Computational and Structural Biotechnology Journal 23 (2024) 3714–3729

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Improving compound-protein interaction prediction by focusing on intra-modality and inter-modality dynamics with a multimodal tensor fusion strategy

Meng Wang [a], Jianmin Wang [b], Jianxin Ji [a], Chenjing Ma [a], Hesong Wang [a], Jia He [a], Yongzhen Song [a], Xuan Zhang [a], Yong Cao [a], Yanyan Dai [a], Menglei Hua [a], Ruihao Qin [a], Kang Li [a,*], Lei Cao [a,*]

[a] Department of Biostatistics, Harbin Medical University, Harbin 150081, China
[b] Department of Integrative Biotechnology, Yonsei University, Incheon 21983, South Korea

ARTICLE INFO

Keywords:
Multimodal fusion framework
Compound-protein interaction
Deep learning
Intra-modality dynamics
Inter-modality dynamics

ABSTRACT

Identifying novel compound–protein interactions (CPIs) plays a pivotal role in target identification and drug discovery. Although the recent multimodal methods have achieved outstanding advances in CPI prediction, they fail to effectively learn both intra-modality and inter-modality dynamics, which limits their prediction performance. To address the limitation, we propose a novel multimodal tensor fusion CPI prediction framework, named MMTF-CPI, which contains three unimodal learning modules for structure, heterogeneous network and transcriptional profiling modalities, a tensor fusion module and a prediction module. MMTF-CPI is capable of focusing on both intra-modality and inter-modality dynamics with the tensor fusion module. We demonstrated that MMTF-CPI is superior to multiple state-of-the-art multimodal methods across seven datasets. The prediction performance of MMTF-CPI is significantly improved with the tensor fusion module compared to other fusion methods. Moreover, our case studies confirmed the practical value of MMTF-CPI in target identification. Via MMTF-CPI, we also discovered several candidate compounds for the therapy of breast cancer and non-small cell lung cancer.

1. Introduction

The development of a novel drug is a time-consuming, expensive and risk-laden process which typically takes 12 years and costs $2.6 billion on average [1]. A major factor contributing to rising costs is the limited success rate in conducting Phase 2 and 3 clinical trials [2,3], which is primarily attributed to suboptimal therapeutic effects and unfavorable off-target toxicities resulting from potential unknown targets of experimental compounds [4]. Therefore, to minimize unexpected costs and maximize of drug effectiveness, identification of the interactions between compounds (drugs) and target proteins plays a crucial role in drug discovery and development.

Although conventional compound-protein interaction (CPI) identification approaches have high reliability [5–7], they are laborious, costly and resource-intensive. Computational approaches have emerged as promising strategies for achieving efficient CPI identification. Given the advantages of deep learning in processing and analyzing biomedical data, many deep learning methods have been widely proposed for predicting CPIs.

A group of approaches focus on utilizing molecular and protein structure data to predict CPIs. For example, Wen et al. used deep belief network (DBN) to extract compound and protein features from fingerprints and descriptors for CPI prediction [8]. Öztürk et al., Yang et al. and Zhao et al. applied two convolutional neural networks (CNNs) to learn compound and protein features from the raw simplified molecular-input line-entry system (SMILES) strings [9] and protein sequences, respectively [10–12]. Then, the learned compound and protein features are concatenated and fed into a fully connected layer to predict CPIs. Tsubaki et al. and Nguyen et al. applied graph neural networks (GNNs) and CNNs to learn compound and protein features, respectively [13,14]. The combination of compound and protein features was subsequently used to predict CPIs. Chen et al. applied a GNN and a

* Corresponding authors.
E-mail addresses: likang@ems.hrbmu.edu.cn (K. Li), caolei@hrbmu.edu.cn (L. Cao).

https://doi.org/10.1016/j.csbj.2024.10.004
Received 13 June 2024; Received in revised form 1 October 2024; Accepted 1 October 2024
Available online 5 October 2024

bidirectional long short-term memory (BiLSTM) for compound and protein representation learning, then the learned representation was employed to predict CPIs [15]. Transformer has been widely applied to encode compound and protein representations from raw SMILES strings and protein sequences for CPI prediction [16–19]. These methods achieved remarkable performance in CPI prediction, but they only focus on single-modal structural representations of compounds and proteins, ignoring the fact that drugs taking effect through therapeutic targets is an intricate biological process involving drug-drug interactions, protein-protein interactions and protein-disease associations.

In recent years, multimodal models integrating structure modality and heterogeneous network modality have been getting lots of attention in CPI prediction tasks. Ye et al. developed a unified framework that integrates heterogeneous information and structural information to enhance feature representation for drug-target interaction (DTI) prediction [20]. Zhou et al. proposed a joint representation framework called MultiDTI, combining the association information of the heterogeneous network and the sequence information of drug and target to predict potential DTIs [21]. Palhamkhani et al. presented a multimodal model that integrates the structure features of compounds and proteins, and the network information for compound-compound and protein-protein interactions [22]. Dehghan et al. introduced DTI-multi-modal, a multimodal approach that fuses drug-drug network, protein-protein network, drug structures and protein sequences to predict DTIs [23]. Zhang et al. combined structure and network information to predict activating/inhibiting mechanisms between drugs and targets [24]. Dong et al. designed a novel multimodal method for drug-protein interactions (DPIs), incorporating both microscopic representations learned from drug SMILES strings and protein sequences, as well as macroscopic representations learned from a single heterogeneous network [25]. Although these multimodal models have promising performance in CPI prediction, they still lack the consideration of transcriptional profiling data, which can provide an unbiased data-driven mechanism for identifying CPIs [26]. To address the issue, Xia et al. proposed MDTips, a multimodal-data-based DTI prediction model, by integrating the knowledge graphs, gene expression profiles, and structural information [27].

Despite the promising performance demonstrated by existing

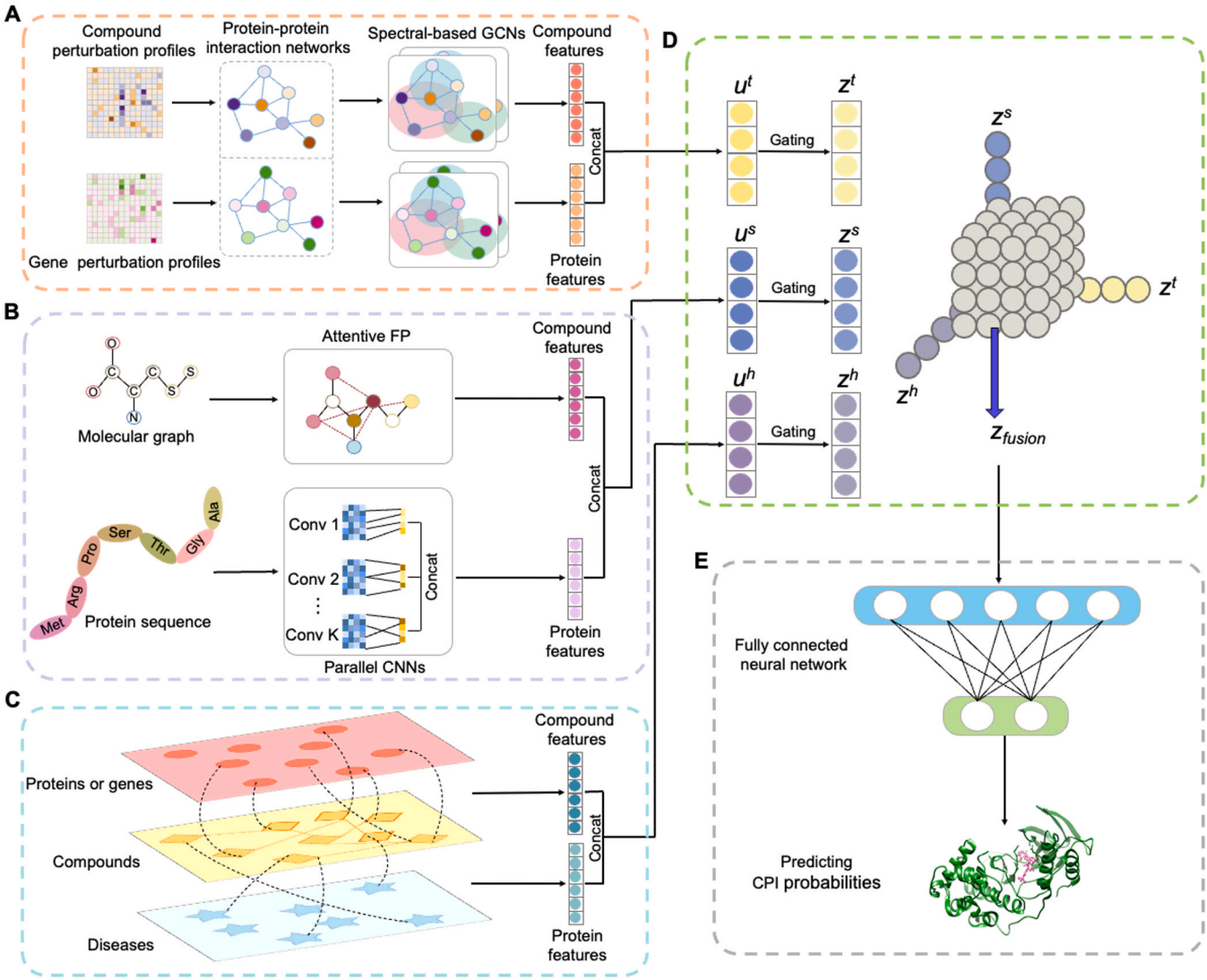

Fig. 1. Overview of the MMTF-CPI framework. (A) Transcriptional profiling modality learning module: two PPI networks are used to learn compound and protein features from raw compound and gene perturbation profiles, respectively; (B) Structure modality learning module: Attentive FP and CNN are used to learn compound and protein features from compound SMILES strings and protein sequences, respectively; (C) Heterogeneous network modality learning module: a heterogeneous network with compound-protein-disease association is applied to learn compound and protein features; (D) Tensor fusion module: the expressiveness of transcriptional profiling, structure and heterogeneous network modalities are controlled and multimodal features are mapped into a common representation; (E) Prediction module: the predictor obtains the probabilities of compound-protein interactions.

multimodal models, they aggregate multiple modalities by simply concatenating multimodal features at the input level [28,29]. There are two challenges for multimodal feature integration in these models. The first challenge is that concatenating multimodal features does not map multiple modalities into a common representation [30], which increases the complexity of inter-modality dynamics, such as the interactions between structure, heterogeneous network and transcriptional profiling modalities. The second challenge is that concatenating multimodal features results in the concatenated feature with collinearity and noise. These challenges make it difficult to efficiently model intra-modality dynamics of a specific modality (unimodal interaction) because the inter-modality dynamics at the input level can be more complex and collinear, leading to model overfitting. Consequently, the existing multimodal models neglect intra-modality dynamics and tend to emphasize more on inter-modality dynamics.

