Acta Pharmaceutica Sinica B 2026;16(3):1233−1249

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

TOOLS

TeroACT: A terpenoid bioactivity landscape and discovery platform

XiaoJuan Shen[a,†], Shijia Yan[a,†], Xu Kang[a], Kangwei Xu[a], Yongxing Jian[a], Tao Zeng[a,b], Guohui Wan[a,*], Ruibo Wu[a,*]

[a]*State Key Laboratory of Anti-Infective Drug Discovery and Development, School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China*
[b]*School of Pharmaceutical Sciences, Hainan University, Haikou 570228, China*

Received 17 February 2025; received in revised form 29 July 2025; accepted 2 September 2025

KEY WORDS

Terpenoids;
Knowledge graph;
Deep learning;
Disease prediction;
Drug reposition;
Deep neural network;
Anti-melanoma;
Anti-inflammatory

Abstract Terpenoids exhibit diverse biological activities and thus have a wide range of pharmacological applications. In modern drug discovery, data-driven deep models play a crucial role in facilitating efficient feature representation and knowledge inference. To explore the uncharted bioactivity space of terpenoids, the construction of a multi-dimensional relational terpenoid database is essential for mapping terpenoid-bioactivity profiles. In this study, we first constructed a large-scale biological knowledge graph by integrating various data types, including terpenoid compounds, protein targets, cellular targets, genes, diseases, and their interrelationships. Subsequently, we developed a network-based disease prediction model, as well as optimized multiple compound-protein interaction prediction tools to extend the framework for activity research. These resources have been deployed on a user-friendly web platform (TeroACT) accessible at: http://terokit.qmclab.com/teroact/. Using *in silico* models within the TeroACT platform, we screened multiple terpenoid molecules for anti-melanoma activity. *In vitro* and *in vivo* animal models further validated the anti-migration and anti-proliferative effects of mollugin and columbianadin in melanoma. Additionally, integrated computational screening and experimental approaches identified numerous terpenoids with anti-inflammatory properties. In this sense, TeroACT fills the gap in terpenoid bioactivity study by providing a comprehensive data resource and AI-driven drug discovery tools.

This article is part of special issue entitled: Machine Learning in Drug Discovery published in Acta Pharmaceutica Sinica B.

*Corresponding authors.
E-mail addresses: wanguoh@mail.sysu.edu.cn (Guohui Wan), wurb3@mail.sysu.edu.cn (Ruibo Wu).
[†]These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.12.036

1. Introduction

Terpenoids are modified terpene compounds with multiple 5-carbon (C5) entities, such as isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which undergo methyl group rearrangements and additional functional modifications[1]. Terpenoids represent the largest class of secondary metabolites produced through a wide variety of plants[2,3], animals[4], and microbes[5,6]. Since Wallach[7] first isolated and analyzed the structure of terpenes in 1884, more than 180,000 terpenoid molecules have been discovered or synthesized[8,9]. They construct an enormous chemical space due to their diverse carbon skeleton and post-modification types, multiple chiral centers and substituents, and multiple rearrangements. In general, they can be classified based on the number of C5 units, such as monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), and triterpenoids (C30). Terpenoids have received much attention since they contribute as a rich reservoir of candidate compounds for drug discovery and industrial application[10-12]. They exhibit a wide range of biological activities and pharmacological potential, including anticancer[13,14], antibacterial[15,16], anti-inflammatory[17], antioxidant[18], anti-allergic[19], and the treatment of major diseases[20,21], such as cardiovascular disease and neurological diseases.

Melanoma, an aggressive cutaneous malignancy, is characterized by rapid progression, high mortality rates, and poor prognosis[22]. Emerging evidence have highlighted the potential of terpenoids in modulating key signaling pathways involved in cell proliferation, apoptosis, and metastasis, offering promising avenues for melanoma treatment[23]. Eucalyptol (1,8-cineole), a monoterpenoid abundantly found in plants, has shown significant anti-metastatic activity against skin cancer cells *in vitro* and *in vivo*[24]. Similarly, β-elemene, a sesquiterpenoid derived from a traditional Chinese herb, has been shown to inhibit proliferation, migration, and tube formation of endothelial cells (ECs), thereby impeding melanoma growth and metastasis[25]. A recent review by Grudzińska et al.[26] compiled the anti-melanoma activities of natural triterpenoids and their semi-synthetic derivatives, including betulin, betulinic acid, ursolic acid, maslinic acid, oleanolic acid, celastrol, and lupeol. Collectively, these findings demonstrate that structurally diverse terpenoids hold considerable promise for melanoma management. However, despite the promising results from preclinical studies, there are no specific clinical trials for terpenoids in the treatment of melanoma to date. Further research is still warranted to elucidate their mechanisms of action, optimize their therapeutic efficacy, and mitigate potential adverse effects to facilitate clinical translation. To expedite the terpenoid-based drug discovery applications, computational approaches are increasingly being employed to explore their chemical space and predict biological activities.

To date, the accumulation of diverse activity data has emerged as a powerful driver for investigating and mining terpenoid bioactivity. For example, TeroKit[27-29], a terpenoid-specific database, has collected more than 150,000 terpenoids along with their associated biological activities, primarily focusing on protein targets and cell line responses. Similarly, TCMBank[30], a traditional Chinese medicine database, also includes a portion of terpenoid structures in its collections. Furthermore, the well-known DrugBank[31] has deposited hundreds of terpenoids with greater pharmaceutical potential as experimental or approved drugs, annotating targets and drug action information. With the growth of experiential data, the utilization of data-driven methodology has gained increasing popularity in the drug discovery pipeline[32,33]. Integrating and analyzing multi-dimensional biological network relationships can aid in predicting potential lead compounds and provide critical insights into the pharmacological mechanisms[34,35]. Knowledge graph (KG) has been leveraged to facilitate infer of biomedical knowledge from heterogeneous data[36]. A biomedical KG is a data structure in which entities can be described as molecules, targets, and diseases. Relations are conceptualized as edges, facts, or links. Recently, KGE models witnessed rapid developments, enabling them to excel in tasks such as link prediction. Specifically, these models embed entities and relations into a continuous vector space and define a scoring function as the objective to predict biological interactions. Therefore, it is feasible to construct a knowledge graph-based platform facilitate terpenoid research for anti-melanoma drug discovery.

In this study, we first established a large-scale biological activity network of terpenoids, which offers comprehensive information, encompassing terpenoid structures, target proteins, genes, phenotypic cell lines, diseases, and their complex associations. Furthermore, we developed a terpenoid−disease association (TerDA) prediction model to facilitate the exploration of terpenoid bioactivity. Our model demonstrated superior performance compared to all baseline models through computational metrics, and it was confirmed experimentally for inferring new disease-therapy of multiple terpenoid compounds. The entire terpenoid-bioactivity knowledge graph and other toolsets are deployed on the TeroACT platform. Finally, we fully utilized this platform to identify multiple anti-melanoma inhibitors. *In vitro* and *in vivo* animal models further confirmed the anti-proliferative and anti-invasive effects of the two compounds (*i.e.*, mollugin and columbianadin) in melanoma.

2. Materials and methods

2.1. Decoder

For the terpenoid−disease binary pair (t_i, d_i), if it has been confirmed to be relevant, we designated its plausibility as 1; if it has not yet been verified, its plausibility was assigned a value of 0. t represented the embedding matrix of terpenoid molecules generated by the feature extraction module. d represented the embedding matrix of disease entities. $t,d \in \mathbb{R}^d$. r can be viewed as a relation vector between diseases and molecular representations in the same d-dimensional vector space. Define a scoring function f to assess the plausibility of the binary pair (t_i,d_i). There were three KGE models as decoders for modeling TerDA. The scoring function $f_r(t,d)$ and the model complexity are as follows.

2.1.1. DistMult

DistMult uses the inner product to model second-order correlations in the real space. Specifically, given the embeddings t, d the scoring function is

$$\mathrm{f_r}(t,d) = t^{\mathrm{T}} W_{\mathbf{r}} d \tag{1}$$

Relationships $W_{\mathbf{r}}$ as translations in the embedding space. Here, each component of molecular entities interacts with each component of the disease entities.

2.1.2. TransE

TransE is a translation-based model, where the functional relation corresponds to a translation of the embeddings. It hypothesizes that $t_i + r_i \approx d_i$ when (t_i, d_i) holds. Here we used the L2 norm to measure their proximity.

$$\mathrm{f_r}(t,d) = -\parallel t + r - d \parallel_2^2 \tag{2}$$

2.1.3. TransH

Compared to TransE, TransH positions the relation-specific translation vector d_{r} in the relation-specific hyperplane w_{r} (the normal vector) rather than in the same space with drug embedding t and disease embedding d. Thus, we defined a scoring function to measure the plausibility of the binary pair.

$$\mathrm{f_r}(t,d) = -\parallel t_\perp + d_{\mathrm{r}} - d_\perp \parallel_2^2 \tag{3}$$

$$t_\perp = t - w_{\mathrm{r}}^{\mathrm{T}} t\, w_{\mathrm{r}} \tag{4}$$

$$d_\perp = d - w_{\mathrm{r}}^{\mathrm{T}} d\, w_{\mathrm{r}} \tag{5}$$

Then the score function is:

$$\mathrm{f_r}(t,d) = -\parallel \left(t - w_{\mathrm{r}}^{\mathrm{T}} t\, w_{\mathrm{r}}\right) + d_{\mathrm{r}} - \left(d - w_{\mathrm{r}}^{\mathrm{T}} d\, w_{\mathrm{r}}\right) \parallel_2^2 \tag{6}$$

2.2. Model training

$y_{(t,d)}$ is the plausibility of all known pairs $\{(t,d) |\, t \in \mathcal{E}, d \in \mathcal{E}\}$, where $\mathcal{E}$ is a set of entities. $\boldsymbol{\Delta}$ denotes the set of all known positive pairs. We regarded all the unknown pairs $y_{(t,\, d')}$ as a negative sample source, and Δ' denotes the set of negative pairs generated by corrupting positive instances[37]. For t_i, the number of negative samples Δ_i' is usually a multiple of the positive instances, but does not exceed the total number of unknown candidates.

Specially,

$$\frac{\left|\Delta_i'\right|}{\left|\Delta_i\right|} = \frac{\sum \left(t_i, d_i'\right)}{\sum \left(t_i, d_i\right)} = 1, 5, 10 \tag{7}$$

For each terpenoid t_i , we restricted the sampling proportion to construct Δ_i' set. Because class imbalance data will bias the learning of the model, we used weight-based loss adjustment methods to perform confidence calibration. We employed the gradient descent method to minimize the following loss function:

$$L = \sum_{(t,d') \epsilon \Delta'} \sum_{(t,d) \epsilon \Delta} p_w \cdot y_{(t,d)} \cdot \log f_r(t,d) + \left(1 - y_{(t,\, d')}\right) \cdot \log\left(1 - f_r(t,d')\right) \tag{8}$$

p_w is a weighting parameter used to increase the loss weight of positive samples. By default, $p_w = \frac{|\Delta'|}{|\Delta|}$.

2.3. Baseline models

To comprehensively study the performance of TerDA, we compared it with four baseline models for disease prediction.

HNet-DNN[38]: This method uses a deep neural network to infer new drug-disease associations based on heterogeneous network features.

REDDA[39]: This method is a drug-disease association prediction model that integrates multiple biological relations into a heterogeneous graph neural network.

LHGCE[40]: The method develops a multi-layer graph convolutional encoder to learn knowledge graphs and biological features for the drug-disease association prediction.

HDGAT[41]: The method proposes a hierarchical and dynamic graph attention network for the prediction of associations between drugs and diseases.

2.4. Evaluation criteria

The F1-based threshold-moving approach was introduced to evaluation metrics to improve confidence calibration under class imbalance data. In this context, the optimal threshold was estimated by maximizing the F1-score to replace the default threshold. There are five classification evaluation metrics including accuracy, AUC (area under receiver operating characteristic curve), AUPR (area under precision recall curve), Precision, and Recall. AUC and AUPR are obtained based on the area under the curve. For AUC, the receiver operating characteristic (ROC) curve gives a pair of x and y values where x is the false-positive rate (FPR) and y is the true-positive rate (TPR). For AUPR, the curve is plotted in a comparable way to the ROC curve but with the x-axis being Recall and the y-axis being Precision. The indicators are defined as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{9}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \tag{10}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{11}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{13}$$

$$\mathrm{F1-score} = \frac{2(\mathrm{Precision} \times \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})} \tag{14}$$

These metrics are related to true positive (TP), true negative (TN), false negative (FN), and false positive (FP).

