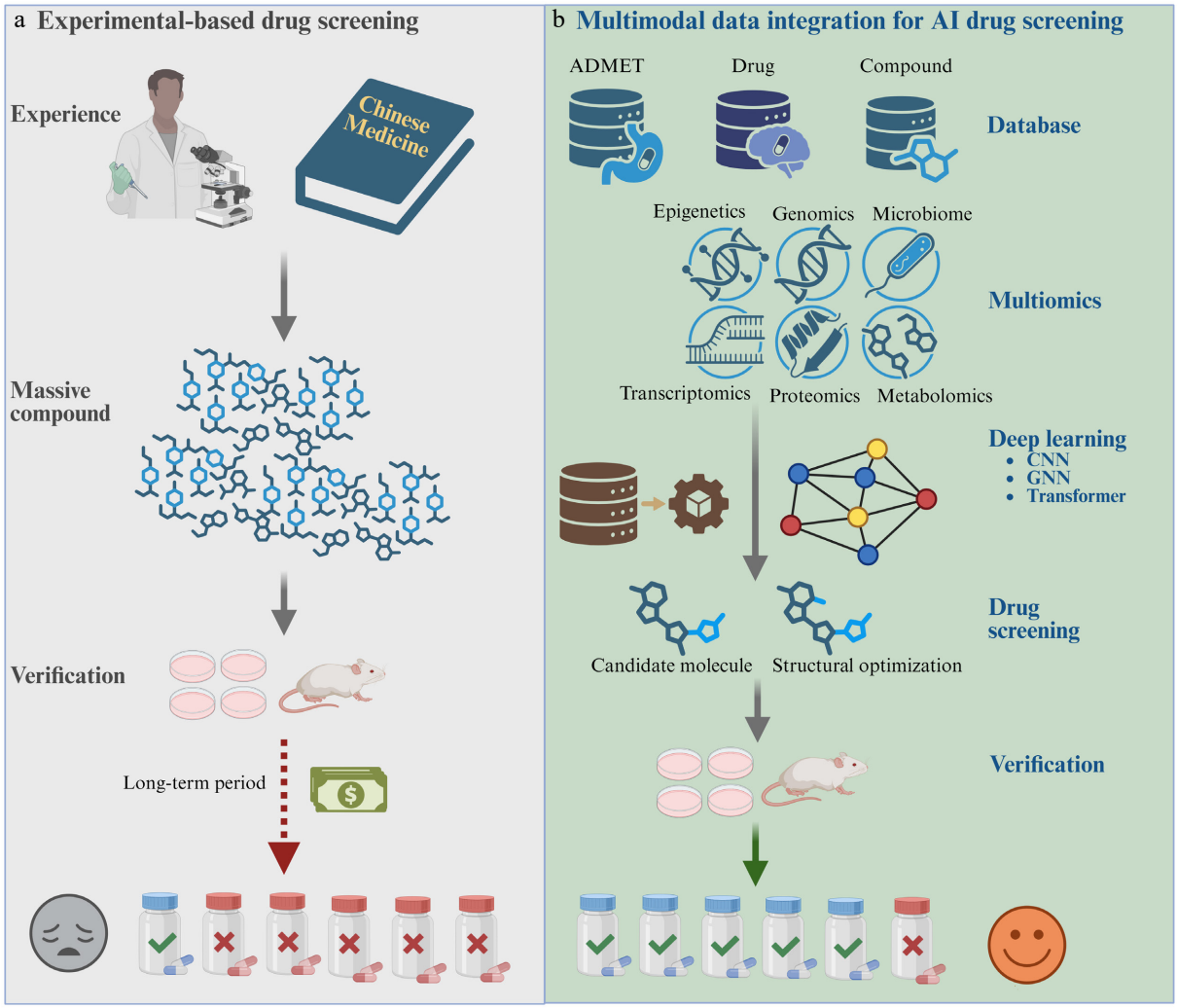

Fig. 4 AI-assisted candidate prioritization as an upstream component of natural product research. (a) Conventional screening uses prior knowledge and large experimental libraries to select compounds, followed by cellular and animal validation. (b) AI-assisted screening integrates chemical, ADMET, target, disease, and multi-omics information using machine-learning or deep-learning models to prioritize candidates before experiments. This workflow reduces the initial search space but does not replace target-engagement or efficacy validation.

active sites, or reliance on protein–protein interactions. Structure prediction and AI-assisted modeling have expanded the set of proteins that can be evaluated computationally, but model-derived pockets and scores still require experimental confirmation[140–143].

CNN- and GNN-based models can encode local pocket geometry, molecular graphs, and protein–ligand interaction patterns. Contrastive models such as DrugCLIP place protein pockets and small molecules in a shared representation space, whereas structure-aware frameworks such as DeepDegradome link binding prediction to ligand or degrader design[144,145]. These methods can accelerate prioritization but remain sensitive to training-set composition and benchmark design.

Natural products such as ginkgetin, oleandrin, berberine, and palmatine have emerged from machine-learning-assisted screens[146,147]. Their inclusion demonstrates that AI can explore natural-product chemical space, but activity prediction should be followed by compound-quality control, direct binding, target-selective perturbation, and disease-relevant functional validation. Table 4 summarizes the principal AI categories.

Data quality, applicability domain, and validation requirements

Model performance is constrained by the evidence used for training. Natural product-target databases overrepresent intensively studied compounds and proteins, contain heterogeneous assay types, and rarely provide reliable negative interactions. Random train-test splits can place closely related scaffolds or homologous proteins in both sets and thereby overestimate generalization. Scaffold-aware, protein-family-aware, temporal, and external validation should therefore be preferred when evaluating target-prediction models.

Applicability domains are especially important for natural products because stereochemistry, glycosylation, tautomerism, covalent reactivity, active metabolites, and multi-component preparations may be incompletely represented. Feature attribution or attention weights do not necessarily provide mechanistic explanations. AI outputs should therefore be reported as ranked hypotheses with uncertainty, provenance, and domain limitations, and should be tested prospectively using orthogonal binding, cellular engagement, genetic perturbation, and rescue experiments.

