OXFORD

Briefings in Bioinformatics, 2023, **24(1)**, 1–12
https://doi.org/10.1093/bib/bbac501
Problem Solving Protocol

GADRP: graph convolutional networks and autoencoders for cancer drug response prediction

Hong Wang†, Chong Dai†, Yuqi Wen, Xiaoqi Wang, Wenjuan Liu, Song He, Xiaochen Bo and Shaoliang Peng

Corresponding authors. Song He, Department of Bioinformatics, Institute of Health Service and Transfusion Medicine, Beijing 100850, China. Tel.: +86 010 66930242; E-mail: hes1224@163.com; Xiaochen Bo, Department of Bioinformatics, Institute of Health Service and Transfusion Medicine, Beijing 100850, China. Tel.: +86 010 66931207; E-mail: boxiaoc@163.com; Shaoliang Peng, College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China. Tel.: +86 0731 88822273; E-mail: slpeng@hnu.edu.cn

†Hong Wang and Chong Dai contributed equally to this work.

Abstract

Drug response prediction in cancer cell lines is of great significance in personalized medicine. In this study, we propose GADRP, a cancer drug response prediction model based on graph convolutional networks (GCNs) and autoencoders (AEs). We first use a stacked deep AE to extract low-dimensional representations from cell line features, and then construct a sparse drug cell line pair (DCP) network incorporating drug, cell line, and DCP similarity information. Later, initial residual and layer attention-based GCN (ILGCN) that can alleviate over-smoothing problem is utilized to learn DCP features. And finally, fully connected network is employed to make prediction. Benchmarking results demonstrate that GADRP can significantly improve prediction performance on all metrics compared with baselines on five datasets. Particularly, experiments of predictions of unknown DCP responses, drug-cancer tissue associations, and drug-pathway associations illustrate the predictive power of GADRP. All results highlight the effectiveness of GADRP in predicting drug responses, and its potential value in guiding anti-cancer drug selection.

Keywords: cancer drug response prediction, graph convolutional networks, autoencoders, multi-omics, drug structure

Introduction

Cancer is still one of the leading causes of death worldwide, and is difficult to treat [1]. Due to differences in genome profiles among cancer patients, similar anticancer drugs often show different therapeutic effects in patients with the same type of cancer [2, 3]. Therefore, it is of great clinical significance to predict the response of each patient to drugs based on their genome profile characteristics. Databases such as the Cancer Cell Line Encyclopedia (CCLE) [4], Genomics of Drug Sensitivity in Cancer (GDSC) [5], Cancer Therapeutics Response Portal (CTRP) [6], and National Cancer Institute 60 Human Cancer Cell Line Screen (NCI-60) [7] provide genome-wide information about many cell lines and their responses to drugs. The emergence of these large-scale, high-throughput screening studies have accelerated the development of personalized cancer therapy and promoted the emergence of computational prediction models for drug response prediction.

For the issue of cancer drug response prediction, the input data are typically various drug features (e.g. linear notations, molecular descriptors, and graph notations) and cell line omics data (e.g. genomics, transcriptomics, epigenomics, proteomic, and metabolomic data), and the output data are drug sensitivity values (e.g. the half-maximal inhibitory concentration (IC50), the half maximal effective concentration (EC50), and the dose-response area under the curve (AUC)) [8, 9]. Some previous predictive models have tended to use single omics data from cell lines to make predictions. For example, Zhao *et al.* [10] and Ahmed *et al.* [11] proposed models using only single omics data, and did not take drug features into account. Later models such as DeepDSC [12], CDRscan [13], and GraphDRP [14] concatenated drug features with individual features of cell lines for prediction, and obtained better results. Recent researches, such as DeepCDR [15], ADRML [16], and GraphCDR [17], have proved that the integration of multi-omics profiles from cancer cell lines can significantly improve performance.

Various computational methods have been proposed for drug response prediction. Machine learning (ML)-based approaches use logistic ridge regression [18], support vector machine (SVM) [19–21], random forest [22], Bayesian multitask multiple kernel learning [23], and matrix factorization [24] models for prediction. Deep learning (DL)-based methods [9, 25] learn latent drug and cell line features from complex data and make accurate predictions using architectures such as deep neural networks (DNNs) [26],

Hong Wang is a Master student in College of Computer Science and Electronic Engineering, Hunan University. Her research interests include bioinformatics and machine learning.
Chong Dai is a Master student in College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China.
Yuqi Wen is a Ph.D. candidate in Beijing Institute of Health Service and Tranfusion Medicine, Beijing, China.
Xiaoqi Wang is a Ph.D. in College of Computer Science and Electronic Engineering, Hunan University. His research interests include bioinformatics, deep learning algorithms for biomedical network data.
Wenjuan Liu is a postdoctoral researcher in College of Computer Science and Electronic Engineering, Hunan University. Her research interests include bioinformatics, big data, parallel computing and artificial intelligence.
Song He is an associate professor in Institute of Health Service and Tranfusion Medicine, Beijing, China.
Xiaochen Bo is a professor in Institute of Health Service and Tranfusion Medicine, Beijing, China.
Shaoliang Peng is a professor in College of Computer Science and Electronic Engineering, Hunan University. His research interests include biomedical big data, computer aided drug discovery and high-performance computing.
Received: July 28, 2022. **Revised:** October 19, 2022. **Accepted:** October 22, 2022

convolutional neural networks (CNNs) [13, 27], autoencoders (AEs) [12], and recurrent neural networks (RNNs) [28]. Graph-based methods treat origin problem as a link prediction problem. The core idea is to construct a homogeneous or heterogeneous network and learn representations of nodes from the graph. GraphDRP [14] and DeepCDR [15] employ graph neural network (GNN) to learn drug features from chemical structure graphs generated from simplified molecular-input line-entry system strings (SMILES) [29]. Ahmed *et al*. utiliz GNN to learn gene representations from gene co-expression networks [11]. MOFGCN [30] construct a bipartitic map between drugs and cell lines, and predict drug sensitivity or resistance of cell lines based on multi-group similarity network fusion, GCNs, and linear correlation coefficient decoders. HNMDRP [31] construct a cell line–drug–target heterogeneous network, and apply a flow-based algorithm for prediction.

Although the above studies have achieved promising results, there are still limitations. First, these methods considered the relationships of drugs and cell lines respectively, constructing drug similarity networks and cell line similarity networks, but ignored the correlation between drug cell line pairs (DCPs). In addition, researches utilizing GCNs did not consider the problem of over-smoothing, that is, the representation of each node becomes increasingly similar as the number of layers increases.

Considering the above constraints, we present a **G**raph convolutional networks and **A**utoencoders based model for cancer **D**rug **R**esponse **P**rediction (GADRP). A sparse DCP network that incorporates similar information about drugs, cell lines, and DCPs is constructed from a combination of the omics features of cell lines after dimension reduction by an AE, features of drugs as the representations of DCP nodes, and the similarity between DCP nodes as edges. Initial residual and layer attention-based GCN (ILGCN) is used to learn the latent embeddings of DCP nodes from a sparse DCP network for prediction. Compared with currently known methods, GADRP achieved state-of-the-art performance in cancer drug response prediction. An ablation study revealed the ability of ILGCN to alleviate over-smoothing, and a feature ablation experiment demonstrated that our DCP network construction method could effectively integrate multivariate information from cell lines and drugs. Experiments with the prediction of unknown DCP responses, drug-cancer tissue association, and drug-pathway association demonstrate the predictive power of GADRP, whose results are consistent with reported biological mechanisms.

Materials and methods

Materials and data preparation

Drug feature data

The physicochemical properties and drug molecular fingerprints of 1,448 drugs were obtained from the PubChem database [32] and open-source cheminformatics software packages. Further details are provided in the Supplemental Text.

Multi-omics data

Based on the drug response data of 499 cell lines provided by the PRISM database, we collected the feature data of the corresponding cell lines from the CCLE database [4], and entered into the model as continuous values. MicroRNA expression data for 470 cancer cell lines, DNA copy number data for 461 cancer cells, gene expression for 462 cancer cell lines and DNA methylation data for 407 cancer cell lines were collected by us. After filtering out cell lines without any of these four features, the final dataset contained 388 cell lines. Further details are provided in the Supplemental Text.