To address these challenges, we propose MMTF-CPI, a novel multimodal tensor fusion framework for CPI prediction based on fusion of structure, heterogeneous network and transcriptional profiling modalities (Fig. 1). The proposed framework is capable of focusing on both intra-modality and inter-modality dynamics simultaneously. First, MMTF-CPI learns unimodal features of transcriptional profiling, structure and heterogeneous network. For transcriptional profiling modality, we constructed two protein-protein interaction (PPI) networks with raw compound and gene perturbation profiles to derive compound and protein embedding features, respectively (Fig. 1**A**). For structure modality, we utilized Attentive FP and CNN to extract compound and protein structure features, respectively (Fig. 1**B**). For heterogeneous network modality, we constructed a heterogeneous network with compound-protein-disease association information, then six schemes of meta-paths are applied to generate compound and protein embedding features from this heterogeneous network (Fig. 1**C**). Second, MMTF-CPI employs a tensor fusion module to reduce the complexity in inter-modality dynamics by mapping multimodal features into a common representation and to reduce the collinearity by controlling the expressiveness information of each modality (Fig. 1**D**). Finally, the fused common representation is sent to a fully connected neural network for predicting probabilities of compound-protein interactions (Fig. 1**E**).

In this study, our contributions are summarized as follows:

(i) We developed a novel multimodal method MMTF-CPI, which fuses structure, heterogeneous network and transcriptional profiling modalities for CPI prediction.
(ii) In MMTF-CPI, we designed a tensor fusion module that can reduce the complexity and collinearity in inter-modality dynamics, which encourages the model to focus on both intra-modality and inter-modality dynamics.
(iii) The tensor fusion module significantly improves the performance of MMTF-CPI, surpassing that of other state-of-the-art multimodal methods.
(iv) We explored the impact of different transcriptional signatures on multimodal performance and illustrated the effectiveness of learning transcriptional profiling and heterogeneous network modalities in MMTF-CPI.
(v) We confirmed the efficacy of MMTF-CPI in target identification and drug discovery.

2. Materials and methods

2.1. Data preparation

We constructed a CPI dataset based on Hetionet knowledge network [31] to evaluate the effectiveness of the proposed model. We collected 39,859 compound-protein interactions from Hetionet, in which 23,146 CPIs have three modalities, including 7,500 compound-upregulated-protein and 15,646 compound-downregulated-protein interactions. Details of the CPI dataset we constructed are shown in Table 1. To obtain a balanced dataset, we randomly sampled unknown compound-protein pairs as negative samples according to a common approach [32], ensuring an equal number to the positive samples. We evaluated the performance of MMTF-CPI using 10-fold cross-validation. First, we randomly partitioned all CPI samples into a training-validation set (90 % of all CPI samples) and a test set (10 % of all CPI samples). Then, for each fold of cross-validation, we randomly divided the training-validation set into a training set (89 % of the training-validation set) and a validation set (11 % of the training-validation set). Finally, such a splitting strategy results in an approximate ratio of 8:1:1 for the training set, validation set and test set. There are no overlapping CPI samples between the training set, validation set and test set. We used the training set to train model, the validation set to tune the hyperparameters, and the test set to evaluate the performance.

The transcriptional profiles of all compounds and proteins/genes are obtained from the phase I L1000 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92742) of the Library of Integrated Network-Based Cellular Signatures (LINCS) program (https://clue.io/) [33], which generates and catalogs gene transcriptional profiles of various cell lines exposed to different perturbing agents in diverse experimental contexts. The profiles in phase I L1000 dataset are produced by L1000 high-throughput gene-expression assay, which involves a set of 978 landmark genes. In this study, we applied the level 5 signature data processed using the moderated z-score (MODZ) (https://clue.io/releases/data-dashboard) and used only real measured expression values of the landmark genes. Based on the previous observation [34,35] that cell type significantly impact the distribution of transcriptional profiles, we evaluated the influence of cell type on transcriptional profiles using gene knockdown induced signatures (trt_sh), gene overexpression induced signatures (trt_oe) and gene expression without any perturbation (ctl_vector) across seven cell lines including MCF7, A375, PC3, HT29, A549, HEPG2, HA1E. As demonstrated in Supplementary Fig. S1, we observed distinct clusters in the trt_sh, trt_oe, and ctl_vector signatures across the seven cell lines, suggesting variability in properties among different cell lines. Thus, it is highly valuable to conduct unbiased estimations of the effectiveness of MMTF-CPI using transcriptional signatures from different cell lines. To impartially evaluate model performance, we divided our CPI dataset into seven datasets using transcriptional signatures from six cancer cell lines (MCF7, A375, PC3, HT29, A549, HEPG2) and one non-cancer cell line (HA1E). All compounds and proteins in each dataset have structural, heterogeneous network, and transcriptional profiling modalities. Details of seven datasets are described in Supplementary Table S1.

2.2. Model architecture

The proposed MMTF-CPI consists of transcriptional profiling modality learning module (Fig. 1**A**), structure modality learning module (Fig. 1**B**), heterogeneous network modality learning module (Fig. 1**C**), tensor fusion module (Fig. 1**D**) and prediction module (Fig. 1**E**).

2.2.1. Transcriptional profiling modality learning

To learn compound and protein transcriptional profiling modality features, two protein-protein interaction (PPI) networks are constructed using compound perturbation and gene perturbation induced transcriptional profiles, respectively. PPIs are essential for numerous biological processes and crucial for the development of human health and disease states [36]. The PPI networks can be represented as graphs, where each node denotes a gene and its property is the corresponding gene expression value in transcriptional profiles. The entire PPI network

Table 1
Statistics of our CPI dataset.

	Drugs	Targets	Interactions
Upregulate	577	892	7,500
Downregulate	716	1,478	15,646

includes 978 nodes corresponding to the landmark genes in transcriptional profiles.

We constructed two spectral-based graph convolutional networks (GCNs) to capture the topological structure information from two PPI networks for extracting compound and protein embedding features, respectively. The input are compound or protein transcriptional profiles and a layer-wise spectral-based GCN can be expressed as:

$$H_l = f\left(U\omega U^T H_{l-1}\right) \tag{1}$$

where U is an orthogonal matrix calculated using the adjacency matrix of the PPI network and its column vector is the eigenvector of the Laplacian matrix. ω denotes a trainable diagonal matrix and $f(\cdot)$ denotes ReLU activation function. H_l is hidden vector at the l-th layer of the spectral-based GCN.

2.2.2. Structure modality learning

2.2.2.1. Compound representation learning. We first collected compound SMILES strings from the DrugBank database (version 5.1.10) [37] and converted them into graphs by RDKit [38]. In molecular graphs, nodes and edges represent atoms and bonds, respectively. Then, we used DGL-LifeSci [39] to initialize atom features of nine types and bond features of four types for characterizing atoms and describing their local surroundings, as listed in Supplementary Tables S2 and S3. Consequently, a node and an edge can be represented by a 39-dimensional and an 11-dimensional feature vector, respectively. Then, we employed Attentive FP to learn compound features with the node and bond feature vectors.

Attentive FP, a graph neural network (GNN), introduces an attention mechanism that captures both the atomic local environment and nonlocal effects at the intramolecular level [40,41]. Attentive FP consists of two main steps: atom embedding and molecule embedding. Atom embedding and molecule embedding use stacked attentive layers to learn local environment and nonlocal effects, respectively. As general GNN [42], the atom embedding of Attentive FP includes a messaging phase and a readout phase.

In the messaging phase, each atom gathers local information from its neighboring atoms and bonds, the information messaging is as follows:

$$C_v^{i-1} = elu\left(\sum_{u\in N(v)} \alpha_{vu}^{i-1} \cdot W \cdot h_v^{i-1}\right) \tag{2}$$

where, $i \in \{1, 2, \ldots, k\}$ is the i-th iteration, $N(v)$ is the set of neighboring atoms of atom v, W is trainable weight matrix, h_v^{i-1} is the state vector of target node v at (i-1)-th iteration. When $i = 1$, h_v^0 is generated by a fully connected layer that includes only the initial atom and bond features. $elu(\cdot)$ is ELU activation function. α_{vu}^{i-1} is the weight of neighbor node u to target node v:

$$\alpha_{vu}^{i-1} = softmax(e_{vu}) = \frac{\exp\left(e_{vu}^{i-1}\right)}{\sum_{u\in N(v)} \exp\left(e_{vu}^{i-1}\right)} \tag{3}$$

e_{vu}^{i-1} is the alignment vector of target node v and neighbor node u, the alignment operation is performed as follow:

$$e_{vu}^{i-1} = leaky\_relu\left(W \cdot \left[h_v^{i-1}, h_u^{i-1}\right]\right) \tag{4}$$

where W is trainable weight matrix and $leaky\_relu(\cdot)$ is leaky ReLU activation function. h_u^{i-1} is the state vector of neighbor node u at (i-1)-th iteration, its initial feature is computed by concatenating initial atom and bond feature, followed by a fully connected layer.

In the readout phase, the state vector h_v^i of target node v is updated by feeding the output of the messaging phase C_v^{i-1} and state vector h_v^{i-1} into a gated recurrent unit (GRU), as follows:

$$h_v^i = GRU\left(C_v^{i-1}, h_v^{i-1}\right) \tag{5}$$

For the molecule embedding, Attentive FP combines the individual atom state vectors into a full-molecule state vector. First, Attentive FP constructs a super virtual node that connects all the nodes of the molecular graph. Second, the super virtual node is embedded using the same attention mechanism in atom embedding. Finally, a state vector for the whole molecule is generated. Thus, the whole molecule can be embedded in the same way as the individual atoms.

2.2.2.2. Protein representation learning. We converted protein sequences collected from the Uniprot database [43] into sequential representations by splitting them into overlapping 3-gram amino acid sequences, which we defined as word sequences. Then, we translated all words into real-valued embeddings using the pre-trained approach Word2vec [44]. We represented protein sequences as real-valued 100-dimensional vectors using a pre-trained Word2vec model [16] on the large corpus constructed from all human protein sequences in UniProt.

