Acta Pharmaceutica Sinica B 2026;16(5):2947−2963

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

DeepICER: A deep learning framework for predicting compound-induced gene expression profiles

Fanbo Meng[a,†], Can Wang[b,c,†], Yue Lin[a], Jing Mo[d], Xunzhi Zhang[a], Zhaotong Cong[b,c], Chi Song[b,c], Sanyin Zhang[b,c], Shilin Chen[b,c], Liang Leng[b,c,*], Wei Chen[a,b,c,*]

[a]*School of Basic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*
[b]*Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*
[c]*Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*
[d]*College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China*

Received 30 July 2025; received in revised form 11 November 2025; accepted 31 January 2026

KEY WORDS

Drug discovery;
Chemical perturbations;
Gene expression profile;
Deep learning;
Bilinear attention mechanism

Abstract Accurate prediction of drug-induced gene expression profiles is crucial for phenotype-based drug discovery. Although computational methods have shown potential, they struggle with the complexities of varying doses and durations. To overcome these limitations, we developed DeepICER, a model that predicts gene expression profiles induced by chemical perturbations across any dose and duration. Utilizing a bilinear attention mechanism, DeepICER captures the interplay between dose, duration, and basal gene expression, enabling accurate predictions for novel compounds and cell lines. DeepICER outperforms existing models with superior flexibility in handling any dose and duration and accuracy, achieving a 45.1% improvement in predictive performance. Experimental validation confirmed that PD-166285, identified by DeepICER, exhibits stronger inhibitory effects on A549 cells compared to paclitaxel. To enhance accessibility, DeepICER is developed as an online platform, providing researchers with a tool to predict gene expression in compound-treated cells, thereby advancing drug repurposing and accelerating drug discovery.

*Corresponding authors.
E-mail addresses: lling@cdutcm.edu.cn (Liang Leng), greatchen@ncst.edu.cn (Wei Chen).
[†]These authors made equal contributions to this work.

https://doi.org/10.1016/j.apsb.2026.01.046

1. Introduction

In the long-term endeavor of drug discovery, target-based strategies have been predominant for a long time[1]. However, they face numerous challenges, such as off-target effects, the inability to deal with diseases lacking a definite target, and high research and development costs along with high failure rates[2]. In recent years, the phenotype-based drug discovery (PBDD) paradigm has emerged as a promising and advanced strategy for drug discovery, which focuses on the effect of a drug candidate on the overall cellular response[3]. Specifically, by monitoring a range of multi-dimensional phenotypic indicators, such as cell viability, changes of cellular morphology and gene expression profiles, PBDD enables the evaluation of the drug's efficacy and potential mechanisms of action. This approach, which does not depend on a pre-defined target[4], broadens the scope for disease therapy exploration, and is especially valuable in the study of diseases without effective therapeutic targets[5].

Gene expression profiles serve as crucial indicators that reflect the functional state of cells and the underlying mechanisms of diseases[6,7], thereby playing an indispensable role in drug discovery processes. The development of high-throughput RNA sequencing technology has enabled the acquisition of large-scale perturbed gene expression profiles. Subsequently, multiple databases, including Connectivity Map (CMap)[8], the Library of Integrated Network-Based Cellular Signatures (LINCS)[9], and Genomics of Drug Sensitivity in Cancer (GDSC)[10], have been established, which provided abundant resources for an in-depth comprehension of the cellular response mechanisms to diverse chemical perturbations. Through the analysis of gene expression profiles, researchers are capable of uncovering the changes of intracellular gene expression in response to drug treatment, thereby facilitating the inference of drug action mechanisms[11-13], screening potential drug targets[14], and predicting drug efficacy[15,16]. However, despite the fact that these data resources provide substantial support for drug discovery[17], the intricate nature of drug-like molecules[18] and the diversity of cell line compositions make it extremely challenging and costly to explore all potential chemical perturbations experimentally. This dilemma has driven researchers to pursue artificial intelligence methods to predict the transcriptional profiles of novel chemicals. The goal is to overcome the limitations of experimental methods and to accelerate the drug discovery process.

The advent of deep learning technologies has presented new prospects for constructing models in drug discovery based on signature gene expression profiles (SGEPs). In recent years, numerous studies have been dedicated to the development of computational models for the prediction of gene expression profiles of cells in response to compound perturbations. For instance, DLEPS[19] is a deep neural network that predicts the gene expression of cells in response to novel chemicals but without cell type specificity. DeepCE[20] and CIGER[21] employed one-hot coding to discriminate among cell types, learning from a diverse range of perturbation profiles. MultiDCP[22] incorporated cellular context to predict gene expression that depend on the context, allowing for predictions of responses in unseen cell lines. TranSiGen[23] was a variational autoencoder-based framework that employs self-supervised representation learning to reduce noise and predicts transcriptional profiles, thereby facilitating the prediction of novel perturbation responses. PRnet[24] was developed as a generative model based on an encoder-decoder structure, aimed at predicting transcriptional responses to chemical perturbations at both bulk and single-cell resolutions. Although existing models are effective in generating transcriptional profiles of chemical perturbations, they have notable limitations, such as not considering compound dose or perturbation duration, and lacking adaptability across different cell types. As a result, their predictive performance could be further improved. Therefore, more advanced models are needed to overcome the limitations of current experimental and computational approaches in exploring novel perturbations. These models are essential for accurately prediction of unknown perturbations and for identification of potential drug candidates.

In this study, we introduce a novel deep learning model, DeepICER, aimed at addressing the limitations of existing methods in chemical perturbation analysis. DeepICER integrates multiple key factors, including compound structure, dose, perturbation duration, and cell line type. It also utilizes a bilinear attention mechanism to prioritize compound structure and identify critical features within gene expression profiles that are closely linked to cellular response, thereby more effectively capturing the complex interactions between compounds and gene expression changes.

Compared with other models, DeepICER demonstrates superior performance in predicting gene expression profiles of cells in response to compound perturbations. Additionally, an interpretable analysis of DeepICER is performed, offering insights into its underlying principles. As a proof-of-principle, we also validate the effectiveness of DeepICER in lung adenocarcinoma (LUAD) drug discovery through biochemical experiments. Furthermore, an online service platform for DeepICER is developed to facilitate its academic use.

2. Materials and methods

2.1. Data processing

The L1000 project generated gene expression signatures for small molecule compounds perturbing various cell lines, including data on gene expression profiles under different doses and durations of treatment. We obtained the signatures from the latest CMAP LINCS 2020 release (https://clue.io/releases/data-dashboard), which includes gene expression data for perturbations caused by 21,870 small molecule compounds. In this study, the level 5 data was used to train the model. The cell line expression data were obtained from the Cancer Cell Line Encyclopedia (CCLE)[25], from which a total of 1479 cell lines were included. We filtered the expression values of 978 landmark genes in each cell line, which were expressed using TPM (transcripts per million) and

$\log_2$-transformed. For model training, we selected cell lines from CCLE that were also covered in the LINCS 2020 dataset, incorporating all expression profiles from compounds that perturbed these cell lines.

2.2. Data splitting

L1000 experiments were typically conducted with three or more biological replicates. To ensure the quality of the data, we split the expression profiles based on the 75th quantile of pairwise Spearman correlations (cc_q75) between replicated signatures in landmark gene space. Experiments with cc_q75 scores below a certain threshold were considered unreliable and excluded from the analysis. In our implementation, this threshold was set to 0.5. High cc_q75 scores indicate better consistency between replicates of the same treatment, ensuring that the model was trained using high-quality data. The resulting high-quality dataset comprised 48,593 signatures, covering 4083 compounds, 83 cell lines, 1293 doses, and 7 durations. We grouped the data by cell line and randomly divided each group into training, validation, and test sets in an 8:1:1 ratio. Additionally, we created training sets using different thresholds for cc_q75 and evaluated their impact on the model's performance to emphasize the importance of data quality in obtaining reliable model results.

2.3. Algorithmic principle of DeepICER

2.3.1. GCN for molecular graph

We converted the isomeric SMILES representation of each compound into a corresponding 2D molecular graph. To encode information about the nodes in the molecule, we initialized each atom node based on the chemical features in the DGL-LifeSci library[26]. Each atom was represented as a 74-dimensional integer vector, containing eight different attributes: atom type, atomic degree, number of implied hydrogen atoms, formal charge, number of electrons in the ground state, hybridization state, total number of hydrogen atoms, and whether the atom is aromatic or not. Since the same compound can be tested under different doses and durations, we jointly coded the compound with the dose and duration in the experiment. The 75th dimension of each atom's feature vector represented the dose, while the 76th dimension represented the duration. In addition, we set the maximum number of molecular nodes to 200, meaning that compounds with fewer nodes were padded with virtual nodes, filled with zeros. Consequently, the node feature matrix for each molecular graph was denoted as $\boldsymbol{M}_{\mathrm{d}} \in \mathbb{R}^{200\times 77}$. To further process these features for downstream tasks, we applied a linear transformation to define $\boldsymbol{X}_{\mathrm{d}} = \boldsymbol{W}_0\boldsymbol{M}_{\mathrm{d}}^T$, which leads to the real-valued dense matrix $\boldsymbol{X}_{\mathrm{d}} \in \mathbb{R}^{200\times D_{\mathrm{d}}}$, as the compound input features, where D_{d} is the latent space dimension.

A three-layer GCN module was used to learn the graph representation of drug compounds. Specifically, we updated atomic feature vectors by aggregating information from neighboring atoms connected by chemical bonds. This propagation method essentially captured the substructural details of the molecule. The node-level representations of the compounds were then preserved for the subsequent explicit learning of local interactions with cell lineage features. The compound encoder is written as:

$$\boldsymbol{H}_{\mathrm{d}}^{(l+1)} = \sigma\left(GCN\left(\tilde{\boldsymbol{A}}, \boldsymbol{W}_{\mathrm{d}}^{(l)}, \boldsymbol{b}_{\mathrm{d}}^{(l)}, \boldsymbol{H}_{\mathrm{d}}^{(l)}\right)\right) \tag{1}$$

where $\sigma(\cdot)$ denotes the non-linear activation function, $\boldsymbol{W}_{\mathrm{d}}^{(l)}$ and $\boldsymbol{b}_{\mathrm{d}}^{(l)}$ are the layer-specific learnable weight matrices and biases of the GCN, $\tilde{\boldsymbol{A}}$ is the adjacency matrix with self-connections added to the molecular graph, and $\boldsymbol{H}_{\mathrm{d}}^{(l)}$ is the representation of the l-th hidden node, where $\boldsymbol{H}_{\mathrm{d}}^{(0)} = \boldsymbol{X}_{\mathrm{d}}$.

2.3.2. Convolution neural networks (CNN) for cell line representation

The cell lines feature encoder consists of three consecutive one-dimensional convolutional layers that transform the input cell lines into a matrix representation in the latent feature space. The order of the 978 landmark genes in the cell line was first fixed, and all GO terms associated with at least three or more landmark genes were extracted using the OntologyX R package[27]. Each gene was described according to the GO terminology library and encoded as a one-hot vector consisting of 1107 features. The expression value of each gene was then multiplied by its corresponding GO feature vector to initialize the feature matrix $\boldsymbol{M}_{\mathrm{d}} \in \mathbb{R}^{978\times 1107}$ for the cell line without compound treatment. A linear transformation was used to define $\boldsymbol{X}_{\mathrm{c}} = \boldsymbol{W}_0\boldsymbol{M}_{\mathrm{c}}^T$, resulting in a real-valued dense matrix $\boldsymbol{X}_{\mathrm{c}} \in \mathbb{R}^{978\times D_{\mathrm{c}}}$ as the input features for the cell line, where D_{c} is the latent space dimension.