Cancer cell line–drug response data

We used IC50 to reflect the drug response of cancer cell lines. A lower IC50 value means a higher degree of drug sensitivity. We downloaded IC50 values from the PRISM Repurposing dataset [33], across 1,448 drugs and 499 cell lines. After excluding cell lines with incomplete omics features, we obtained 249,801 IC50 values between 1,448 drugs and 388 cell lines. We converted these values to $\log_{10}IC50(\mu M)$. Later, 15,958 outliers were deleted using a box-plot, leaving 233,852 values across 1,448 drugs and 388 cell lines. Finally, these values were normalized by using min-max normalization. In addition, we classified the cell lines in the PRISM database according to different tissues, and selected five tissues containing the largest number of DCPs to create the *PRISM-lung*, *PRISM-skin*, *PRISM-ovary*, *PRISM-pancreas* and *PRISM-central nervous system* datasets for tissue-specific experiments.

Methods

Overview of GADRP framework

In this study, a GCN- and AE-based model, GADRP, is proposed for cancer drug response prediction, as shown in Figure 1. The two previously mentioned drug features and four cell line features are used as input variables, and the IC50 value is the output response. GADRP has five steps: (i) A stacked deep AE is employed to learn the low-dimensional characteristics of cell lines in the feature extraction stage; (ii) A network with DCPs as nodes and node similarities as edges is constructed based on the similarity between drugs and cell lines; (iii) The K nearest neighbors (K-NN) method is used to select the k most similar neighbors for each DCP and generate a sparse DCP network; (iv) The features of each DCP from the network of step (iii) is extracted by applying ILGCN; (v) Some fully connected layers are applied to accurately predict the IC50 value for each DCP.

Our model has hyperparameters of two types. Some, such as K and α, are set based on previous publications, while others, such as the layers of ILGCN and the learning rate, are set based on performance comparisons. The hyperparameter settings are listed in Supplementary Tables S1 and S2. The pseudocode of GADRP is described in Algorithm 1.

AE-based dimension reduction

Multi-omics profiles of cancer cell lines, such as copy number and gene expression data, inevitably have high dimensionality. To overcome this problem, AEs has been extensively studied to learn a mapping from high-dimensional features to low-dimensional representations [34, 35]. We compared the prediction performance of AE with the traditional dimensionality reduction method principal components analysis (PCA), as shown in Table S1, and then selected the better performing AE method to learn low dimensional representations of cell lines. Thus proposed GADRP uses a stacked deep AE to extract features from cell line features.

The inputs of AE are 23,316-dimensional DNA copy number data and 48,392-dimensional gene expression data, and we use the output of encoder as the low-dimensional representations of DNA copy number data and gene expression data, i.e. the result after dimensionality reduction is used for the construction of the DPC network. In order to let the dimension of cell line features comparable to that of drug features, we reduced all the high-dimensional features of the cell lines to 400 dimensions, which means that there are 400 neural units in the reduced space. Both the encoder and decoder of AE have two hidden layers.

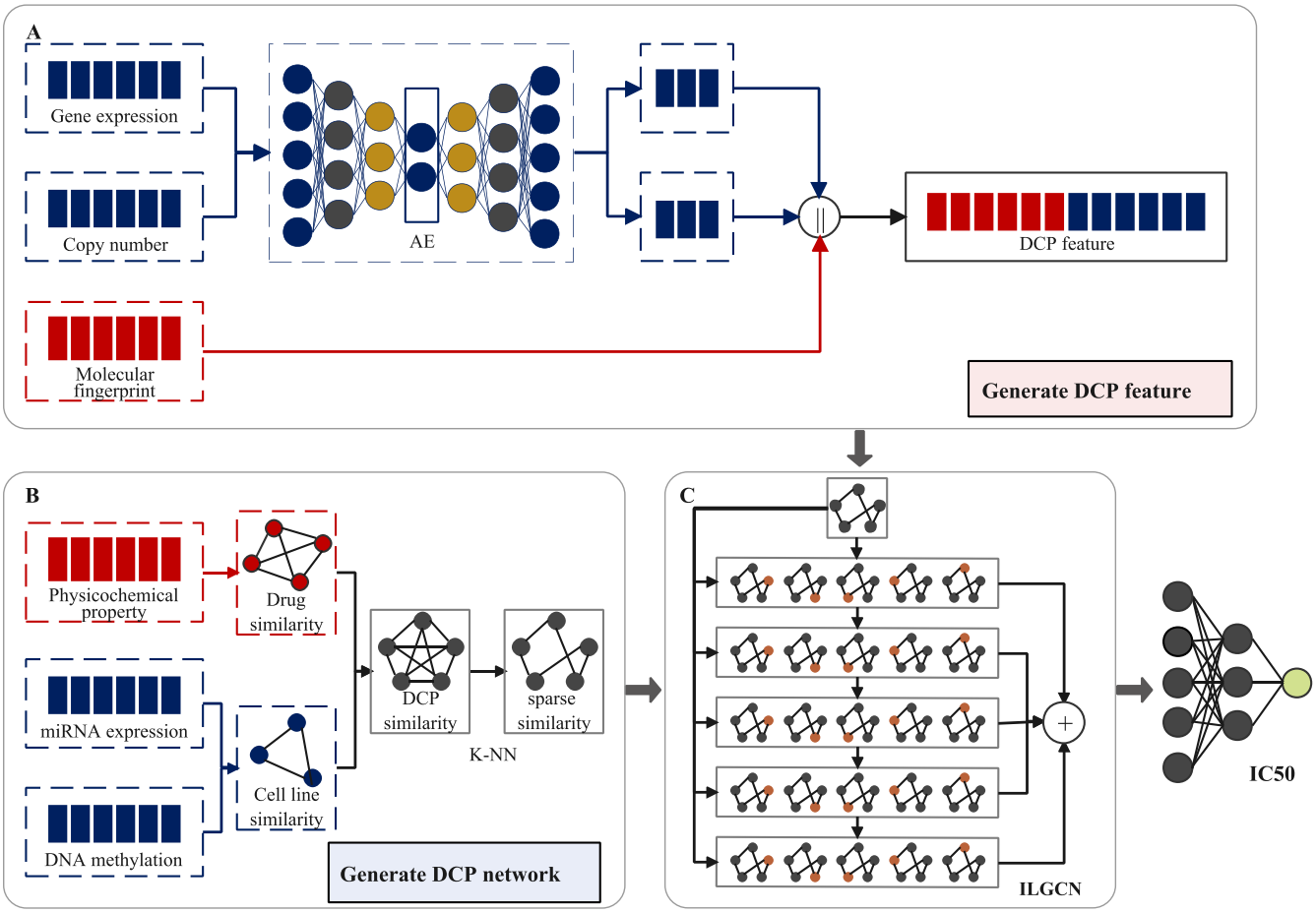


Figure 1. The framework of GADRP. (**A**) Performing AE to learn low-dimensional representations of cell lines from DNA copy number and gene expression data, and then using Concat method to connect cell line features with drug molecular fingerprints to construct the features of nodes in DCP network. (**B**) DCP network based on drug similarity network and cell line similarity network. The physicochemical properties of drugs are used to contact the drug similarity network, and the microRNA expression data and DNA methylation data of cell lines are used to contact the cell line similarity network; (**C**) Latent representations of DCP nodes are learned using five layers of ILGCN.

The encoder has 2,048 and 1,024 neural units, and those in the decoder have a reverse order from the encoder. The output layer of the encoder has 400 neural units with a Sigmoid activation function, and the 400-dimensional vector in the range of (0,1) is the result of dimensionality reduction. Other layers use the scaled exponential linear units (SELU) activation function. It is noted that the AE is only used in the feature-extraction phase to generate low-dimensional representations of cell line omics characteristics. In the training phase, GADRP directly uses low-dimensional representations of cell lines for predictions.