Practical integration of AI and single-cell multi-omics

The central technical problem in AI-assisted single-cell multi-omics is not simply dimensionality reduction, but the integration of sparse measurements with chemical and biological context. A useful model should specify its inputs, representation strategy, fusion

https://doi.org/10.48130/targetome-0026-0036

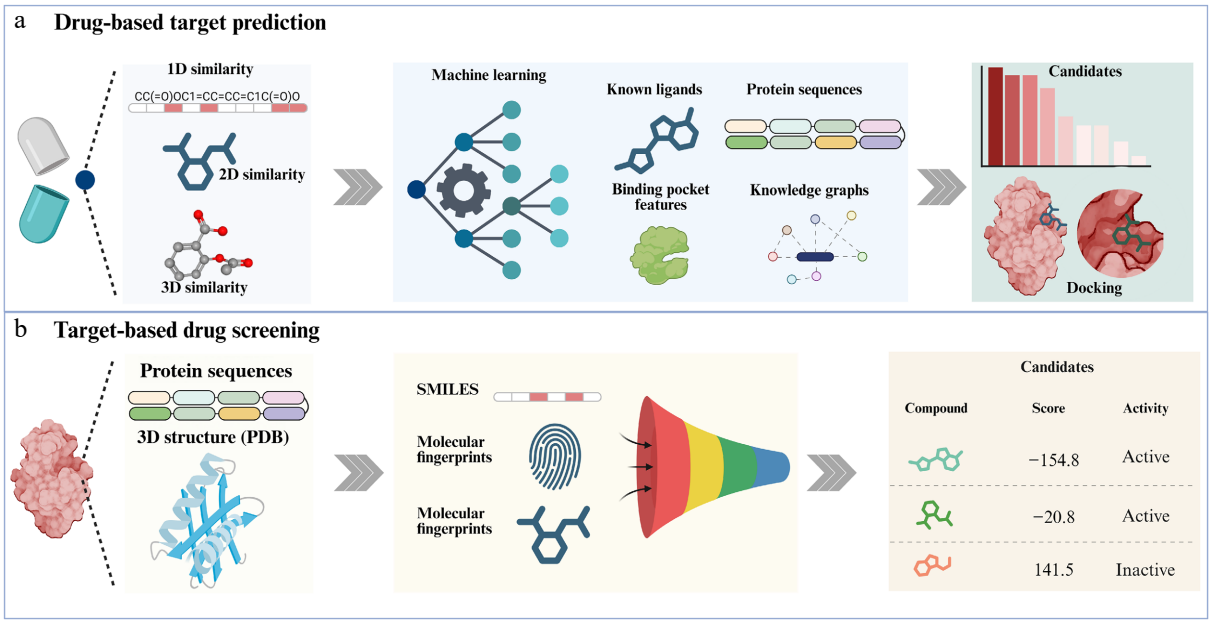

Fig. 5 AI-assisted drug-target identification and target-based drug screening. (a) Drug-based target prediction, which integrates 1D, 2D, and 3D similarity information of drugs together with multi-source data such as known ligands, protein sequences, binding-pocket features, and knowledge graphs to identify potential targets, followed by molecular docking to validate candidate hits. (b) Target-based drug screening, which uses protein sequence and three-dimensional structural information, combined with SMILES representations, molecular fingerprints, and other features, to build predictive models that score and rank candidate compounds, thereby enabling efficient virtual screening and prioritization of promising drug candidates.

Table 4. Task-oriented comparison of AI approaches for natural product target research.

AI category	Typical inputs and outputs	Strength for natural products	Main limitation and validation requirement
Ligand-based prediction	Fingerprints, SMILES, molecular graphs, known compound-target pairs→ranked targets	Fast reverse screening and off-target nomination when related ligands are annotated	Weak extrapolation beyond known chemical space; use scaffold-aware validation and direct-binding assays
Structure-based prediction	Protein structures or pockets and ligand conformers→docking poses, scores, or target ranks	Can suggest binding sites and rationalize stereochemical interactions	Protein flexibility and scoring errors; validate affinity, pose-dependent mutations, and cellular engagement
GNN/molecular representation learning	Molecular and interaction graphs→learned embeddings and interaction probabilities	Captures nonlinear structural and network features	Data leakage and opaque features; require external or prospective testing
Knowledge-graph/network reasoning	Compound-target-disease-pathway relations→mechanistic paths or candidate targets	Integrates sparse, heterogeneous evidence and supports polypharmacology hypotheses	Database popularity bias and correlation without causality; trace evidence and validate each edge experimentally
Perturbation/omics modeling	Drug-response, CRISPR, transcriptomic, proteomic, or single-cell signatures→target or pathway ranking	Links compounds to context-specific cell states and phenotypes	May prioritize downstream effectors rather than binders; combine with target-fishing and engagement assays
Multimodal/foundation models	Chemical, structural, omics, imaging, and text data→joint representations and multiple predictions	Potential to integrate natural-product structure with cell-context and disease knowledge	Modality imbalance, interpretability, and domain shift; benchmark each output and perform prospective validation

point, output, uncertainty, and experimental endpoint. Depending on the question, inputs may include gene-count matrices, chromatin-accessibility peaks, protein abundances, spatial coordinates, histological images, treatment labels, compound structures, protein sequences or structures, ligand-receptor priors, and pathway networks[148,149].

Data inputs, model architectures, and feature integration

Autoencoders and variational autoencoders learn latent cellular states from sparse high-dimensional data; graph neural networks represent relationships among compounds, targets, genes, cells, and diseases; CNNs extract features from histological or spatial images; and transformer or attention-based models learn long-range and cross-modal dependencies. The architecture should be selected according to the data-generating process rather than by model novelty alone.

Feature integration can occur at several stages. Early fusion concatenates normalized features before modeling but is vulnerable to scale imbalance and missing modalities. Intermediate fusion encodes each modality separately and aligns or combines latent representations, as illustrated by contrastive and graph-based approaches such as scMDCF and scMAGCA[150,151]. Late fusion combines independently trained predictions and can be more robust when modalities are unmatched. Graph-based integration preserves explicit biological relationships, whereas contrastive learning aligns matched cells, samples, or perturbations across modalities.

Expected outputs include treatment-responsive cell states, compound-cell-type associations, candidate-target rankings, gene regulatory programs, ligand-receptor networks, spatially restricted response regions, and predicted perturbation outcomes. A model that outputs a cell state or pathway does not automatically identify a direct target; linking chemical features or target priors to the output is necessary for target-focused inference.

Analytical outputs, natural-product applications, and validation

Existing frameworks illustrate distinct output types. CellNavi models factors that drive cell-state transitions; SIDISH and DEGAS connect single-cell states with clinical risk; scRank infers drug-responsive cell types from target-perturbed regulatory networks; and scFOCAL predicts sensitive or resistant tumor-cell populations[152–156]. These are valuable methodological precedents, but most were not developed specifically for direct natural-product target identification.