A three-layer CNN module was used to learn the cell lines representation from the feature matrix. The first convolutional layer was used to capture local features with a kernel size of 3, while the subsequent two layers progressively expand the receptive field to learn more abstract features. The cell lines encoder is described as follows:

$$\boldsymbol{H}_{\mathrm{c}}^{(l+1)} = \sigma\left(CCN\left(\boldsymbol{W}_{\mathrm{c}}^{(l)}, \boldsymbol{b}_{\mathrm{c}}^{(l)}, \boldsymbol{H}_{\mathrm{c}}^{(l)}\right)\right) \tag{2}$$

where $\boldsymbol{W}_{\mathrm{c}}^{(l)}$ and $\boldsymbol{b}_{\mathrm{c}}^{(l)}$ are the learnable weight matrices (filters) and bias vectors of the l-th CNN layer. $\boldsymbol{H}_{\mathrm{c}}^{(l)}$ is the l-th hidden representation and $\boldsymbol{H}_{\mathrm{c}}^{(0)} = \boldsymbol{X}_{\mathrm{c}}$.

2.3.3. Pairwise interaction learning

The Bilinear Attention Network (BAN) uses a bilinear attention map to learn interaction representations by considering pairs of input channels[28]. BAN can provide rich joint information while maintaining computational efficiency and preserving the same size. For the feature matrices characterizing compounds and cells, we employed a BAN-based pairwise interaction module to learn their local interactions in pairs. This module consists of two layers: (1) a bilinear interaction map capturing pairwise attentional weights, and (2) a bilinear pooling layer extracting the joint representation of compound-cell interactions from the interaction map.

Following the separate GCN and CNN encoders, compounds and cell lines are represented as $\boldsymbol{H}_{\mathrm{d}}^{(3)} = \left\{\boldsymbol{h}_{\mathrm{d}}^1, \boldsymbol{h}_{\mathrm{d}}^1, \cdots, \boldsymbol{h}_{\mathrm{d}}^{200}\right\}$ and $\boldsymbol{H}_{\mathrm{c}}^{(3)} = \left\{\boldsymbol{h}_{\mathrm{c}}^1, \boldsymbol{h}_{\mathrm{c}}^1, \cdots, \boldsymbol{h}_{\mathrm{c}}^M\right\}$, respectively, where $\boldsymbol{M}$ represents the dimension encoded by the cell line. The bilinear interaction map produces the single-head pairwise interaction matrix $\boldsymbol{I} \in \mathbb{R}^{200\times M}$, capturing interactions between compound and cell line features

$$\boldsymbol{I} = \left(\left(\mathbf{1}\cdot\boldsymbol{q}^T\right) \circ \sigma\left(\left(\boldsymbol{H}_d^{(3)}\right)^T\boldsymbol{U}\right)\right)\cdot\sigma\left(\boldsymbol{V}^T\boldsymbol{H}_c^{(3)}\right) \tag{3}$$

where $\mathbf{1} \in \mathbb{R}^N$ is a fixed all-ones vector, $\boldsymbol{q} \in \mathbb{R}^K$ is a learnable weight vector, $\boldsymbol{U} \in \mathbb{R}^{D_{\mathrm{d}}\times K}$ and $\boldsymbol{V} \in \mathbb{R}^{D_{\mathrm{c}}\times K}$ are learnable weight matrices for the compound and cell line representations, and $\circ$ denotes the Hadamard product. The elements in $\boldsymbol{I}$ denote the

strength of the interaction of the compound-cell line pairs, which are subsequently. To provide an intuitive understanding of bilinear interactions, the element $\boldsymbol{I}_{i,j}$ in Equation (3) can also be written as:

$$\boldsymbol{I}_{i,j} = \boldsymbol{q}^T\left(\sigma\left(\boldsymbol{U}^T\boldsymbol{h}_{\mathrm{d}}^i\right) \circ \sigma\left(\boldsymbol{V}^T\boldsymbol{h}_{\mathrm{c}}^j\right)\right) \tag{4}$$

where $\boldsymbol{h}_{\mathrm{d}}^i$ is the i-th column of $\boldsymbol{H}_{\mathrm{d}}^{(3)}$ and $\boldsymbol{h}_{\mathrm{c}}^j$ is the j-th column of $\boldsymbol{H}_{\mathrm{c}}^{(3)}$, denoting the i-th substructure of a compound and the j-th dimensional representation of a cell line, respectively. Thus, the bilinear interaction can be viewed as first mapping $\boldsymbol{h}_{\mathrm{d}}^i$ and $\boldsymbol{h}_{\mathrm{c}}^j$ to a shared feature space with weight matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, followed by learning the interaction through the Hadamard product and the weight vector $\boldsymbol{q}$. Finally, this process captures pairwise interactions between potential representations of compounds and cell lines.

We then obtained the joint compound-cell lineage representation $f' \in \mathbb{R}^K$ by introducing a bilinear pooling layer on the interaction map $\boldsymbol{I}$. The formula for the k-th element of f' is:

$$\begin{aligned} \boldsymbol{f}_k' &= \sigma\left(\left(\boldsymbol{H}_{\mathrm{d}}^{(3)}\right)^T\boldsymbol{U}\right)_k^T \cdot \boldsymbol{I} \cdot \sigma\left(\left(\boldsymbol{H}_{\mathrm{c}}^{(3)}\right)^T\boldsymbol{V}\right)_k \\ &= \sum_{i=1}^{200}\sum_{j=1}^{M}\boldsymbol{I}_{i,j}\left(\boldsymbol{h}_{\mathrm{d}}^i\right)^T\left(\boldsymbol{U}_k\boldsymbol{V}_k^T\right)\boldsymbol{h}_{\mathrm{c}}^j \end{aligned} \tag{5}$$

where $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are the k-th column of weight matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively. The weight matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are shared with the previous interaction layers to reduce the number of parameters and alleviate overfitting. In addition, a sum pooling operation is applied to the joint representation vector to obtain a compact feature map:

$$\boldsymbol{f} = SumPool(\boldsymbol{f}', s) \tag{6}$$

where $SumPool(\cdot)$ is a 1D and non-overlapping pooling operation with stride s. To extend the single pairwise interactions to a multi-head form, the fusion characterization between a compound and a cell line is computed by summing the individual heads.

Finally, the feature $\boldsymbol{f}$ generated by the BAN was used as input to predict SGEPs of landmark genes through a five-layer dense network. Each layer of the dense network incorporates a dropout layer with a dropout rate of 0.25. The activation function for the first three layers is ReLU, while the fourth layer employs the tanh function, and the fifth layer uses a linear activation.

The model is trained using the mean square error (MSE) as the objective function, which measures the difference between the predicted and actual SGEPs, calculated as follows:

$$\boldsymbol{loss}_{DeepICER}(\Theta) = \frac{1}{N \times 978}\sum_{i=1}^{N}\sum_{j=1}^{978}\left(z_{i,j} - y_{i,j}\right)^2 \tag{7}$$

where $\boldsymbol{\Theta}$ is the set of parameters in the model; N denotes the total number of gene expression profiles in the dataset; $z_{i,j}$ and $y_{i,j}$ are the true and predicted SGEPs of the j-th landmark gene in the i-th gene expression profile, respectively.

2.4. Performance evaluation metrics

To evaluate model performance, we primarily used the Pearson correlation coefficient, which provides an unbiased measure for models optimized using the mean squared error (MSE). The average Pearson correlation for the dataset was computed as follows:

$$r = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{M978}\left(z_{i,j} - \overline{z_i}\right)\left(y_{i,j} - \overline{y_i}\right)}{\sqrt{\sum_{j=1}^{978}\left(z_{i,j} - \overline{z_i}\right)^2}\sqrt{\sum_{j=1}^{978}\left(y_{i,j} - \overline{y_i}\right)^2}} \tag{8}$$

where $z_{i,j}$, and $y_{i,j}$, are the true and predicted signature of the j-th gene in the i-th gene expression profile, $\overline{z_i}$, and $\overline{y_i}$ are the averages of the true and predicted signatures of the i-th gene expression profile.

Spearman correlation was also used to assess how well the models ranked gene expressions. This metric is particularly useful for evaluating rankings rather than absolute values. In addition, we used Precision@k to evaluate the model's ability to prioritize the most significantly up- and down-regulated genes. Specifically, we assessed Positive Precision@100 for the top 100 up-regulated genes and Negative Precision@100 for the top 100 down-regulated genes.

2.5. Phenotype-based drug repurposing for LUAD

2.5.1. Gene signature of approved drugs

We obtained a list of approved LUAD therapeutics from https://www.cancer.gov/about-cancer/treatment/drugs/lung#1. Among them, the LINCS 2020 dataset has 28 gene signatures from two LUAD cell lines treated with eight different drugs. These profiles were used for screening anti-LUAD compounds.

2.5.2. Differential gene expression profile of disease

RNA-seq expression data for LUAD tumor and normal samples were downloaded from The Cancer Genome Atlas (TCGA)[29] using the TCGABiolinks package (v2.18.4)[30]. Differential gene expression (DEGs) was identified using the DESeq2[31] package. DEGs were filtered based on the criteria: $|\log_2\text{FoldChange}| > 1$, P-value <0.05, and false discovery rate (FDR) < 0.25. This yielded 2880 up-regulated genes and 9859 down-regulated genes.

2.5.3. Inferring perturbation gene signatures of compounds

We used the BROAD repurposing dataset[32], which comprises 6394 compounds, for the screening of anti-LUAD drugs. This dataset was obtained from https://repo-hub.broadinstitute.org/repurposing#download-data. Compounds with true expression profiles in the LINCS 2020 dataset were removed, resulting in a dataset containing 6389 compounds. DeepICER was then employed to infer gene signatures for 978 landmark genes related to these compounds in A549 cells.

2.6. Connectivity score

We measured the relationship between gene signatures by the connectivity score obtained from gene set enrichment analysis[33]. A query (q) consists of genes corresponding to a biological state of interest, with each gene flagged as either up- or down-regulated. This produces two mutually exclusive gene lists (q_{up}, q_{down}) for each query. The connectivity score assesses the degree to which up-regulated query genes (q_{up}) appear at the top and down-regulated query genes (q_{down}) appear at the bottom of the reference signature (positive connectivity), or vice versa (negative connectivity).

First, the enrichment score (ES) evaluates how well the query gene list is enriched at the top or bottom of the reference

signature. The enrichment scores for the top (a) and bottom (b) sets of genes are calculated as follows:

$$a = \max_{j=1\ to\ t}\left[\frac{j}{t} - \frac{V(j)}{n}\right] \tag{9}$$

$$b = \max_{j=1\ to\ t}\left[\frac{V(j)}{n} - \frac{(j-1)}{t}\right] \tag{10}$$

$$ES = \begin{cases} a, if\ a > b \\ -b, if\ b > a \end{cases} \tag{11}$$

where n denotes the number of genes in the reference signature gene list, t denotes the number of genes in the query gene list, $V(j)$ denotes the rank of a specific gene in the rank list, and j denotes the index of the gene ranging from 1 to t.