We also adopted the all-rank criteria, which alleviated the evaluation bias introduced by negative sampling. More specially, the predicted score for every binary pair (t_i, d_i) s computed and the entities d_i are sorted in decreasing order by the predicted scores. There are three rank evaluation metrics, including MR (Mean Rank of correct entities), MRR (Mean Reciprocal Rank of correct entities), and Top-k Hit-Ratio (the accuracy of correct entities ranking in top k). The measures are all reported in the raw and filtered settings[42]. In the filtered setting, all metrics are computed

after excluding all other positive pairs that appear in the training, validation, or test sets from the ranking. Whereas in the raw setting, these are not removed. The rationale behind the filtered setting is that ranking a positive pair after another one should not be considered a mistake.

$$\mathrm{MR}=\frac{1}{|s|}\sum_{i=1}^{|s|}\mathrm{Rank}_i \tag{15}$$

$$\mathrm{MMR}=\frac{1}{|s|}\sum_{i=1}^{|s|}\frac{1}{\mathrm{Rank}_i} \tag{16}$$

$$\mathrm{Hit@}K=\frac{1}{|s|}\sum_{i=1}^{|s|}\P(\mathrm{Rank}_i\le K) \tag{17}$$

s denotes the disease association dataset. $\P(\cdot)$ is the indicator function (if the condition is true then the function takes the value of 1, otherwise it is 0).

ILS (Intra-list similarity) is the average pairwise similarity of the items in a list. In this work, ILS@20 is calculated as the average pairwise similarity between the predicted compounds and molecular entities in the current knowledge graph related to "melanoma" based on molecular fingerprints. For a query molecule t_i, $S(t_i,t_j)$ represents the similarity between t_i and another molecular entity t_j. K refers to taking the top k entities with the highest similarity into consideration.

$$\mathrm{ILS@}K=\frac{1}{K}\sum_{1\le j\le K}S(t_i,t_j) \tag{18}$$

2.5. Cell lines and cell culture

A375, B16-F10, 293T, Hacat and NCM460 cells were cultured in Dulbecco's modified Eagle's medium (11965092, Gibco, USA) supplemented with 10% fetal bovine serum (FBS-BE500, Newzerum, New Zealand) and 1% penicillin/streptomycin (15140122, Gibco, USA) at 37 °C in a humidified incubator containing 5% CO_2 atmosphere. Mollugin was purchased from D&B with a purity of at least 98% in HPLC analysis. All cell lines were tested for mycoplasma-free by Mycoplasma Stain Assay Kit (C0296, Beyotime, China). Cell line authentication has been validated by the STR analysis using the GenePrint 10 System according to the manufacturer's instructions (B9510, Promega, USA).

2.6. Measurement of cell viability by CCK8 assay

A375, B16-F10, 293T, Hacat and NCM460 cells were seeded at a density of 8×10^3 cells per well in 96-well plates with fresh medium and cultured overnight. After incubation, the cells were treated with various concentrations of mollugin for 24−48 h. The supernatant was removed and the cells were labeled with Cell Counting Kit-8 (C0005, TargetMol, USA) solution according to the manufacturer's instructions. The incubation was carried out at 37 °C for 3 h before absorbance was read at 450 nm using a microplate reader (Multiskan GO, ThermoFisher, USA).

2.7. Apoptosis assays

Apoptosis assays were performed by the method previously described[43]. Annexin V staining was conducted using the Annexin V-FITC Apoptosis Detection Kit (40302ES50, Yeasen, China) according to the manufacturer's instructions. Briefly, A375 and B16-F10 cells were collected after incubation with various concentrations of compounds and washed with phosphate-buffered solution (PBS). Afterward, the cells were centrifuged and incubated with Annexin V-FITC and 2 μg/mL propidium iodide in binding buffer for 15 min at room temperature in the dark. The apoptotic cells were subsequently analyzed by flow cytometry.

2.8. EdU labeling and immunofluorescence

Cells were cultured in 12-well plates and treated with 30−60 μmol/L mollugin. After 36 h incubation, the cells were labeled with 5-ethynyl-2′-deoxyuridine for 3 h and stained with Alexa Fluor 555 following the instructions of the EdU Cell Proliferation Kit (C0075S, Beyotime, China). Briefly, cells were first fixed with paraformaldehyde/PBS (4%) for 30 min and then permeabilized in Triton X-100/PBS (0.5%) for 20 min. Next, cells were washed with PBS, blocked with PBS supplemented with 2% BSA for 30 min, and stained with dye mix for 30 min. Then cells were washed three times with staining buffer for 5 min, and the cellular nuclear were stained by Hochest in darkness for 10 min, followed by observation under a Nikon microscopy (Nikon, Japan). The data was measured using the Nikon fluorescence microscope.

2.9. Colony formation assay

B16-F10 cells were seeded at a density of 2000 viable cells in 12-well plates for 24 h. Then, the cells were treated with various concentrations of the mollugin and cultured for 1−2 weeks until cell colonies were visible and countable. The colonies were fixed with 4% formaldehyde in PBS and stained with 0.02% crystal violet Staining Solution (C0121, Beyotime, China) for observation.

2.10. Wound-healing assay

For the wound-healing assay, cells were seeded into 6-well plates and cultured until a 90% confluent monolayer was formed. Cells were then scratched by a sterile pipette tip and treated with indicated concentration of compounds in the FBS-free medium. Cell migration distances into the scratched area were measured in 3 randomly chosen fields under a microscope.

2.11. Transwell assays

For transwell assay, 2×10^5 cells were planted in 200 μL of serum-free medium to a polycarbonate filter (8-μm pore size) (3422, Corning, USA) in the insert of a 24-well plate. DMEM medium containing 10% FBS and inhibitors was added to the bottom chamber. After incubation for 18 h, the migrated cells in the lower filters were fixed with 4% polyoxymethylene, and stained in crystal violet staining solution, and the images were collected with Nikon microscopy (Nikon, Japan).

2.12. Quantitative real-time PCR and Western blot analysis

Quantitative real-time PCR and Western blot analysis were performed by the method previously described[43]. In brief, total RNAs were extracted from cells by the Fast Cell RNA Extraction kit (400-100, Googie, China), adhering to the manufacturer-provided guidelines. Complementary DNAs (cDNAs) were synthesized from equal amounts of RNA samples using the Evo M-MLV RT Master Mix (AG11706, Accurate Biology, China), and real-time

reverse-transcription PCR was carried out subsequently by SYBR Green Pro Taq HS Premix (AG11701, Accurate Biology, China) in CFX96 Real-Time PCR Detection System (Bio-rad, USA) with a melting-curve analysis performed. β-Actin was used as an internal control for normalization.

For the Western blot assay, protein extraction from cells was lysed on ice for 20 min by RIPA buffer (P0013B, Beyotime, China) and then centrifuged at 4 °C for 15 min. Proteins were separated by 10%−12% SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membrane (LC2005, Invitrogen, USA), and blotted with related antibodies in 4 °C overnight with appropriate concentrations of primary antibodies. Peroxidase-conjugated secondary antibodies were subsequently labeled in room temperature for 1 h. The antibody−protein complexes were visualized with an enhanced chemiluminescence Western blotting kit (#FD8000, Fude Bio. Tech. Co., China) and captured by ChemiDocTouch visualizer (Bio-rad, USA).

The following antibodies were used in this study: anti-phospho-NF-κB-p65(Ser536) (#3033, 1:1000, Cell Signaling Technology, USA), anti-phospho-IκBα (Ser32/36) (F0197, 1:1000, Selleck, China), anti-GAPDH (5174, 1:1000, Cell Signaling Technology, USA) and Goat Anti-Rabbit IgG H&L (HRP) (ab205718, 1:5000, Abcam, USA). GAPDH was included as a protein loading control.

2.13. Surface plasmon resonance

Surface plasmon resonance experiment was determined using a Biacore 8K instrument (GE Healthcare, Uppsala, Sweden). The recombinant FABP5 protein was immobilized on a sensor chip (CM5) using the amine-coupling method according to standard protocols. FABP5 protein was diluted in sodium acetate buffer, pH 4.0. Various concentrations of the compound Mollugin were subsequently injected as analytes. To estimate the affinity, the binding assay was examined at 25 °C at a flow rate of 30 L/min using PBS buffer. The affinity constants of binding were obtained using a 1:1 Langmuir binding model *via* BIA evaluation software.

2.14. Animal experiments

The entire surgical procedures and animal care were conducted following the guidelines of the Institutional Animal Care and Use Committee (IACUC). 8-Week-old specific-pathogen-free C57 mice were randomly grouped ($n = 5$ per group) for each candidate compounds. In the subcutaneous implantation mouse model, each mouse received a subcutaneous injection of 0.1 mL containing 1×10^6 B16-F10 cells/mL. When the tumor became palpable, mice were orally administered a suspension of Mollugin/Columbianadin in saline/corn oil every other day. Tumor sizes were measured daily using calipers to analyze tumor growth. Tumor volume was calculated based on Eq. (19):

$$\text{Tumor voume} = \left(\text{Length} \times (\text{Width})^2\right) / 2 \quad (19)$$

At the experimental endpoint, tumor tissues were harvested. All animal care and experiments were approved by the Institutional Animal Care and Use Committee of Sun Yat-sen University, Guangzhou, China, and the study complied with all relevant ethical regulations concerning animal research. Mice were euthanized according to institutional euthanasia criteria based on tumor size and overall health condition, following the National Institute of Health Guide for the Care and Use of Laboratory Animals.

2.15. HE staining

Tissue samples were fixed with 4% paraformaldehyde at room temperature for 24 h and embedded in paraffin. Sections were cut for HE staining.

For HE staining, tissue specimen sections were washed with dewaxing solution (G1128, Servicebio, China), rehydrated through an alcohol series (anhydrous ethanol, alcohol, and then distilled water sequentially). The sections were then stained with hematoxylin for 5 min, followed by washing under running tap water for 5−6 min to allow nuclei staining. Next, the sections were differentiated in 1% acid alcohol for a few seconds and rinsed in distilled water. For cytoplasmic staining, sections were counterstained with eosin for 1−2 min, followed by dehydration through an ascending ethanol series (70%, 80%, 95%, and 100%). Images were collected with Nikon microscopy (Nikon, Japan).

2.16. Statistical analysis

All results were expressed as mean ± standard deviation (SD) and analyzed using GraphPad Prism 8.0. Group comparisons were performed with one-way ANOVA and Tukey's multiple comparison tests. Statistical significance was accepted for P-values of <0.05.

3. Results and discussion

3.1. Terpenoid bioactivity landscape and knowledge graph construction

Our previously established TeroKit has collected detailed information on terpenoid substructures, protein targets, cellular targets, and their associations between molecules and targets. Herein, we further identified and extracted both approved small-molecule drugs and herbal ingredients classified as terpenoids from the DrugBank and TCMbank databases, respectively. The associations between cellular targets and diseases encompassing human rare diseases were collected from Cellosaurus[44]. Besides, the interaction data between terpenoid compounds and protein targets are a vital component of the activity network as well as essential data sources for computational research. Therefore, we have additionally established a more standardized storage repository specifically designed for CPI information using the PostgreSQL[45] database software. This database not only contains detailed information on CPIs but also categorizes protein targets by disease type using genes as bridges. This is accomplished by integrating the CTD[46], GeneCards[47], and TTD[48] database, to map with the TeroKit database. Finally, a biological activity network on terpenoids was constructed, which contains 11,653 unduplicated terpenoids (including derivatives), 1321 target proteins, 975 cell lines, 1178 gene targets, diseases with three types of identifiers, namely International Classification of Diseases 11th Revision (ICD-11), National Cancer Institute Thesaurus (NCIt), and Drphanet Rare Diseases Ontology (ORDO). Their relationships are illustrated in Fig. 1A.