In recent years, convolutional neural networks (CNNs) have gained more popularity in medical image analysis [45,46] as well as protein sequence processing [10,47]. We applied multiple parallel 1D-convolutional layers with different kernel sizes to enhance the learning of protein features from different perspectives [48]. Using 100-dimensional protein embeddings as input, the output hidden vectors is calculated by convolutional layers as follows:

$$c_k^l = f\left(\sigma_{bn}\left(W_k \cdot c_k^{l-1} + b_k\right)\right) \tag{6}$$

where $k \in \{1, 2, \ldots, K\}$ is the number of different convolution kernels. $l \in \{1, 2, \ldots, L\}$ is the number of convolutional layers. $W_k \in \mathbb{R}^{100\times d}$ and b_k are weight matrix and bias in the l-th convolutional layer, respectively. d is the dimension of the hidden vector. σ_{bn} stands for BatchNorm operation. $f(\cdot)$ is ReLU activation function. We obtained the final protein features by concatenating the hidden vectors:

$$p = < c_1^L, c_2^L, \ldots, c_K^L > \tag{7}$$

2.2.3. Heterogeneous network modality learning

We first constructed a heterogeneous network with compound-protein-disease associations collected from Hetionet. The heterogeneous network includes 1,435 compounds, 4,739 proteins and 91 diseases. The heterogeneous network can be represented as a graph $G = (V, E, T)$, where $v_i \in V$ denotes node i, $e \in E$ denotes edge, T_V and T_E denote the sets of node and edge types in the heterogeneous network. Then, we used meta-path-based random walks in heterogeneous networks to represent compounds and proteins. A meta-path scheme S is denoted in the form of $V_1 \longrightarrow^{E_1} V_2 \longrightarrow^{E_2} V_3 \cdots \longrightarrow^{E_{t-1}} V_t \rightarrow^{E_t} V_{t+1}$, where $V \in T_V$ and $E \in T_E$. Essentially, a meta-path describes various composite relations in different types of nodes. Given a heterogeneous network $G = (V, E, T)$ and a meta-path scheme S, the definition of transition probability at step i is as follows:

$$p\left(v_{i+1}|v_i^t\right) = \begin{cases} \dfrac{1}{N^{t+1}\left(v_i^t\right)}, & \left(v_{i+1}, v_i^t\right) \in E; S(v_{i+1}) = V_{t+1} \\ 0, & \left(v_{i+1}, v_i^t\right) \in E; S(v_{i+1}) \neq V_{t+1} \\ 0, & \left(v_{i+1}, v_i^t\right) \notin E \end{cases} \tag{8}$$

where $v_i^t \in V_t$ represents node i of type t, v_{i+1} is the neighboring node of v_i^t, $N^{t+1}\left(v_i^t\right)$ represents the V_{t+1} type of neighboring nodes of node v_i^t. $S(v_{i+1})$ represents the type of node v_{i+1} under the given meta-path scheme S.

We defined six meta-path schemes, including protein-compound-protein, protein-compound-disease-compound-protein, protein-compound-compound-protein, compound-compound, compound-protein-compound and compound-disease-compound. In each schema,

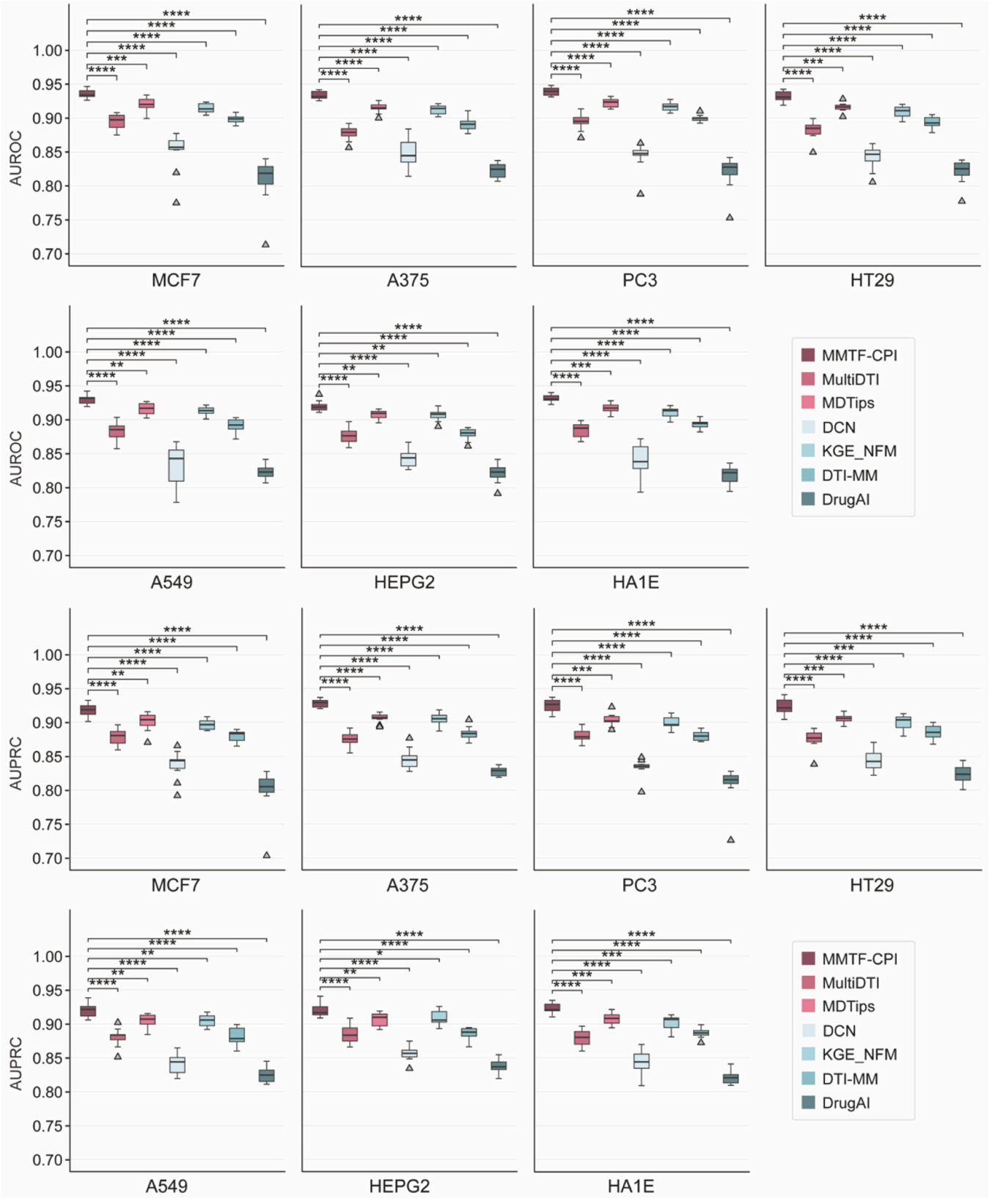

Fig. 2. Comparison of the performance between MMTF-CPI and state-of-the-art multimodal CPI prediction methods on seven datasets.

all compound or protein nodes are individually utilized as the initial node to ensure structural integrity, facilitating feature learning of relations. Finally, a skip-gram model [44] is applied to learn embedding representations of compounds and proteins.

2.2.4. *Multimodal tensor fusion*

After learning the features of structure, heterogeneous network and transcriptional profiling modalities of compounds and proteins, the tensor fusion module maps the three unimodal features into a common

Table 2
The performance of MMTF-CPI and unimodal methods.

	Methods	MCF7	A375	PC3	HT29	A549	HEPG2	HA1E
AUROC	MMTF-CPI	**0.944**	**0.940**	**0.945**	**0.934**	**0.933**	**0.921**	**0.930**
	S_model	0.903	0.900	0.907	0.897	0.900	0.888	0.897
	H_model	0.881	0.880	0.882	0.879	0.874	0.877	0.879
	T_model	0.914	0.911	0.915	0.911	0.908	0.898	0.910
	DeepDTA	0.909	0.904	0.911	0.904	0.905	0.894	0.903
	CPI_GNN	0.860	0.864	0.893	0.846	0.855	0.877	0.846
	NeoDTI	0.897	0.900	0.896	0.834	0.875	0.907	0.915
AUPRC	MMTF-CPI	**0.926**	**0.940**	**0.934**	**0.923**	**0.924**	**0.917**	**0.919**
	S_model	0.883	0.893	0.888	0.889	0.891	0.892	0.888
	H_model	0.857	0.870	0.859	0.866	0.863	0.880	0.866
	T_model	0.895	0.903	0.895	0.903	0.898	0.900	0.902
	DeepDTA	0.888	0.896	0.892	0.893	0.894	0.896	0.892
	CPI_GNN	0.837	0.851	0.870	0.834	0.843	0.877	0.833
	NeoDTI	0.870	0.888	0.871	0.817	0.858	0.904	0.894

representation using Kronecker Product, which reduces the complexity of inter-modality dynamics and captures important interactions across these modalities by modeling pairwise feature interactions. To prevent excessively focusing on inter-modality dynamics and neglecting intra-modality dynamics, we appended the extra constant dimension with value 1 to each unimodal feature representation to preserve unimodal features when capturing important interactions across the three modalities. This definition is shown the equation below:

$$z_{fusion} = \begin{bmatrix} z^s \\ 1 \end{bmatrix} \otimes \begin{bmatrix} z^h \\ 1 \end{bmatrix} \otimes \begin{bmatrix} z^t \\ 1 \end{bmatrix} \tag{9}$$

where $z^s \in \mathbb{R}^{40\times 1}$ represents gated structure modality features, $z^h \in \mathbb{R}^{16\times 1}$ represents gated heterogeneous network modality features, and $z^t \in \mathbb{R}^{25\times 1}$ represents gated transcriptional profiling modality features. $\otimes$ is the Kronecker Product and $z_{fusion} \in \mathbb{R}^{41\times 17\times 26}$ is the common representation that forms in a 3D Cartesian space. In this computation, every neuron in z^s is multiplied by every other neuron in z^h, and subsequently multiplied with every other neuron in z^t. Finally, the common representation z_{fusion} captures inter-modality dynamics and intra-modality generated with the extra constant dimension.

To reduce the impact of collinearity and noise in the fusion of multimodal features, we additionally applied a gating-based attention mechanism before the Kronecker Product. The gating-based attention mechanism can control the expressiveness of each modality [49]. The gating-based attention mechanism in the tensor fusion module is defined as:

$$g^n = softmax\left(W^{sht\to n} \cdot < u^s, u^h, u^t >\right) \tag{10}$$

$$z^n = g^n * u^n, \forall n \in \{s, h, t\} \tag{11}$$

where $W^{sht\to n}$ are learnable weight matrices for feature gating, u^s, u^h, u^t is the unimodal features from structure, heterogeneous network, transcriptional profiling modality learning modules, respectively, $< \cdot >$ represents the concatenation operation. For modality n, the attention weight g^n can be learned as shown in Eq. (10), which is the relative importance of each unimodal feature to the modality. Then, the gated representation z^n is obtained by taking the element-wise product of attention weight g^n and unimodal features u^n, as presented in Eq. (11). Following the tenor fusion module, we propagated the common representation through a hidden layer of size 256, which is subsequently supervised using a cross-entropy-based loss function for predicting CPIs.

3. Experiment setup

3.1. Evaluation metrics

We used AUROC (area under the receiver operating characteristics curve) and AUPRC (area under the precision-recall curve) to measure the performance of MMTF-CPI.

3.2. Comparison to baseline methods

In the comparative analysis, we compared our MMTF-CPI with following state-of-the-art multimodal models: MultiDTI [21], MDTips [27], DeepCompoundNet [22], KGE_NFM [20], DTI-multi-modal [23], DrugAI [24]. The description of comparison methods is shown in Supplementary Experiment Setup.