Next, the enrichment scores of up-regulated and down-regulated genes in the query gene list against the reference signature gene list were computed using the above formulas to obtain ES_{up} and ES_{down}, respectively. Finally, the connectivity score of the query q relative to the reference signature gene list was computed by combining these two enrichment scores for the given pairs of query gene list (q_{up}, q_{down}) and reference signature gene list r:

$$S_{q,r} = \begin{cases} \left(ES_{\text{up}} - ES_{\text{down}}\right)/2, if\ sgn\left(ES_{\text{up}}\right) \neq sgn(ES_{\text{down}}) \\ 0,\ otherwise \end{cases} \tag{12}$$

where ES_{up} is the enrichment of q_{up} in r and ES_{down} is the enrichment of q_{down} in r. $S_{q,r}$ is between -1 and 1. It will be positive for positively related signatures, negative for negatively related signatures, and close to zero for unrelated signatures. When the signs of ES_{up} and ES_{down} are the same, a zero (0) score is assigned.

We defined the Drug Connectivity Score as the connectivity score between the SGEPs of candidate compounds and the gene signatures of known drugs. Similarly, the DEG Connectivity Score measures the connectivity score between the SGEPs of candidate compounds and the DEG profiles of the disease state.

2.7. Experimental setting of validating screened compounds for LUAD

2.7.1. Compound

For screening potential LUAD drugs, the reference gene signature was obtained by DeepICER prediction for all compounds. Gene signatures of approved drugs for LUAD were used to identify compounds exhibiting positive connectivity scores, while DEGs for LUAD were used to identify compounds with negative connectivity scores. We selected five drugs for validation: the compound with the largest positive connectivity score to the gene signature of the approved drugs for LUAD, the compound with the largest negative connectivity score to the DEGs for LUAD, and the three compounds with the highest combined positive and negative connectivity scores. In addition, paclitaxel was used as a positive control, and compounds with maximum negative connectivity scores to the gene signature of approved drugs for LUAD were used as a negative control. All seven compounds were purchased from MedChemExpress and their details were given in Supporting Information Table S1.

2.7.2. Cell culture

The human LUAD cell line A549 were purchased from the Institute of Biochemistry and Cell Biology of the Chinese Academy of Sciences (Shanghai, China). The cells were cultured in RPMI 1640 medium (GIBCO) supplemented with 10% fetal bovine serum, 100 U/mL penicillin and 100 U/mL streptomycin, and incubated at 37 °C in a humidified atmosphere with 5% CO_2.

2.7.3. Cell viability assays

We evaluated cell viability using the CCK8 assay. A549 cells were seeded at a density of 5000 per well in a 96-well plate. Subsequent to the cells adhering to the well surface, compounds were added to each well with eight concentration gradients (0, 0.25, 0.5, 1, 2.5, 5, 10, and 25 μmol/L followed by incubation for 24 h. Subsequently, 10 μL of CCK8 solution was added to each well, thoroughly shaken, and then placed in the incubator. Following a 2-h incubation period, the 96-well plate was placed in a preheated microplate reader. The absorbance values (optical density, OD) were measured at 450 nm in each well.

2.8. Atlas of perturbational signatures

A total of 3 compound libraries, namely FDA-approved compounds library (n = 959), BROAD Repurposing compounds library (n = 6384), and natural compounds library (n = 30,440)[34] were used to generate perturbational signatures in this study. Data cleaning was conducted based on the following criteria: the molecular graphs of compounds were required to have fewer than 200 nodes, and the molecule SMILES had to be successfully parsed using RDKit (version 2024.3.2). Ultimately, DeepICER was employed to predict perturbation signatures for each compound across 80 cell lines at two durations (24 and 48 h) and three doses (1, 10, and 20 μmol/L).

2.9. Statistics and reproducibility

A two-sided t-test was used for statistical analysis of model performance, with detailed descriptions of specific tests provided in the figure legends. The significance levels were set as follows: ****$P < 0.0001$; ***$0.0001 < P \leq 0.001$; **$0.001 < P \leq 0.01$; *$0.01 < P \leq 0.05$ and ns, $0.05 < P \leq 1.0$.

3. Results

3.1. Architecture of DeepICER

The cellular response to chemical perturbations is influenced by a variety of factors, including compound structure, dose, duration of perturbation, and cell line type. To account for these factors comprehensively, we developed the DeepICER model to predict the SGEPs of chemical perturbation. The architecture of DeepICER consists of four components, including a compound coding module, a cell line coding module, a BAN module, and a prediction module (Fig. 1A). Given a compound, DeepICER generates its molecular graph based on the Simplified Molecular Input Line Entry System (SMILES) strings by using DGL-LifeSci[26] and initializes the features of each atomic node (**see 2.3.1**). The features of these atomic nodes were augmented with the dose and duration of compound to generate the initial graph structure representation. Subsequently, we employed Graph Convolutional Networks (GCN) to learn the numerical representations of the compounds as well as the perturbation dose and duration from the graph structure. For cell line representations, we generated gene descriptors for each landmark gene based on their Gene Ontology (GO) terms. These descriptors were scaled by their

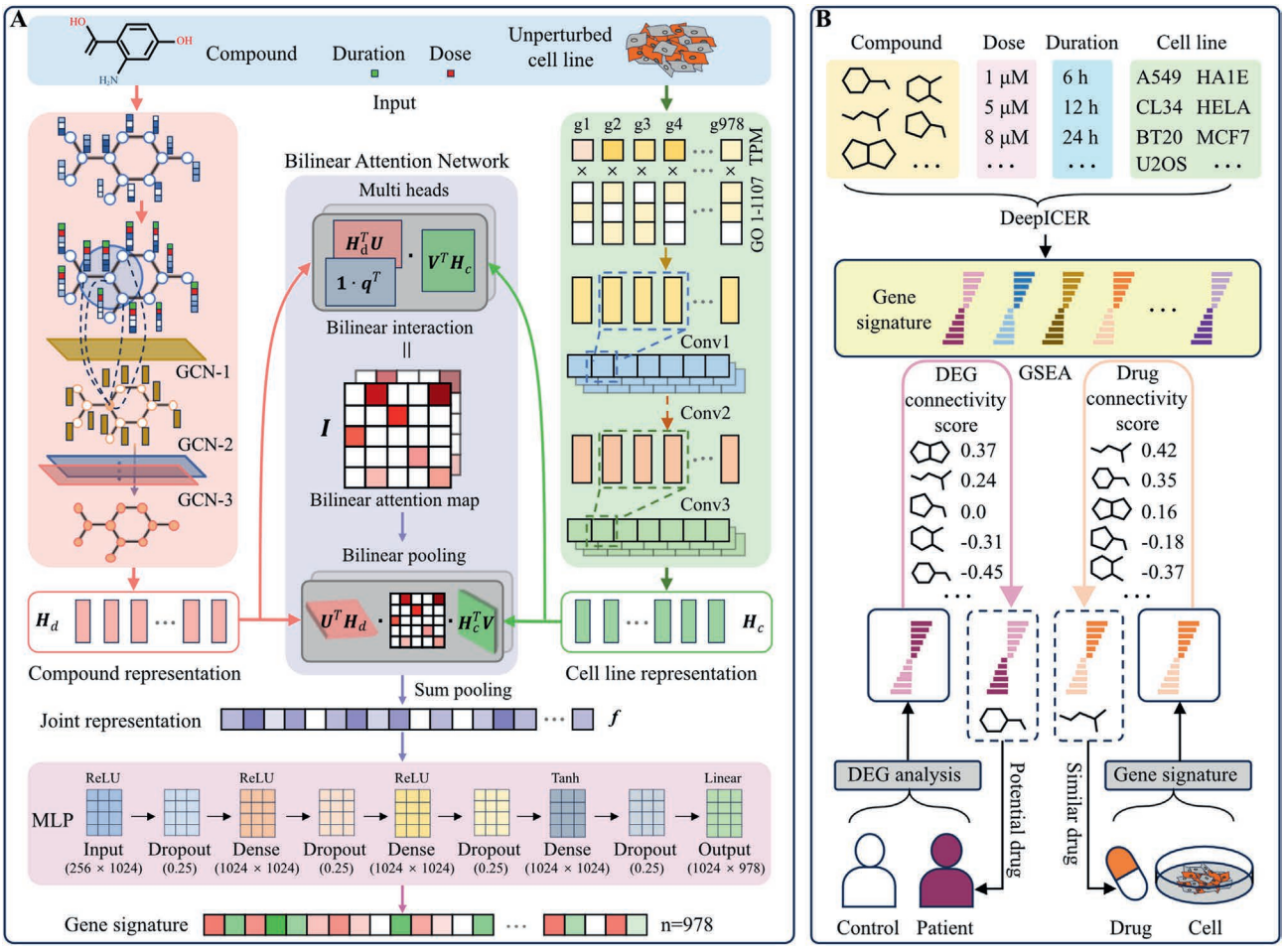


Figure 1 Schematic diagram of DeepICER's architecture and application workflow. (A) DeepICER consists of four core components, with their functions and data flows described as follows. Converts isomeric SMILES into 2D molecular graphs, embedding atomic features along with dose and duration information into node vectors. A three-layer GCN extracts a joint representation that integrates molecular structure with treatment conditions. Cell line feature encoder: Integrates basal gene expression and GO annotations *via* element-wise multiplication to generate a fusion matrix. A three-layer 1D CNN captures cell line-specific features and biological context. Computes attention-weighted interactions between compound and cell line features to form a joint representation vector $\boldsymbol{f}$. Prediction network: A multi-layer perceptron decodes $\boldsymbol{f}$ to predict the post-treatment SGEP. (B) Screening of candidate compounds using DeepICER: DeepICER generates SGEPs of a specific cell line stimulated by compounds under different dose and duration conditions. Given the gene signature of a specific disease or drug, gene set enrichment analysis is utilized to calculate the connectivity score of the query compounds to assess whether they can reverse the disease gene signature or mimic the effect of a known drug.

expression values to form the initial representation of the cell lines and were used as the inputs of a 1D convolutional neural network (1D-CNN). Next, a BAN module was employed to capture local interactions between the encoded compounds and the cell line representation. The BAN operates in two phases: bilinear attention and bilinear pooling, which collectively produce the joint representation. Finally, the prediction module takes the output from the interaction module to predict the SGEPs corresponding to the compound, dose, duration, and cell type-specific contexts. In the training process, we adopted a learning rate of 5e-5, set the number of training epochs to 1000, and used a batch size of 128. Additionally, a dropout rate of 0.2 was applied in the BAN module.

Based on the CMap concept, which assumes that gene signatures serve as indicators of the underlying mechanisms of disease, DeepICER predicts therapeutic candidates by identifying compounds that reverse disease signatures (Fig. 1B). The application of DeepICER in an *in silico* screening task consists of two steps. First, DeepICER predicts the gene signatures of a specific cell line perturbed by a compound library at user-defined concentrations and durations. Next, given a query gene profile for a specific disease or a known sensitive compound, gene set enrichment analysis (GSEA) is used to calculate enrichment scores, which are then converted into connectivity scores (**see 2.6**). Finally, compounds are ranked based on these connectivity scores to identify potential drug candidates.

3.2. Performance evaluation of DeepICER for predicting SGEPs

Both Pearson correlation and Spearman correlation are used to evaluate the model's performance. The Pearson correlation quantifies the direct numerical fit between the predicted and actual gene expression values, providing an accurate measure of

prediction accuracy. In contrast, the Spearman correlation assesses the consistency in gene rankings, capturing how well the predicted SGEPs align with the actual rankings of gene expression, thereby offering a broader evaluation of model performance. DeepICER demonstrated impressive performance on the test dataset, achieving an average Pearson correlation coefficient of 0.729, which reflects the model's high accuracy in numerical predictions. Moreover, the average Spearman correlation coefficient reached 0.703, highlighting the model's robust ability to maintain consistency in the ranking of gene expression values. Additionally, we assessed the precision of gene rankings using top-K precision, which is crucial for downstream tasks such as calculating connectivity scores. DeepICER achieved prediction precisions of 0.530 for the top 100 up-regulated genes and 0.583 for the top 100 down-regulated genes, further confirming its strong performance in identifying key gene perturbations.