Terpenoids in the activity network exhibit diverse structural types along with a wide range of significant biological activities (Supporting Information Fig. S1). On one hand, a total of 2526 structurally rich terpenoids were identified in TCMBank through

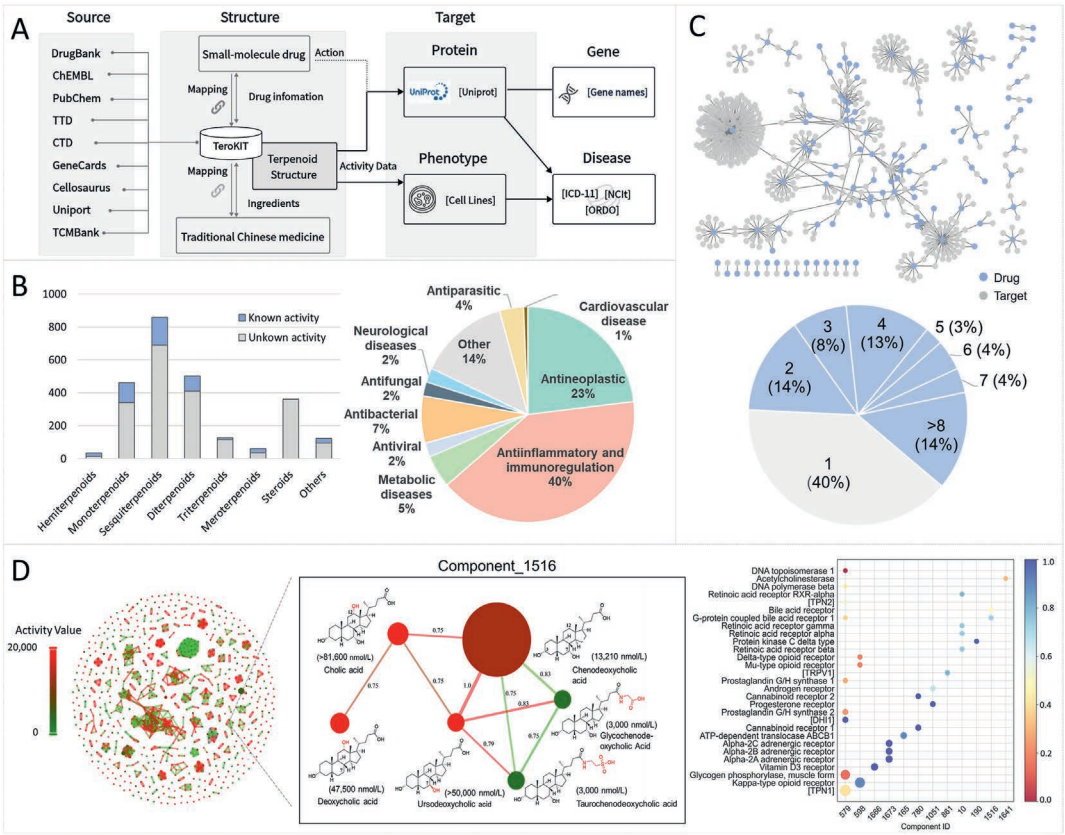

Figure 1 Overview of the constructing protocol and the comprehensive biological activities network towards terpenoids. (A) Activity-based data structure of the network. (B) Information on terpenoids as herbal ingredients. left: Distribution of structural categories and the proportion of known activities. right: The pharmacological activity distribution from known active annotations. (C) Information of terpenoids as small molecule drugs. above: Drug−target interaction network. down: Distribution of the number of drug targets. (D) Profiling of CPIs in activity network. left: The bioactive chemical space networks constructed based on the similarity of terpenoids in CPIs. The similarity between nodes is quantified using the Tanimoto coefficient, with a minimum threshold set at 0.7. Each connected subgraph within the network is referred to as a component such as Component_1516. Node size correlates with the number of bioassays, while node color signifies the average bioactivity value derived from the bioassay. Edge width is proportional to molecular similarity, and edge color corresponds to the active cliff. right: Bubble chart of selected CPIs. The horizontal axis corresponds to component IDs, whereas the vertical axis represents protein targets. The size of each data point is directly proportional to the total number of bioassays associated with terpenoid compounds within each component. The color of each data point reflects the ratio of active compounds within the respective component. More comprehensive CPI activity information and visualization can be found by visiting: http://terokit.qmclab.com/activity/visual_scatter/.

structure matching. These molecules are sourced from herbal ingredients, of which only 19% have known active annotations. They are mostly validated in the realms of anti-inflammation, immunomodulation, and antineoplastic (Fig. 1B). On the other hand, there are 238 drug entries identified as terpenoids in DrugBank, and among them, 111 small-molecule drugs exhibit explicit interactions with specific targets. Terpenoids, according to the pharmacological actions' statistics, mainly act as ligands within the crystal structure, as well as inhibitors and agonists in that order (Supporting Information Fig. S2). As shown in the drug-target interaction network (Fig. 1C), one of the most extensively researched terpenoid drugs is artemisinin, a sesquiterpene compound renowned for its role as an anti-malarial agent[49]. In terms of therapeutic indications, aside from its use as an anti-malarial drug, artemisinin has been repurposed for the treatment of systemic lupus erythematosus[50,51] and the COVID-19[52,53] pandemics and is being explored for the treatment of other malignancies and inflammatory diseases[54,55]. In addition, about 60% of terpenoids as small-molecule drugs interact with multiple protein targets. It demonstrates that terpenoids have great therapeutic potential across multiple target-related diseases.

Furthermore, we have visualized the bioactive chemical space networks[56] of CPIs to display the bioactive profiles of similar terpenoids (Fig. 1D). Within each component in the network, compounds have essentially the same ring skeleton, facilitating structure-activity relationship analysis. For example, one class of compounds in "component_1516" interacts with bile acid receptors. While chenodeoxycholic acid is an effective natural ligand for the bile acid receptor, its epimer, ursodeoxycholic acid, with a 7β-OH group, does not activate the receptor *in vivo*. This observation highlights the presence of an activity cliff. Additionally, the bubble chart visualizes the current research status of similar compound clusters with all proven targets. For example, the tyrosine-protein phosphatase non-receptor (TPN1) has the highest amount of bioassay data among compounds within Component_579. This dataset consists of 393 test measurements obtained from 59 distinct compounds. The substantial volume of bioactivity test data indicates significant research efforts in this

context. Such a background investigation would give the inspiration that terpenoid compounds sharing the same carbon skeleton as compound_579 are likely to exhibit potential biological activity towards the TPN1. Hence, the bioactivity network can assist researchers in providing intuitive and guidable insights for terpenoid drug discovery based on existing activity knowledge.

3.2. Overview of the prediction model towards terpenoid−disease associations

To accelerate terpenoid-based drug discovery, we constructed a terpenoid-bioactivity knowledge graph (KG) by extracting a subnetwork comprising terpenoid−disease binary pairs. Other network nodes (*e.g.*, cell lines) were treated as invisible intermediate bridges. The resulting KG comprised quantities of documented terpenoid−disease associations. In this KG, terpenoid molecules were considered as head entities, disease types as tail entities, and their associations were represented as relations. All experimentally validated terpenoid−disease pairs were regarded as positive samples. Building upon this KG, we developed the TerDA prediction model to systematically investigate terpenoid bioactivities. The model synergistically combines deep learning-based molecular representations with knowledge graph-based network inference to enhance the predictive performance. Here, KG embedding was employed to predict new links between entities in KG, *i.e.*, the potential terpenoid−disease associations. As illustrated in Fig. 2, the TerDA architecture consists of two modules.

In the molecular feature fusion module, we computed ten molecular descriptors for each terpenoid molecule and applied a conditional numerical transformation to all values, generating a descriptor feature matrix. This matrix was then processed by a convolutional neural network (CNN), which output a feature vector representing the molecule's physicochemical properties.

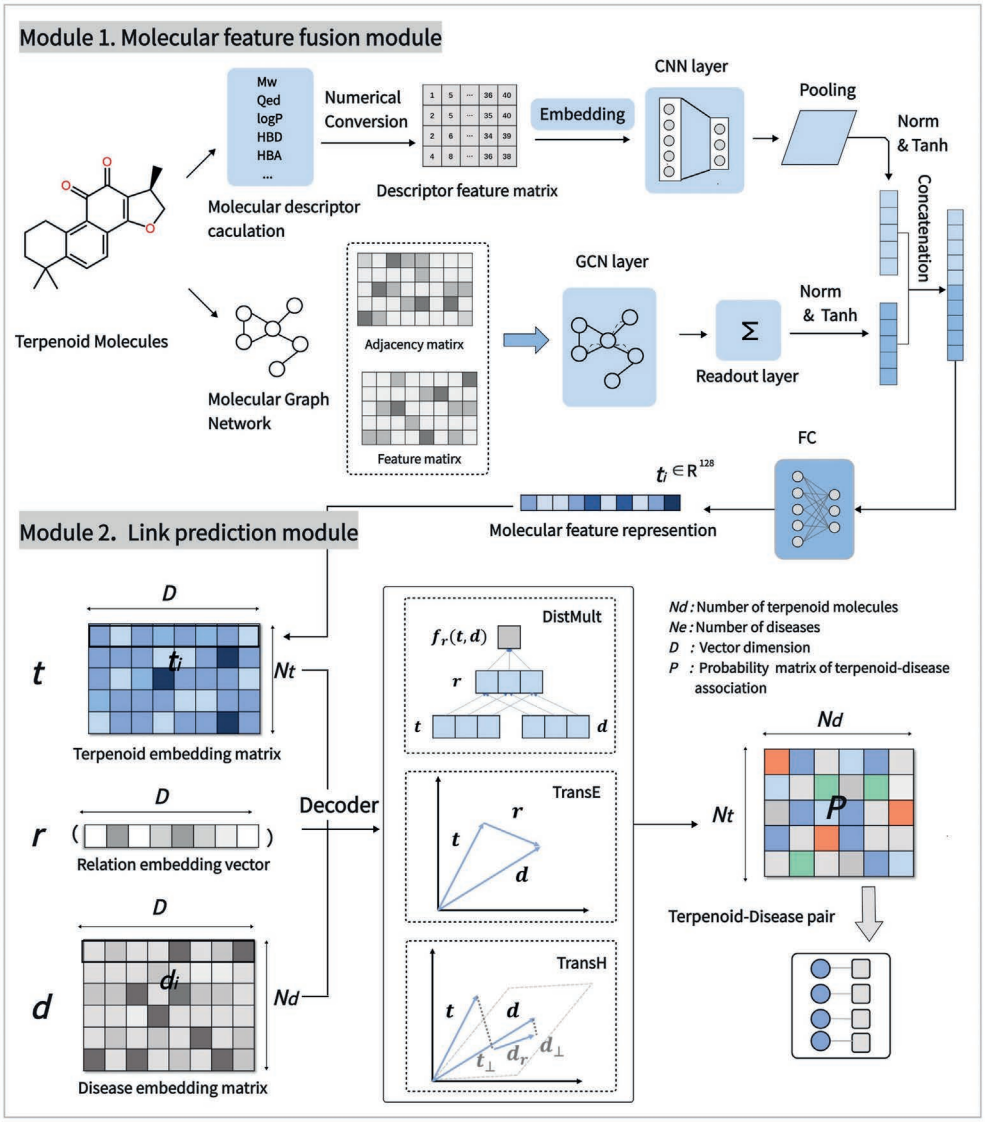

Figure 2 The architecture of the TerDA prediction model. There are 10 molecular descriptors in the molecular feature fusion module, including Wildman−Crippen partition coefficient (log*P*), topological polar surface area (TPSA), molecular weight (MW), drug-likeness (QED), molar refractivity (MR), number of hydrogen-bond acceptors (HBA) and donors (HBD), number of aromatic rings and rotatable bonds (RB), and FractionCSP3, calculated by the RDKit Python library.

Simultaneously, we treated the molecular structure as a graph network to extract topological and atom-level features. These features were passed through a graph convolutional network (GCN) layer, followed by a readout layer to produce a global graph-based feature vector. To integrate these two feature streams, the module employs an early fusion strategy. The CNN-derived (physicochemical) and GCN-derived (topological) feature vectors are concatenated. The combined features are then projected into a joint representation space *via* a Multilayer Perceptron (MLP), yielding a unified, high-level molecular feature representation.