4. Results and discussion

4.1. Model performance

To provide insight into the prediction performance of MMTF-CPI, we compared against recent state-of-the-art multimodal models, i.e. MultiDTI [21], MDTips [27], DeepCompoundNet (DCN) [22], KGE_NFM [20], DTI-multi-modal (DTI-MM) [23], DrugAI [24] in seven datasets (MCF7, A375, PC3, HT29, A549, HEPG2, HA1E). Comparison results are displayed in Fig. 2. MMTF-CPI achieved the best AUROC and AUPRC values, significantly outperforming all competitive approaches in all datasets. We observed that MMTF-CPI and MDTips, the models integrating structure, heterogeneous network, and transcriptional profiling modalities, exhibited better prediction performance than MultiDTI, DCN, KGE_NFM, DTI-MM, DrugAI, which only include structure and heterogeneous network modalities. Our results indicate that the transcriptional profiling modality improves CPI prediction performance, which is consistent with the findings of Xia et al. [27]. It is noteworthy that MMTF-CPI significantly outperforms MDTips in all datasets, which is attributed to our multimodal tensor fusion module. Our model not only focuses on inter-modality dynamics but also emphasizes intra-modality dynamics during model training, enhancing the prediction performance of the model. In contrast, MDTips only concatenates multimodal features at input level, which may be a limitation in predicting CPIs.

To verify the effectiveness of unimodal information, we also compared MMTF-CPI with its unimodal modules: the structure modality learning module (S_model), the heterogeneous network modality learning module (H_model), the transcriptional profiling modality learning module (T_model), as well as other unimodal methods, including two structure modality based methods (DeepDTA and CPI_GNN) [10,13] and one heterogeneous network modality based method (NeoDTI) [50]. As shown in Table 2, we observed that MMTF-CPI achieves the best performance compared with other unimodal methods in seven datasets, demonstrating that the features of structure, heterogeneous network, and transcriptional profiling modalities all contribute to the model performance.

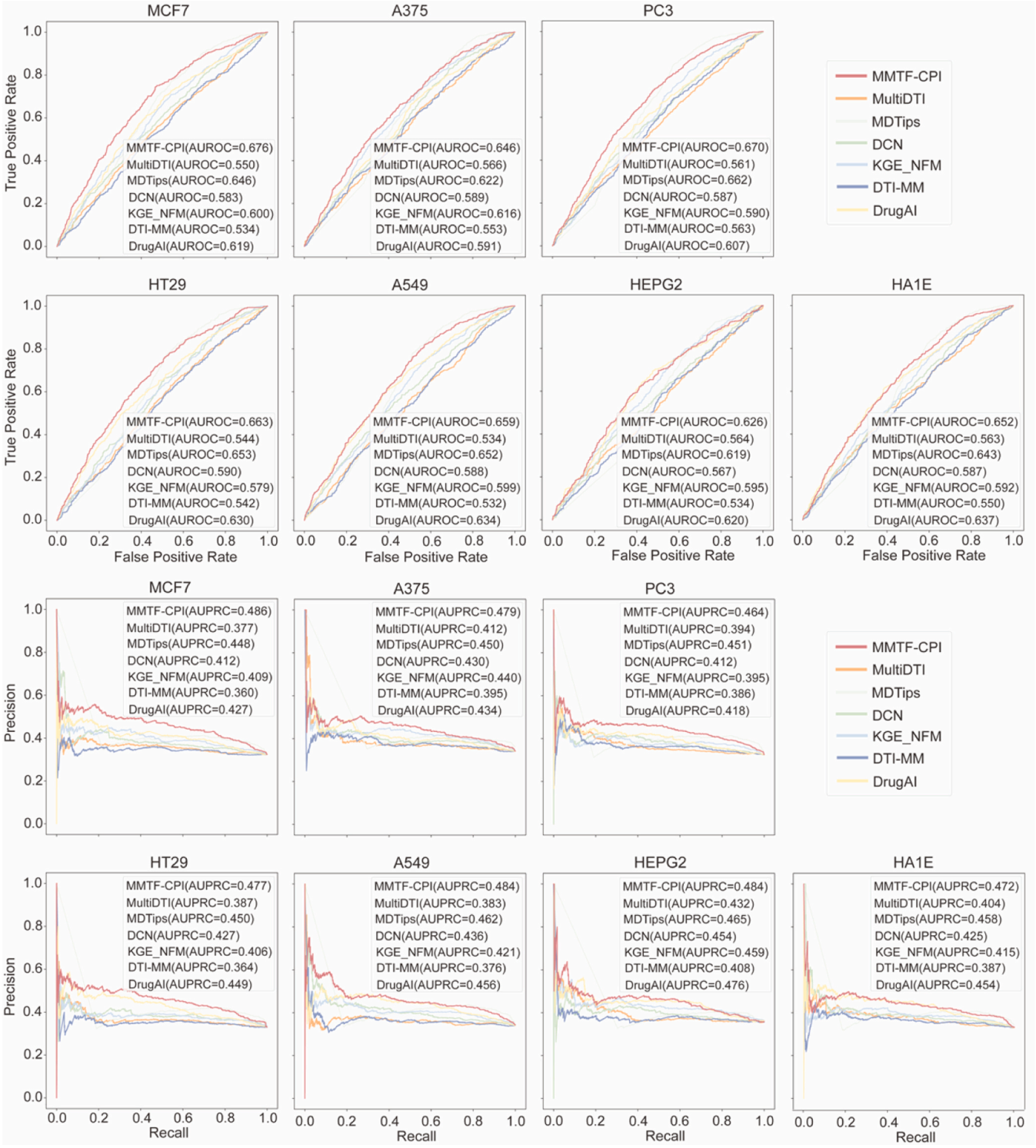

Fig. 3. Evaluation of the generalizability between MMTF-CPI and state-of-the-art multimodal methods in extra independent test sets of seven datasets.

4.2. Generalizability to real-world data

It is crucial to ensure the generalizability of CPI prediction models to real-world data in the practice of drug discovery. To evaluate the generalizability of MMTF-CPI, we constructed extra independent test sets from seven datasets to simulate real-world conditions for predicting CPIs. The independent test set is unbalanced, with a positive-to-negative sample ratio of 1:2, and both the compounds and proteins in the independent test set are unseen in the training set and validation set. Additionally, the compound protein interactions in the independent test set are completely filtered from the heterogeneous networks. The training set and validation set are balanced, with a positive-to-negative sample ratio of 1:1. The training set is used to train the model, validation set to tune the parameters, independent test set to evaluate the generalizability of the model. As depicted in Fig. 3, the evaluation results obtained in seven datasets are encouraging, indicating MMTF-CPI generalized well in real-world data compared to other multimodal models.

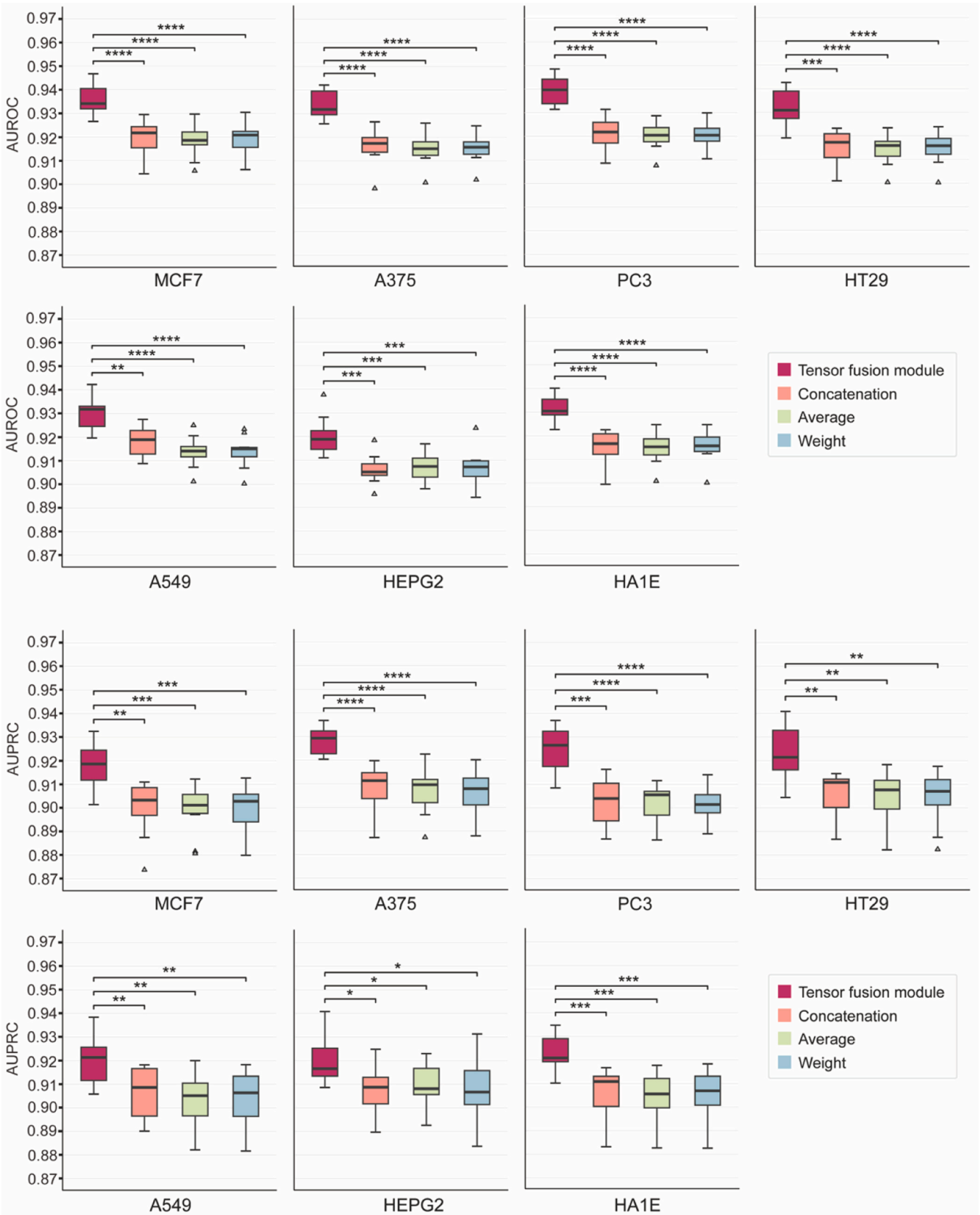

Fig. 4. Performance comparison of MMTF-CPI with different feature fusion methods.