3.3. Comparison with existing models in inferring SGEPs

We also compared the performance of DeepICER for predicting SGEPs against existing models, namely DeepCE, MultiDCP, and TranSiGen, measured by Pearson correlation, Spearman correlation, Positive Precision@100, and Negative Precision@100. DeepCE was only able to predict expression profiles for seven cell lines, perturbed by compounds at six doses and a duration of 24 h. MultiDCP, while capable of predicting expression profiles for novel cell lines, was similarly restricted to six doses and a 24-h treatment duration. TranSiGen, on the other hand, could only predict the expression profiles for 164 cell lines, perturbed by compounds at a concentration of 10 μmol/L and a duration of 24 h. To ensure a fair comparison, we created independent test sets for each model, selecting perturbations that matched each model's constraints (Supporting Information Table S2).

DeepICER outperformed all three models across all metrics (Fig. 2A). Compared to DeepCE, DeepICER achieved a 94.9% improvement in Pearson correlation and an 87.6% improvement in Spearman correlation on the test set. Additionally, DeepICER demonstrated significant gains in top-K precision, with a 128.6% higher precision for predicting the top 100 up-regulated genes and an 82.8% higher precision for predicting the top 100 down-regulated genes. In comparison with MultiDCP, DeepICER achieved a 66.5% improvement in Pearson correlation, a 59.5% improvement in Spearman correlation, and respective improvements of 77.23% and 65.4% in Positive and Negative Precision@100. While TranSiGen is a state-of-the-art model for predicting chemical-induced differential gene expression, DeepICER outperformed it in all four metrics. The Pearson correlation coefficient for DeepICER was 45.1% higher than TranSiGen (0.711 *vs* 0.490), demonstrating superior agreement with the true chemical transcriptomics data. Similarly, in terms of Spearman correlation, DeepICER outperformed TranSiGen by 44%, with a value of 0.684 compared to TranSiGen's 0.475. In terms of top-K precision, DeepICER also showed a significant improvement. The precision for predicting the top 100 up-regulated genes and top 100 down-regulated genes within the test dataset was improved by 67.9% and 47.0%, respectively, compared to TranSiGen. Additionally, DeepICER achieved the lowest mean squared error (MSE) among all models, reducing MSE by 50.48%, 47.63%, and 60.82% relative to DeepCE, MultiDCP, and TranSiGen, respectively.

These results underscore DeepICER's superior accuracy in both prioritizing key gene perturbations and minimizing overall prediction error. Compared to existing methods, DeepICER achieves stronger performance, which can be attributed to several innovations in its architectural design. Unlike traditional approaches that process molecular structures and experimental conditions independently, DeepICER embeds intrinsic atomic features along with dose and duration information directly into node vectors when constructing 2D molecular graphs from isomeric SMILES. A three-layer GCN is then used to jointly learn structural and contextual representations. This design allows the model to capture nonlinear interactions between molecular structure and dynamic experimental settings, enabling robust predictions across arbitrary dose-time combinations. For cell line representation, DeepICER moves beyond reliance on raw expression data alone. It constructs a fusion matrix by element-wise multiplication of gene expression values and their corresponding GO functional annotations. A three-layer CNN is applied to this matrix to extract biologically informed, cell line-specific representations. This approach preserves the quantitative characteristics of gene expression while integrating functional biological meaning, substantially improving the model's capacity to capture cellular heterogeneity. At the core of the model, the BAN module serves as a fusion mechanism, learning dynamic global-local interaction patterns between compound and cell line representations *via* an attention weight matrix. This module not only enables the modeling of complex cross-modal associations but also provides model interpretability through attention heatmap visualization.

3.4. DeepICER can predict SGEPs for unseen compounds and cell lines

Furthermore, we evaluated the performance of DeepICER on unseen data under four challenging experimental setups, *i.e.*, chemical-blind splitting, cell-blind splitting, dose-blind splitting, and duration-blind splitting. Due to the heterogeneity of the LINCS data, some compounds yielded many SGEPs, while others had only a few SGEPs; a similar situation was observed for cell lines and doses. To ensure sufficient data for model training, we counted the number of SGEPs in the dataset by compound, cell, and dose, respectively. Subsequently, we randomly selected 10% of the compounds from the 90% of compounds possessing fewer SGEPs to constitute the test set, and the remaining data were combined with the 10% of compounds having a greater number of SGEPs prior to being randomly partitioned into the training and validation sets in a ratio of 9:1. In the partitioning of cell line and dose dimensions, we adopted the identical core strategy as that for compounds. Specifically, 10% of cell lines and 10% of dose conditions were randomly selected from the 90% of data with fewer SGEPs to construct the test set. These cell lines and doses were strictly excluded from the training and validation sets, ensuring that the model had no prior exposure to the characteristics of these cell lines or dose levels during training. Given that there were merely 7 distinct durations in our screened dataset, we designated the data with durations of 3 and 48 h as the test set, and the remainder of the data were randomly partitioned into the training and validation sets in a ratio of 9:1. Independent test sets used for model performance comparisons under chemical-blind and cell-blind splitting strategies are provided in Supporting Information Tables S3 and S4.

To assess the robustness of DeepICER in handling various types of unseen data, we first compared the performance of DeepICER with DeepCE, MultiDCP, and TranSiGen under the

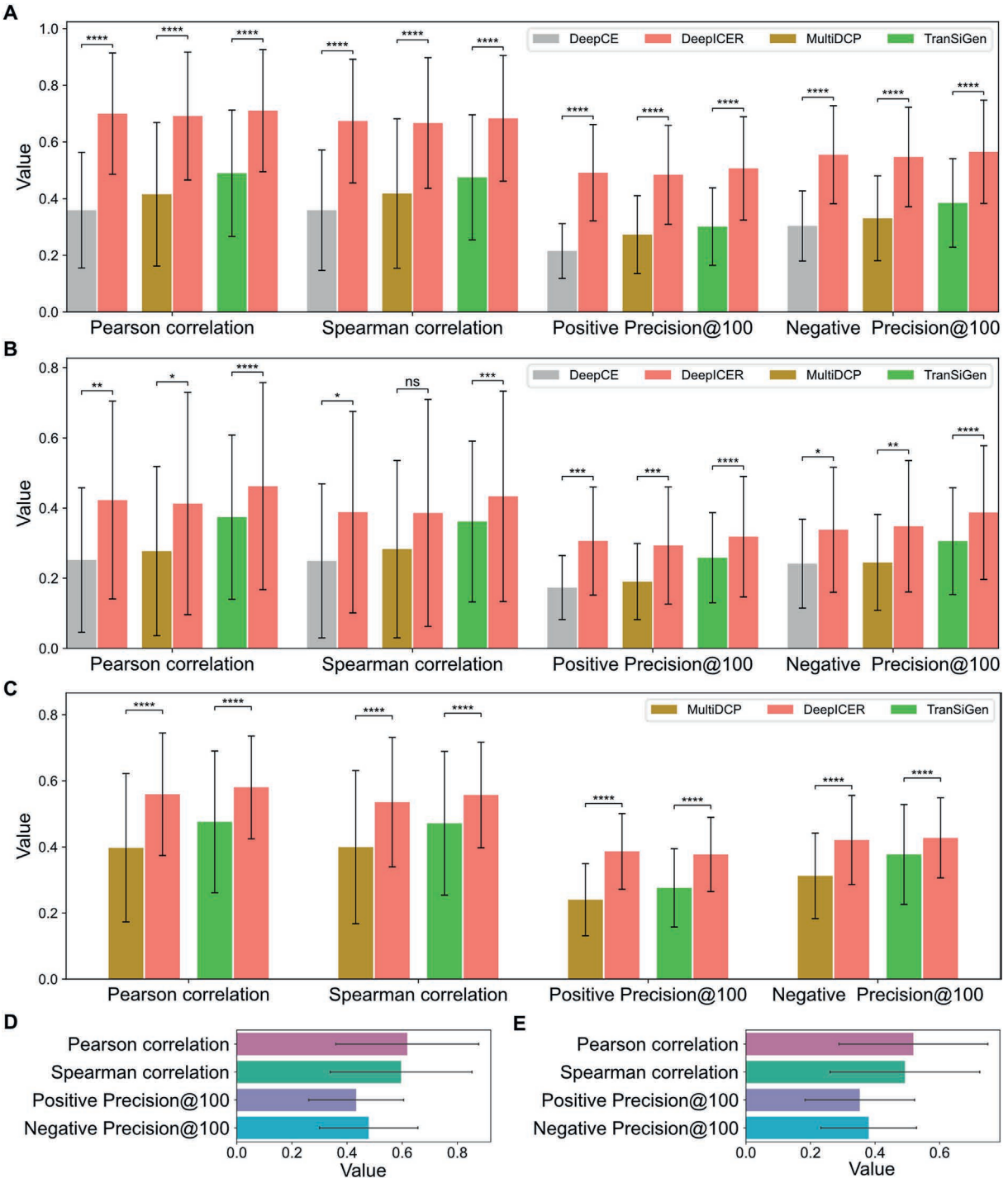


Figure 2 Comparison of SGEP prediction performance for compound perturbations. (A) Comparison of DeepICER with DeepCE, MultiDCP, and TranSiGen in the corresponding independent test set. (B) Comparison of the performance of DeepICER with DeepCE, MultiDCP, and TranSiGen for the chemical-blind splitting case. (C) Comparison of the performance of DeepICER with MultiDCP and TranSiGen in cell-blind splitting case. (D) The performance of DeepICER in dose-blind splitting. (E) The performance of DeepICER in duration-blind splitting. PCC: Pearson correlation coefficient. Error bars represent the mean ± standard deviation. Two-sided *t*-test was applied to compare the models. (****$P < 0.0001$; ***$0.0001 < P \leq 0.001$; **$0.001 < P \leq 0.01$; *$0.01 < P \leq 0.05$; ns, $0.05 < P \leq 1.0$).

chemical-blind splitting condition. DeepICER showed superior performance in predicting SGEPs for unseen compounds (Fig. 2B). It achieved higher average Pearson correlation and Spearman correlation coefficients compared to the other three models. In addition, DeepICER outperformed the other models in key metrics such as Positive Precision@100 and Negative Precision@100, which emphasize the identification of the most significantly regulated genes.

Since DeepCE could not predict responses for unseen cell lines, we only compared the performance of DeepICER with MultiDCP and TranSiGen under cell-blind splitting condition. The results showed that DeepICER significantly outperformed the other two models in all four metrics (Fig. 2C). These results underscore the advantage of DeepICER in capturing complex gene expression patterns across different conditions.

DeepCE, MultiDCP, and TranSiGen all lack the ability to account for variations in dose and duration of compound perturbations. In contrast, DeepICER was able to evaluate performance under both dose-blind and duration-blind splitting scenarios. In the dose-blind splitting setup, DeepICER achieved average Pearson and Spearman correlation coefficients of 0.618 and 0.596, respectively, along with Positive Precision@100 and Negative Precision@100 values of 0.433 and 0.478 (Fig. 2D). Similarly, in the duration-blind splitting scenario, despite the limited representation of durations, DeepICER maintained robust performance, achieving Pearson and Spearman correlation coefficients of 0.519 and 0.493, and Positive Precision@100 and Negative Precision@100 values of 0.353 and 0.380, respectively (Fig. 2E).