Construction of DCP network

Nodes of DCP network: To define the nodes set in the DCP network, the drug and cell line set are denoted as $D = \{d_1, d_2, \ldots, d_n\}$ and $C = \{c_1, c_2, \ldots, c_m\}$, respectively. As mentioned above, each DCP contains a drug and a cell line, and thus each DCP can be defined as $V = \{v_{1,1}, v_{1,2}, v_{1,3}, \ldots v_{n,m}\}$. Then the number of DCP nodes is $n*m$.

Edges of DCP network: Assuming that the interaction between DCPs $v_{i1,j1}$ and $v_{i2,j2}$ can be derived from drug–drug and cell line–cell line similarity networks, and these similarity networks can be established based on existing drug and cell line features. Networks constructed with different characteristics are compared to select the construction method for optimization performance. In this study, the physicochemical properties of drugs are used for the drug–drug network, and the microRNA expression data and CpG islands of cell lines are used for the cell line–cell line similarity network.

The similarity matrix of physicochemical properties of drugs is SIM_{pc}, and the similarity matrices of CpG islands and microRNA expression data for cell lines are represented as SIM_{CpG} and SIM_{miRNA}, respectively. They are uniformly defined as the absolute value of Pearson's correlation coefficient between corresponding features. The cell line similarity matrix is obtained as

$$S_c = \frac{SIM_{CpG} + SIM_{miRNA}}{2}, \quad (1)$$

and the DCP similarity matrix can be calculated as

$$S(v_{i1,j1}, v_{i2,j2}) = \frac{S_d(d_{i1}, d_{i2}) + S_c(c_{j1}, c_{j2})}{2}, \quad (2)$$

where $S(v_{i1,j1}, v_{i2,j2})$ denotes the interaction between DCPs $v_{i1,j1}$ and $v_{i2,j2}$, S is the DCP similarity and adjacency matrix of the DCP network, and $S_d = SIM_{pc}$ is the drug similarity matrix.

DCP feature representation: After getting the DCP network, the features of DCP are represented by the molecular substructure fingerprint feature F_{fp} of drugs, DNA copy number data F'_{cn}, and gene expression data F'_{exp} of cell lines after dimension reduction. We compared two different feature fusion methods, namely Concat and Add [36] as listed in Table S2, and then selected the

Algorithm 1 GADRP

Input: microRNA expression data F_{miRNA}, DNA methylation data (CpG islands) F_{CpG}, gene expression data F_{exp}, and DNA copy number data F_{cn} of cell lines; physicochemical properties F_{pc} and molecular fingerprints F_{fp} of drugs; drug set D; cell line set C; training iterations $epoch$.

Output: IC50

/* Feature extracting */

$F^{'}_{cn} = Autoencoder(F_{cn})$ $F^{'}_{exp} = Autoencoder(F_{exp})$

/* Construct DCP network */

Nodes:

$V \leftarrow combine(D, C)$

Adjacency matrix:

$A \leftarrow K - NN(generate(F_{pc}, F_{CpG}, F_{miRNA}))$

Features matrix:

$X \leftarrow concatenate(F_{fp}, F^{'}_{cn}, F^{'}_{exp})$

$G = (V, E, X)$

/* Training */

$H^{(0)} = X$

for $e \in [0, epoch]$ **do**

  for $l \in [0, 5]$ **do**

    $H^{(1+1)} = \sigma(((1-\alpha)\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)} + \alpha H^{(0)})W^{(l)})$

  end for

  $H = \sum_{l=1}^{5} a^{(l)} H^{(l)}$

  $F_{DCP} = H$

  $output = Linear(F_{DCP})$

  $IC50 = Sigmoid(output)$

end for

better performing Concat method to fuse the features of drugs and cell lines. The features of each DCP are concatenation of characteristics of its corresponding drug and cell line. X_d is the feature matrix of the drug $X_d = F_{fp}$, and X_c is the feature matrix of the cell line, which is derived from the characteristics of two cell lines,

$$X_c = F^{'}_{cn} || F^{'}_{exp}, \tag{3}$$

where X_c and X_d are directly concatenated as the feature of DCP, with a dimension of 1,681, which is too high when training on a DCP network. Consequently, a fully connected layer is used to reduce the characteristics of both the drug and cell line to 200 dimensions, and the feature of each DCP can be calculated as

$$X_{d_i,c_j} = X_{d_i} || X_{c_j}, \tag{4}$$

where X_{d_i} and X_{c_j} are respectively the i_{th} row of X_d and j_{th} row of X_c, || represents concatenation, and X_{d_i,c_j} is the feature of $v_{i,j}$.

K-NN based graph sampling

There are 561,824 DCP nodes and 561,824*561,824 edges in the DCP network, as mentioned above. Learning graph features on such a huge, dense network will consume much time and storage [37]. Moreover, a dense graph will produce noise and ultimately affect the performance of the model. To solve these problems requires some special methods to select an appropriate number of neighbors for each DCP and generate a sparse DCP network. Some recent studies [38, 39] have used K-NN to generate a sparse network to somewhat mitigate time and storage consumption.

We generate such a sparse network with the assumption that DCPs with similar drugs and cell lines tend to have similar labels. The K neighbors with the highest similarity are selected for each DCP, and the adjacency matrix of sparse DCP network G is generated as

$$A_{u,v} = \begin{cases} S_{u,v} & if\ v \in N_u \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where N_u is the set of k neighbors of node u, and A is the adjacency matrix of the sparse DCP network G. We set $K = 10$ in this work.

ILGCN-based feature representation

Traditional GCN has the problem of over-smoothing, especially in a higher layer model. To solve this problem, we add initial residual connection [40] and layer attention [41] to original GCN, and propose ILGCN. The core operation of ILGCN is to connect each layer of GCN with the initial representation of DCP nodes and combine representations from different layers.

ILGCN is only suitable for undirected graphs, as original GCN [42,43]. The sparse DCP network is written as $G = (V, E, X)$, where V represents the DCP node set, E denotes the set of edges, and X is the node feature matrix. To satisfy the condition of an undirected graph, the symmetric matrix A' of symmetric graph G' is expressed as

$$A^{'}_{u,v} = A^{sym}_{u,v}, \tag{6}$$

where the superscript "sym" denotes the extended symmetric matrix, and the edge between nodes u and v means $v \in N_u$ or $u \in N_v$.

The propagation rule of original GCN is [42]

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}), \tag{7}$$

where $H^{(l)}$ is the output of the l_{th} layer, $H^{(0)} = X$, $\hat{D}$ is the degree matrix of $\hat{A}$, $\hat{D} = diag(\sum_j^n \hat{A}_{i,j})$, $W^{(l)}$ is the weight of the l_{th} layer, and σ is a nonlinear activation function, which is set to *ReLU*. Since all diagonal elements of adjacency matrix A' are 1, $\hat{A} = A'$. The initial representation information is then added to the input in each layer, and the $(l+1)_{th}$ layer of ILGCN is

$$H^{(l+1)} = \sigma(((1-\alpha)\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)} + \alpha H^{(0)})W^{(l)}), \tag{8}$$

where α is the proportion of initial representation, which is set to 0.1 [40].

Finally, the representations from different layers are combined as

$$H = \sum_{l=1}^{L} a^{(l)} H^{(l)}, \tag{9}$$

where L is the number of layers of ILGCN, which is set to 5, and $a^{(l)}$ is auto-learned by neural networks.

Evaluation metrics

We evaluate the performance of GADRP by metrics: root mean squared error (*RMSE*), Pearson's correlation coefficient (*PCC*), Spearman's correlation coefficient (*SCC*) and coefficient of determination (R^2). *RMSE* is the deviation between the observed and true values. *PCC* represents the correlation between them.

SCC is a nonparametric correlation coefficient that describes *PCC* between variables after permutation. R^2 indicates the fitness accuracy between them. It is noted that a better prediction effect is indicated by a higher *PCC*, *SCC*, and R^2, and a smaller *RMSE*. These metrics are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}, \tag{10}$$

$$PCC = \frac{\mathrm{cov}(Y, \tilde{Y})}{\sigma_Y \sigma_{\tilde{Y}}}, \tag{11}$$

$$SCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}, \tag{13}$$

where N is the size of the test data, y_i is the true value, $\tilde{y}_i$ is the predicted value, d_i is the rank difference between y_i and $\tilde{y}_i$, and $\bar{y}_i$ is the average of all y_i.