In the link prediction module, we employed KGE to model relation patterns and predict new associations within the KG. Disease embeddings were derived from a randomly broken adjacency matrix, while relation embeddings were initialized as random vectors of the same dimensions. All the embeddings of entities and relations were fed into the KGE model together. We introduced three representative models, including DistMult[57], TransE[42], and TransH[58], where the entities (diseases and terpenoids) and relations were transformed into a continuous vector space. The factual terpenoid−disease pairs were expected to receive higher predicted scores than unproven associations. During the training phase, terpenoid−disease pairs were sampled at various positive-to-negative ratios, and the model parameters were optimized by minimizing the cross-entropy loss. This training procedure ultimately generated a predictive matrix P of terpenoid−disease associations. The KGE approach effectively addressed the incompleteness of the knowledge graph by inferring potential missing relationships. Detailed descriptions of the loss functions and training procedures are provided in the "Experimental" section.

Table 1 Performance comparison between TerDA and all baselines in classification metrics.

Model	Pos: Neg	Accuracy	AUC	AUPR	Precision	Recall
HNet-DNN	1:01	0.928	0.979	0.970	**0.933**	0.924
	1:05	0.919	0.960	0.869	0.822	0.718
	1:10	0.937	0.952	0.772	0.807	0.538
REDDA	1:01	**0.994**	**0.991**	0.969	0.903	0.901
	1:05	**0.975**	0.883	0.724	0.821	0.651
	1:10	0.975	0.660	0.374	0.724	0.319
LHGCE	1:01	0.923	0.970	0.943	0.900	0.918
	1:05	0.960	0.982	0.912	0.870	0.872
	1:10	0.977	0.985	0.900	0.868	0.853
HDGAT	1:01	0.821	0.911	0.908	0.761	0.937
	1:05	0.871	0.910	0.713	0.596	0.710
	1:10	0.916	0.909	0.594	0.536	0.563
TerDA (TransH)	1:01	0.932	0.973	0.971	0.871	0.907
	1:05	0.950	0.978	0.851	0.836	0.873
	1:10	0.942	0.976	0.806	0.807	0.842
TerDA (TransE)	1:01	0.932	0.978	**0.973**	0.881	0.890
	1:05	0.948	0.979	0.887	0.858	0.868
	1:10	0.948	0.974	0.820	0.820	0.830
TerDA (DistMult)	1:01	0.929	0.975	0.972	0.906	**0.957**
	1:05	0.966	**0.989**	**0.957**	**0.907**	**0.908**
	1:10	**0.978**	**0.990**	**0.943**	**0.905**	**0.881**

All models were evaluated based on 5-fold cross validation. The best results under the three different negative sampling ratio settings are highlighted in bold.
AUC, area under receiver operating characteristic curve; AUPR, area under precision recall curve.

3.3. Performance of TerDA on cross-validation

The 5-fold cross-validation procedure was conducted to compare the prediction performance of TerDA with that of four state-of-the-art disease prediction models. Considering the real-world scenarios, we set varying ratios of negative to positive data to train models. As shown in Table 1, all models exhibited declining classification performance with increasing negative data, particularly HNet-DNN, REDDA, and HDGAT, which showed marked decreases in AUPR, Precision, and Recall metrics. In contrast, both LHGCE and our models demonstrated remarkable robustness, maintaining most metrics above 80%. Among the KGE approaches, DistMult demonstrated superior performance compared to TransE and TransH. TerDA with DistMult achieved superior classification performance overall, ranking first in the majority of evaluation metrics. We hypothesize that this advantage stems from the inherent symmetry of DistMult's scoring function. DistMult's bilinear interaction mechanism enables more effective capture of fine-grained semantic information, thereby exhibiting better generalization capability for rare terpenoid−disease pairs. Therefore, we selected TerDA with the DistMult as the KGE module for subsequent applications. We further compared the performance of our model with LHGCE across common ranking measures in recommender systems (Table 2). In this context, higher prediction scores correspond to more prominent rankings. Our model outperformed the LHGCE model on all ranking metrics across three scenarios, demonstrating its excellent recommendation capabilities.

3.4. Performance of TerDA in drug repositioning task

We designed two computational schemes to further assess the model's ability for global and local drug repositioning, respectively. The primary objective of this task is to promote the rank of all true relations at the top of the recommendation list. For the global drug repositioning, we progressively and randomly reduced the true relations in the knowledge graph by 20%, 40%, 60%, and 80%, and then predicted the rank of the disease entities within the entire set of candidates. The information on data scale and sparsity for the four experimental settings is provided in Supporting Information Table S1. We also used LHGCE for performance comparison. From the results, the distribution of prediction ranks for both models tended to shift lower overall as the known knowledge decreased (Fig. 3A). However, in comparison, LHGCE model exhibited a wider distribution in the ranks for these masked true relations than TerDA. Further statistical analysis showed that our model consistently achieved lower MR and higher MRR (Fig. 3B). Additionally, our model demonstrated an advantage in top-k recommendations for these true relations based on Top-k Hit-Ratio (Fig. 3C).

In the local drug repositioning task, we randomly selected a set of 10 terpenoid compounds for testing. From the results, it is evident that our model had a significantly higher proportion of masked potential relations with ranks less than or equal to 40 compared to LHGCE (Fig. 3D). While it is understandable that the performance of disease association prediction would decline as the model retained less knowledge in the training data, ranging from 80% down to 20%, TerDA demonstrated a notable exception (Fig. 3E). The average rank did not decline significantly but exhibited minor fluctuations around 50. This indicates that our model has fully leveraged the rich background knowledge embedded in the vast knowledge graph, particularly its complex

Table 2 Performance comparison between TerDA and LHGCE in ranking metrics.

Model	Pos: Neg	MR[a]	MRR	Hit@1	Hit@3	Hit@10
LHGCE	1:01	36.2	0.252	0.141	0.286	0.466
	1:05	19.4	0.418	0.281	0.482	0.706
	1:10	14.4	0.542	0.408	0.624	0.807
TerDA	1:01	21.7	0.401	0.279	0.446	0.652
	1:05	10.6	0.693	0.600	0.748	0.872
	1:10	9.8	0.765	0.690	0.815	0.899

All models were evaluated based on 5-fold cross validation. MR, MRR, and Top-*k* Hit-Ratio (Hit@*K*, $K = 1, 3, 10$) values are reported in the filtered settings.
MR, Mean Rank of correct entities; MRR, Mean Reciprocal Rana of correct entities.
[a]↓ represents the lower the better.

topological features. Consequently, it compensates for the lack of relation information for local entities, maintaining a robust ability to infer diseases for individual molecular entities. Additionally, the Top-*k* Hit-Ratio (k = 40, 50, 60, 70) metrics of our model surpass those of the benchmark model LHGCE in all scenarios (Fig. 3F). This positive outcome underscores the reliability of our model in recommending disease indications in the drug repositioning task.

3.5. *Ablation study and hyperparameter settings*

To study the effectiveness of the key components, we performed an ablation study over two variants of TerDA that removed semantic features generated from the molecular graph and the molecular descriptors, respectively. The results show that the ablated model has a worse performance in all classification metrics and ranking metrics (Supporting Information Table S2). This indicates

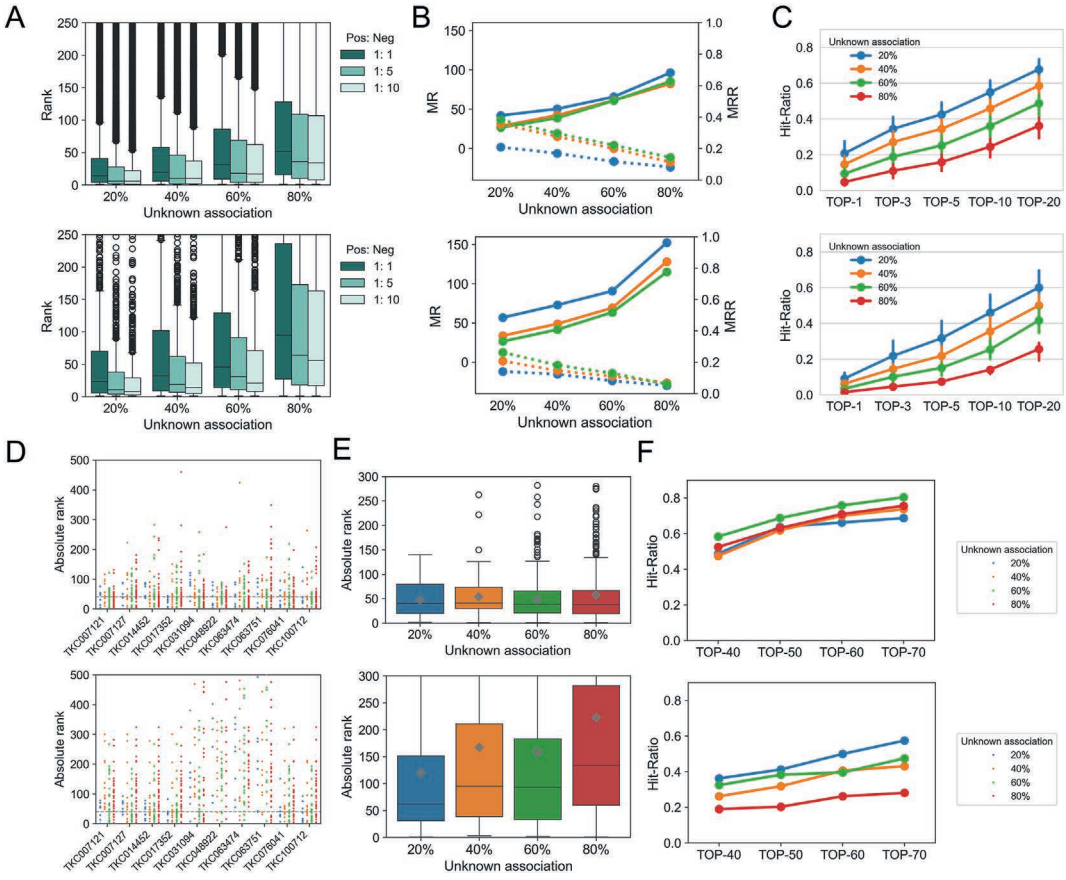

Figure 3 Performance comparison in drug repositioning task. (A−C) Model prediction results of TerDA (upper) and LHGCE (lower) at the global level of drug repositioning. The global drug repositioning scheme was designed to progressively reduce the terpenoid-disease associations by 20%, 40%, 60%, and 80% within the total training data, subsequently employing the trained model to predict them. There are statistics of the prediction results for the two models under varying masked knowledge. (A) Distribution of prediction ranks. (B) Line plot of MR (solid) and MMR (dotted). (C) Line plot of Top-*k* Hit-Ratio. These measures are reported in the filtered settings. (D−F) Model prediction results of TerDA (upper) and LHGCE (lower) at the local level of drug repositioning. Ten molecular entities were randomly selected from the knowledge graph, each associated with 40 diseases. The training data was modified by removing 20%, 40%, 60%, and 80% of the disease associations for these 10 compounds, respectively. (D) Scatter plot of prediction ranks of ten molecular entities. A horizontal line across the plot at the data coordinate 40. (E) Distribution of prediction ranks with various masked knowledge. The four gray dots represent the average predicted ranks. (F) Line plot of Top-*k* Hit-Ratio. All models are trained under the condition of Pos:Neg = 1:5. The measures are reported based on absolute rank in the raw settings.

that the fusion strategy effectively captures the more semantic knowledge of molecules. Additionally, the injection of rich molecular semantic information enables our TerDA to provide more accurate and robust disease inference guidance.

Additionally, we also rigorously evaluated alternative fusion methods, including dual-channel attention fusion, gated fusion, cross-interaction fusion, residual fusion, and convolution fusion (Supporting Information Table S3 and Fig. S3). Based on five-fold cross-validation, the early fusion approach achieved superior performance in both classification and ranking metrics. We hypothesize that this is because early fusion preserves the complete information of the original features. Although it may introduce some redundancy, it is particularly suitable for scenarios where the two feature types exhibit significant differences.

We implemented our models using PyTorch. We used the Adam optimizer for parameter learning with an initial learning rate and exponential decay after 10 epochs. We tuned the initial learning rate in {0.01, 0.03, 0.05, 0.07}. We searched for a decay rate ranging from 0.97 to 0.996. For the dimension of latent space, we tuned the size of entity and relation embeddings from 8 to 512. We also chose the number of GCN layers in the range of {1, 2, 3}, and the number of CNN layers in {1, 2, 3}. The detailed observations corresponding to different hyperparameter settings are provided in Supporting Information Fig. S4.