Collectively, these analyses underscore the remarkable performance of DeepICER in predicting SGEPs. It surpassed existing state-of-the-art models, demonstrating a strong ability to generalize across challenging scenarios involving unseen compounds, cell lines, doses, and durations.

3.5. Impact of data quality and feature ablation on DeepICER performance

The SGEPs in LINCS were typically generated from three or more biological replicates. However, the transcriptional profiles across different replicates can vary significantly. A previous study demonstrated that using unreliable data with weak correlations can have a substantial negative impact on the prediction outcomes[20]. To investigate how noisy profiles affect the prediction performance of DeepICER, we trained the model on various training sets created by filtering out unreliable gene expression profiles. This filtering was based on different 75th percentile thresholds of pairwise Spearman correlation coefficients, spanning from −1 (the original training set) to 0.5 (the high-quality training set). The model's performance improved on the test set as the filtering threshold increased, with notable improvements starting at a threshold of 0.2 (Fig. 3A). A significant performance boost was observed for thresholds between 0 and 0.1, likely due to the larger size of the training dataset at these lower thresholds, suggesting that increasing the training data volume helps enhance the model's predictive power. To ensure that the model received sufficient high-quality data for better generalization, we selected the training data where the 75th percentile of pairwise Spearman correlation coefficients exceeded 0.5.

To assess the contribution of additional features (compound dose, perturbation duration, and cell line initial expression) to the model's performance, we conducted ablation experiments. These experiments involved creating variants of DeepICER, including DeepICER-E (without the initial gene expression features of the cell line), DeepICER-T (without perturbation duration), DeepICER-D (without compound dose), DeepICER-DT (without both compound dose and perturbation duration), and DeepICER-BAN (without BAN module). All variants were trained with consistent hyperparameters and strategies. As shown in Fig. 3B,

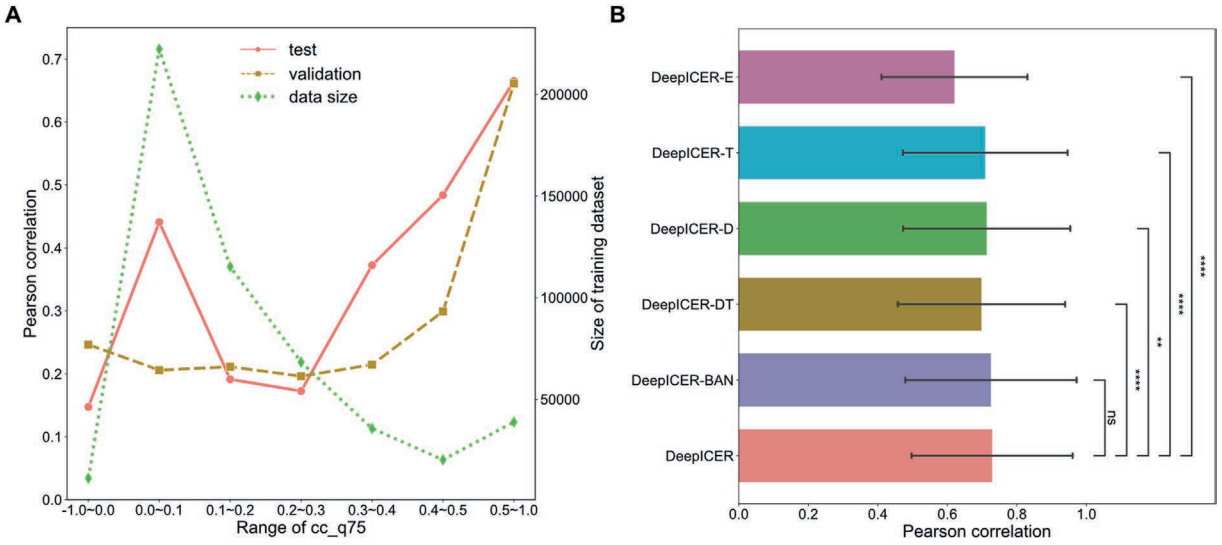


Figure 3 Impact of data quality and feature ablation on DeepICER performance. (A) The size of the training set in the dataset filtered by different cc_q75 thresholds and the Pearson correlation scores of the models trained on them. The Pearson correlation scores of these models were obtained on a high-quality test set. cc_q75: the 75th quantile of pairwise Spearman correlations between replication signatures in landmark gene space. (B) Performance comparison of DeepICER and its variants: DeepICER-E (no initial gene expression signature of the cell line), DeepICER-T (no compound perturbation duration), DeepICER-D (no compound dose), DeepICER-DT (no compound dose and perturbation duration), and DeepICER-BAN (without BAN module). Error bars represent the mean ± standard deviation. Statistical significance was assessed using a two-sided *t*-test (****$P < 0.0001$; ***$0.0001 < P \leq 0.001$; **$0.001 < P \leq 0.01$).

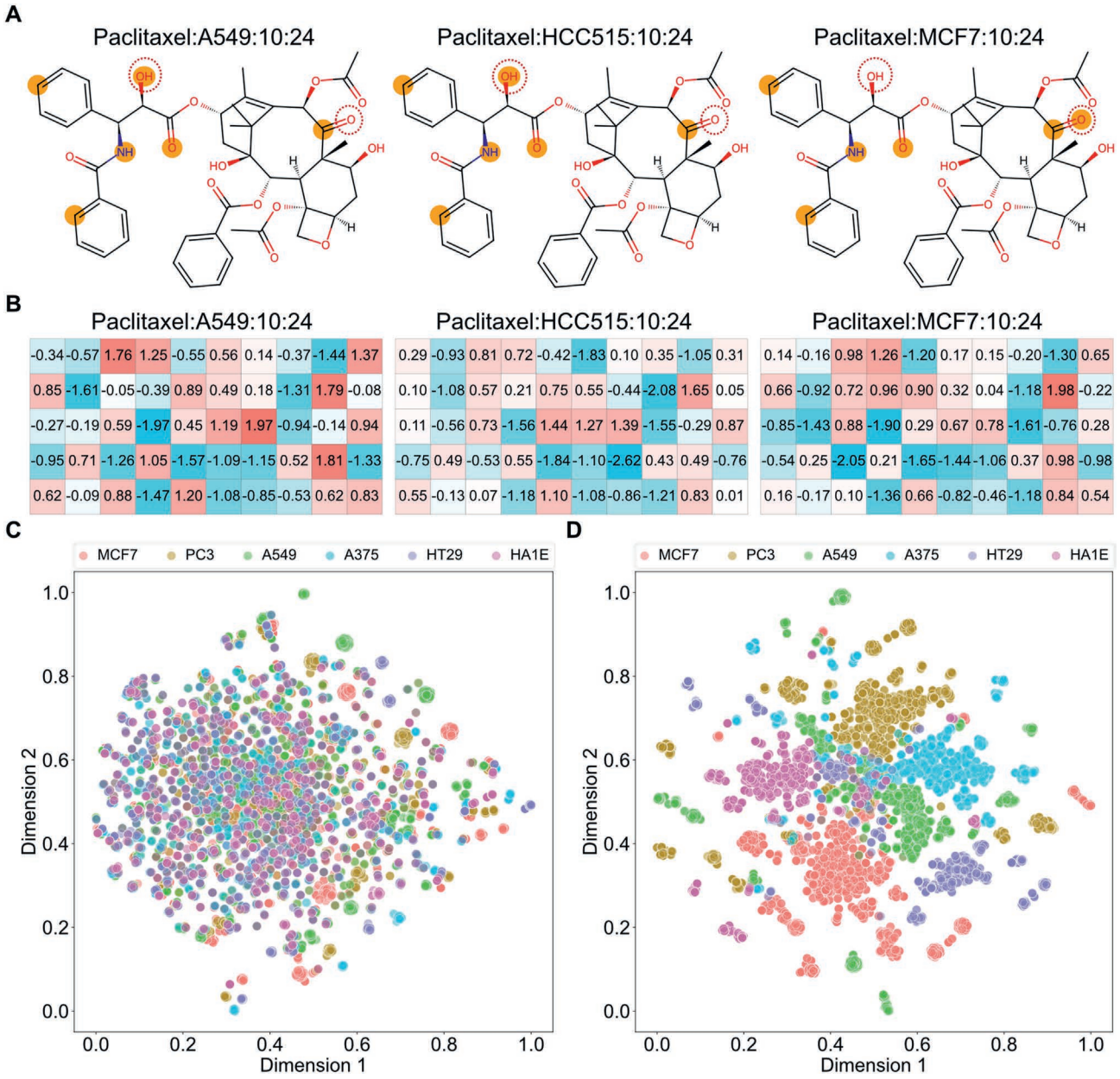
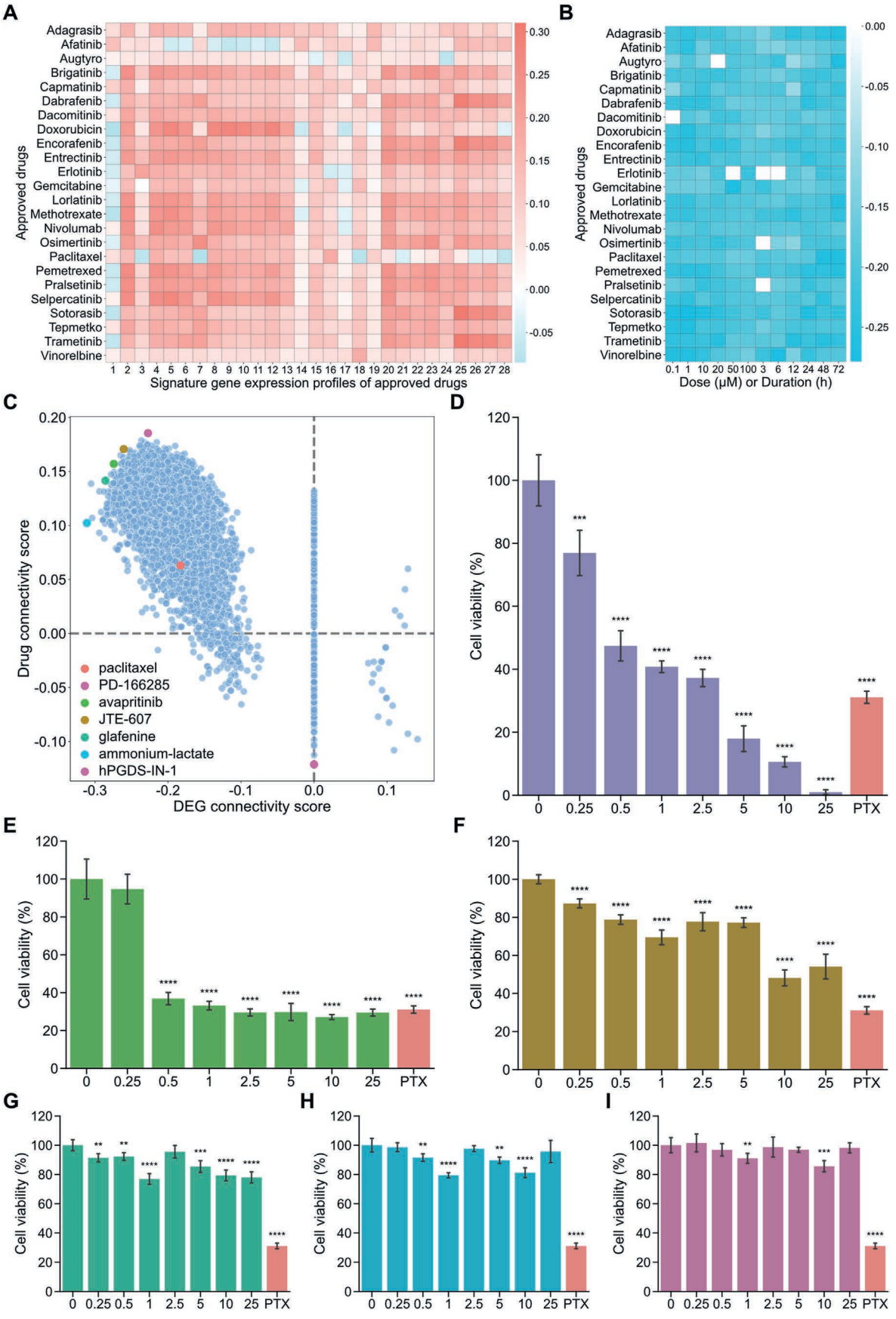