Result

Performance of GADRP

Aiming to evaluate the performance of GADRP on cancer drug response prediction, we compared our method with two traditional machine learning methods and seven state-of-the-art deep learning methods including: Ridge Regression, Random Forest, MOLI [26], CDRscan [13], tCNNs [27], DeepDSC [12], DeepCDR [15], MAOFGCN [30], and DeepTTA [44] based on the same dataset and metrics, and using the default or optimal performance parameters for each model. All baseline models shared the same input features and preprocessing methods as GADRP. we used 5-fold cross-validation based on stratified sampling to evaluate the performance of the model. Specifically, we divided the instances of each cell line type into five parts, and then each part is chosen as the testing set and the remaining four parts are used as the training set. The average value of each indicator at each fold is used to measure different models.

The experimental results are shown in Table 1, indicating that GADRP achieved state-of-the-art performance compared with baseline models. Compared to MOLI, CDRscan, tCNNs, DeepDSC, DeepCDR, and MOFGCN, GADRP achieves the lowest *RMSE* (0.0895) indicates the most accurate prediction, and the highest *PCC* (0.8610), *SCC* (0.8342) and R^2 (0.6643) show strong agreement between the recorded and predicted IC50 values. Although the R^2 metric of DeepTTA is a little higher than our model, the performance of DeepTTA on the other three metrics, as well as the time and space complexity listed in Table S3, is much worse than our model. Besides, aiming to measure the complexity of GADRP, we used the total number of floating point operations (*FLOPs*) and the total number of learnable parameters (*Params*) to measure the time-complexity and space-complexity of GADRP. The results and analysis as shown in the Supplemental Text and Table S3.

Effectiveness of ILGCN

To verify the ability of ILGCN to alleviate over-smoothing, we compared ILGCN with traditional GCN and residual learning based GCN (ResGCN) [45]. The performance comparison of GNN models with different layers is plotted in Figure 2.

It is noteworthy that these methods achieve the same effect at the first layer, because neither the residual nor the initial residual connection works in this case. The performance of the original GCN decreases with the increase of network depth due to over-smoothing, but both ResGCN and ILGCN alleviate this problem and improve performance to some extent. In addition, ILGCN performs much better than ResGCN with more than two layers. These observations illustrate that ILGCN is superior to GCN and ResGCN on all four metrics, and it can better alleviate over-smoothing than ResGCN. The experimental results show that the performance of GADRP increases with the number of ILGCN layers. However, since the results after the fifth layer cannot be measured, we define the number of layers of ILGCN as five in this study.

Tissue-specific experiment

To assess the non-tissue specificity of GADRP, we evaluated the performance of the model with baseline methods on five datasets based on tissue partitioning. The 388 cell lines in our dataset belong to 23 incompatible tissues, so the corresponding labeled DCPs can be divided into 23 categories. These DCPs are not distributed evenly across tissues, as shown in Figure 3A. The five tissues with the largest numbers of DCPs are lung, skin, ovary, pancreas, and central nervous system, which account for 49.5% of all DCPs.

The experimental steps are as follows. DCPs belonging to these five tissues were selected from the PRISM dataset, and named as *PRISM-lung*, *PRISM-skin*, *PRISM-ovary*, *PRISM-pancreas*, and *PRISM-central nervous system*. We tested GADRP on these five datasets and compared its performance with seven deep learning benchmark models. Details of the results based on four metrics are shown in Figure 3B and the Supplementary Tables S6–S10.

We find that GADRP performs better than the other seven methods on all five tissues. For example, the *PCC* values for each dataset are 0.8633, 0.8542, 0.8523, 0.8368 and 0.8433, respectively, and the average *PCC* values of these eight methods on different datasets are 0.8516, 0.8394, 0.8354, 0.8173 and 0.8215. The experimental results show that our model can identify tissue-specific drug responses, and can achieve consistent and excellent performance on different tissues.

Feature ablation experiments

Feature ablation experiments were conducted to verify the validity of the features selected in this model. We selected different feature combinations to construct the DCP similarity network, and used the remaining features as representations of DCP nodes. The results are shown in Figure 4. It can be seen that, compared with other network construction methods, our model achieves the most objective effect on all four metrics. Hence, the DCP network construction method is effective in integrating multivariate information from cell lines and drugs, thereby contributing to the state-of-the-art predictive performance of our model. It is not difficult to tell from these results that only using gene expression data or DNA copy number data to construct DCP features has an impact on the global performance of the model, and that gene expression data contribute more to model performance than DNA

Table 1. Performance comparison of different algorithms based on four metrics

Method	RMSE↓	PCC↑	SCC↑	R^2 ↑
Ridge Regression	0.1267±0.0004	0.6937±0.0034	0.6568±0.0034	–0.0755±0.0135
Random Forest	0.1063±0.0006	0.7990±0.0019	0.7641±0.0024	0.5301±0.0055
MOLI	0.0938±0.0003	0.8461±0.0009	0.8211±0.0028	0.6047±0.0036
CDRscan	0.0931±0.0003	0.8490±0.0010	0.8226±0.0020	0.6241±0.0064
tCNNs	0.0934±0.0004	0.8478±0.0006	0.8194±0.0018	0.6224±0.0034
DeepDSC	0.0952±0.0007	0.8423±0.0018	0.8118±0.0034	0.6195±0.0098
DeepCDR	0.0939±0.0029	0.8457±0.0050	0.8096±0.0136	0.5627±0.0403
MOFGCN	0.0944±0.0002	0.8430±0.0002	0.8152±0.0010	0.6109±0.0017
DeepTTA	0.0966±0.0004	0.8352±0.0010	0.8085±0.0022	**0.6940±0.0061**
GADRP	**0.0895±0.0004**	**0.8610±0.0014**	**0.8342±0.0018**	0.6643±0.0079

Note: Bolded numbers represent the best performance ('↑' means the larger the better, and '↓' means the smaller the better).

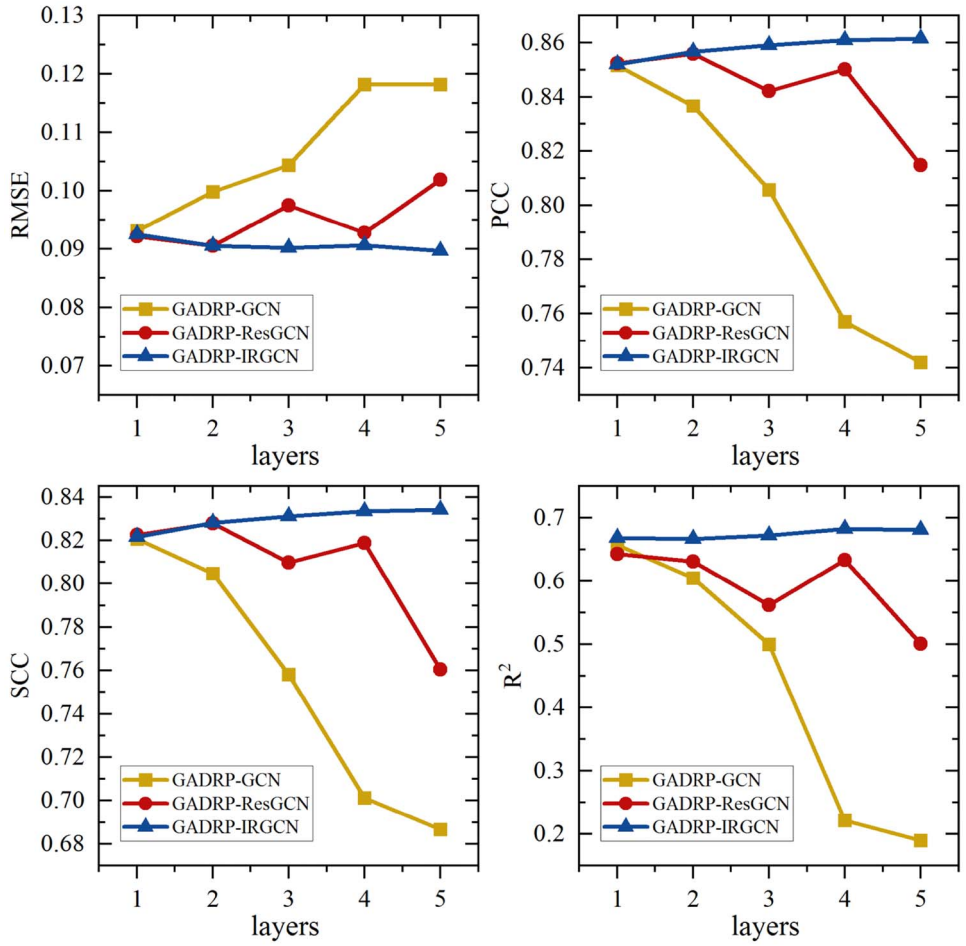


Figure 2. Performance comparison of different GCN-based models. Performance comparison results of GADRP-GCN, GADRP-ResGCN and GADRP-IRGCN based on four metrics.

copy number data, which is consistent with results reported in previous literatures [9,23,46].