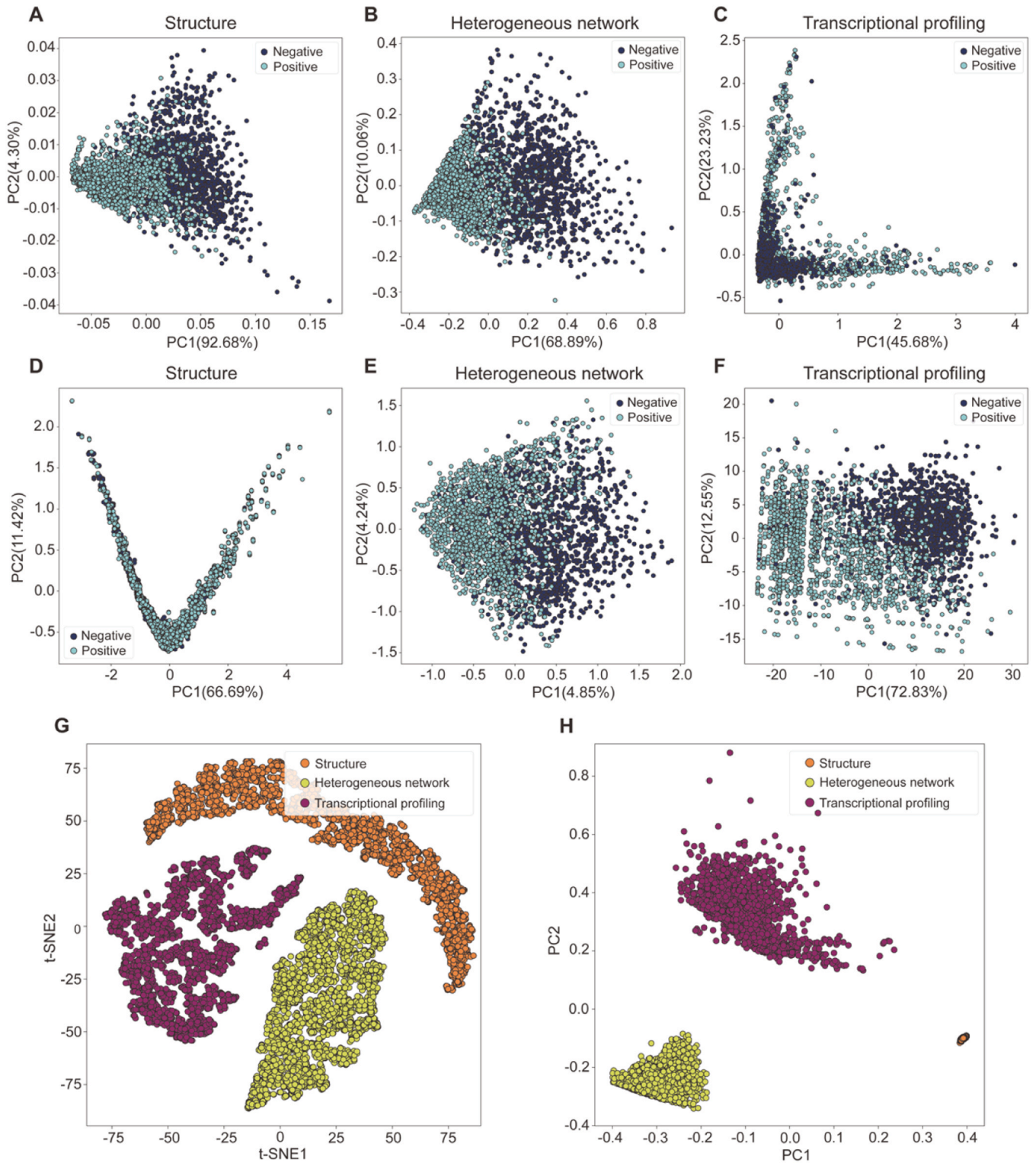

Fig. 5. Feature visualization in PCA and t-SNE. (A-C) PCA plots for structure, heterogeneous network, transcriptional profiling modality features in MMTF-CPI with the tensor fusion module; (D-F) PCA plots for structure, heterogeneous network, transcriptional profiling modality features in MMTF-CPI with concatenation; (G) t-SNE plot for three unimodal features; (H) PCA plot for three unimodal features.

4.3. Efficiency of the tensor fusion module

4.3.1. Tensor fusion module improves prediction performance

To validate the ability of the tensor fusion module for improving multimodal performance, we evaluated model performance with Concatenation, Average and Weight as feature fusion methods, which are currently used in other multimodal methods. Concatenation represents that we directly concatenated the three learned unimodal features from structure, heterogeneous network, and transcriptional profiling modality learning modules for CPI prediction, and the dimension of the concatenated feature is the sum of the sizes of each unimodal feature. Average denotes that the final prediction result is the average of the

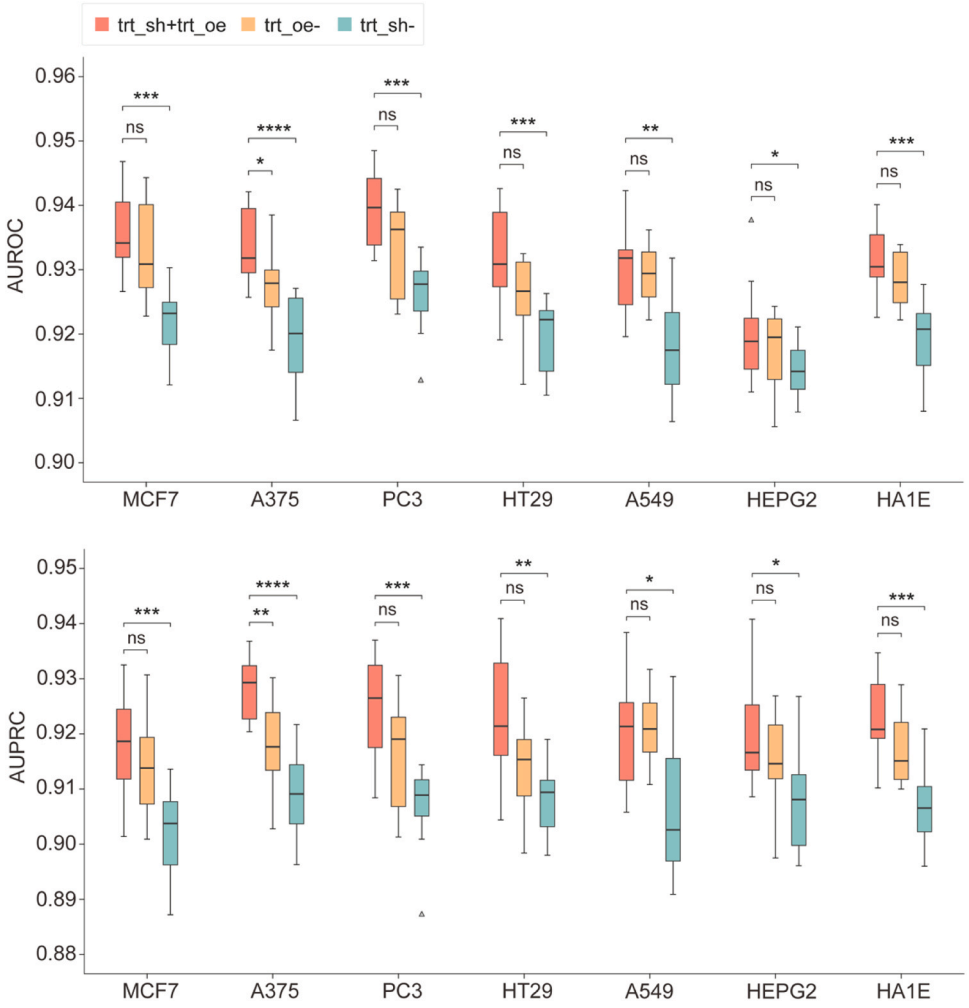

Fig. 6. Performance comparison of MMTF-CPI with different signatures in seven datasets. Trt_oe+trt_sh denotes both the signatures induced by gene knockdown and gene overexpression; Trt_oe- denotes only the signatures induced by gene knockdown; Trt_sh- denotes only the signatures induced by gene overexpression.

prediction results based on three unimodal features, respectively. Weight indicates that each prediction result, based on three unimodal features, is assigned a weight, and the final result is computed as the sum of all prediction results multiplied by their corresponding weights. Fig. 4 illustrates the comparison of multimodal performance using different fusion strategies in seven datasets. We observed that MMTF-CPI using the tensor fusion module for feature fusion demonstrated the best performance in all seven datasets, confirming the effectiveness of the tensor fusion module in improving multimodal performance of MMTF-CPI.

4.3.2. Tensor fusion module enhances the learning of intra-modality dynamics

To further explore whether the tensor fusion module focuses on intra-modality dynamics, we visualized the three unimodal features using principal component analysis (PCA). Different from the tensor fusion module and Concatenation, Average and Weight combine multiple prediction results from various models constructed based on different modules rather than a unified model. In this section, we emphasized analyzing the visualization of unimodal features in MMTF-CPI with the tensor fusion module and concatenation. As displayed in Fig. 5, the dark blue and cyan plots represent the unimodal features labeled as negative and positive CPIs, respectively. From Figs. 5**A** and 5**D**, we observed that the tensor fusion module promotes MMTF-CPI to focus more on structure modality features than concatenation. Figs. 5**B** and 5**E** also demonstrate that the tensor fusion module significantly improves the concentration of heterogeneous network modality features. Although the PCA conducted on the transcriptional profiling modality features in MMTF-CPI with concatenation (Fig. 5**F**) obtains a better result than the tensor fusion module (Fig. 5**C**), it is mainly attributed to the complexity of the transcriptional modality learning module containing a large number of parameters (4,421,740), which makes MMTF-CPI with concatenation overfit the transcriptional modality features. These results indicate that MMTF-CPI with the tensor fusion module can balance attention across all three intra-modality dynamics more effectively than with concatenation. To clearly display the features of each modality, we also provided t-SNE and PCA plots for the three unimodal features (Figs. 5**G** and 5**H**).

4.4. The impact of different transcriptional signatures on prediction performance

Although the previous study of Xia et al. has demonstrated the potential improvement in multimodal performance by incorporating transcriptional profiling, the importance of different types of transcriptional signatures cross various cell lines in CPI prediction has not been analyzed in detail. To investigate the importance of different types of transcriptional signatures in CPI prediction, we used the signatures induced by gene knockdown (trt_sh), the signatures induced by gene overexpression (trt_oe), and the combination of trt_sh and trt_oe from seven cell lines to predict CPIs (Fig. 6). The performance of MMTF-CPI with both trt_sh and trt_oe (trt_sh+trt_oe) is superior to that with only trt_oe (trt_sh-), indicating that the signatures induced by gene