3.6. Screening and identification campaign for the discovery of anti-melanoma agents

Terpenoids and their derivatives constitute a vast source of bioactive compounds with a potential antimelanoma effect. To explore this, we employed the established TeroACT platform for *in silico* drug screening to identify novel anti-melanoma agents. The computational screening process utilized a multi-model ensemble approach, systematically calculating plausibility scores for all terpenoid−disease pairs in the knowledge graph using the TerDA model. The predictive results inferred a set of potential terpenoid compounds with potential anti-melanoma characteristics (Table 3). From these predictions, we selected six high-ranking terpenoids that showed significant cytotoxic and antiproliferative effects against A375 melanoma cells. As a control, 7 terpenoids with low-ranking associations were tested but exhibited minimal anti-melanoma activity (Supporting Information Fig. S5 and Table S4). These findings confirm the recommendation capabilities of the TerDA model in identifying novel compound-disease associations.

Furthermore, we visualized the similarity distribution between these active terpenoids and known melanoma-associated molecular entities in the knowledge graph (Supporting Information Fig. S6). The ILS@20 metric was used to assess the structural novelty of the compounds. Due to the high similarity of some compounds to known anti-melanoma molecules, such as docetaxel and 7-epitaxol, we selected columbianadin and mollugin, which had no similarity above 0.7 to known active molecules, for further laboratory validation of their anti-melanoma effects.

Columbianadin was known for its anti-inflammatory properties and suppression of colorectal cancer cell growth[59,60]. Meanwhile, previous biological and pharmacological studies have reported mollugin's anti-tumor activities against colorectal cancer[61] and hepatocarcinoma[62]. However, the association between these two terpenoids and melanoma remains unexplored. Therefore, we conducted a series of detailed assays to explore the pharmacological potential of mollugin and columbianadin against melanoma. Cell viability assays indicated that both compounds significantly inhibited the proliferation of A375 and B16-F10 cells in a concentration-dependent manner (Fig. 4A). Mollugin exhibited IC_{50} values of 25 and 59 μmol/L for A375 and B16-F10 cells, respectively, while the IC_{50} of columbianadin was 32 and 69 μmol/L for A375 and B16-F10 cells. Clonogenic assays demonstrated the compounds' long-term anti-proliferative effects, with significant suppression of B16-F10 cellular growth (Fig. 4B). Furthermore, EdU incorporation assays confirmed that both terpenoids inhibit melanoma cell proliferation (Fig. 4C). These results collectively illustrate that these two candidate terpenoids show marked *in vitro* antineoplastic effects against melanoma.

Metastasis is another prevalent issue in melanoma, as the proportion of newly diagnosed advanced melanoma cases

Table 3 Summary of prediction ranks and experimental activity results of 13 terpenoids (including derivatives) against melanoma.

ID	Name	CAS	Rank	IC_{50} (μmol/L)	ILS@20↓[a]
TKC219645	Columbianadin	5058-13-9	1	31.94	0.65
TKC014820	7-Epitaxol	105454-04-4	4	0.0038	0.86
TKC117556	Periplogenin	514-39-6	4	3.27	0.86
TKC214908	Mollugin	55481-88-4	5	24.78	0.58
TKC002399	Pseudolaric acid A	82508-32-5	7	2.00	0.77
TKC136632	Arteannuin B	50906-56-4	7	9.99	0.87
TKC012927	Phorbol	17673-25-5	28	−	0.75
TKC123088	Betulinic acid	472-15-1	37	−	0.78
TKC032634	Linderane	13476-25-0	92	−	0.73
TKC214917	Citral	5392-40-5	127	−	0.41
TKC068533	Transcrocetin	27876-94-4	142	−	0.29
TKC004481	Columbin	546-97-4	339	−	0.73
TKC044550	Artemisinic acid	80286-58-4	443	−	0.65

"Rank" is calculated by predicting each compound's association with the "melanoma" disease using multiple TerDA models with an ensemble strategy. Antiproliferative activity is reported as the average IC_{50} values of three independently repeated cell viability assays on A375 cells.
ILS, Intra-list similarity.
[a]↓ represents the lower the better.

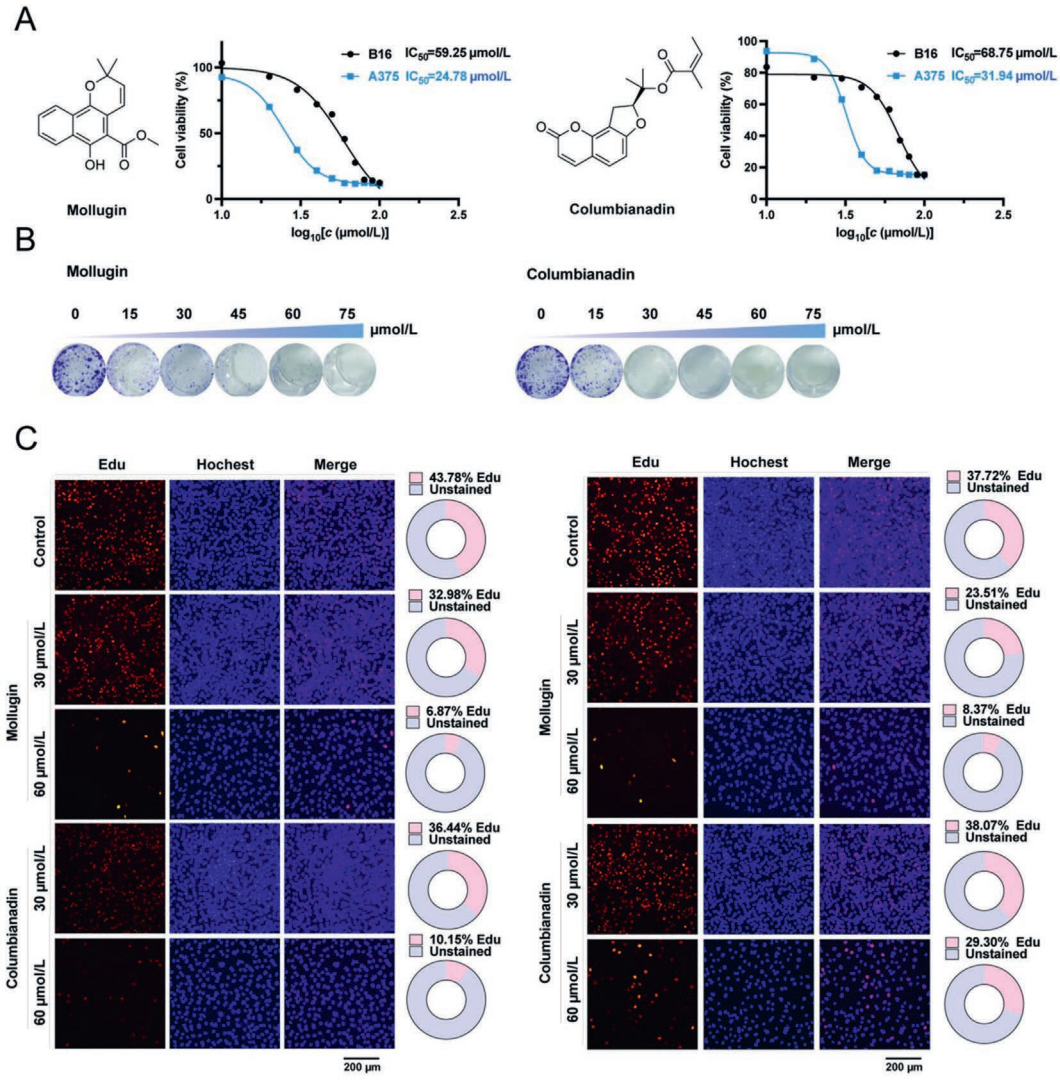

Figure 4 Validation of mollugin and columbianadin's anti-proliferative effects in melanoma *in vitro*. (A) Determination of IC_{50} (half maximal inhibitory concentration) for mollugin and columbianadin in melanoma cell lines. The IC_{50} of mollugin for A375 and B16-F10 cells was approximately 25 and 59 μmol/L, respectively, while the IC_{50} of columbianadin for A375 and B16-F10 cells was 32 and 69 μmol/L, respectively. (B) Evaluation of anti-proliferative effects using colony formation assays in melanoma treated with mollugin and columbianadin. (C) Capacity of mollugin and columbianadin on prohibiting melanoma proliferation *in vitro* shown by EdU assay (left: A375; right: B16-F10). All images were obtained with a 20 × objective. Scale bars: 200 μm.

presenting with liver metastasis has increased in recent years[63]. Metastatic progression markedly diminishes clinical treatment efficacy, highlighting the urgent need for therapeutic breakthroughs on advanced melanoma[64]. Given the critical role of invasive and metastatic capabilities in assessing melanoma malignancy, we investigated the inhibitory effects of mollugin and columbianadin on melanoma cell migration. Quantitative wound healing assays revealed significant inhibition of horizontal cell migration, while transwell migration assays demonstrated substantial reduction in vertical invasive capacity (Fig. 5). Furthermore, mollugin and columbianadin were illustrated to induce apoptosis, which is consistent with previous studies (Fig. 6).

Given their significant *in vitro* anti-tumor activity, these compounds present promising therapeutic applications for melanoma treatment. Subsequent *in vivo* studies further revealed that mollugin treatment can significantly inhibit tumor growth in a melanoma-CDX model, as indicated by decreased tumor size, with no significant weight loss observed (Fig. 7, Supporting Information Fig. S7). Unexpectedly, columbianadin did not exhibit notable anti-tumor effects despite its antineoplastic effects on B16-F10 cell line *in vitro*. Previous pharmacokinetic studies reported that columbianadin can be detected in all of the selected tissues after *i.v.* administration and metabolized to columbianetin (CBT) in most detected tissues in rats[65]. Unlike Columbianadin (CBN), columbianetin (CBT) exhibits little cytotoxic effects on melanoma cells (Supporting Information Fig. S8). Rapid clearance of columbianadin *in vivo* (Supporting Information Table S5) may explain the failure of columbianadin *in vivo*.

To elucidate the anticancer mechanisms of mollugin, we employed the CPI prediction toolset within the TeroACT platform. The SPVec_Tero model, a SPVec[66]-based model optimized using terpenoid-protein target interaction data, identified the top five potential molecular targets of mollugin (Supporting Information Table S6). Among these, fatty acid-binding protein 5 (FABP5)

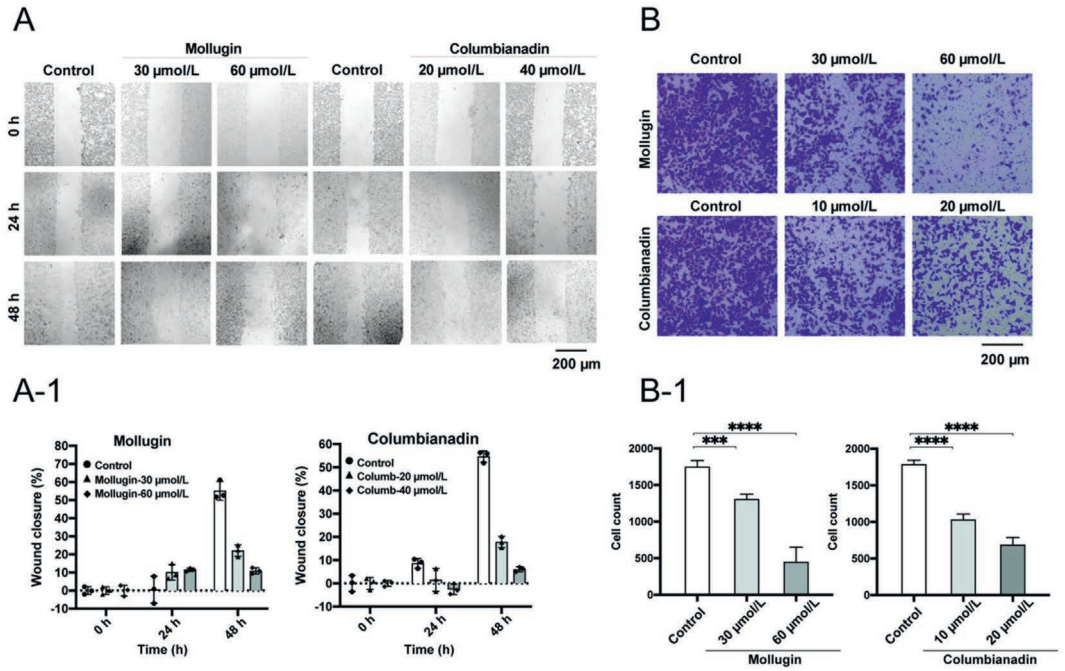

Figure 5 *In vitro* anti-migration and anti-invasive effects of mollugin and columbianadin on melanoma cell lines. (A) Wound healing assay illustrated mollugin and columbianadin effectively restrain melanoma cell migration. Quantification of fold change was shown in A-1. (B) Inhibition of melanoma cell invasiveness confirmed by transwell migration assays. Quantification of fold change was shown in B-1. All images were obtained with 10 × objective. Scale bars: 200 μm.