Figure 4 DeepICER distinguishes compound substructures and different cell lines. (A) Two-dimensional structures of paclitaxel with highlighted atoms (orange) that contribute most to the model when perturbing A549, HCC15, and MCF7 at a dose of 10 μmol/L for 24 h. All structures were visualized using RDKit. Red circles indicate atoms of paclitaxel with different attention in different cell lines. (B) Heatmap of attention values for the top 50 positions in the feature of A549, HCC15, and MCF7 cell lines. The value in the heatmap is the *z*-score of the attentional value at that position among the attentional values at all positions. (C) Dimensionality reduction visualization of different cell lines based on the joint representation of BAN outputs. (D) Dimensionality reduction visualization of different cell lines using the penultimate layer of the prediction module.

the variant excluding the cell line initial expression feature (DeepICER-E) performed the worst, with a Pearson correlation coefficient of 0.620, indicating that the initial gene expression feature contributes more informative value than the other attributes. The performance also significantly declined in the variants without the compound dose and/or perturbation duration, highlighting the importance of these features for the model. When the BAN module was removed, the average Pearson correlation coefficient on the test set decreased marginally from 0.729 to 0.725, a reduction of approximately 0.5%, which was not statistically significant. These findings indicate that while the BAN module contributes minimally to predictive accuracy, it enhances the interpretability of the model by facilitating the elucidation of compound−cell line interactions.

3.6. Bilinear attention enables DeepICER to capture cell line-specific features

One of the strengths of DeepICER lies in its ability to use bilinear attention mapping to capture cell line-specific features, which


A
Adagrasib
Afatinib
Augtyro
Brigatinib
Capmatinib
Dabrafenib
Dacomitinib
Doxorubicin
Encorafenib
Entrectinib
Erlotinib
Gemcitabine
Lorlatinib
Methotrexate
Nivolumab
Osimertinib
Paclitaxel
Pemetrexed
Pralsetinib
Selpercatinib
Sotorasib
Tepmetko
Trametinib
Vinorelbine
Approved drugs
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Signature gene expression profiles of approved drugs
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
B
Approved drugs
0.1 1 10 20 50 100 3 6 12 24 48 72
Dose (μM) or Duration (h)
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
C
Drug connectivity score
DEG connectivity score
paclitaxel
PD-166285
avapritinib
JTE-607
glafenine
ammonium-lactate
hPGDS-IN-1
D
Cell viability (%)
0 0.25 0.5 1 2.5 5 10 25 PTX
E
Cell viability (%)
F
Cell viability (%)
G
Cell viability (%)
H
Cell viability (%)
I
Cell viability (%)

contribute to the final predicted outcome. Bilinear attention allows DeepICER to focus on substructures and features that vary between different cell lines, providing understandings of how each cell line responds to perturbations. To demonstrate this, we first predicted the SGEPs of two lung adenocarcinoma (LUAD) cell lines (A549 and HCC515) and one breast cancer cell line (MCF7) under the perturbation of paclitaxel (10 μmol/L for 24 h). Bilinear attention maps were then visualized to highlight the model's focus on the top 10% nodes based on molecular graph attention and the top 50 positions based on cellular features. For LUAD cell lines, the atoms with the highest contribution to the model were the same (Fig. 4A). However, there were differences in the contributing atoms when compared to the breast cancer cell lines. In terms of cellular features, identical positions across all three cell lines showed variations in their contribution to the model (Fig. 4B). We further applied the model to predict the SGEPs of ABMT0280 (immortalized astrocyte cells, non-cancer cell line) and HEPG2 (hepatocellular carcinoma cell line) following perturbation with the multi-target kinase inhibitor sorafenib (10 μmol/L, 24-h treatment), and visualized the resulting bilinear attention maps (Supporting Information Fig. S1). The attention maps revealed that the model allocated differential attention to key atoms within the sorafenib molecule and exhibited distinct feature weight distributions between the two cell lines. These results demonstrate that the model can effectively distinguish between different combinations of features and focus on the specific ones that vary across cell lines.

We also visualized the joint representation of compounds and cell lines, as well as the output of the penultimate fully connected layer of the model, through dimensionality reduction. Fig. 4C and D displays the low-dimensional (t-SNE) representations of features from these two layers, with points color-coded by cell type. In the initial joint representation, only a limited number of cell clusters were apparent, with substantial overlap between cell types. In contrast, the penultimate layer's output revealed more distinct cell clusters, with clearer separation between different cell types. These results suggest that the model's final representation is more effective at discriminating among diverse cell types.

3.7. Screening for anti-LUAD drugs using DeepICER

The SGEPs generated by DeepICER can be used with profiles from compound-treated and untreated cases to screen candidate compounds for potential therapeutic applications. The Drug Connectivity Score was used to assess the relationship between the SGEPs of candidate compounds and those of known drugs, while the DEG Connectivity Score was employed to evaluate the relationship between the SGEPs of candidate compounds and those of disease states (**see 2.6**). A positive Drug Connectivity Score indicates that the candidate compound exhibits similar efficacy to known drugs, while a negative DEG Connectivity Score suggests that the candidate compound may reverse the SGEP of the disease state, potentially exerting therapeutic effects.

To investigate the effectiveness of DeepICER, we designed a drug screening process for LUAD treatment using A549 cell line at dose of 10 μmol/L and a duration of 24 h. Firstly, SGEPs of approved LUAD drugs without transcriptional profiles in LINCS were generated using DeepICER. Subsequently, their Drug Connectivity Scores and DEG Connectivity Scores were calculated. The results showed that most of these approved drugs had positive Drug Connectivity Scores relative to the drugs with existing expression profiles (Fig. 5A). When evaluated using the average Drug Connectivity Score, all of these drugs exhibited positive scores (Supporting Information Table S5). We also calculated DEG Connectivity Scores for approved LUAD drugs at various doses and durations. The results showed that the DEG Connectivity Scores at different doses and durations were mostly negative (Fig. 5B). Notably, the calculation of DEG Connectivity Scores does not require chemical-treated expression profiles, thereby overcoming the challenges posed by the lack of known therapeutic drugs for the disease. These results suggest that SGEPs generated by DeepICER can successfully identify approved LUAD drugs. Additional *in silico* validation on breast cancer (MCF7) and colon cancer (HT29) showed that the model accurately predicted gene expression changes and reversed disease-specific DEGs, demonstrating its cross-disease generalization potential (Supporting Information Fig. S2).

Furthermore, we performed phenotype-based screening using the BROAD Repurposing Compound Library and validated the top candidate compounds experimentally *in vitro*. Specifically, we used DeepICER to predict the SGEPs of 6389 compounds at a dose of 10 μmol/L, perturbing the A549 LUAD cell line for 24 h. These compounds were ranked based on their Drug Connectivity Score and DEG Connectivity Score. From the screening results (Fig. 5C), we selected five positive compounds (PD-166285, avapritinib, JTE-607, ammonium-lactate, and glafenine) and one negative compound (hPGDS-IN-1), which were then assessed for activity against A549 cells using the CCK8 assay. Additionally, paclitaxel (PTX) was included as a reference positive compound (see 2.7.1).

PD-166285 is a protein tyrosine kinase inhibitor with a broad spectrum of activity that effectively inhibits many kinase-mediated cellular functions[35,36]. Our experiments showed that PD-166285 can significantly inhibit A549 cell proliferation, with a half maximal inhibitory concentration (IC_{50}) of 0.752 μmol/L (Fig. 5D). Avapritinib, which targets PDGFRA D842V-mutant has shown unprecedented and durable clinical benefits in gastrointestinal stromal tumors[37]. We found that its inhibitory effect on

Figure 5 Screening for anti-LUAD drugs using DeepICER. (A) Heatmap of drug connectivity scores for LUAD drugs. The horizontal axis represents the SGEPs of 28 known approved drugs from the LINCS database, while the vertical axis lists the names of drugs without expression profiles. The heatmap values indicate the drug connectivity scores of each drug relative to the 28 approved drugs. (B) Heatmap of DEG connectivity scores for approved drugs used in LUAD treatment, shown for different doses and drug perturbation durations. The horizontal axis represents the drug doses or perturbation durations, the vertical axis lists the names of drugs without available expression profiles, and the heatmap values indicate the DEG connectivity scores of each drug with the differentially expressed genes in LUAD. (C) Scatterplot of connectivity scores for compounds in the BROAD repositioning database for LUAD. Compounds highlighted in different colors are candidates for *in vitro* experimental validation. (D−I) The viability of A549 cells after treatment with PD-166285 (D, PTX = paclitaxel, 10 μmol/L), avapritinib (E, PTX = paclitaxel, 10 μmol/L), JTE-607 (F, PTX = paclitaxel, 10 μmol/L), glafenine (G, PTX = paclitaxel, 10 μmol/L), ammonium-lactate (H, PTX = paclitaxel, 10 μmol/L) and hPGDS-IN-1 (I, PTX = paclitaxel, 10 μmol/L) ($n = 6$ replicates). Error bars represent the mean ± standard deviation. Statistical significance was assessed using a two-sided *t*-test (****$P < 0.0001$; ***$0.0001 < P \leq 0.001$; **$0.001 < P \leq 0.01$).