Prediction of unknown drug cell line pair responses

We built a GADRP model using all known DCPs, and predicted the responses of unknown DCPs. We divided known DCPs into training and test sets at an 8:2 ratio, and used the trained model to predict unknown DCP responses (approximately 58.4% of all pairs).

All predictions were grouped by drugs and sorted by median predicted IC50 values across all missing cell lines. A raincloud plot of the 10 drugs with the highest and lowest median IC50 values is depicted in Figure 5. This shows that our model can effectively distinguish sensitive and resistant drugs. Due to the absence of the true values of these unknown DCPs, we verified the results according to relevant literatures. Specifically, TW-37 has the lowest median predicted IC50 values. This outstanding performance is consistent with previous studies that indicate that TW-37 can inhibit a variety of cancers, including pancreatic, ovarian, rectal and oral cancers, by binding to Bcl-2 [47–52]. More importantly, Niclosamide is one of the drugs that we predicted to have strong anti-cancer activity, with research showing that it inhibits multiple signaling pathways and has shown anti-proliferative activity in a broad spectrum of cancer cells, including hematologic

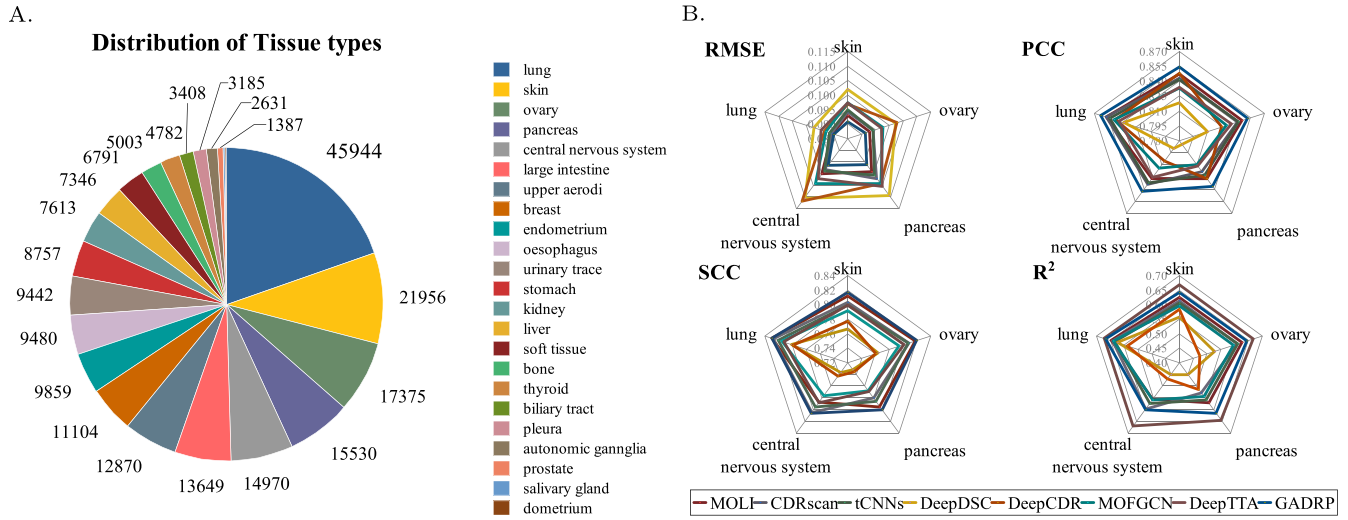


Figure 3. Comparison of GADRP with baseline methods on five tissues datasets (**A**) Distribution of DCP on 23 tissues; five issues with highest percentage are lung, skin, ovary, pancreas and central nervous system, accounting for 19.6% (45,944), 9.3% (21,956), 7.4% (17,375), 6.6% (15,530) and 6.4% (14,970), respectively, of all DCPs; (**B**) Model performance comparison of MOLI, CDRscan, tCNNs DeepDSC, DeepCDR, DeepTTA, MOFGCN and GADRP based on four metrics.

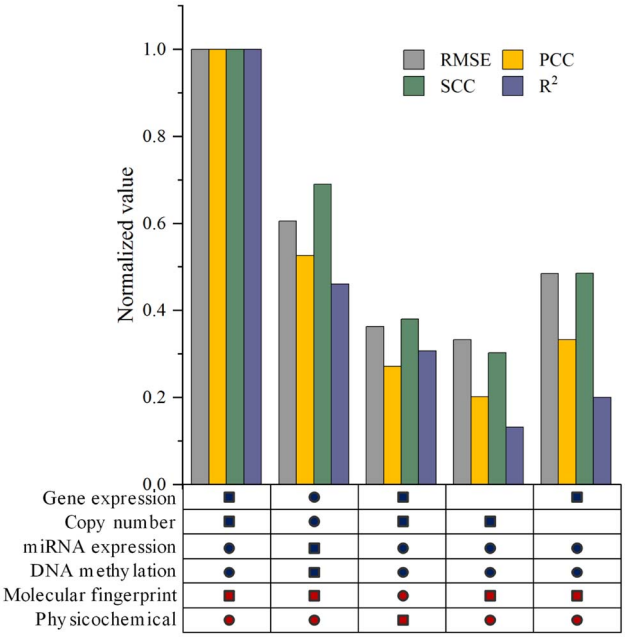


Figure 4. Results of feature ablation experiments. Blue and red circles represent cell line and drug features, respectively, used to construct features of DCP nodes. Blue and red boxes represent cell line and drug features, respectively, used to construct DCP similarity matrix.

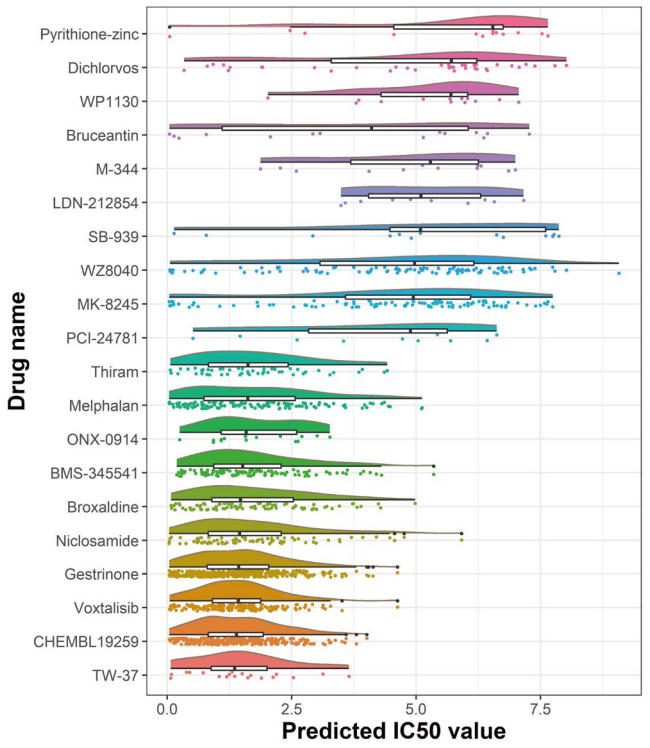


Figure 5. Predicted IC50 values of unknown DCPs grouped by drugs. Drugs are sorted by median predicted IC50 values across all missing cell lines. First 10 drugs with highest median IC50 values are least effective; last 10 drugs with lowest median IC50 values are probably most effective.

cancer cells (e.g. acute myeloid leukemia) and solid tumor cells (e.g. colon, breast, and prostate cancer) [53,54]. Also, Melphalan, at number nine, has been proved to be an effective chemotherapy agent for many types of cancer [55,56]. And the poor performance of Dichlorvos is supported by findings [57], which highlight that, as a organophosphate pesticide, it can increase cancer risk.