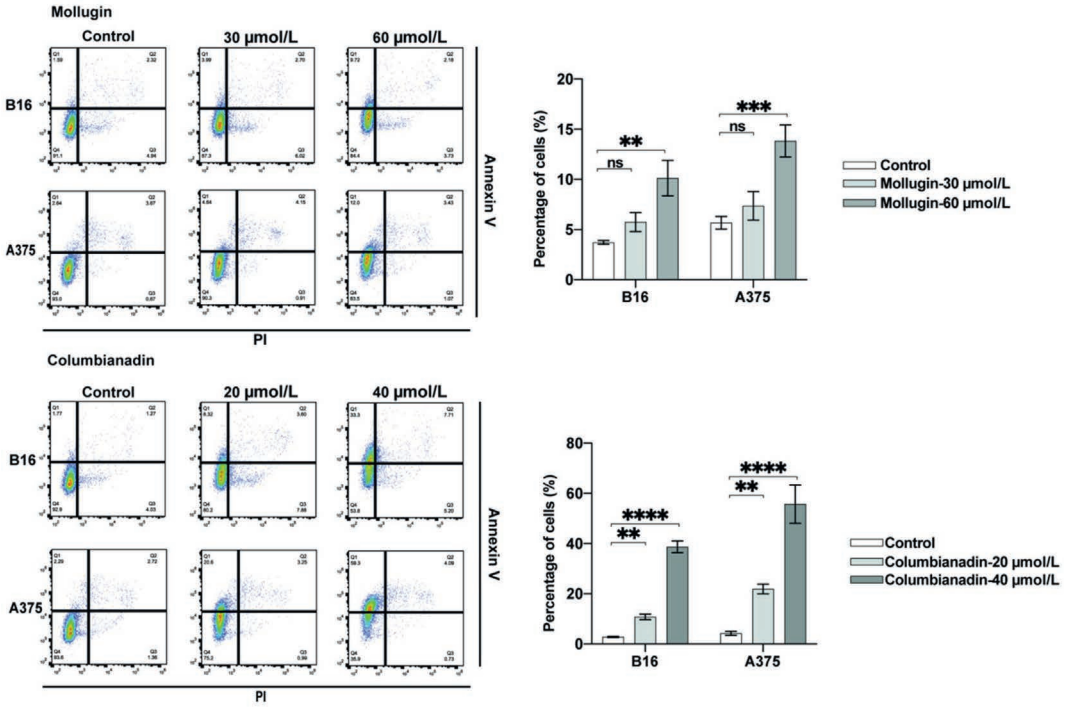

Figure 6 Apoptosis induced by mollugin and columbianadin in melanoma cell. Annexin-V^+ PI^- and Annexin-V^+ PI^+ represents early apoptotic cells and late apoptotic cells, respectively. In all relevant panels, $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. ns, not significant. Data are presented as mean ± SD ($n = 3$).

emerged as a high-ranking candidate (ranked third). FABP5 is a 15-kDa cytoplasmic protein involved in fatty acid uptake, transport, and intracellular metabolism. Studies indicate that FABP5 plays significant roles in metabolic syndrome and cancer progression[67]. Given these findings, we selected FABP5 as a strategic target to investigate mollugin's anti-melanoma mechanism.

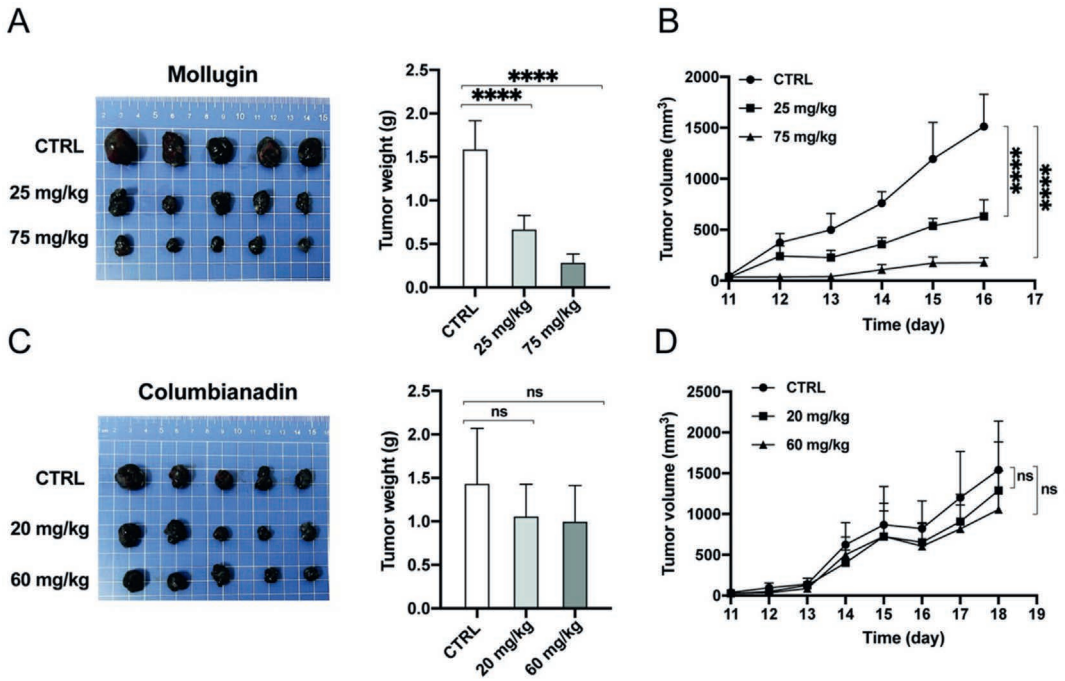

Figure 7 *In vivo* anti-tumor activity of mollugin and columbianadin in a melanoma-CDX model. (A, C) Measurement of tumor weight from collected specimens. (B, D) Tumor growth curves in the melanoma-CDX model. Data information: in all relevant panels, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$. ns, not significant. Data are presented as mean ± SD ($n = 5$).

Molecular dynamics simulations provided key structural insights into the FABP5-mollugin interaction. The simulations revealed that the carbonyl oxygen of mollugin forms a hydrogen bond with Arg109, which are also observed in native fatty acid binding[68] (Supporting Information Fig. S9). Furthermore, mollugin's tricyclic scaffold exhibits extensive shape complementarity with the hydrophobic portal formed by Phe17, Met21, Pro39, Phe58, Ile105, and Val116. These hydrophobic portal residues also contribute to the binding of the native substrate fatty acids. Surface plasmon resonance (SPR) experiments quantitatively confirmed this interaction, with steady-state approximation yielding a dissociation constant (K_D) of 1.03 μmol/L (Supporting Information Fig. S10), indicating moderate binding affinity. Subsequent quantitative RT-PCR analysis demonstrated that mollugin significantly influences lipid metabolic regulation in melanoma cells, showing coordinated downregulation of key metabolic regulators across multiple pathways. Specifically, the compound downregulated lipogenic enzymes (ACC and FASN), β-oxidation mediators (CPT1A), and master transcriptional controllers (SREBPs and PPARγ) (Supporting Information Fig. S11). These results suggest that mollugin disrupts tumor bioenergetic homeostasis through multi-faceted inhibition of lipid metabolic pathways, potentially compromising energy provisioning in melanoma cells.

Through systematic *in vitro* and *in vivo* investigations, we have conclusively validated the anti-melanoma efficacy of both mollugin and columbianadin. Toxicity evaluations revealed no significant cytotoxic effects in non-cancerous cell lines within the anti-tumor concentration range (Fig. 8). Furthermore, histological examination of internal organs from CDX models treated with compounds revealed favorable toxicological profiles, confirming their systemic safety at therapeutic doses. These findings not only establish mollugin and columbianadin as promising therapeutic candidates but also propose novel treatment strategies for melanoma treatment.

3.7. In silico screening and experimental validation of anti-inflammatory terpenoids

Inflammatory response remains one of the most prevalent and serious complications of disease. As important sources of anti-inflammatory agents, terpenoids exert their therapeutic effects primarily through modulation of critical pathological pathways. To assess model generalizability, we conducted *in silico* screening of anti-inflammatory terpenoids using the TerDA platform. Initial *in silico* predictions suggest that 23 terpenoid compounds may possess broad-spectrum anti-inflammatory potential. Subsequent experimental validation in LPS-stimulated Raw264.7 macrophage models confirmed their varying degrees of anti-inflammatory activity (Supporting Information Table S7). Quantitative analysis demonstrated that the majority of tested terpenoids significantly reduced pro-inflammatory cytokine levels (IκBα, CXCL2, IL-1β, and IL-6), while exhibiting minimal impact on TNF-α (Supporting Information Fig. S12). Among the terpenoids, ethisterone, incensole, and Euphorbia factor L3 (5,15-diacetyl-3-benzoyllathyrol) shows anti-inflammatory effect by inhibiting LPS-induced NF-κB activation. Further comparative analysis revealed distinct cytokine expression profiles between prediction-based groups. The high-rank group (rank $\leq$20) showed significantly reduced expression of TNF-α, IκBα, and IL-1β compared to the low-rank group (rank $>$20) ($P < 0.05$; Supporting Information Fig. S13A), while the high-score group (score $\geq$0.95) exhibited markedly lower levels of TNF-α, IκBα, and CXCL2 relative to the low-score group (score $<$0.95) ($P < 0.05$; Fig. S13B). These results confirm the predictive

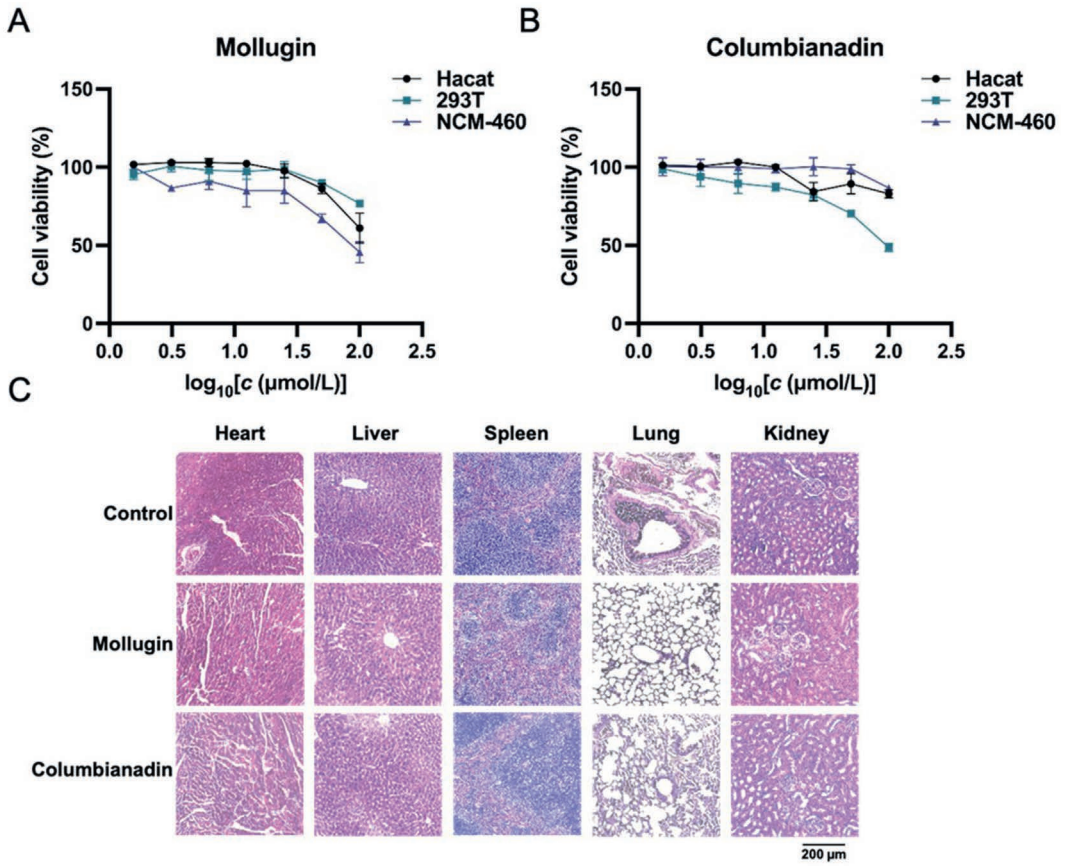

Figure 8 Toxic evaluation of mollugin and columbianadin *in vitro and in vivo*. (A, B) Measurement of toxic effects in normal (non-cancerous) cell lines. The IC_{50} (half maximal inhibitory concentration) of mollugin and columbianadin in non-cancerous cell lines were over 100 μmol/L (C) H&E staining of organs in mice bearing B16-F10 xenografts treated with mollugin or columbianadin.