For natural products, a practical workflow is to use single-cell or spatial data to define the responsive cell compartment, combine chemical and target information to rank proteins active in that compartment, and then test those candidates using target-fishing and engagement assays. Independent datasets should be used to reproduce the cell state, while sorted-cell experiments, spatial colocalization, target knockdown or knockout, binding-deficient mutants, and rescue studies should test the inferred mechanism.

AI virtual-cell models may eventually predict how a natural product shifts molecular and cellular states[157]. At present, however, their utility is limited by incomplete perturbation maps, uneven cell-type coverage, uncertain cross-study generalization, and limited prospective validation. The near-term priority is therefore not an unrestricted virtual cell, but transparent models whose predictions are linked to experimentally measurable target-engagement and cell-state endpoints.

An evidence-informed closed-loop framework for natural product target discovery

Building on our previous framework linking target identification with single-cell multi-omics[1], we propose a closed-loop strategy in which experimental and computational evidence are connected through explicit validation gates (Fig. 6). The framework is intended to guide study design rather than imply that AI or single-cell analysis can replace direct target confirmation.

Phenotypic and disease-context data provide the starting point. Single-cell and spatial analyses identify responsive cell populations and regulatory programs, while experimental target fishing and AI-assisted prediction generate candidate proteins. The first validation gate establishes direct binding or cellular target engagement using competition, affinity, structural, or stability-based assays. The second gate establishes functional causality using genetic perturbation, binding-site mutation, rescue, organoid, and *in vivo* experiments. Validated targets and cell-state responses are then fed back into model refinement and subsequent experimental design. Dashed or predictive connections should be interpreted as hypotheses until they pass these gates.

A practical implementation can begin either from a compound or from a disease-associated cell state. In a compound-first design, chemical structure and phenotype are used for AI prediction and experimental target fishing, followed by engagement and causal validation; single-cell or spatial data then define the responsive cell

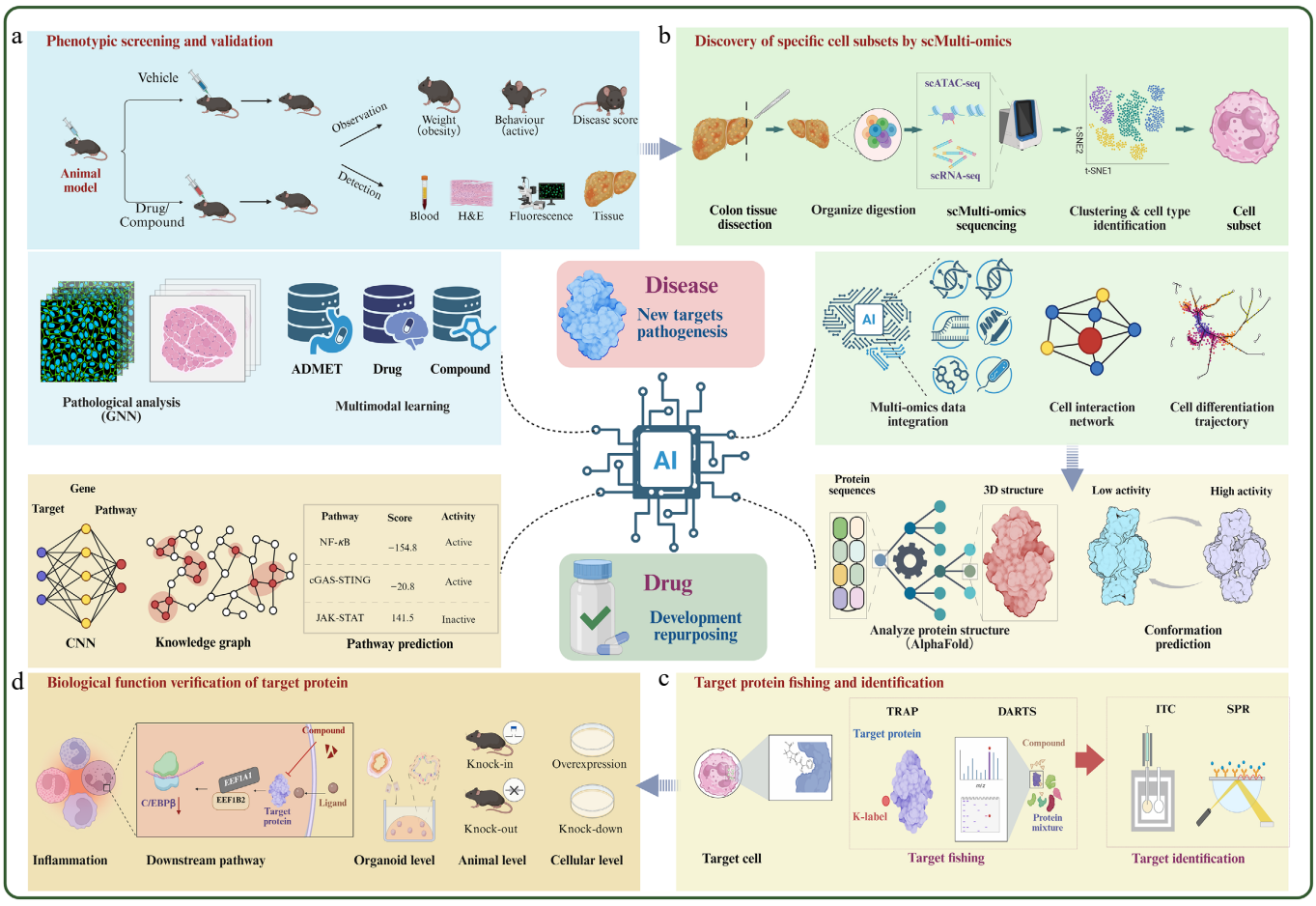

Fig. 6 Evidence-informed closed-loop framework for natural product target discovery.

compartment and downstream network. In a cell-state-first design, scRNA-seq, scATAC-seq, or spatial analysis identifies a treatment-responsive population, after which target fishing is performed in the relevant cellular context and AI integrates chemical, structural, and regulatory evidence. In both designs, the final claim should distinguish direct target, required pathway component, and secondary response marker.

The framework also clarifies the role of negative evidence. Failure of a predicted target to bind, lack of cellular engagement, or persistence of the phenotype after target depletion should feed back to revise the model rather than be omitted. Conversely, concordance across independent prediction, target fishing, direct binding, mutation, and rescue provides progressively stronger evidence. No single study currently implements every element of this framework at full scale. The framework should therefore be used as an evidence map: it shows which claims have passed nomination, engagement, causality, and cell-context gates and which remain provisional.