Association between predicted drugs and specific cancer tissues

To further verify that the predicted results of the GADRP model are biologically significant, we selected the top 1% of predicted DCP data based on IC50 values, called it *Pre-DCP* (Supplementary Dataset S1), classified and counted *Pre-DCP* according to cancer tissues, and selected lung, skin, endometrium, and breast cancer tissue for analysis (Figure 6). Based on these data, we performed three experiments. First, we performed a gene ontology (GO) biological process enrichment analysis of target genes for all drugs and specific cancer tissue drugs in *Pre-DCP* to verify whether these target genes could be significantly enriched in important biological processes in cancer. Second, based on the theory of

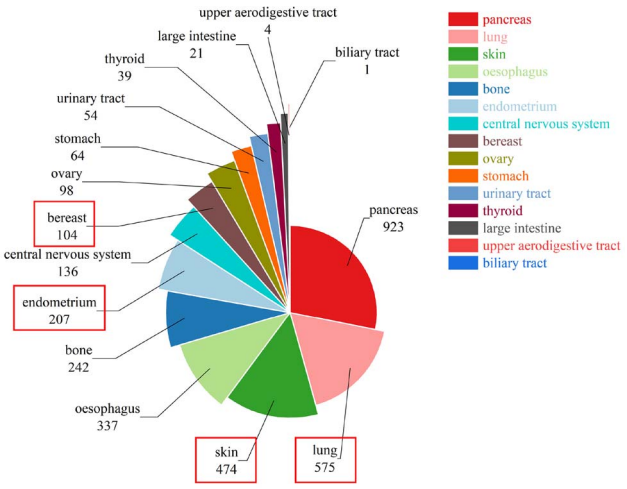


Figure 6. Cancer tissues classification. The predicted top 1% drug-cell pairs were classified according to cancer tissue, and lung, skin, endometrium and breast cancer tissues were selected for analysis.

synthetic lethality (SL), we explored whether the target genes of drugs in *Pre-DCP* could form SL pairs with the related genes of specific cancer tissues that the cell lines belong to. Finally, we used a network proximity measurement to quantify the relationship between specific cancer tissues and drugs in the human PPI network.

GO enrichment analysis

According to drug target gene data provided by Corsello *et al.* [33], we obtained the target genes involved in all drugs in *Pre-DCP*, and performed GO biological process enrichment based on these. The top 20 enrichment results are shown in Figure 7. Significantly, GO enrichment analysis confirm multiple cancer-related processes, such as that the target genes of drugs are mainly enriched in the regulation of membrane potential term, which is in good agreement with existing literature [58]. Furthermore, the regulation of membrane potential is closely related to the proliferation and differentiation of cancer cells [59,60]. In particular, the regulation of calcium ion is a key regulator of processes associated with tumor progression [61]. We find that target genes are also enriched in the response to the xenobiotic stimuli term, which also verifies the accuracy of the predicted drugs. Several review articles have summarized the important role of tyrosine phosphate on the metabolism of cancer cell lines and its significance for the study of targeted cancer therapies [62–64]. In addition, the enriched mitogen-activated protein kinase (MAPK)-related processes control the growth and survival of a broad spectrum of human tumors, and studies have shown that targeted MAPK cascade can be used to treat cancer [65]. These results demonstrate that the drugs with anti-cancer activity predicted by the GADRP model match the biological mechanisms of clinical cancer drugs, thus further demonstrating the accuracy of GADRP predictions. Then, we performed a GO enrichment analysis on the target genes of the above four tissue drugs. The enrichment results are shown in the supplementary material (Supplementary Figures S1–S4).

Mechanisms of SL for drug sensitivity

SL describes the relationship between two genes, where the abnormal expression of either gene can retain normal cell survival, while their combined abnormal expression can lead to cell death [66]. In the practical application of cancer treatment and drug discovery, SL can provide an approach for targeted therapy and a new idea for expanding the drug target space [67–69]. SL-based treatments have been regarded among the most effective anticancer therapies in the last decade [70]. To this end, we further verify the relationship between the predicted drug target gene and specific cancer tissue genes. Based on the SL gene-pair data provided by the SynLethDB database [71], we found SL pairs in breast, lung, and skin cancers, as shown in Supplementary Dataset S2.

Network-based proximity between cancers and drugs

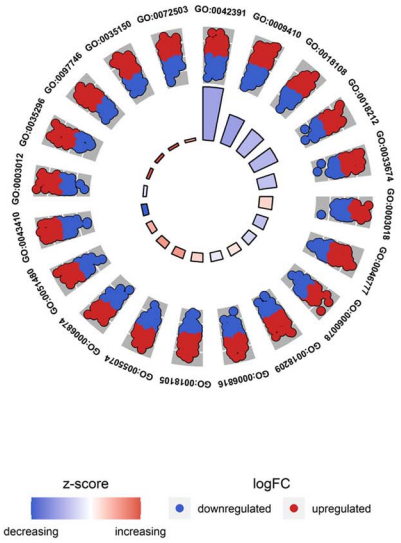
To examine drug effects on specific cancers, we used a network-based proximity measurement that quantifies the relationship between cancer-specific disease modules and drug targets in the human protein–protein interaction (PPI) network (Supplementary Text). First, based on the cancer type-specific disease-associated genes provided in GPSNet [72], we obtained genes associated with cancer tissues of lung, skin, endometrium, and breast diseases, and calculated the network proximity of the genes associated with these cancer tissues to the target gene set for each cancer tissue drugs. A negative z-score means that the drug has a significant association with the cancer, and a smaller negative value indicates a stronger association. Importantly, we found that most of our predicted results met this definition. The calculations show that the proportion of negative values is more than 60% for all four types of cancer tissue (Supplementary Dataset S3). This evidence supports that GADRP could accurately predict cancer drug responses. We then conducted a literature search based on these negative values, and found that many drugs have been shown to have effects on specific cancer diseases as listed in Table 2.

Taken together, these evidences support that GADRP could accurately predict cancer drug responses. Furthermore, GADRP could find potentially valuable and highly sensitive drugs that match our current knowledge, thereby accelerating the process of drug entry into the clinic.

Association between predicted drugs and pathways

To examine the significance of predicted results, we investigated the relationship between predicted drug responses and pathway activity scores. We screened 152 pathways with at least 90% gene expression from MSigDB [73], following the method of Yassaee Meybodi *et al.* [2]. The association between drug d_i and pathway p_i can be represented by the *PCC* value of the predicted IC50 value vector of drug d_i and the activity score vector of pathway p_i (Supplementary Text). A negative association between drug d_i and pathway p_i means p_i is sensitive to d_i, while a positive correlation means p_i is resistant to d_i. Figure 8 presents a heatmap depicting the correlations of the 10 drugs with the lowest mean predicted IC50 values, as mentioned before (see section on prediction of unknown drug cell line pair responses), with all pathways (Supplementary Dataset S4).