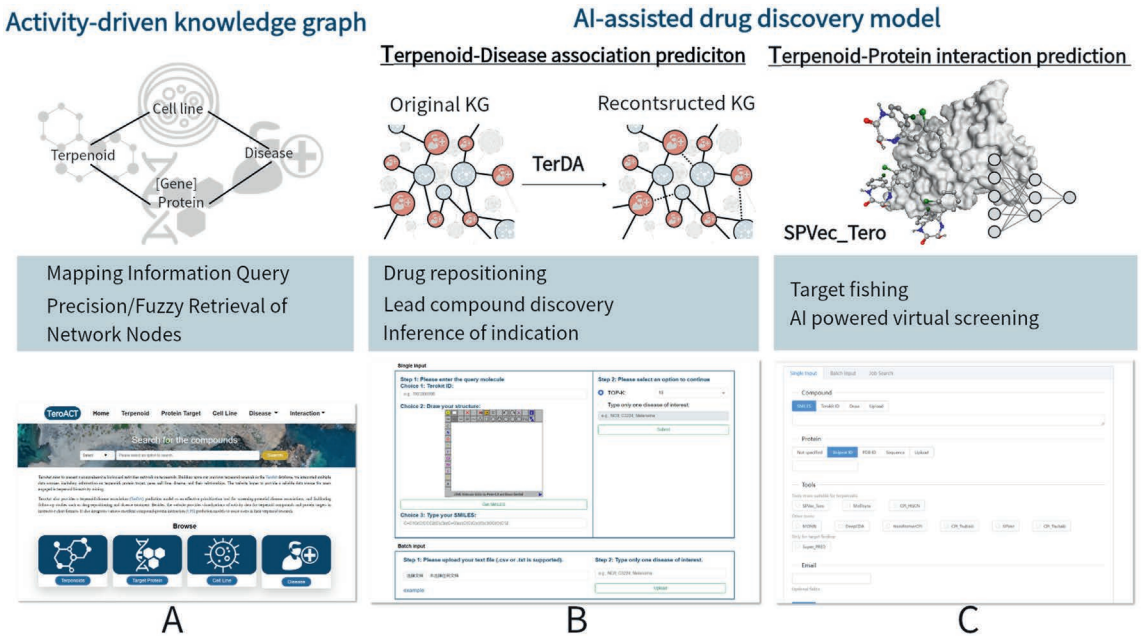

Figure 9 The overall framework and functional profile of TeroACT webserver. (A) Homepage. (B) Entry page of TerDA prediction. (C) Entry page of CPI prediction.

capability of the TerDA model for identifying anti-inflammatory terpenoid compounds. More importantly, these studies demonstrate the platform's versatility beyond oncology applications.

To further validate the discriminative capability of the TerDA model, we systematically curated experimentally verified terpenoid-bioactivity associations from recent literature as an independent test set (Supporting Information Table S8 and Fig. S14). Statistical analysis demonstrated significantly higher prediction rankings for experimentally confirmed active compounds compared to inactive controls, thereby robustly validating the model's discriminative power.

3.8. TeroACT web server

We offer a user-friendly web server, namely TeroACT (http://terokit.qmclab.com/teroact/), that allows users to easily access the comprehensive biological activities knowledge graph on terpenoids. The website synergizes with our previous work Terokit in its application, notably by retaining the same identifiers of terpenoid molecules (*e.g.*, TKC000006). In detail, users can access multi-source information, including terpenoids, protein targets, genes, cell lines, diseases, and their relationships. TeroACT's homepage is designed for easy navigation with an introduction, a search box, and a navigation bar with different browsing categories (Fig. 9A). The search box supports precise and fuzzy retrieval of entities of interest, and users can navigate to the detailed interface of the mapping information centered on that query entity.

The TeroACT platform has embedded various AI-assisted drug discovery models to empower research on terpenoid activity mining. First, TeroACT provides access to our TerDA prediction model (Fig. 9B). The user can enter the molecular ID associated with active knowledge included in the profile to perform drug repositioning, as demonstrated in the melanoma case we have shown. Then there is a selection to either generate a top-k recommendation list or obtain prediction scores and ranks specific to a custom-defined disease. Besides, the model supports batch input of molecules to score and rank them for the interested disease, facilitating the discovery of lead compounds.

TeroACT also integrates many excellent CPI prediction methods, including SPVec_Tero, DeepCDA[69], CPI_GNN[70], MONN[71], CPI_HGCNc[72], MolTrans[73], and TransformerCPI[74]. The development of these web-based toolsets aims to meet researchers 'needs for target fishing and virtual screening for terpenoid molecules against a target. CPI prediction tools enable screen and prioritize terpenoids with high binding potential. They also provide valuable computational support for investigating molecular targets and elucidating mechanistic pathways. Through synergistic integration, the TeroACT platform establishs a comprehensive "terpenoid-target-disease" prediction framework. It not only predicts potential therapeutic diseases but also infers probable pharmacological mechanisms by analyzing compound-target interaction patterns.

The performance of all methods in the CPI-predicted task can be found in Supporting Information Tables S9−S11 and Supporting Information Figs. S15−S19. The entry and output pages of CPI prediction are displayed in Fig. 9C. For a given compound-protein target pair, the prediction results will be presented in the form of comments indicating 'active' or 'inactive'. From the results, SPVec_Tero outperforms these benchmark models on classification tasks.

4. Conclusions

In this study, we introduced TeroACT—a dedicated platform for terpenoid activity discovery, supported by a terpenoid-centric knowledge graph and equipped with computational tools for systematic terpenoid bioactivity exploration. Among these, the TerDA model leverages pre-extracted molecular semantic features and knowledge inference to accurately predict terpenoid−disease associations, demonstrating superior recommendation performance over baseline methods. Additionally, TeroACT integrates optimized machine learning approaches for compound−protein interaction (CPI) prediction tailored to terpenoid target identification. The platform not only provides a comprehensive chemical-bioactivity dataset for efficient information retrieval but also offers valuable insights for the terpenoid-based therapeutic discovery. By combining the platform with experimental validation, we conducted an anti-melanoma agent screening campaign, identifying six high-priority terpenoids with significant cellular anti-tumor efficacy. Among these, two structurally novel compounds, mollugin and columbianadin, further exhibited anti-proliferative and anti-invasive effects in melanoma through a series of *in vitro* and *in vivo* experiments. Looking forward, this study highlights the utility of TeroACT as a scalable and predictive framework for accelerating terpenoid pharmaceutical research. It paves the way for future studies in terpenoid drug repurposing and the inference of new therapeutic indications.

Compliance with ethics requirements

All procedures for animal experiments were approved and carried out in accordance with the Institutional Animal Care and Use Committee of Sun Yat-sen University (SYSU-IACUC-2020-B1041).

Acknowledgments

This study was partly supported by the National Natural Science Foundation of China (22473118, 82430108, 82122069, and 82073869), Guangdong Basic and Applied Basic Research Foundation (2021B1515020004 and 2024B1111140001, China), Guangzhou Science and Technology Planning Program (SL2023A04J01660, China), Fundamental Research Funds for the Central Universities (23yxqntd001 and 24ykzy003, China), and Opening Project of Guangdong Provincial Key Laboratory of New Drug Design and Evaluation (2020B1212060034, China). We are thankful for the Top-Notch Young Talents Program of China. This study was grateful to the National Supercomputer Center in Guangzhou, specifically the "Tianhe-2" supercomputer system, for generously providing us with high-performance computing resources.

Author contributions

XiaoJuan Shen and Shijia Yan analyzed the data and wrote the paper. Kangwei Xu and Xu Kang collaboratively established the TeroACT web server. Yongxing Jian integrated CPI prediction models and developed a web-based prediction service. Tao Zeng supervised data collection. Ruibo Wu and Guohui Wan designed and supervised the project and revised the paper. All authors

revised the article critically for important content. All of the authors have read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Code availability

The constructed terpenoid activity network is deployed on TeroACT. Moreover, the code of TerDA is publicly available from the GitHub repository: https://github.com/shenxj9/TerDA. TerDA is also available at https://zenodo.org/records/12699106. All tools are available as a web server, which is free for academic users without registration at http://terokit.qmclab.com/teroact/.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2025.12.036.