Representative studies mapped to the framework

The celastrol-LRP1 study in psoriasis illustrates a cell-context-first route. Single-cell analyses identified a fibroblast-macrophage communication circuit; biochemical target studies then linked celastrol to the LRP1 β-chain, and disruption of the nuclear LRP1-c-Jun interaction connected engagement to reduced CCL2-dependent signaling[91]. The bufalin-ERα study illustrates a prediction-first route: computational prioritization was followed by SPR, biotin pull-down, CETSA, and cellular localization, converting a ranked target into an experimentally supported interaction[139].

The ginkgetin-STING study similarly combines graph-based nomination with biochemical and cellular testing[140]. These studies validate different segments of Fig. 6 rather than the complete loop. In each case, the remaining evidential gap, such as binding-site mutation, cell-type-specific rescue, or prospective model testing, should be stated explicitly. Their comparison supports the central conclusion of this review: confidence increases when computational prioritization, orthogonal engagement evidence, causal perturbation, and disease-relevant cellular context converge.

Conclusions and future prospects

Natural product target research should be evaluated as an evidence chain rather than as a competition among technologies. Experimental target fishing generates candidates through distinct physicochemical readouts; biophysical and cellular assays determine whether engagement is direct and occurs in the relevant context; genetic and pharmacological perturbations establish functional causality; and single-cell or spatial analyses identify the cell states and tissue niches in which the validated target acts.

AI can accelerate this process by reducing the candidate space, integrating chemical and biological evidence, and predicting target-cell-state relationships. Its outputs nevertheless remain dependent on data quality, applicability domain, and model design. Similarly, single-cell multi-omics resolves heterogeneity but does not independently prove physical binding. Both should therefore be embedded in a closed-loop workflow with prospective experimental validation and explicit reporting of negative or discordant results.

The most urgent priorities are natural-product-specific benchmark datasets with assay provenance and reliable negatives; scaffold-, protein-family-, and time-aware model evaluation; interpretable multimodal models linking compounds to target engagement and cell states; spatially resolved pharmacology; and blinded prospective studies. Progress in these areas will determine whether integrated AI, target-fishing, and single-cell strategies move from retrospective association toward reproducible and causal natural product pharmacology.

Ethical statements

Not applicable.

Author contributions

The authors confirm their contributions to the paper as follows: conception and design: Sun Y, Du H; draft manuscript preparation: Du H, Chen N; project administration: Sun Y. All authors reviewed the results 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 work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2025ZD0546900) and the National Natural Science Foundation of China (No. U24A20794). We gratefully acknowledge ChatGPT for its assistance in polishing the language and revising this manuscript. All figures were prepared using a licensed version of BioRender (www.biorender.com/).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 21 June 2026; Revised 24 July 2026; Accepted 31 July 2026; Published online 18 August 2026

References

[1] Zhu Y, Ouyang Z, Du H, Wang M, Wang J, et al. 2022. New opportunities and challenges of natural products research: when target identification meets single-cell multiomics. *Acta Pharmaceutica Sinica B* 12:4011–4039

[2] Li G, Shi Q, Wu Q, Sui X. 2025. Target identification of natural products in cancer with chemical proteomics and artificial intelligence approaches. *Cancer Biology & Medicine* 22:549–597

[3] Liu TT, Zeng KW. 2025. Recent advances in target identification technology of natural products. *Pharmacology & Therapeutics* 269:108833

[4] Bordukova M, Makarov N, Rodriguez-Esteban R, Schmich F, Menden MP. 2024. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opinion on Drug Discovery* 19:33–42

[5] You Y, Lai X, Pan Y, Zheng H, Vera J, et al. 2022. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduction and Targeted Therapy* 7:156

[6] Rozera T, Pasolli E, Segata N, Ianiro G. 2025. Machine learning and artificial intelligence in the multi-omics approach to gut microbiota. *Gastroenterology* 169:487–501

[7] Lee S, Kim J, Jung HU, Kim D, Cho E, et al. 2026. Artificial intelligence and multiomics integration for Parkinson's disease drug development. *Molecules and Cells* 49:100343

[8] Corrales F, Cardinale V. 2026. Multiomics- and artificial intelligence-powered research platforms for enhancing understanding and