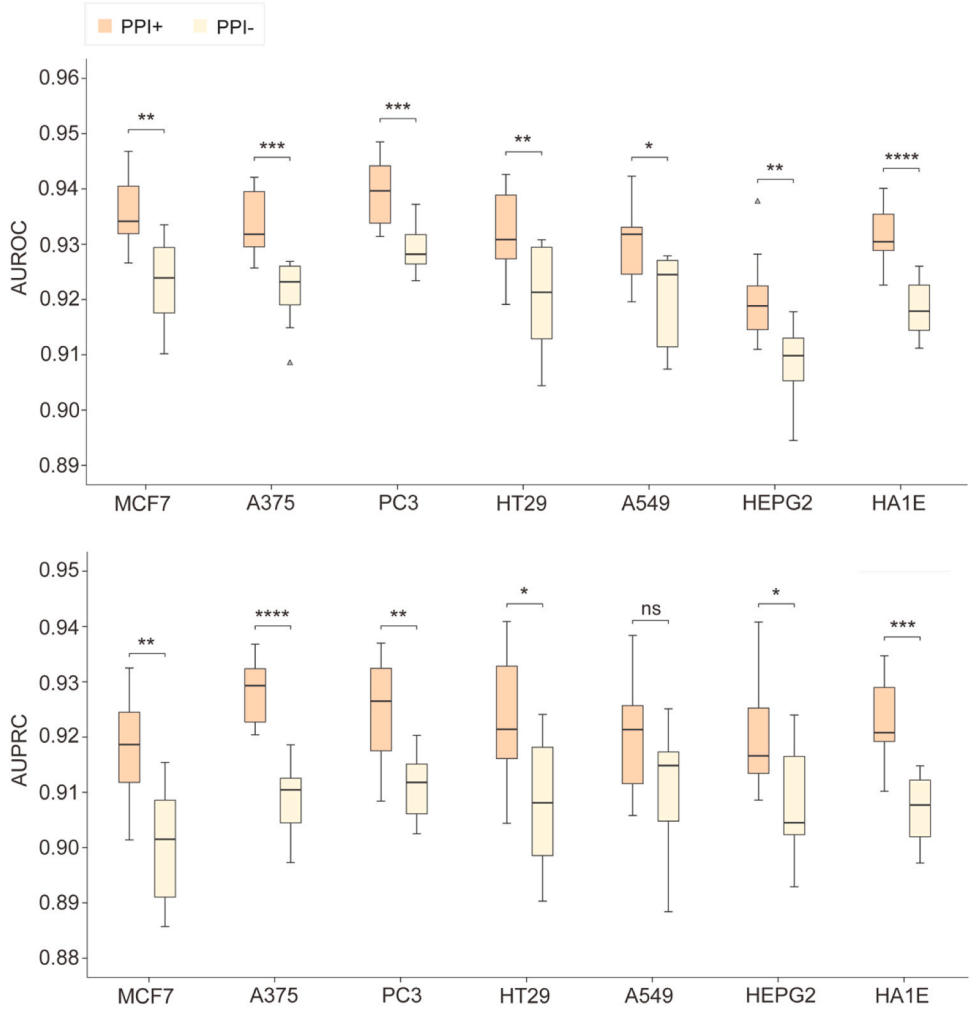

Fig. 7. Ablation experiments of the transcriptional profiling modality learning. PPI+ denotes the construction of two PPI networks for compound and protein transcriptional profiles, followed by the application of spectral-based GCNs to extract compound and protein embeddings, respectively. PPI- denotes that the original transcriptional profiles are directly used in feature fusion.

knockdown from all seven cell lines significantly enhance the prediction performance. We found that the performance of MMTF-CPI using both trt_sh and trt_oe significantly outperforms that of MMTF-CPI using only trt_sh (trt_oe-) in A375 dataset. This result illustrates that the signatures induced by gene overexpression from the A375 cell line can significantly enhance model performance. However, the gene overexpression signatures from other cell lines (MCF7, PC3, HT29, A549, HEPG2, HA1E) do not effectively improve the multimodal performance in CPI prediction. This may be because the melanoma cell line A375 has a higher genetic mutation burden compared to other cell lines. Mutations of genes (such as BRAF, NRAS, TP53, CDKN2A, and RB1), which are included in the PPI networks of MMTF-CPI, frequently occur in melanoma [51], leading to changed gene expression. These changes of gene expression result in gene overexpression providing more important information for CPI prediction. Therefore, predicting CPIs in A375 dataset needs to use signatures induced by both gene knockdown and overexpression.

4.5. Importance of transcriptional profiling and heterogeneous network modality learning strategies

We further conducted a series of ablation experiments in seven datasets to explore the importance of transcriptional profiling modality learning strategy for in improving the performance of MMTF-CPI. In MMTF-CPI, two PPI networks are built with compound perturbation and gene perturbation induced transcriptional profiles. Subsequently, two spectral-based GCNs are employed to capture the topological structure information of two PPI networks for compound and protein feature extraction, and we denoted the model configuration as PPI+ in this section. Next, we constructed PPI- that PPI networks and spectral-based GCNs are removed from MMTF-CPI. Specifically, the original transcriptional profiles are directly fused with the features from structure and heterogeneous network modality learning modules. We compared PPI+ with PPI- in seven datasets, the results are as shown in Fig. 7. We observed that the proposed model MMTF-CPI with PPI networks and spectral-based GCNs achieved superior performance in all the seven datasets, indicating that the transcriptional profiling modality learning strategy we designed significantly improved the performance of MMTF-CPI.

Next, to validate the effectiveness of the heterogeneous network modality learning strategy, we evaluated the performance of MMTF-CPI with different heterogeneous network modality learning strategies. The heterogeneous network learning strategies include three network embedding methods: DeepWalk [52], LINE [53], Node2vec [54], and two knowledge graph embedding methods: DistMult [55], TransE [56]. As displayed in Fig. 8, despite some fluctuations, the model performance with meta-path consistently achieved the best performance in most datasets compared to with other learning strategies. The results demonstrate that our learning strategy effectively captures structural correlations from the heterogeneous network. Moreover, the latent feature learned from the strategy contribute to improving the prediction

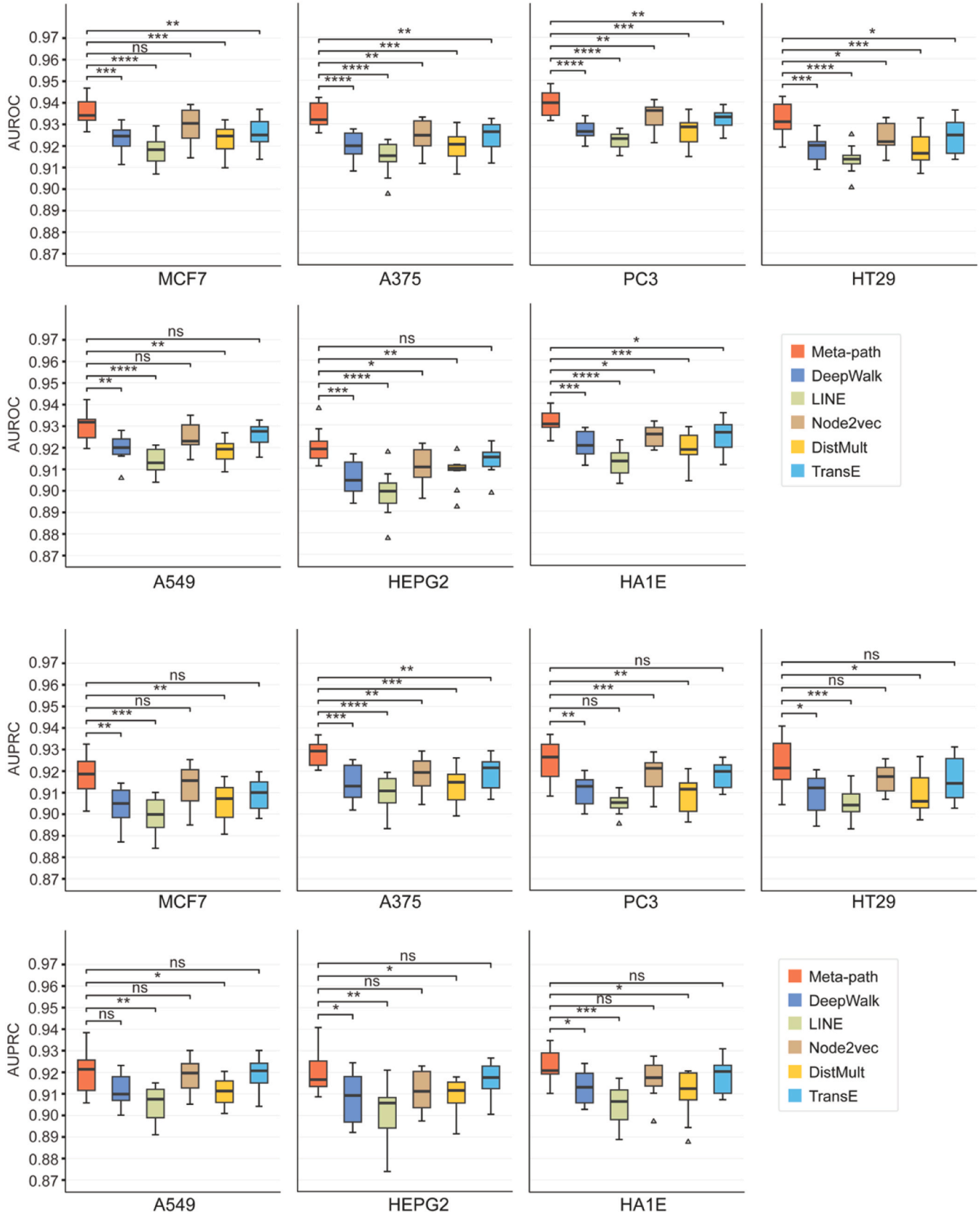

Fig. 8. Comparison performance of MMTF-CPI with different heterogeneous network modality learning strategies.

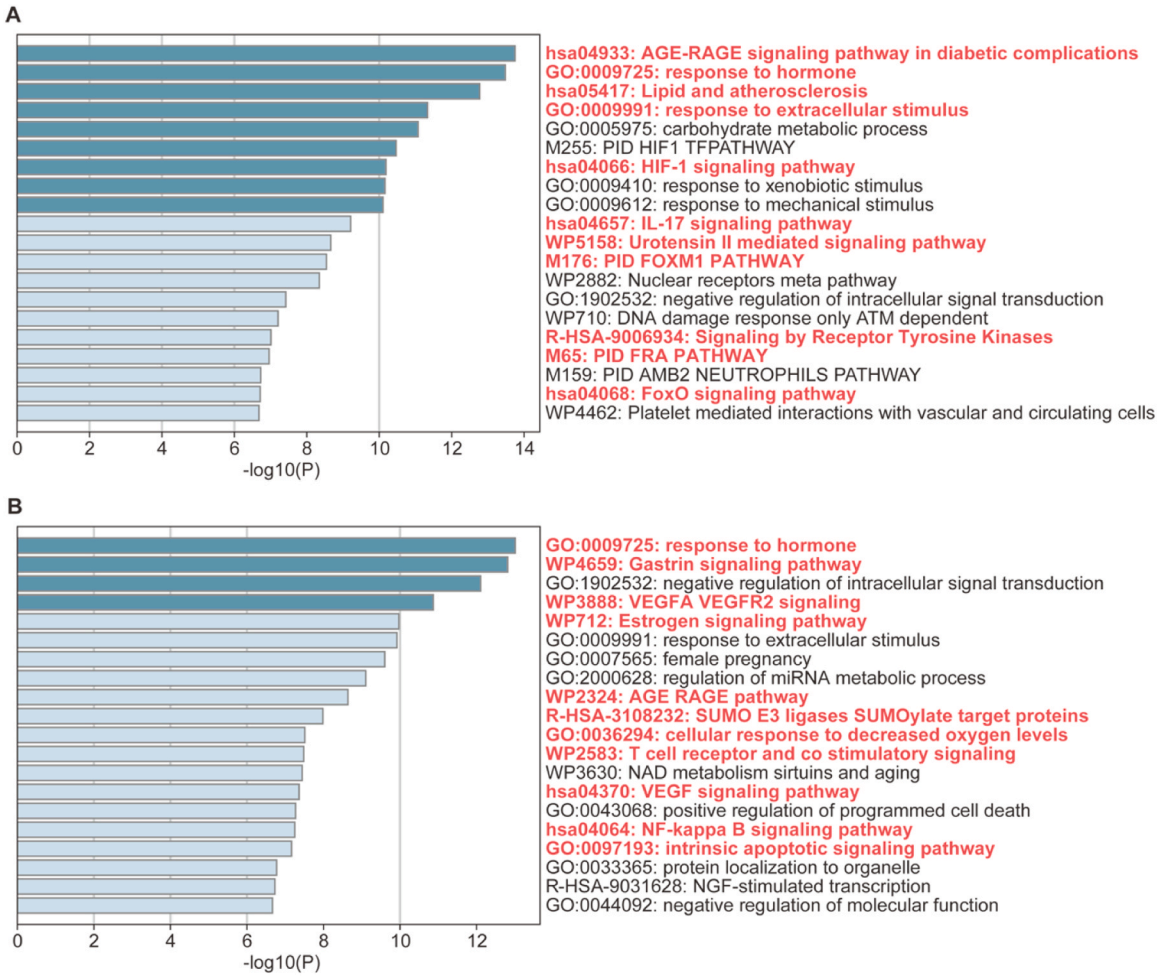

Fig. 9. Ontology enrichment analyses on identified targets for approved drugs. (A) The enrichment analysis on identified targets for breast cancer drugs; (B) The enrichment analysis on identified targets for NSCLC drugs.

performance of MMTF-CPI.