As expected, we found that in the pharmacogenomic space, the relationship between drugs and crutial pathways can be identified based on pathway activity scores, which also indicates that our model is able to distinguish and predict well between sensitive and resistant drugs. In addition, we confirmed several instances from existing literatures. For example, Thiram, as an inhibitor of angiogenesis and inflammation, can activate the NF-kappaB pathway and enhance the expression of ICAM-1 in human microvascular endothelial HMEC-1 cells [74]. This is consistent



ID	Description
GO:0042391	regulation of membrane potential
GO:0009410	response to xenobiotic stimulus
GO:0018108	peptidyl-tyrosine phosphorylation
GO:0018212	peptidyl-tyrosine modification
GO:0033674	positive regulation of kinase activity
GO:0003018	vascular process in circulatory system
GO:0046777	protein autophosphorylation
GO:0060078	regulation of postsynaptic membrane potential
GO:0018209	peptidyl-serine modification
GO:0006816	calcium ion transport
GO:0018105	peptidyl-serine phosphorylation
GO:0055074	calcium ion homeostasis
GO:0006874	cellular calcium ion homeostasis
GO:0051480	regulation of cytosolic calcium ion concentration
GO:0043410	positive regulation of MAPK cascade
GO:0003012	muscle system process
GO:0035296	regulation of tube diameter
GO:0097746	blood vessel diameter maintenance
GO:0035150	regulation of tube size
GO:0072503	cellular divalent inorganic cation homeostasis

Figure 7. Gene Ontology (GO) biological process enrichment analysis.

Table 2. Case studies

Cancer tissue	Drug	MOA	PMID
Breast	Vemurafenib	Inhibitor of mutated BRAF kinase	26264150, 33976643
	NVP-AUY922	Inhibitor of HSP90	18430202, 29396630
	Navitoclax	inhibitor of BCL2 family	32911681
Lung	Cabazitaxel	Inhibitor of semi-synthetic microtubule	27607471
	Alectinib	Inhibitor of anaplastic lymphoma kinase (ALK)	28586279
	Rociletinib	Inhibitor of epidermal growth factor receptor (EGFR)	25923550
Endometrial	NVP-AEW541	Inhibitor of insulin-like growth factor-I receptor (IGF-IR)	21295335
	Olaparib	Inhibitor of poly (ADP-ribose) polymerase (PARP)	24625059, 32988624
Skin	Axitinib	Inhibitor of vascular endothelial-derived growth factor (VEGF) 1, 2, and 3	25867272, 21976544
	Apatinib	Inhibitor of tyrosine kinase	30056366

Note: MOA, mechanism of action; PMID: PubMed unique identifier.

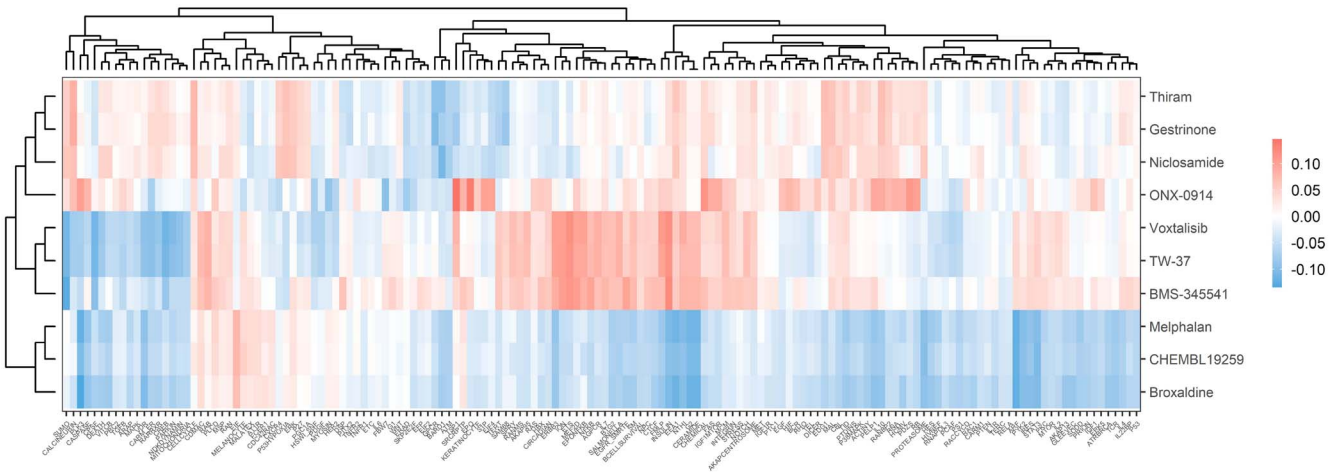


Figure 8. Heatmap of association between predicted drug response and pathway activity scores. Rows and columns represent drugs and pathways, respectively. Red represents assistant association between drug and pathway, and blue expresses resistance association.

with the observation of a positive correlation between Thiram and the NF-kappaB pathway. Moreover, the negative PCC value between the predicted IC50 value of ONX-0914 and the activity score of the ERK pathway matches the finding that ONX-0914, an immunoproteasome (IP) inhibitor, can impair T and B cell activation by inhibiting ERK signaling and proteostasis [75]. In addition, the sensitive association between Niclosamide and the mTOR pathway was confirmed by a previous study [76], which emphasized that Niclosamide can significantly inhibit mTOR signaling pathway in cervical cancer cells. Also, the predicted IC50 value of Gestrinone, a progesterone receptor antagonist, is positively correlated with the activity score of the P38MAPK

pathway, which is consistent with studies showing that Gestrinone can activate the P38MAPK pathway [77]. The observation that the activity score of the mTOR pathway is negatively correlated with predicted IC50 vectors of Voxtalisib was also confirmed by a previous study [78], which certificated that Voxtalisib inhibited PI3K/mTOR signaling and tumor growth in human glioblastoma xenograft models. The assistant association is observed between the TW-37 and ERK pathway, where a study [79] highlighted that TW-37, as an inhibitor of the Bcl-2 protein family with potential antitumor activities, can combine with ABT-263 to activate the ERK signaling pathway.

Discussion and conclusion

In this paper, we presented GADRP, a GCN- and AE-based model for cancer drug response prediction. Different from previous approaches to construct drug and cell line separation networks, we constructed a sparse DCP network that incorporates similar information about drugs, cell lines, and DCPs by using two drug features (drug physicochemical properties and drug molecular fingerprints) and four types of omics data (microRNA expression, DNA methylation, gene expression, and DNA copy number data) of cell lines. Then ILGCN, which connects each layer of GCN with the initial representations of DCP nodes and combine representations from different layers to alleviate the over-smoothing problem of traditional GCN, was proposed for representation learning. Finally, the features of DCPs were fed into fully connected layers for prediction.

Experimental results indicated that GADRP can realize sensitivity prediction with high accuracy of cancer drug response prediction compared with baseline methods. Results from different tissues illustrated that GADRP achieves consistent accuracy on diverse tissue types. A model ablation study revealed the ability of ILGCN to alleviate over-smoothing, and a feature ablation experiment demonstrated that our DCP network construction method is able to effectively integrate multivariate information from cell lines and drugs, and an effective data dimensionality reduction and feature fusion methods can help the model achieve better performance. Moreover, experiments on the prediction of unknown DCP responses, drug–cancer tissue associations, and drug–pathway associations demonstrated the predictive power of GADRP, whose results are consistent with reported biological mechanisms.

GADRP is a model for oncology drug response prediction with personalized therapy potential. Although proposed for cancer drug response prediction, thereby accelerating drug screening and helping to find potential anticancer drugs. GADRP can also be used for tasks such as the prediction of drug–target interaction, drug–target affinity and drug–disease interaction. However, there are limitations. First, due to the size of the DCP network and the limitations of computer storage capacity, ILGCN can only be controlled within five layers. Second, GADRP, as a deep learning model, has a certain degree of inexplicability, and biological entities such as targets and diseases are not involved in our model. Incorporating more entities and associations for cancer drug response prediction deserves consideration. Furthermore, despite the powerful predictive ability of GADRP, it is trained based on in vitro data, and its application in the clinic remains a major challenge. In the future work, we will try to train a model on a large number of cell lines, and then applying the idea of transfer learning to learn different practical scenarios such as patient derived tumor xenograft (PDTC), patient derived tumor xenograft (PDX) and even clinical applications [80]. In this way, we can make full use of the limited patient data and design more accurate computational models to better achieve precision medicine.

Key Points

- By integrating two features of drugs and four types of omics data of cell lines, we construct a sparse DCP network that effectively incorporates information from drugs, cell lines, and DCP similarity for cancer drug response prediction.
- GADRP clearly alleviates the over-smoothing problem of traditional GCNs by utilizing ILGCN to learn the latent embeddings of DCP nodes.
- Experimental results on multiple datasets indicate that GADRP outperforms state-of-the-art approaches for cancer drug response prediction.