prediction of the cholangiocarcinoma patient journey. *Gut* 75:1272−1274
[9] Liu Y, Jiang JJ, Du SY, Mu LS, Fan JJ, et al. 2024. Artemisinins ameliorate polycystic ovarian syndrome by mediating LONP1-CYP11A1 interaction. *Science* 384:eadk5382
[10] Wang Y, Zhang Y, Luo H, Wei W, Liu W, et al. 2024. Identification of USP2 as a novel target to induce degradation of KRAS in myeloma cells. *Acta Pharmaceutica Sinica B* 14:5235−5248
[11] Hueber W, Kidd BA, Tomooka BH, Lee BJ, Bruce B, et al. 2005. Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis & Rheumatism* 52:2645−2655
[12] Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, et al. 2004. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proceedings of the National Academy of Sciences of the United States of America* 101:16594−16599
[13] Zhao MM, Ren TT, Wang JK, Yao L, Liu TT, et al. 2025. Endoplasmic reticulum membrane remodeling by targeting reticulon-4 induces pyroptosis to facilitate antitumor immune. *Protein & Cell* 16:121−135
[14] Zhang XW, Feng N, Liu YC, Guo Q, Wang JK, et al. 2022. Neuroinflammation inhibition by small-molecule targeting USP7 noncatalytic domain for neurodegenerative disease therapy. *Science Advances* 8:eabo0789
[15] 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
[16] Siriwongsup S, Schmoker AM, Ficarro SB, Marto JA, Kim J. 2024. Bioorthogonally activated reactive species for target identification. *Chem* 10:1306−1315
[17] Wang X, Liew SS, Huang J, Hu Y, Wei X, et al. 2024. Dual-locked enzyme-activatable bioorthogonal fluorescence turn-on imaging of senescent cancer cells. *Journal of the American Chemical Society* 146:22689−22698
[18] 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
[19] Gao P, Wang J, Qiu C, Zhang H, Wang C, et al. 2024. Photoaffinity probe-based antimalarial target identification of artemisinin in the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *iMeta* 3:e176
[20] Martín-Acosta P, Meng Q, Klimek J, Reddy AP, David L, et al. 2022. A clickable photoaffinity probe of betulinic acid identifies tropomyosin as a target. *Acta Pharmaceutica Sinica B* 12:2406−2416
[21] Li F, Cai C, Wang F, Zhang N, Zhao Q, et al. 2025. 20(S)-ginsenoside Rg3 suppresses gastric cancer cell proliferation by inhibiting E2F-DP dimerization. *Phytomedicine* 141:156740
[22] Peng W, Shi D, Xu D, Wang X, Cai Y, et al. 2026. Identification of Bruceine A as a novel HSP90AB1 inhibitor for suppressing hepatocellular carcinoma growth. *Journal of Advanced Research* 82:863−879
[23] Lin C, Wan Y, Huo Q, Liu D, Liu X, et al. 2025. Chemoproteomics reveals ailanthone directly binds to PKM2 to inhibit the progression of hepatocellular carcinoma. *Phytomedicine* 143:156886
[24] Wu Y, Li Y, Huang Y, Li Q, Li Z, et al. 2025. Nobiletin promotes ferroptosis in breast cancer through targeting AKR1C1-mediated ubiquitination and degradation of GPX4. *Phytomedicine* 146:157074
[25] Wu Y, Yang Y, Wang W, Sun D, Liang J, et al. 2022. PROTAC technology as a novel tool to identify the target of lathyrane diterpenoids. *Acta pharmaceutica Sinica B* 12:4262−4265
[26] Ni Z, Shi Y, Liu Q, Wang L, Sun X, et al. 2024. Degradation-based protein profiling: a case study of celastrol. *Advanced Science* 11:2308186
[27] Lomenick B, Hao R, Jonai N. 2010. Target identification using drug affinity responsive target stability (DARTS). *Science-Business eXchange* 3:71
[28] Lomenick B, Jung G, Wohlschlegel JA, Huang J. 2011. Target identification using drug affinity responsive target stability (DARTS). *Current Protocols in Chemical Biology* 3:163−180
[29] Guo W, Zhou H, Wang J, Lu J, Dong Y, et al. 2024. Aloperine suppresses cancer progression by interacting with VPS4A to inhibit autophagosome-lysosome fusion in NSCLC. *Advanced Science* 11:e2308307
[30] Hu J, Liu W, Zou Y, Jiao C, Zhu J, et al. 2024. Allosterically activating SHP2 by oleanolic acid inhibits STAT3−Th17 axis for ameliorating colitis. *Acta Pharmaceutica Sinica B* 14:2598−2612
[31] 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
[32] Miettinen TP, Björklund M. 2014. NQO2 is a reactive oxygen species generating off-target for acetaminophen. *Molecular Pharmaceutics* 11:4395−4404
[33] Ji H, Lu X, Zhao S, Wang Q, Liao B, et al. 2023. Target deconvolution with matrix-augmented pooling strategy reveals cell-specific drug-protein interactions. *Cell Chemical Biology* 30:1478−1487.e7
[34] Liu R, Zhang Y, Zou H, Zhang M, Yang Z, et al. 2026. Ginkgolic acid targets HSPA8 to trigger ferroptosis in hepatocellular carcinoma *via* chaperone-mediated autophagy-dependent GPX4 degradation. *Pharmaceutical Biology* 64:514−535
[35] Yang A, Zeng K, Huang H, Liu D, Song X, et al. 2023. Usenamine A induces apoptosis and autophagic cell death of human hepatoma cells *via* interference with the Myosin-9/actin-dependent cytoskeleton remodeling. *Phytomedicine* 116:154895
[36] Li Y, Dong M, Qin H, An G, Cen L, et al. 2025. Mulberrin suppresses gastric cancer progression and enhances chemosensitivity to oxaliplatin through HSP90AA1/PI3K/AKT axis. *Phytomedicine* 139:156441
[37] Li K, Chen S, Wang K, Wang Y, Xue L, et al. 2025. A peptide-centric local stability assay enables proteome-scale identification of the protein targets and binding regions of diverse ligands. *Nature Methods* 22:278−282
[38] Strickland EC, Geer MA, Tran DT, Adhikari J, West GM, et al. 2013. Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation. *Nature protocols* 8:148−161
[39] Ogburn RN, Jin L, Meng H, Fitzgerald MC. 2017. Discovery of tamoxifen and N-desmethyl tamoxifen protein targets in MCF-7 cells using large-scale protein folding and stability measurements. *Journal of Proteome Research* 16:4073−4085
[40] Tian Y, Wan N, Zhang H, Shao C, Ding M, et al. 2022. Chemoproteomic mapping of glycolytic targetome in cancer cells. *Nature Chemical Biology* 19:1480−1491
[41] Yan W, Wang D, Wan N, Wang S, Shao C, et al. 2022. Living cell-target responsive accessibility profiling reveals silibinin targeting ACSL4 for combating ferroptosis. *Analytical Chemistry* 94:14820−14826
[42] Yi J, Ye Z, Xu H, Zhang H, Cao H, et al. 2024. EGCG targeting STAT3 transcriptionally represses PLXNC1 to inhibit M2 polarization mediated by gastric cancer cell-derived exosomal miR-92b-5p. *Phytomedicine* 135:156137
[43] Ni H, Zhang Z, Lu Y, Liu Y, Zhou Y, et al. 2025. Trace component fishing strategy based on offline two-dimensional liquid chromatography combined with PRDX3-surface plasmon resonance for *Uncaria* alkaloids. *Journal of Pharmaceutical Analysis* 15:101244
[44] Tan QM, Li M, Zhu JM, Liao BZ, Kong LY, et al. 2025. Surface plasmon resonance guided identification of quinolone alkaloids from the fruits of *Tetradium ruticarpum* as FSP1 inhibitors. *Journal of Natural Products* 88:1919−1927
[45] Wei J, Zhang J, Hu F, Zhang W, Wu Y, et al. 2024. Anti-psoriasis effect of 18β-glycyrrhetinic acid by breaking CCL20/CCR6 axis through its vital active group targeting GUSB/ATF2 signaling. *Phytomedicine* 128:155524
[46] Huang W, Xie W, Liu H, Chen H, Ling Y, et al. 2026. Gut microbiota-derived xanthohumol protects against heatstroke by inhibiting macrophage pyroptosis in mice. *Journal of Advanced Research* 82:967−980
[47] Zhang X, Wang Q, Li Y, Ruan C, Wang S, et al. 2020. Solvent-induced protein precipitation for drug target discovery on the proteomic scale. *Analytical Chemistry* 92:1363−1371
[48] Zhang X, Wang K, Wu S, Ruan C, Li K, et al. 2022. Highly effective identification of drug targets at the proteome level by pH-dependent protein precipitation. *Chemical Science* 13:12403−12418