4.6. Case studies

Breast cancer is the commonest cause of cancer-related deaths among women worldwide and the incidence rates are increasing [57]. Lung cancer leads to about 1.6 million deaths worldwide each year [58], with approximately 85 % of lung cancer patients diagnosed as non-small cell lung cancer (NSCLC) [59]. However, the overall survival and cure rates for patients with NSCLC are still low. Thus, there is an urgent need for rapid exploration and discovery of drugs to treat and improve the prognosis of breast cancer and NSCLC. In this study, we mainly focus on identifying targets for approved drugs used to treat breast cancer and NSCLC, as well as discovering experimental drugs with the potential to treat both cancers.

4.6.1. Enrichment analyses for identified targets of approved drugs

To confirm the practical applicability of MMTF-CPI, we applied a pre-trained MMTF-CPI to identify targets for approved drugs used in treating breast cancer and NSCLC, and subsequently conducted ontology enrichment analyses on these targets.

First, we collected approved small molecule drugs for treating breast cancer and NSCLC, which have structure, heterogeneous network, and transcriptional profiling modalities. The collected approved drugs are listed in Supplementary Table S4. Second, we constructed training sets based on MCF7 and A549 datasets to pretrain MMTF-CPI, respectively. The MMTF-CPI pre-trained on MCF7 and A549 datasets were respectively used to identify targets for the approved drugs used in the treatment of breast cancer and NSCLC. To mitigate bias towards popular known targets, we adopted a balanced data sampling strategy for constructing training sets. Specifically, in the training set, each target occurred in equal frequency in both positive and negative compound-protein pairs. We also excluded the compounds with more than five targets from training sets to decrease the impact of polypharmacology. Additionally, to demonstrate the generalization performance of MMTF-CPI, the collected approved drugs were removed from the training set. Third, we predicted the targets for the approved drugs. Finally, we conducted ontology enrichment analyses using Metascape [60] on the top 50 ranked identified targets for each drug (Fig. 9).

As shown in Fig. 9, the targets identified for breast cancer and NSCLC drugs using MMTF-CPI are significantly enriched in breast cancer-related (Fig. 9**A**) and NSCLC-related (Fig. 9**B**) processes. It is noteworthy that the pathways highlighted in red are closely associated with the occurrence, development, and metastasis of breast cancer and NSCLC, as evidenced by previous studies (Supplementary Tables S5 and S6). Our findings suggest that MMTF-CPI has immense potential in identifying drug targets.

4.6.2. Discovery of anticancer drugs

To demonstrate the ability of MMTF-CPI in discovering experimental drugs for cancer therapy, we predicted the interactions between experimental drugs and cancer-related targets. First, we collected experimental small molecule drugs with structure, heterogeneous network, and transcriptional profiling modalities from the DrugBank

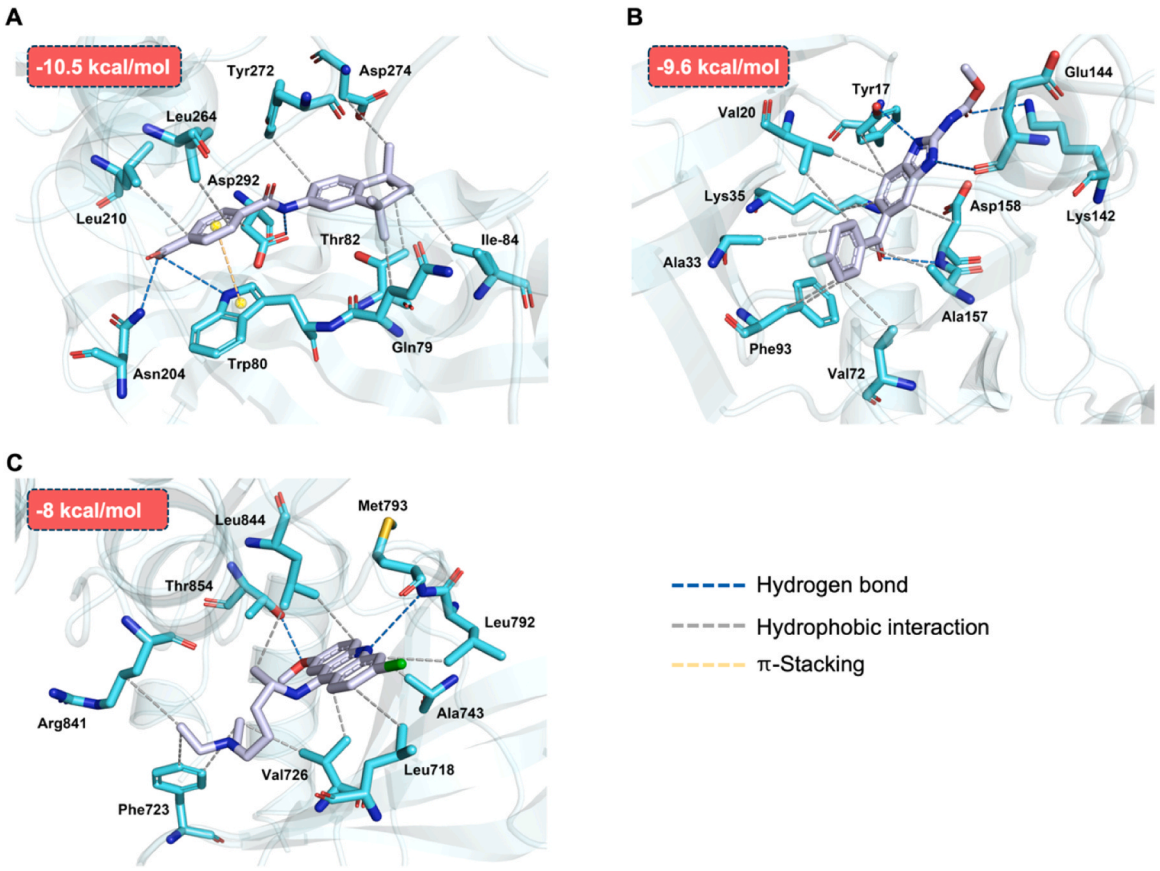

Fig. 10. Visualization of molecular docking. (A) The binding mode of co-crystal structure of AKT1 with Tamibarotene; (B) The binding mode of co-crystal structure of CDK4 with Flubendazole; (C) The binding mode of co-crystal structure of EGFR with Quinacrine.

database [37]. Then, we obtained breast cancer-related and NSCLC-related therapeutic targets with the three modalities from the TTD database [61]. Finally, we utilized the pre-trained MMTF-CPI to discover experimental drugs for treating breast cancer and NSCLC, respectively. Furthermore, to confirm the reliability of the prediction results by MMTF-CPI, we conducted molecular docking with Autodock Vina [62] on the compound-protein pairs with higher predicted probabilities. The results of molecular docking visualization are displayed in Fig. 10.

For breast cancer, the serine threonine kinase Akt1 (AKT1) and cyclin D/cyclin-dependent kinases 4 (CDK4) play a key role in the proliferation of breast cancer cells [63,64]. In our prediction results, we found that Tamibarotene and Flubendazole have higher probability values of 0.921 and 0.933 acting with AKT1 and CDK4, respectively. The binding free energy of Tamibarotene to the co-crystal structure of AKT1 (PDB: 7NH5) [65] is −10.5 kcal/mol. As shown in Fig. 10**A**, Tamibarotene forms hydrogen bonding interactions with residues Trp80, Asn204 and Asp292, as well as a strong π-stacking interaction with residue Trp80. Additionally, it forms hydrophobic interactions with other key residues including Gln79, Thr82, Ile84, Leu210, Leu264, Tyr272 and Asp274. Similarly, there is a lower binding free energy of −9.6 kcal/mol between Flubendazole and the co-crystal structure of CDK4 (PDB: 7SJ3). From Fig. 10**B**, Flubendazole forms strong hydrogen bonding interactions with residues Tyr17, Lys142, Glu144 and Asp158, and hydrophobic interactions with other key residues (Tyr17, Val20, Ala33, Lys35, Val72, Phe93, Ala157, Asp158). Moreover, we found that Tamibarotene and Flubendazole exhibit lower binding free energies when interacting with TOP2A and ERBB2, important therapeutic targets for breast cancer, respectively. The visualization of molecular docking can be found in Supplementary Fig. S2.

For non-small cell lung cancer (NSCLC), epidermal growth factor receptor (EGFR) is one of the crucial therapeutic targets, expressed on the majority of NSCLC cells. Our MMTF-CPI infers that Quinacrine can targeting EGFR. Then molecular docking result shows the binding free energy of −8 kcal/mol between Quinacrine and the co-crystal structure of EGFR (PDB:4I22) [66]. As displayed in Fig. 10**C**, Quinacrine forms strong hydrogen bonds with residues Met793 and Thr854 at the distances of less than 3 Å, and hydrophobic interactions with other key residues, such as Leu718, Phe723, Val726, Ala743, Leu792, Arg841 and Leu844. These results indicate the tremendous practical significance of MMTF-CPI in discovering and developing anti-tumor agents.

5. Conclusion

Altogether, we presented a novel multimodal CPI prediction framework MMTF-CPI that fuses structure, heterogeneous network and transcriptional profiling modalities. MMTF-CPI employs different feature extractors to learn unimodal features from the three modalities, respectively. We designed a multimodal tensor fusion module that can reduce the complexity and collinearity in inter-modality dynamics, simultaneously focusing on both intra-modality and inter-modality dynamics. MMTF-CPI is the first multimodal CPI prediction framework that effectively learns both intra-modality and inter-modality dynamics.