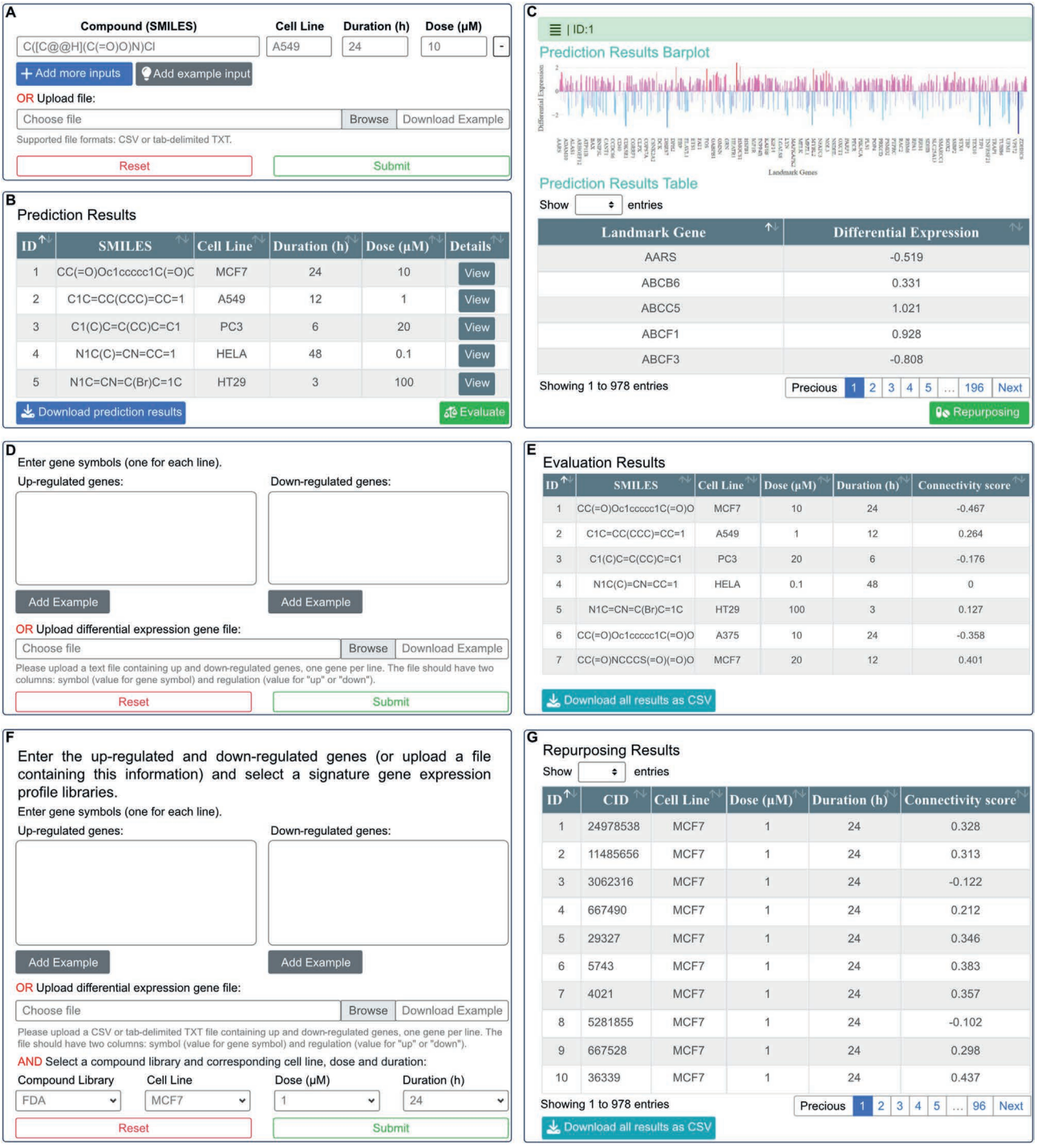


Figure 6 Schematic overview of the online service platform interface and functional workflow. The platform features a modular design with interconnected pages, enabling a seamless workflow from input submission to result analysis. (A) "Predict" page. Serves as the input interface for compound and experimental parameters. Users can manually enter individual entries (compound SMILES, cell line, treatment duration, and dose) or upload batch files (CSV/TSV format with columns "smiles," "cell_line," "duration," and "dose"). Upon clicking the "Submit" button, the system automatically transitions to the prediction results summary page to initiate SGEP calculations. (B) Prediction results summary page. Displays an overview of all prediction tasks in a tabular format. Key operations include downloading all results (CSV format) *via* the "Download prediction results" button; viewing detailed gene expression profiles of individual entries by clicking "View ", and initiating connectivity score analysis for all results *via* the "Evaluate" button. (C) Detailed prediction results page. Presents individual prediction outcomes with visualization and tabular data. A bar chart shows SGEPs, accompanied by a table listing differential expression values for all signature genes. The "Repurposing" button allows direct transition to the drug repurposing page, with the current compound's SGEPs pre-loaded. (D) "Evaluate" page. Facilitates connectivity score calculation between user-provided gene sets (upregulated/downregulated genes) and predicted SGEPs. Genes can be entered manually (one per line) or uploaded *via* file (columns "up" and "down"). (E) Evaluation results page. Displays connectivity scores in a table,

A549 cell viability was comparable to that of paclitaxel (Fig. 5E). JTE-607, known for its inhibitory effects on acute myelogenous leukemia cells[38], exhibited a moderate inhibition of A549 cell viability, particularly at 10 μmol/L, where cell activity decreased to 48.1% (Fig. 5F). Ammonium-lactate, which has mild antimicrobial properties and is used in xerosis studies[39,40], showed the smallest DEG connectivity score, but the inhibitory effect on A549 cells was not strong (Fig. 5G). Glafenine, a potent ABCG2 inhibitor[41], has been withdrawn by the FDA, showed weak and unstable effects on A549 cell viability in the experimental results (Fig. 5H). hPGDS-IN-1, which was used as a negative control drug, demonstrated no significant effect on cell viability even at a concentration of 25 μmol/L (Fig. 5I). Additionally, from the natural compound dataset, we identified pisiferic acid, a compound previously unreported for anticancer activity. *In vitro* experiments confirmed that pisiferic acid inhibited A549 cell proliferation in a dose-dependent manner (Supporting Information Fig. S3), further demonstrating the model's capability to uncover novel bioactive compounds. These results highlight the effectiveness of the SGEPs generated by DeepICER in identifying potential compound candidates.

Furthermore, we treated cells with the three compounds (PD-166285, avapritinib, and JTE-607) and performed transcriptomics sequencing. Connectivity scores between the experimentally derived DEGs and the model-predicted SGEPs were 0.42, 0.52, and 0.42, respectively, indicating strong concordance between predicted and observed gene expression trends and validating the model's predictive reliability. Additionally, analysis of SGEPs in control and drug-treated groups (Supporting Information Fig. S4A) revealed clustering results consistent with those based on all DEGs (Fig. S4B), further confirming the utility of signature genes for effective drug screening.

3.8. Online service platform of DeepICER

To facilitate the use of the DeepICER model and the large-scale perturbation profiles, we have developed a web service platform accessible at https://cbcb.cdutcm.edu.cn/deepicer/. This platform provides a straightforward workflow for drug recommendation based on the DeepICER model (Fig. 6). The process begins by generating an SGEP for a specific cell line using user-provided inputs, including the SMILES, dose, and perturbation duration of the compound (Fig. 6A−C). Next, a DEG connectivity score is calculated based on the generated SGEP and the disease-specific DEG input by the user (Fig. 6D and E). This score evaluates whether the compound can reverse the gene signature of the disease state. Alternatively, a Drug Connectivity Score can be computed by comparing the generated SGEP with the gene signature of a known drug, allowing users to assess whether the compound mimics the action of the known drug. Additionally, the model training codes together with training environment are available on GitHub (https://github.com/GreatChenLab/DeepICER), enabling users to localize and apply DeepICER in their own studies.

In addition to providing an easy-to-use platform, we applied DeepICER to generate a large-scale dataset of compound perturbations. Specifically, DeepICER was used to generate over 18 million SGEPs for compounds in three libraries, each perturbed at three doses (1, 10, and 20 μmol/L) and two durations (24 and 48 h) across 80 different cell lines. The three compound libraries include: (1) FDA-approved drug dataset, encompassing compounds from 959 FDA-approved drugs; (2) BROAD drug repositioning dataset, comprising compounds from 6384 active compounds; and (3) natural compound dataset, containing 30,440 natural compounds. Detailed information on all cell lines and compounds is provided in Supporting Information Supporting Information Tables S6−S9. These large-scale perturbation profiles can be used for drug repositioning, enabling the identification of drugs for specific diseases based on gene signatures. Accordingly, these profiles have been integrated into the platform (Fig. 6F). Users can calculate connectivity scores using the large-scale perturbation profiles, simply by inputting the gene signature of a specific disease or the post-perturbation gene signature of a drug. Based on these connectivity scores, the platform ranks compounds within the screening library to identify potential candidate drugs (Fig. 6G).

4. Discussion

The field of drug discovery is undergoing a transformative shift, driven by the growing involvement of artificial intelligence (AI), which is fundamentally reshaping the landscape of pharmaceutical research and development[42,43]. Traditionally, target-based approaches have dominated drug discovery; however, these methods face several significant limitations. These include the lack of well-defined and accessible protein targets for many diseases and the challenge of achieving sufficient cell permeability to ensure therapeutic efficacy. These obstacles have sparked renewed interest in phenotype-based screening, which provides deeper insights into disease mechanisms by assessing the broader effects of drugs on cellular phenotypes. This approach opens new avenues for discovering therapeutic mechanisms by evaluating how compounds influence the overall cell state. Despite its promise, the full experimental exploration of the vast combinatorial space of novel drugs and diverse cell lines remains practically unfeasible. Therefore, the development of computational methods that can predict and analyze these complex interactions is essential to advancing drug discovery[44].

This study introduces DeepICER, a computational platform designed to predict SGEPs for compound-perturbed cell lines, thus enabling a deeper exploration of the chemical and cellular landscape in drug discovery. DeepICER addresses several critical limitations in current computational phenotypic screening methods by functioning independently of factors such as dose, duration, and cell line type. With a Pearson correlation coefficient of 0.729 on the test dataset, DeepICER surpasses existing state-of-the-art models. Notably, DeepICER is capable of predicting perturbed expression profiles across arbitrary doses and treatment

including parameters (compound SMILES, cell line, dose, duration) and connectivity scores. Results can be downloaded as CSV *via* the "Download all results as CSV" button. (F) "Repurposing" page. Enables drug repositioning by comparing user-provided gene sets with precomputed SGEPs from compound libraries. Gene input format matches that of "Evaluate" page, and users select libraries from a dropdown menu. Clicking "Submit" triggers connectivity score calculation and transitions to the repurposing results page. (G) Repurposing results page. Displays drug repositioning results in tabular format, with an option to download the results.

durations by co-characterizing both dose and time alongside compounds. Furthermore, the model incorporates basal gene expression profiles from a diverse array of cell lines as input features. This innovative architecture allows DeepICER to leverage data from over 1000 cell lines in the CCLE dataset, enhancing its ability to learn hidden vector representations specific to different cell types. By utilizing gene expression profile encoders to represent cell lines, DeepICER can effectively use unlabeled basal gene expression data, enabling it to make context-specific predictions for novel, unseen cell lines. A key strength of DeepICER is its use of a BAN module, which processes compounds and cell lines separately and integrates their features to provide molecular-level insights and interpretations. Although ablation experiments show that BAN has a modest effect on predictive performance, it enhances interpretability by revealing drug-cell interactions, making it a valuable tool for mechanism-informed drug design.

To validate the efficacy of DeepICER, we used the model to identify novel compound candidates for A549 cells. The experimental results confirmed the activity of these predicted compounds, demonstrating DeepICER's ability to guide experimental drug screening and significantly reduce both time and costs. Furthermore, DeepICER generated large-scale perturbation profiles for three compound libraries, and we developed a web service platform that enables researchers to access the model and perform compound screenings. This platform offers a user-friendly interface, making DeepICER a valuable tool for gene-based therapeutic screening and drug discovery.