[49] Xu M, Moresco JJ, Chang M, Mukim A, Smith D, et al. 2018. SHMT2 and the BRCC36/BRISC deubiquitinase regulate HIV-1 Tat K63-ubiquitylation and destruction by autophagy. *PLoS Pathogens* 14:e1007071
[50] Steinhart Z, Pavlovic Z, Chandrashekhar M, Hart T, Wang X, et al. 2017. Genome-wide CRISPR screens reveal a Wnt–FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nature Medicine* 23:60–68
[51] Myszka DG, Rich RL. 2000. Implementing surface plasmon resonance biosensors in drug discovery. *Pharmaceutical Science & Technology Today* 3:310–317
[52] Rich RL, Myszka DG. 2000. Advances in surface plasmon resonance biosensor analysis. *Current Opinion in Biotechnology* 11:54–61
[53] Ward WH, Holdgate GA. 2001. Isothermal titration calorimetry in drug discovery. *Progress in Medicinal Chemistry* 38:309–376
[54] Chaires JB. 2008. Calorimetry and thermodynamics in drug design. *Annual Review of Biophysics* 37:135–151
[55] Uri A, Nonga OE. 2020. What is the current value of fluorescence polarization assays in small molecule screening? *Expert Opinion on Drug Discovery* 15:131–133
[56] Hall MD, Yasgar A, Peryea T, Braisted JC, Jadhav A, et al. 2016. Fluorescence polarization assays in high-throughput screening and drug discovery: a review. *Methods and Applications in Fluorescence* 4:022001
[57] Degorce F, Card A, Soh S, Trinquet E, Knapik GP, et al. 2009. HTRF: a technology tailored for drug discovery – a review of theoretical aspects and recent applications. *Current Chemical Genomics* 3:22–32
[58] Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S. 2010. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications* 1:100
[59] Seidel SA, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, et al. 2013. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* 59:301–315
[60] Schenone M, Dančík V, Wagner BK, Clemons PA. 2013. Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* 9:232–240
[61] Maveyraud L, Mourey L. 2020. Protein X-ray crystallography and drug discovery. *Molecules* 25:1030
[62] 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
[63] Lei B, Zhang M, Shi X, Feng N, Yin J, et al. 2025. Ganoderic acid T, a novel activator of pyruvate carboxylase, exhibits potent anti-liver cancer activity. *Metabolism* 170:156321
[64] Moore JD. 2015. The impact of CRISPR–Cas9 on target identification and validation. *Drug Discovery Today* 20:450–457
[65] Gao K, Zhang W, Xu D, Zhao M, Tao X, et al. 2025. Chikusetsusaponin IVa targeted YAP as an inhibitor to attenuate liver fibrosis and hepatic stellate cell activation. *Chinese Medicine* 20:36
[66] Zhao J, Tang Z, Selvaraju M, Johnson KA, Douglas JT, et al. 2022. Cellular target deconvolution of small molecules using a selection-based genetic screening platform. *ACS Central Science* 8:1424–1434
[67] Sinha S, Sinha N, Perales M, Tarrab A, Nguyen T, et al. 2025. DeepTarget predicts anti-cancer mechanisms of action of small molecules by integrating drug and genetic screens. *npj Precision Oncology* 9:340
[68] Li D, Yang C, Zhu JZ, Lopez E, Zhang T, et al. 2022. Berberine remodels adipose tissue to attenuate metabolic disorders by activating sirtuin 3. *Acta Pharmacologica Sinica* 43:1285–1298
[69] Ito T, Ando H, Suzuki T, Ogura T, Hotta K, et al. 2010. Identification of a primary target of thalidomide teratogenicity. *Science* 327:1345–1350
[70] Lopez-Girona A, Mendy D, Ito T, Miller K, Gandhi AK, et al. 2012. Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia* 26:2326–2335
[71] Krönke J, Udeshi ND, Narla A, Grauman P, Hurst SN, et al. 2014. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* 343:301–305
[72] Lu G, Middleton RE, Sun H, Naniong M, Ott CJ, et al. 2014. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* 343:305–309
[73] Heffner CS, Herbert Pratt C, Babiuk RP, Sharma Y, Rockwood SF, et al. 2012. Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. *Nature Communications* 3:1218
[74] Kim H, Kim M, Im SK, Fang S. 2018. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Laboratory Animal Research* 34:147–159
[75] Tabana Y, Babu D, Fahlman R, Siraki AG, Barakat K. 2023. Target identification of small molecules: an overview of the current applications in drug discovery. *BMC Biotechnology* 23:44
[76] Zou M, Zhou H, Gu L, Zhang J, Fang L. 2024. Therapeutic target identification and drug discovery driven by chemical proteomics. *Biology* 13:555
[77] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, et al. 2019. Comprehensive integration of single-cell data. *Cell* 177:1888–1902.e21
[78] Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* 16:1289–1296
[79] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. 2018. Deep generative modeling for single-cell transcriptomics. *Nature Methods* 15:1053–1058
[80] Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, et al. 2020. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* 21:111
[81] Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, et al. 2021. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature Methods* 18:1352–1362
[82] Browaeys R, Saelens W, Saeys Y. 2020. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* 17:159–162
[83] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, et al. 2021. Inference and analysis of cell-cell communication using CellChat. *Nature Communications* 12:1088
[84] Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, et al. 2022. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology* 40:661–671
[85] Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. 2021. Single-cell chromatin state analysis with Signac. *Nature Methods* 18:1333–1341
[86] Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, et al. 2021. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics* 53:403–411
[87] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* 6:377–382
[88] Liao Y, Liu Z, Zhang Y, Lu P, Wen L, et al. 2023. High-throughput and high-sensitivity full-length single-cell RNA-seq analysis on third-generation sequencing platform. *Cell Discovery* 9:5
[89] Wang Y, Lu H, Cheng L, Guo W, Hu Y, et al. 2024. Targeting mitochondrial dysfunction in atopic dermatitis with trilinolein: a triacylglycerol from the medicinal plant Cannabis fructus. *Phytomedicine* 132:155856
[90] Zhu Y, Zhao L, Yan W, Ma H, Zhao W, et al. 2025. Celastrol directly targets LRP1 to inhibit fibroblast-macrophage crosstalk and ameliorates psoriasis progression. *Acta Pharmaceutica Sinica B* 15:876–891
[91] Hu J, Shi Q, Xue C, Wang Q. 2024. Berberine protects against hepatocellular carcinoma progression by regulating Intrahepatic T cell heterogeneity. *Advanced Science* 11:e2405182
[92] Chen J, Zhang Q, Guo J, Gu D, Liu J, et al. 2024. Single-cell transcriptomics reveals the ameliorative effect of rosmarinic acid on diabetic nephropathy-induced kidney injury by modulating oxidative stress and inflammation. *Acta Pharmaceutica Sinica B* 14:1661–1676
[93] Sun X, Zhou L, Wang Y, Deng G, Cao X, et al. 2023. Single-cell analyses reveal cannabidiol rewires tumor microenvironment *via* inhibiting alternative activation of macrophage and synergizes with anti-PD-1 in colon cancer. *Journal of Pharmaceutical Analysis* 13:726–744
[94] Wang M, Yin F, Li P, Han J, Zheng Y, et al. 2026. Spatially resolved multi-omics reveals that paeoniflorin ameliorates IgA nephropathy *via* Oat, Aco1 and Fh-mediated metabolic reprogramming and tubuloimmune crosstalk. *Phytomedicine* 156:158263