We highlighted several improvements of MMTF-CPI compared with existing state-of-the-art multimodal CPI prediction methods. (i) MMTF-CPI significantly outperforms several state-of-the-art multimodal frameworks in all datasets. (ii) The performance of MMTF-CPI is more significantly improved with the tensor fusion module than other feature fusion methods. (iii) The visualization of unimodal features demonstrates that MMTF-CPI with the tensor fusion module exhibits superior ability to focus on intra-modality dynamics across three modalities compared to concatenation. (iv) We conducted a series of detailed

analyses on the impact of different types of transcriptional signatures from various cell lines on multimodal performance of CPI prediction. In A375 dataset, both the signatures induced by gene knockdown (trt_sh) and gene overexpression (trt_oe) significantly enhance prediction performance. However, in MCF7, PC3, HT29, A549, HEPG2 and HA1E datasets, only trt_sh contributes to improving the multimodal performance of CPI prediction. Hence, the enhancement in multimodal performance varies based on different transcriptional signatures across various cell lines, suggesting that merely incorporating transcriptional signatures does not always improve multimodal performance of CPI prediction. (v) Our experimental results highlight the importance of transcriptional profiling and heterogeneous network modality learning strategies in MMTF-CPI for enhancing model performance. (vi) The case studies indicate that MMTF-CPI not only accurately identifies drug targets but also discovers anticancer drugs. Furthermore, the molecular docking exhibits lower binding free energies between predicted drugs and therapeutic targets, confirming that the drugs predicted by MMTF-CPI indeed possess therapeutic potential for breast cancer and non-small cell lung cancer. Therefore, we believe that the proposed novel multimodal CPI prediction framework, MMTF-CPI, can serve as a powerful tool for accelerating drug discovery and development for various cancers.

Although MMTF-CPI achieves satisfactory performance in enhancing CPI prediction, there are several directions for future improvement in MMTF-CPI. Our current MMTF-CPI framework does not include three-dimensional (3D) structural information, which may result in insufficient learning of structural information. We will integrate more 3D structural information of compounds, proteins and binding pockets into MMTF-CPI to better capture the interactions between compounds and proteins. Another potential direction for future work could be to incorporate the interactions between drugs and side effects into the heterogeneous network, focusing on better extracting the features of heterogeneous network modality.

Code and data availability

The source codes are available at https://github.com/wangmeng-code/MMTF-CPI. Data in this study is availale at https://drive.google.com/drive/folders/1lfVZNHlpgdlBhozK_oeS1upU8NvpBOx6.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 82304250 and 82273734.

CRediT authorship contribution statement

Lei Cao: Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization. **Ruihao Qin:** Writing – review & editing, Data curation. **Kang Li:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Yanyan Dai:** Investigation, Data curation. **Menglei Hua:** Writing – review & editing, Data curation. **Xuan Zhang:** Writing – review & editing. **Yong Cao:** Writing – review & editing, Investigation. **Jia He:** Writing – review & editing, Data curation. **Yongzhen Song:** Formal analysis. **Chenjing Ma:** Writing – review & editing, Formal analysis, Data curation. **Hesong Wang:** Writing – review & editing, Formal analysis. **Jianmin Wang:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Jianxin Ji:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Meng Wang:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no competing interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.10.004.

References

[1] Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci 2019;40:592–604.
[2] Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. Nat Rev Drug Discov 2011;10:428–38.
[3] Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B 2022;12:3049–62.
[4] Rautio J, Meanwell NA, Di L, Hageman MJ. The expanding role of prodrugs in contemporary drug design and development. Nat Rev Drug Discov 2018;17:559–87.
[5] Szardenings K, Li B, Ma L, Wu M. Fishing for targets: novel approaches using small molecule baits. Drug Discov Today: Technol 2004;1:9–15.
[6] Bantscheff M, Drewes G. Chemoproteomic approaches to drug target identification and drug profiling. Bioorg Med Chem 2012;20:1973–8.
[7] Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nat Chem Biol 2009;5:616–24.
[8] Wen M, Zhang Z, Niu S, Sha H, Yang R, et al. Deep-learning-based drug–target interaction prediction. J Proteome Res 2017;16:1401–9.
[9] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28:31–6.
[10] Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics 2018;34:i821–9.
[11] Yang Z, Zhong W, Zhao L, Chen CY-C. ML-DTI: mutual learning mechanism for interpretable drug–target interaction prediction. J Phys Chem Lett 2021;12:4247–61.
[12] Zhao Q, Zhao H, Zheng K, Wang J. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. Bioinformatics 2022;38:655–62.
[13] 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.
[14] Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics 2021;37:1140–7.
[15] Chen W, Chen G, Zhao L, Chen CY-C. Predicting drug–target interactions with deep-embedding learning of graphs and sequences. J Phys Chem A 2021;125:5633–42.
[16] Chen L, Tan X, Wang D, Zhong F, Liu X, 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.
[17] Cheng Z, Zhao Q, Li Y, Wang J. IIFDTI: predicting drug–target interactions through interactive and independent features based on attention mechanism. Bioinformatics 2022;38:4153–61.
[18] Huang L, Lin J, Liu R, Zheng Z, Meng L, et al. CoaDTI: multi-modal co-attention based framework for drug–target interaction annotation. Brief Bioinforma 2022;23:bbac446.
[19] Huang K, Xiao C, Glass LM, Sun J. MolTrans: molecular interaction transformer for drug–target interaction prediction. Bioinformatics 2021;37:830–6.
[20] Ye Q, Hsieh C-Y, Yang Z, Kang Y, Chen J, et al. A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. Nat Commun 2021;12:6775.
[21] Zhou D, Xu Z, Li W, Xie X, Peng S. MultiDTI: drug–target interaction prediction based on multi-modal representation learning to bridge the gap between new chemical entities and known heterogeneous network. Bioinformatics 2021;37:4485–92.
[22] Palhamkhani F, Alipour M, Dehnad A, Abbasi K, Razzaghi P, Ghasemi JB. DeepCompoundNet: enhancing compound-protein interaction prediction with multimodal convolutional neural networks. J Biomol Struct Dyn 2023:1–10.
[23] Dehghan A, Razzaghi P, Abbasi K, Gharaghani S. TripletMultiDTI: multimodal representation learning in drug-target interaction prediction with triplet loss function. Expert Syst Appl 2023;232:120754.
[24] Zhang S, Yang K, Liu Z, Lai X, Yang Z, et al. DrugAI: a multi-view deep learning model for predicting drug–target activating/inhibiting mechanisms. Brief Bioinforma 2023;24:bbac526.
[25] Dong W, Yang Q, Wang J, Xu L, Li X, et al. Multi-modality attribute learning-based method for drug–protein interaction prediction based on deep neural network. Brief Bioinforma 2023;24:bbad161.
[26] Noh H, Shoemaker JE, Gunawan R. Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection. Nucleic Acids Res 2018;46. e34-e34.

[27] Xia X, Zhu C, Zhong F, Liu L. MDTips: a multimodal-data-based drug–target interaction prediction system fusing knowledge, gene expression profile, and structural data. Bioinformatics 2023;39:btad411.
[28] Morency L.-P., Mihalcea R., Doshi P. (2011) Towards multimodal sentiment analysis: Harvesting opinions from the web. p. 169–176.
[29] Pérez-Rosas V., Mihalcea R., Morency L.-P. (2013) Utterance-level multimodal sentiment analysis. p. 973–982.
[30] Snoek C.G.M., Worring M., Smeulders A.W.M. (2005) Early versus late fusion in semantic video analysis. p. 399–402.
[31] Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife 2017;6:e26726.
[32] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 2018;34:i457–66.
[33] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017;171:1437–52.
[34] Wei Z, Zhu S, Chen X, Zhu C, Duan B, Liu Q. DrSim: similarity learning for transcriptional phenotypic drug discovery. Genom, Proteom Bioinforma 2022;20:1028–36.
[35] Jiang L, Qu S, Yu Z, Wang J, Liu X. MOASL: Predicting drug mechanism of actions through similarity learning with transcriptomic signature. Comput Biol Med 2024;169:107853.
[36] Wang J, Chu Y, Mao J, Jeon H-N, Jin H, et al. De novo molecular design with deep molecular generative models for PPI inhibitors. Brief Bioinforma 2022;23:bbac285.
[37] Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008;36:D901–6.
[38] Landrum G. (2006) RDKit: Open-source cheminformatics. 2006. Google Scholar.
[39] Li M, Zhou J, Hu J, Fan W, Zhang Y, et al. Dgl-lifesci: An open-source toolkit for deep learning on graphs in life science. ACS Omega 2021;6:27233–8.
[40] Xiong Z, Wang D, Liu X, Zhong F, Wan X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem 2019;63:8749–60.
[41] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[42] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. PMLR; 2017. p. 1263–72.
[43] UniProt C. UniProt: a hub for protein information. Nucleic Acids Res 2015;43:D204–12.
[44] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 2013;26.
[45] Wang S-H, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. J Med Syst 2018;42:1–11.
[46] Qayyum A, Anwar SM, Awais M, Majid M. Medical image retrieval using deep convolutional neural network. Neurocomputing 2017;266:8–20.
[47] Hou J, Adhikari B, Cheng J. DeepSF: deep convolutional neural network for mapping protein sequences to folds. Bioinformatics 2018;34:1295–303.
[48] Pakhrin SC, Pokharel S, Pratyush P, Chaudhari M, Ismail HD, Kc DB. LMPhosSite: a deep learning-based approach for general protein phosphorylation site prediction using embeddings from the local window sequence and pretrained protein language model. J Proteome Res 2023;22:2548–57.
[49] Arevalo J., Solorio T., Montes-y-Gómez M., González F.A. (2017) Gated multimodal units for information fusion. arXiv preprint arXiv:1702.01992.
[50] Wan F, Hong L, Xiao A, Jiang T, Zeng J. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. Bioinformatics 2019;35:104–11.
[51] Akbani R, Akdemir Kadir C, Aksoy BA, Albert M, Ally A, et al. Genomic classification of cutaneous melanoma. Cell 2015;161:1681–96.
[52] Al-Rfou R., Skiena S., Perozzi B. (2014) Deepwalk: Online learning of social representations.
[53] Tang J., Qu M., Wang M., Zhang M., Yan J., Mei Q. (2015) Line: Large-scale information network embedding. p. 1067–1077.
[54] Grover A., Leskovec J. (2016) node2vec: Scalable feature learning for networks. p. 855–864.
[55] Yang B., Yih W.-t, He X., Gao J., Deng L. (2014) Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575.
[56] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. Adv Neural Inf Process Syst 2013;26.
[57] Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. lancet 1997;349:1269–76.
[58] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
[59] Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 2008;83:584–94.
[60] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
[61] Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. Nucleic Acids Res 2002;30:412–5.
[62] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. J Chem Inf Model 2021;61:3891–8.
[63] Ju X, Katiyar S, Wang C, Liu M, Jiao X, et al. Akt1 governs breast cancer progression in vivo. Proc Natl Acad Sci 2007;104:7438–43.
[64] Pernas S, Tolaney SM, Winer EP, Goel S. CDK4/6 inhibition in breast cancer: current practice and future directions. Ther Adv Med Oncol 2018;10.1758835918786451.
[65] Quambusch L, Depta L, Landel I, Lubeck M, Kirschner T, et al. Cellular model system to dissect the isoform-selectivity of Akt inhibitors. Nat Commun 2021;12:5297.
[66] Gajiwala KS, Feng J, Ferre R, Ryan K, Brodsky O, et al. Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. Structure 2013;21:209–19.