References

1. McGarvey DJ, Croteau R. Terpenoid metabolism. *Plant Cell* 1995;**7**: 1015−26.
2. Roberts SC. Production and engineering of terpenoids in plant cell culture. *Nat Chem Biol* 2007;**3**:387−95.
3. Bergman ME, Kortbeek RWJ, Gutensohn M, Dudareva N. Plant terpenoid biosynthetic network and its multiple layers of regulation. *Prog Lipid Res* 2024;**95**:101287.
4. Schramm K, Skopec M, Dearing D. Metabolomic evidence of independent biotransformation pathways for terpenes in two specialist Mammalian herbivores (genus Neotoma). *Integr Zool* 2023;**19**:143−55.
5. Beekwilder J, Schempp FM, Styles MQ, Zelder O. Microbial synthesis of terpenoids for human nutrition—An emerging field with high business potential. *Curr Opin Biotechnol* 2024;**87**:103099.
6. Ru L, Hui T, Tiangang L. Genome deep mining boosts the discovery of microbial terpenoids. *Synth Biol J* 2024;**5**:507−26.
7. Wallach O. *Terpene und campher: zusammenfassung eigener untersuchungen auf dem Gebiet der alicyclischen Kohlenstoffverbindungen*. Leipzig: Veit; 1914.
8. Zeng T, Liu Z, Liu H, He W, Tang X, Xie L, et al. Exploring chemical and biological space of terpenoids. *J Chem Inf Model* 2019;**59**:3667−78.
9. Zeng T, Hess J, Andes Bernard, Zhang F, Wu R. Bio-inspired chemical space exploration of terpenoids. *Briefings Bioinf* 2022;**23**:bbac197.
10. Wang G, Tang W, Bidigare RR. Terpenoids as therapeutic drugs and pharmaceutical agents. In: Zhang L, Demain AL, editors. *Natural products: drug discovery and therapeutic medicine*. Humana Totowa: Humana Press; 2005. p. 197−227.
11. Kim WS, Shalit ZA, Nguyen SM, Schoepke E, Eastman A, Burris TP, et al. A synthesis strategy for tetracyclic terpenoids leads to agonists of ERβ. *Nat Commun* 2019;**10**:2448.
12. Siddiqui T, Sharma V, Khan M, Gupta K. Terpenoids in essential oils: chemistry, classification, and potential impact on human health and industry. *Phytomed Plus* 2024;**4**:100549.
13. Chopra B, Dhingra AK, Dhar KL, Nepali K. Emerging role of terpenoids for the treatment of cancer: a review. *Mini Rev Med Chem* 2021;**21**:2300−36.
14. Yoon YE, Jung YJ, Lee SJ. The anticancer activities of natural terpenoids that inhibit both melanoma and non-melanoma skin cancers. *Int J Mol Sci* 2024;**25**:4423.
15. Guimarães AC, Meireles LM, Lemos MF, Guimarães MCC, Endringer DC, Fronza M, et al. Antibacterial activity of terpenes and terpenoids present in essential oils. *Molecules* 2019;**24**:2471.
16. Prathapan P. Magic bullets, magic shields, and antimicrobials in between. *Pharm Sci Adv* 2023;**1**:100002.
17. Prottay AAS, Bappi MH, Akbor MS, Asha AI, Bhuia MS, Shafin AA, et al. Sclareol exerts an anti-inflammatory effect, possibly through COXs inhibition pathway: *in vivo* and *in silico* studies. *Pharm Sci Adv* 2024;**2**:100029.
18. Chen Y, Wang Y, He L, Wang L, Zhao J, Yang Z, et al. Supercritical CO_2 extraction of terpenoids from indocalamus latifolius leaves: optimization, purification, and antioxidant activity. *Foods* 2024;**13**: 1719.
19. Yang T, Li C, Xue W, Huang L, Wang Z. Natural immunomodulating substances used for alleviating food allergy. *Crit Rev Food Sci Nutr* 2023;**63**:2407−25.
20. Roy DN. *Terpenoids against human diseases*. Boca Raton: CRC Press; 2019.
21. Lahyaoui M, Bouyahya A, El Menyiy N, Bakrim S, Dakka N. Chapter: neuroprotection induced by terpenes: focus on Alzheimer's disease and Parkinson's disease. In: De Oliveira MR, editor. *Natural molecules in neuroprotection and neurotoxicity*. Amsterdam: Elsevier; 2024. p. 1669−78.
22. Guo W, Wang H, Li C. Signal pathways of melanoma and targeted therapy. *Signal Transduct Targeted Ther* 2021;**6**:424.
23. Yoon YE, Jung YJ, Lee SJ. The anticancer activities of natural terpenoids that inhibit both melanoma and non-melanoma skin cancers. *Int J Mol Sci* 2024;**25**:4423.
24. Rahaman A, Chaudhuri A, Sarkar A, Chakraborty S, Bhattacharjee S, Mandal DP. Eucalyptol targets PI3K/Akt/mTOR pathway to inhibit skin cancer metastasis. *Carcinogenesis* 2022;**43**:571−83.
25. Yan S, Wang Y, Zeng J, Liu D, Qin H, Du C, et al. The pharmacological research of traditional chinese medicine and herbal remedies in the treatment of melanoma. *Pharmacogn Mag* 2025;**21**:738−50.
26. Grudzińska M, Stachnik B, Galanty A, Sołtys A, Podolak I. Progress in antimelanoma research of natural triterpenoids and their derivatives: mechanisms of action, bioavailability enhancement and structure modifications. *Molecules* 2023;**28**:7763.
27. Zeng T, Liu Z, Zhuang J, Jiang Y, He W, Diao H, et al. TeroKit: a database-driven web server for terpenome research. *J Chem Inf Model* 2020;**60**:2082−90.
28. Zeng T, Chen Y, Jian Y, Zhang F, Wu R. Chemotaxonomic investigation of plant terpenoids with an established database (TeroMOL). *New Phytol* 2022;**235**:662−73.
29. Chen N, Zhang R, Zeng T, Zhang X, Wu R. Developing TeroENZ and TeroMAP modules for the terpenome research platform TeroKit. *Database* 2023;**2023**:baad020.
30. Lv Q, Chen G, He H, Yang Z, Zhao L, Zhang K, et al. TCMBank-the largest TCM database provides deep learning-based Chinese−Western medicine exclusion prediction. *Signal Transduct Targeted Ther* 2023; **8**:127.
31. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res* 2024;**52**:D1265−75.
32. Young AL, Oxtoby NP, Garbarino S, Fox NC, Barkhof F, Schott JM, et al. Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. *Nat Rev Neurosci* 2024;**25**:111−30.
33. Klinke DJ, Fernandez A, Deng W, Razazan A, Latifizadeh H, Pirkey AC. Data-driven learning how oncogenic gene expression locally alters heterocellular networks. *Nat Commun* 2022;**13**:1986.
34. Soh J, Park S, Lee H. HIDTI: integration of heterogeneous information to predict drug-target interactions. *Sci Rep* 2022;**12**:3793.
35. Ma M, Lei X, Zhang Y. A review of drug-related associations prediction based on artificial intelligence methods. *Curr Bioinf* 2024;**19**: 530−50.
36. Shen X, Yan S, Zeng T, Xia F, Jiang D, Wan G, et al. TarIKGC: a target identification tool using semantics-enhanced knowledge graph completion with application to CDK2 inhibitor discovery. *J Med Chem* 2025;**68**:1793−809.
37. Socher R, Chen D, Manning CD, Ng A. *Reasoning with neural tensor networks for knowledge base completion*. *Proceedings of the 27th*

international conference on neural information processing systems - volume 1. Nevada: Lake Tahoe; 2013. p. 926−34.
38. Liu H, Zhang W, Song Y, Deng L, Zhou S. HNet-DNN: inferring new drug−disease associations with deep neural network based on heterogeneous network features. *J Chem Inf Model* 2020;**60**:2367−76.
39. Gu YW, Zheng S, Yin QJ, Jiang R, Li J. REDDA: integrating multiple biological relations to heterogeneous graph neural network for drug−disease association prediction. *Comput Biol Med* 2022;**150**: 106127.
40. Ghasemi S, Lakizadeh A. Drug-disease association prediction based on end-to-end multi-layer heterogeneous graph convolutional encoders. *Inform Med Unlocked* 2023;**37**:101177.
41. Huang S, Wang M, Zheng X, Chen J, Tang C. Hierarchical and dynamic graph attention network for drug−disease association prediction. IEEE J Biomed Health Inform 2024;**28**:2416−27.
42. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. *Proceedings of the 26th international conference on neural information processing systems*. Nevada: Lake Tahoe; 2013. p. 2787. 79.
43. Sun L, Wan A, Zhou Z, Chen D, Liang H, Liu C, et al. RNA-binding protein RALY reprogrammes mitochondrial metabolism *via* mediating miRNA processing in colorectal cancer. *Gut* 2021;**70**:1698−712.
44. Bairoch A. The cellosaurus, a cell-line knowledge resource. *J Biomol Tech* 2018;**29**:25−38.
45. Douglas K, Douglas S. *PostgreSQL: a comprehensive guide to building, programming, and administering PostgresSQL databases*. Sams Publishing; 2003.
46. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wiegers J, Wiegers TC, et al. Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res* 2021;**49**:D1138−43.
47. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards version 3: the human gene integrator. *Database* 2010; **2010**:baq020.
48. Zhou Y, Zhang Y, Zhao D, Yu X, Shen X, Zhou Y, et al. TTD: therapeutic target database describing target druggability information. *Nucleic Acids Res* 2024;**52**:D1465−77.
49. Chaturvedi D, Goswami A, Saikia PP, Barua NC, Rao PG. Artemisinin and its derivatives: a novel class of anti-malarial and anti-cancer agents. *Chem Soc Rev* 2010;**39**:435−54.
50. Li WD, Dong YJ, Tu YY, Lin ZB. Lupus symptom of BXSB mice by inhibiting production of TNF-alpha and blocking the signaling pathway NF-κB translocation. *Int Immunopharmacol* 2006;**6**: 1243−50.
51. Tao CY, Shang J, Chen T, Yu DH, Jiang YM, Liu D, et al. Impact of antimalarial (AM) on serum lipids in systemic lupus erythematosus (SLE) patients: a systematic review and meta-analysis. *Medicine* 2019;**98**:e15030.
52. Gendrot M, Duflot I, Boxberger M, Delandre O, Jardot P, Le Bideau M, et al. Antimalarial artemisinin-based combination therapies (ACT) and COVID-19 in Africa: inhibition of SARS-CoV-2 replication by mefloquine-artesunate. *Int J Infect Dis* 2020;**99**:437−40.
53. Li GM, Yuan M, Li HH, Deng CS, Wang Q, Tang YX, et al. Safety and efficacy of artemisinin-piperaquine for treatment of COVID-19: an open-label, non-randomised and controlled trial. *Int J Antimicrob Agents* 2021;**57**:106216.
54. Efferth T, Oesch F. The immunosuppressive activity of artemisinin-type drugs towards inflammatory and autoimmune diseases. *Med Res Rev* 2021;**41**:3023−61.
55. Yu R, Jin GH, Fujimoto M. Dihydroartemisinin: a potential drug for the treatment of malignancies and inflammatory diseases. *Front Oncol* 2021;**11**:722331.
56. Vogt M, Stumpfe D, Maggiora GM, Bajorath J. Lessons learned from the design of chemical space networks and opportunities for new applications. *J Comput Aided Mol Des* 2016;**30**:191−208.
57. Yang B, Yih WT, He X, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases. *International conference on learning representations*. 2014.
58. Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: *Proceedings of the AAAI conference on artificial intelligence*; 2014.
59. Kang JI, Hong JY, Choi JS, Lee SK. Columbianadin inhibits cell proliferation by inducing apoptosis and necroptosis in HCT116 colon cancer cells. *Biomol Ther* 2016;**24**:320.
60. Lim HJ, Lee JH, Choi JS, Lee SK, Kim YS, Kim HP. Inhibition of airway inflammation by the roots of *Angelica decursiva* and its constituent, columbianadin. *J Ethnopharmacol* 2014;**155**:1353−61.
61. Wang W, Zhou L, Zhang X, Li Z. Mollugin suppresses proliferation and drives ferroptosis of colorectal cancer cells through inhibition of insulin-like growth factor 2 mRNA binding protein 3/glutathione peroxidase 4 axis. *Biomed Pharmacother* 2023;**166**: 115427.
62. Jin X, Tang S, Chen Q, Zou J, Zhang T, Liu F, et al. Furazolidone induced oxidative DNA damage *via* up-regulating ROS that caused cell cycle arrest in human hepatoma G2 cells. *Toxicol Lett* 2011;**201**: 205−12.
63. Samur M. Prognosis of metastatic melanoma: where are the details?. *J Clin Oncol* 1999;**17**:3856−60.
64. Tasdogan A, Faubert B, Ramesh V, Ubellacker JM, Shen B, Solmonson A, et al. Metabolic heterogeneity confers differences in melanoma metastatic potential. *Nature* 2020;**577**:115−20.
65. Zhang YB, Yang XW. Tissue distribution study of columbianadin and its active metabolite columbianetin in rats. *Biomed Chromatogr* 2016; **30**:256−62.
66. Zhang YF, Wang X, Kaushik AC, Chu Y, Shan X, Zhao M-Z, et al. SPVec: a Word2vec-inspired feature representation method for drug-target interaction prediction. *Front Chem* 2020;**7**:895.
67. O'Sullivan SE, Kaczocha M. FABP5 as a novel molecular target in prostate cancer. *Drug Discov Today* 2020;**25**:2056−61.
68. Armstrong EH, Goswami D, Griffin PR, Noy N, Ortlund EA. Structural basis for ligand regulation of the fatty acid-binding protein 5, peroxisome proliferator-activated receptor β/δ (FABP5−PPARβ/δ) signaling pathway. *J Biol Chem* 2014;**289**:14941−54.
69. Abbasi K, Razzaghi P, Poso A, Amanlou M, Ghasemi JB, Masoudi-Nejad A. DeepCDA: deep cross-domain compound−protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics* 2020;**36**:4633−42.
70. Tsubaki M, Tomii K, Sese J. Compound−protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* 2019;**35**:309−18.
71. Li S, Wan F, Shu H, Jiang T, Zhao D, Zeng J. MONN: a multi-objective neural network for predicting compound−protein interactions and affinities. *Cell Syst* 2020;**10**:308−22.
72. Bui-Thi D, Rivière E, Meysman P, Laukens K. Predicting compound-protein interaction using hierarchical graph convolutional networks. *PLoS One* 2022;**17**:e0258628.
73. Huang K, Xiao C, Glass LM, Sun J. MolTrans: molecular interaction transformer for drug−target interaction prediction. *Bioinformatics* 2021;**37**:830−6.
74. Chen L, Tan X, Wang D, Zhong F, Liu X, Yang T, et al. TransformerCPI: improving compound−protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* 2020;**36**:4406−14.