[95] Orsburn BC, Yuan Y, Bumpus NN. 2022. Insights into protein post-translational modification landscapes of individual human cells by trapped ion mobility time-of-flight mass spectrometry. *Nature Communications* 13:7246
[96] Bennett HM, Stephenson W, Rose CM, Darmanis S. 2023. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nature Methods* 20:363−374
[97] Gatto L, Aebersold R, Cox J, Demichev V, Derks J, et al. 2023. Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments. *Nature Methods* 20:375−386
[98] Huffman RG, Leduc A, Wichmann C, Di Gioia M, Borriello F, et al. 2023. Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nature Methods* 20:714−722
[99] Budnik B, Levy E, Harmange G, Slavov N. 2018. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology* 19:161
[100] Schoof EM, Furtwängler B, Üresin N, Rapin N, Savickas S, et al. 2021. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nature Communications* 12:3341
[101] Woo J, Williams SM, Markillie LM, Feng S, Tsai CF, et al. 2021. Author Correction: High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nature Communications* 12:7075
[102] Derks J, Leduc A, Wallmann G, Huffman RG, Willetts M, et al. 2023. Increasing the throughput of sensitive proteomics by plexDIA. *Nature Biotechnology* 41:50−59
[103] Thielert M, Itang EC, Ammar C, Rosenberger FA, Bludau I, et al. 2023. Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth *via* a reference channel. *Molecular Systems Biology* 19:e11503
[104] Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, et al. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* 14:865−868
[105] Gerritsen JS, White FM. 2021. Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells. *Expert Review of Proteomics* 18:661−674
[106] Blair JD, Hartman A, Zenk F, Wahle P, Brancati G, et al. 2025. Phospho-seq: integrated, multi-modal profiling of intracellular protein dynamics in single cells. *Nature Communications* 16:1346
[107] Chen X, Liu J, Lu P, Zhou J, Jiang L, et al. 2025. Unraveling traditional Chinese medicine with single-cell RNA sequencing: current applications and future frontiers. *Phytomedicine* 149:157556
[108] Wang Y, Meng L, Su S, Zhao Y, Hu X, et al. 2025. Artemisia annua-derived extracellular vesicles reprogram breast tumor immune microenvironment *via* altering macrophage polarization and synergizing recruitment of T lymphocytes. *Chinese Medicine* 20:149
[109] Zhao FJ, Wang F, Qin C, Ye LL. 2026. Single cell profiling of ER stress in coronary artery disease and therapeutic mechanisms of *Ginkgo biloba* extract. *Scientific Reports* 16:14508
[110] Parolo S, Mariotti F, Bora P, Carboni L, Domenici E. 2023. Single-cell-led drug repurposing for Alzheimer's disease. *Scientific Reports* 13:222
[111] Ali MS, Alqahtani T, Shmrany HA, Gupta G, Goh KW, et al. 2026. Artificial Intelligence in drug discovery and development: transforming pharmaceutical innovation. *Drug Development Research* 87:e70281
[112] Mak KK, Pichika MR. 2019. Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today* 24:773−780
[113] Niazi SK, Mariam Z. 2025. Artificial intelligence in drug development: reshaping the therapeutic landscape. *Therapeutic Advances in Drug Safety* 16:20420986251321704
[114] Asfand-e-yar M, Hashir Q, Ali Shah A, Malik HAM, Alourani A, et al. 2024. Multimodal CNN-DDI: using multimodal CNN for drug to drug interaction associated events. *Scientific Reports* 14:4076
[115] Perdomo-Quinteiro P, Belmonte-Hernández A. 2024. Knowledge Graphs for drug repurposing: a review of databases and methods. *Briefings in Bioinformatics* 25:bbae461
[116] Sarkar C, Das B, Rawat VS, Wahlang JB, Nongpiur A, et al. 2023. Artificial intelligence and machine learning technology driven modern drug discovery and development. *International Journal of Molecular Sciences* 24:2026
[117] 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
[118] Zhang K, Yang X, Wang Y, Yu YF, Huang N, et al. 2025. Artificial intelligence in drug development. *Nature Medicine* 31:45−59
[119] Goldstein I, Lue TF, Padma-Nathan H, Rosen RC, Steers WD, et al. 2002. Oral sildenafil in the treatment of erectile dysfunction. *Journal of Urology* 167:1197−1203
[120] Shim JS, Liu JO. 2014. Recent advances in drug repositioning for the discovery of new anticancer drugs. *International Journal of Biological Sciences* 10:654−663
[121] Shagufta, Ahmad I. 2018. Tamoxifen a pioneering drug: an update on the therapeutic potential of tamoxifen derivatives. *European Journal of Medicinal Chemistry* 143:515−531
[122] Davies C, Pan H, Godwin J, Gray R, Arriagada R, et al. 2013. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *The Lancet* 381:805−816
[123] Zielińska A, Fornalik M, Szczepaniak M, Gimla M, Lemańska A, et al. 2026. Artificial intelligence in drug research and development: a review of methods and applications in drug repurposing. *Briefings in Bioinformatics* 27:bbag203
[124] Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, et al. 2024. A foundation model for clinician-centered drug repurposing. *Nature Medicine* 30:3601−3613
[125] Zeng X, Song X, Ma T, Pan X, Zhou Y, et al. 2020. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *Journal of Proteome Research* 19:4624−4636
[126] Arora S, Mittal A, Duari S, Chauhan S, Dixit NK, et al. 2025. Discovering geroprotectors through the explainable artificial intelligence-based platform AgeXtend. *Nature Aging* 5:144−161
[127] Dong Y, Xiao X, Zhuang XX, Wu W, Wang ZY, et al. 2026. DeepDrugDiscovery identifies blood-brain barrier permeable autophagy enhancers for Alzheimer's disease. *Nature Biomedical Engineering*
[128] Sun Y, Liu S, Chen L, Zhou Z, Yin X, et al. 2025. AI-driven discovery of dual antiaging and anti-AD therapeutics *via* PROTAC target deconvolution of a super-enhancer-regulated axis. *Science Advances* 11:eadz9283
[129] Xing J, Tan M, Leshchiner D, Sun M, Abdelgied M, et al. 2026. Deep-learning-based *de novo* discovery and design of therapeutics that reverse disease-associated transcriptional phenotypes. *Cell* 189:2556−2572.e19
[130] Talkington AM, Cao Y, Kearsley AJ, Lai SK. 2025. Opportunities for machine learning and artificial intelligence in physiologically-based pharmacokinetic (PBPK) modeling. *Advanced Drug Delivery Reviews* 227:115716
[131] Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, et al. 2023. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* 15:1916
[132] Mavroudis PD, Teutonico D, Abos A, Pillai N. 2023. Application of machine learning in combination with mechanistic modeling to predict plasma exposure of small molecules. *Frontiers in Systems Biology* 3:1180948
[133] Habiballah S, Reisfeld B. 2023. Adapting physiologically-based pharmacokinetic models for machine learning applications. *Scientific Reports* 13:14934
[134] Jia X, Teutonico D, Dhakal S, Psarellis YM, Abos A, et al. 2025. Application of machine learning and mechanistic modeling to predict intravenous pharmacokinetic profiles in humans. *Journal of Medicinal Chemistry* 68:7737−7750
[135] Chou WC, Chen Q, Yuan L, Cheng YH, He C, et al. 2023. An artificial intelligence-assisted physiologically-based pharmacokinetic model to predict nanoparticle delivery to tumors in mice. *Journal of Controlled Release* 361:53−63