Although DeepICER has proven effective in predicting drug candidates, there are several areas for future improvement. While we represented compounds as molecular graphs based on isomeric SMILES, such representations may not fully capture important molecular features like 3D geometries, conformational flexibility, or dynamic behaviors, which are crucial for understanding complex molecular interactions. In future work, we plan to explore alternative encoding methods, such as MOL/SDF files, which can better represent atomic coordinates and bond connections, offering a more comprehensive molecular depiction. Although the incorporation of bilinear attention mechanisms offers a preliminary level of interpretability, the inherent 'black-box' nature of deep learning models still limits their transparency. Future efforts will focus on leveraging advanced interpretability techniques-such as gradient-based attribution, concept activation vectors, or model distillation-to better align learned features with biologically meaningful patterns. In addition, integrating prior biological knowledge, such as protein-protein interaction networks and signaling pathways, could enable mapping the model-predicted gene expression changes to established biological frameworks. This would not only enhance interpretability but also strengthen the biological credibility of model outputs. Another limitation lies in DeepICER's cell line encoder, which uses GO coding and gene expression data. This encoder faces challenges in accounting for technical biases across datasets[45]. Incorporating multi-omics data (*e.g.*, genomics, proteomics, metabolomics) into the model could provide a more holistic view of cellular responses and potentially enhance the model's predictive accuracy[46]. Integrating multi-omics data such as genomic and proteomic profiles poses several significant challenges, primarily due to data heterogeneity and imbalance. Different omics modalities differ not only in data structure and scale, but also in biological meaning and measurement noise. Naïve fusion of such heterogeneous data types can disrupt the feature space and compromise model performance. To address these challenges, future work will focus on designing robust fusion strategies that preserve modality-specific information (*e.g.*, *via* omics-specific encoders), enable adaptive weighting of different omics inputs (*e.g.*, through attention mechanisms), and capture complex cross-modal relationships using hierarchical or graph-based integration frameworks. Such approaches are expected to enhance DeepICER's capacity to model intricate biological systems in a more comprehensive and interpretable manner. Additionally, leveraging newer Transformer or knowledge graph-based models could further improve the accurate modeling of cell line characteristics and compound−cell interactions by incorporating extensive biomedical knowledge and long-range dependencies. Lastly, DeepICER relies on the expression data of 978 landmark genes from the L1000 database. While this is a widely used subset of genes, previous studies by Subramanian et al. highlighted that restricting analysis to landmark genes may result in the loss of approximately 20% of important gene associations, which could omit key genes critical to understanding disease mechanisms[9]. To address this, future versions of DeepICER will aim to include a broader set of genes, improving its capacity to predict drug responses more comprehensively.

5. Conclusions

DeepICER represents a powerful and flexible tool for drug repurposing and the identification of novel therapeutic candidates. While there is room for improvement, the platform's ability to generate large-scale perturbation profiles, predict drug efficacy, and facilitate the exploration of complex compound-cell interactions holds great promise for accelerating drug discovery in both oncology and other therapeutic areas.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 32470710), Natural Science Foundation of Sichuan (No. 2024ZDZX0019), and the 'Xinglin Scholar' Discipline Talent Research Promotion Program of Chengdu University of TCM (No. MPRC2021036).

Author contributions

Conceptualization, Wei Chen and Shilin Chen; methodology, Fanbo Meng, Yue Lin, and Jing Mo; Investigation, Fanbo Meng, Yue Lin, and Jing Mo and Xunzhi Zhang; writing-original draft, Fanbo Meng and Wei Chen; writing-review & editing, Can Wang, Zhaotong Cong, Chi Song, Sanyin Zhang, Liang Leng, Shilin Chen, Liang Leng, and Wei Chen; supervision, Wei Chen and Liang Leng. All of the authors have read and approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability

All relevant data supporting the main findings of this study are included within the article and its Supporting Information files. The raw CMap LINCS Resource 2020 can be accessed at https://clue.io/data/CMap2020#LINCS2020. The row FDA-approved

library and natural compound library were downloaded from OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix/release/OMIX006910). The raw BROAD Repurposing dataset is available at https://repo-hub.broadinstitute.org/repurposing#download-data, and the raw LUAD cohort from TCGA can be found at https://xenabrowser.net/. All data can be requested from the corresponding author. Additionally, predicted signature results can be accessed *via* the website (https://cbcb.cdutcm.edu.cn/deepicer/).

Code availability

The code for DeepICER is available on Github https://github.com/GreatChenLab/DeepICER.

Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2026.01.046.

References

1. Jia ZC, Yang X, Wu YK, Li M, Das D, Chen MX, et al. The art of finding the right drug target: emerging methods and strategies. *Pharmacol Rev* 2024;**76**:896−914.
2. Sadri A. Is target-based drug discovery efficient? Discovery and "off-target" mechanisms of all drugs. *J Med Chem* 2023;**66**:12651−77.
3. Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat Rev Drug Discov* 2022;**21**:899−914.
4. Shigemizu D, Hu Z, Hung JH, Huang CL, Wang Y, DeLisi C. Using functional signatures to identify repositioned drugs for breast, myelogenous leukemia and prostate cancer. *PLoS Comput Biol* 2012;**8**:e1002347.
5. Kuusanmäki H, Leppä AM, Pölönen P, Kontro M, Dufva O, Deb D, et al. Phenotype-based drug screening reveals association between venetoclax response and differentiation stage in acute myeloid leukemia. *Haematologica* 2020;**105**:708−20.
6. Sithara S, Crowley TM, Walder K, Aston-Mourney K. Gene expression signature: a powerful approach for drug discovery in diabetes. *J Endocrinol* 2017;**232**:R131−9.
7. Vengoji R, Atri P, Macha MA, Seshacharyulu P, Perumal N, Mallya K, et al. Differential gene expression-based connectivity mapping identified novel drug candidate and improved temozolomide efficacy for glioblastoma. *J Exp Clin Cancer Res* 2021;**40**:335.
8. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929−35.
9. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437−52.e17.
10. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;**41**:D955−61.
11. Wei G, Twomey D, Lamb J, Schlis K, Agarwal J, Stam RW, et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 2006;**10**:331−42.
12. Xia Y, Li X, Bie N, Pan W, Miao YR, Yang M, et al. A method for predicting drugs that can boost the efficacy of immune checkpoint blockade. *Nat Immunol* 2024;**25**:659−70.
13. Wei Z, Zhu S, Chen X, Zhu C, Duan B, Liu Q. DrSim: similarity learning for transcriptional phenotypic drug discovery. *Genom Proteom Bioinform* 2022;**20**:1028−36.
14. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 2017;**8**:16022.
15. Ahmed F, Ho SG, Samantasinghar A, Memon FH, Rahim CSA, Soomro AM, et al. Drug repurposing in psoriasis, performed by reversal of disease-associated gene expression profiles. *Comput Struct Biotechnol J* 2022;**20**:6097−107.
16. Tian S, Liao X, Cao W, Wu X, Chen Z, Lu J, et al. GSFM: a genome-scale functional module transformation to represent drug efficacy for *in silico* drug discovery. *Acta Pharm Sin B* 2024;**15**:133−50.
17. Zhao Y, Chen X, Chen J, Qi X. Decoding connectivity map-based drug repurposing for oncotherapy. *Briefings Bioinf* 2023;**24**:bbad142.
18. Hoffmann T, Gastreich M. The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discov Today* 2019;**24**:1148−56.
19. Zhu J, Wang J, Wang X, Gao M, Guo B, Gao M, et al. Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat Biotechnol* 2021;**39**:1444−52.
20. Pham TH, Qiu Y, Zeng J, Xie L, Zhang P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nat Mach Intell* 2021;**3**:247−57.
21. Pham TH, Qiu Y, Liu J, Zimmer S, O'Neill E, Xie L, et al. Chemical-induced gene expression ranking and its application to pancreatic cancer drug repurposing. *Patterns* 2022;**3**:100441.
22. Wu Y, Liu Q, Qiu Y, Xie L. Deep learning prediction of chemical-induced dose-dependent and context-specific multiplex phenotype responses and its application to personalized Alzheimer'S disease drug repurposing. *PLoS Comput Biol* 2022;**18**:e1010367.
23. Tong X, Qu N, Kong X, Ni S, Zhou J, Wang K, et al. Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery. *Nat Commun* 2024;**15**:5378.
24. Qi X, Zhao L, Tian C, Li Y, Chen ZL, Huo P, et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat Commun* 2024;**15**:9256.
25. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald 3rd ER, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature* 2019;**569**:503−8.
26. Li M, Zhou J, Hu J, Fan W, Zhang Y, Gu Y, et al. DGL-LifeSci: an open-source toolkit for deep learning on graphs in life science. *ACS Omega* 2021;**6**:27233−8.
27. Greene D, Richardson S, Turro E. OntologyX: a suite of R packages for working with ontological data. *Bioinformatics* 2017;**33**:1104−6.
28. Kim JH, Jun J, Zhang BT. Bilinear attention networks. In: *32nd conference on neural information processing systems (NeurIPS 2018)*; 2018.
29. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;**45**:1113−20.
30. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;**44**:e71.
31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
32. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The drug repurposing hub: a next-generation drug library and information resource. *Nat Med* 2017;**23**:405−8.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545−50.
34. Fang S, Dong L, Liu L, Guo J, Zhao L, Zhang J, et al. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res* 2021;**49**:D1197−206.

35. Panek RL, Lu GH, Klutchko SR, Batley BL, Dahring TK, Hamby JM, et al. *In vitro* pharmacological characterization of PD 166285, a new nanomolar potent and broadly active protein tyrosine kinase inhibitor. *J Pharmacol Exp Therapeut* 1997;**283**:1433−44.
36. Lang JD, Hendricks WPD, Orlando KA, Yin H, Kiefer J, Ramos P, et al. Ponatinib shows potent antitumor activity in small cell carcinoma of the ovary hypercalcemic type (SCCOHT) through multikinase inhibition. *Clin Cancer Res* 2018;**24**: 1932−43.
37. Jones RL, Serrano C, von Mehren M, George S, Heinrich MC, Kang YK, et al. Avapritinib in unresectable or metastatic PDGFRA D842V-mutant gastrointestinal stromal tumours: long-term efficacy and safety data from the NAVIGATOR phase I trial. *Eur J Cancer* 2021;**145**:132−42.
38. Tajima N, Fukui K, Uesato N, Maruhashi J, Yoshida T, Watanabe Y, et al. JTE-607, a multiple cytokine production inhibitor, induces apoptosis accompanied by an increase in p21waf1/cip1 in acute myelogenous leukemia cells. *Cancer Sci* 2010;**101**:774−81.
39. Lavker RM, Kaidbey K, Leyden JJ. Effects of topical ammonium lactate on cutaneous atrophy from a potent topical corticosteroid. *J Am Acad Dermatol* 1992;**26**:535−44.
40. Ademola J, Frazier C, Kim SJ, Theaux C, Saudez X. Clinical evaluation of 40% urea and 12% ammonium lactate in the treatment of xerosis. *Am J Clin Dermatol* 2002;**3**:217−22.
41. Zhang Y, Byun Y, Ren YR, Liu JO, Laterra J, Pomper MG. Identification of inhibitors of ABCG2 by a bioluminescence imaging-based high-throughput assay. *Cancer Res* 2009;**69**:5867−75.
42. Chen W, Liu X, Zhang S, Chen S. Artificial intelligence for drug discovery: resources, methods, and applications. *Mol Ther Nucleic Acids* 2023;**31**:691−702.
43. Chen W, Yu Z, Leng L, Sun D, Liu H, Gong RZ, et al. Artificial intelligence-curated repository of gene-encoded natural diverse components from herbal medicines. *Innovation* 2025;**6**:101011.
44. Musa A, Ghoraie LS, Zhang SD, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. *Briefings Bioinf* 2018;**19**:506−23.
45. Pacini C, Dempster JM, Boyle I, Gonçalves E, Najgebauer H, Karakoc E, et al. Integrated cross-study datasets of genetic dependencies in cancer. *Nat Commun* 2021;**12**:1661.
46. He D, Xie L. A cross-level information transmission network for hierarchical omics data integration and phenotype prediction from a new genotype. *Bioinformatics* 2021;**38**:204−10.