[136] Daina A, Zoete V. 2024. Testing the predictive power of reverse screening to infer drug targets, with the help of machine learning. *Communications Chemistry* 7:105
[137] Rao M, McDuffie E, Sachs C. 2023. Artificial intelligence/machine learning-driven small molecule repurposing *via* off-target prediction and transcriptomics. *Toxics* 11:875
[138] Jiang S, Liu K, Jiang T, Li H, Wei X, et al. 2025. Harnessing artificial intelligence to identify Bufalin as a molecular glue degrader of estrogen receptor alpha. *Nature Communications* 16:7854
[139] Liu Y, Ye J, Fan Z, Wu X, Zhang Y, et al. 2025. Ginkgetin alleviates inflammation and senescence by targeting STING. *Advanced Science* 12:e2407222
[140] Sun Q, Wang H, Xie J, Wang L, Mu J, et al. 2025. Computer-aided drug discovery for undruggable targets. *Chemical Reviews* 125:6309–6365
[141] Alipourgivi F, Su J, Lu T. 2025. Cracking PRMT5: mechanistic insights, clinical advances, and AI-driven strategies. *Cancer Letters* 634:218075
[142] Abramson J, Adler J, Dunger J, Evans R, Green T, et al. 2024. Addendum: accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 636:E4
[143] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589
[144] Hu Q, Cao Y, Ren P, Zhang X, Li F, et al. 2026. DeepDegradome: a structure-aware deep learning framework for PROTAC and ligand generation against protein targets. *Proceedings of the National Academy of Sciences of the United States of America* 123:e2518248123
[145] Jia Y, Gao B, Tan J, Zheng J, Hong X, et al. 2026. Deep contrastive learning enables genome-wide virtual screening. *Science* 391:eads9530
[146] Smer-Barreto V, Quintanilla A, Elliott RJR, Dawson JC, Sun J, et al. 2023. Discovery of senolytics using machine learning. *Nature Communications* 14:3445
[147] Yoo H, Han SJ, Lee JE, Cho C, Hong D, et al. 2026. Discovery of natural RORγt inhibitor using machine learning, virtual screening, and *in vivo* validation. *Journal of Advanced Research* 84:1059–1071
[148] Liu Y, Zhu K, Peng W, Liu Z, Mao X. 2026. Multi-omics and artificial intelligence for precision drug discovery and potential clinical applications. *Signal Transduction and Targeted Therapy* 11:210
[149] Sabit H, Yadav AK, Salimy S, Sakr A, Abdel-Ghany S, et al. 2026. Integrating multi-omics and artificial intelligence for personalized breast cancer management: a guide to clinicians. *Cancer Letters* 649:218468
[150] Cheng Y, Su Y, Fan Y, Yang Y, Chen X, et al. 2026. Aligned cross-modal integration and regulatory heterogeneity characterization of single-cell multiomic data with deep contrastive learning. *Genome Medicine* 18:10
[151] Wang Y, Zhou Z, Liu W, Zhang P, Cheng Y, et al. 2026. Interpretable modality-aware mapping of gene regulation in single-cell multiomics with scMAGCA. *Nature Communications* 17:6459
[152] Johnson TS, Yu CY, Huang Z, Xu S, Wang T, et al. 2022. Diagnostic Evidence GAuge of Single cells (DEGAS): a flexible deep transfer learning framework for prioritizing cells in relation to disease. *Genome Medicine* 14:11
[153] Jolasun Y, Song K, Zheng Y, Wang J, Fonseca GJ, et al. 2025. SIDISH integrates single-cell and bulk transcriptomics to identify high-risk cells and guide precision therapeutics through *in silico* perturbation. *Nature Communications* 16:11271
[154] Li C, Shao X, Zhang S, Wang Y, Jin K, et al. 2024. scRank infers drug-responsive cell types from untreated scRNA-seq data using a target-perturbed gene regulatory network. *Cell Reports Medicine* 5:101568
[155] Suter RK, Jermakowicz AM, Veeramachaneni R, D'Antuono M, Zhang L, et al. 2026. Drug and single-cell gene expression integration identifies sensitive and resistant glioblastoma cell populations. *Nature Communications* 17:99
[156] Wang T, Pan Y, Ju F, Zheng S, Liu C, et al. 2025. CellNavi predicts genes directing cellular transitions by learning a gene graph-enhanced cell state manifold. *Nature Cell Biology* 27:1863–1874
[157] Bunne C, Roohani Y, Rosen Y, Gupta A, Zhang X, et al. 2024. How to build the virtual cell with artificial intelligence: priorities and opportunities. *Cell* 187:7045–7063