Supplementary Data

The datasets and source code are freely available at https://github.com/flora619/GADRP. Supplementary data are available online at http://bib.oxfordjournals.org/.

Authors' contributions statement

S.P., S.H., and X.B., conceived, directed, and supervised the study, and ultimately revised the manuscript. H.W., C.D., Y.W., X.W. and W.L. wrote the manuscript. C.D. and Y.W. acquired the data and conducted exploratory analysis. H.W. and X.W. designed the GADRP model and comparative experiments. H.W. implemented the model and carried out comparative experiments. C.D. conducted a biological analysis of the experimental results.

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions. This work was supported by NSFC Grants U19A2067; Science Foundation for Distinguished Young Scholars of Hunan Province (2020JJ2009); National Key R&D Program of China 2022YFC3400404; Science Foundation of Changsha (Z2020694 20652, kq2004010; JZ20195242029, JH20199142034); The Funds of State Key Laboratory of Chemo/Biosensing and Chemometrics, the National Supercomputing Center in Changsha (http://nscc.hnu.edu.cn/), and Peng Cheng Lab.

References

1. Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**(3):209–49.
2. Meybodi FY, Eslahchi C. Predicting anti-cancer drug response by finding optimal subset of drugs. *Bioinformatics* 2021;**37**(23):4509–16.
3. Li Q, Shi R, Liang F. Drug sensitivity prediction with high-dimensional mixture regression. *PloS One* 2019;**14**(2):e0212108.
4. Barretina J, Caponigro G, Stransky N, *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**(7391):603–7.
5. Yang W, Soares J, Greninger P, *et al.* Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2012;**41**(D1):D955–61.

6. Basu A, Bodycombe NE, Cheah JH, *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 2013;**154**(5):1151–61.
7. Shoemaker RH. The nci60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;**6**(10):813–23.
8. An X, Chen X, Yi D, *et al.* Representation of molecules for drug response prediction. *Brief Bioinform* 2022;**23**(1):bbab393.
9. Baptista D, Ferreira PG, Rocha M. Deep learning for drug response prediction in cancer. *Brief Bioinform* 2021;**22**(1):360–79.
10. Zhao Z, Li K, Toumazou C, *et al.* A computational model for anti-cancer drug sensitivity prediction. In: *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2019, 1–4.
11. Ahmed KT, Park S, Jiang Q, *et al.* Network-based drug sensitivity prediction. *BMC Med Genomics* 2020;**13**(11):1–10.
12. Li M, Wang Y, Zheng R, *et al.* Deepdsc: a deep learning method to predict drug sensitivity of cancer cell lines. *IEEE/ACM Trans Comput Biol Bioinform* 2019;**18**(2):575–82.
13. Chang Y, Park H, Yang H-J, *et al.* Cancer drug response profile scan (CDRSCAN): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep* 2018;**8**(1):1–11.
14. Nguyen T, Nguyen GTT, Nguyen T, *et al.* Graph convolutional networks for drug response prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**(1):146–54.
15. Liu Q, Zhiqiang H, Jiang R, *et al.* Deepcdr: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 2020;**36**(Supplement_2):i911–8.
16. Moughari FA, Eslahchi C. Adrml: anticancer drug response prediction using manifold learning. *Sci Rep* 2020;**10**(1):1–18.
17. Liu X, Song C, Huang F, *et al.* Graphcdr: a graph neural network method with contrastive learning for cancer drug response prediction. *Brief Bioinform* 2022;**23**(1):bbab457.
18. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol* 2014;**15**(3):1–12.
19. Dong Z, Zhang N, Li C, *et al.* Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* 2015;**15**(1):1–12.
20. Wang Y, Fang J, Chen S. Inferences of drug responses in cancer cells from cancer genomic features and compound chemical and therapeutic properties. *Sci Rep* 2016;**6**(1):1–11.
21. Huang C, Mezencev R, McDonald JF, *et al.* Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PLoS One* 2017;**12**(10):e0186906.
22. Rahman R, Matlock K, Ghosh S, *et al.* Heterogeneity aware random forest for drug sensitivity prediction. *Sci Rep* 2017;**7**(1):1–11.
23. Costello JC, Heiser LM, Georgii E, *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 2014;**32**(12):1202–12.
24. Emdadi A, Eslahchi C. Dsplmf: a method for cancer drug sensitivity prediction using a novel regularization approach in logistic matrix factorization. *Front Genet* 2020;**11**:75.
25. Ballester PJ, Stevens R, Haibe-Kains B, *et al.* Artificial intelligence for drug response prediction in disease models. *Brief Bioinform* 2022;**23**(1):bbab450.
26. Sharifi-Noghabi H, Zolotareva O, Collins CC, *et al.* Moli: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 2019;**35**(14):i501–9.
27. Liu P, Li H, Li S, *et al.* Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinformatics* 2019;**20**(1):1–14.
28. Oskooei A, Born J, Manica M, *et al.* Paccmann: prediction of anticancer compound sensitivity with multi-modal attention-based neural networks. arXiv preprint arXiv:1811.06802. 2018.
29. O'Boyle NM. Towards a universal smiles representation-a standard method to generate canonical smiles based on the inchi. *J Chem* 2012;**4**(1):1–14.
30. Peng W, Chen T, Dai W. Predicting drug response based on multi-omics fusion and graph convolution. *IEEE J Biomed Health Inform* 2021;**26**(3):1384–93.
31. Zhang F, Wang M, Xi J, *et al.* A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Sci Rep* 2018;**8**(1):1–9.
32. Wang Y, Xiao J, Suzek TO, *et al.* Pubchem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009;**37**(suppl_2):W623–33.
33. Corsello SM, Nagari RT, Spangler RD, *et al.* Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nature cancer* 2020;**1**(2):235–48.
34. Tschannen M, Bachem O, Lucic M. Recent advances in autoencoder-based representation learning. arXiv preprint arXiv:1812.05069. 2018.
35. Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. *JMLR Workshop and Conference Proceedings*, 2012, p. 37–49.
36. Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction. *BMC Bioinformatics* 2021;**22**(1):1–17.
37. Liu X, Yan M, Deng L, *et al.* Sampling methods for efficient training of graph convolutional networks: a survey. *IEEE/CAA J Autom Sin* 2021;**9**(2):205–34.
38. Han P, Yang P, Zhao P, *et al.* Gcn-mf: disease-gene association identification by graph convolutional networks and matrix factorization. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, 705–13.
39. Chu Y, Wang X, Dai Q, *et al.* Mda-gcnftg: identifying miRNA-disease associations based on graph convolutional networks via graph sampling through the feature and topology graph. *Brief Bioinform* 2021;**22**(6):bbab165.
40. Chen M, Wei Z, Huang Z, *et al.* Simple and deep graph convolutional networks. In: *International Conference on Machine Learning*. PMLR, 2020, 1725–35.
41. Yu Z, Huang F, Zhao X, *et al.* Predicting drug–disease associations through layer attention graph convolutional network. *Brief Bioinform* 2021;**22**(4):bbaa243.
42. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. 2016.
43. Tong Z, Liang Y, Sun C, *et al.* Directed graph convolutional network. arXiv preprint arXiv:2004.13970. 2020.
44. Jiang L, Jiang C, Xinyu Y, *et al.* Deeptta: a transformer-based model for predicting cancer drug response. *Brief Bioinform* 2022;**23**(3):bbac100.
45. Li G, Muller M, Thabet A, *et al.* Deepgcns: Can gcns go as deep as cnns? In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 2019, 9267–76.
46. Iorio F, Knijnenburg TA, Vis DJ, *et al.* A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;**166**(3):740–54.
47. Ziedan N, Kadri H, Westwell A. The development of pro-apoptotic cancer therapeutics. *Mini-Rev Med Chem* 2008;**8**(7):711–8.

48. Zeitlin BD, Spalding AC, Campos MS, *et al.* Metronomic small molecule inhibitor of bcl-2 (tw-37) is antiangiogenic and potentiates the antitumor effect of ionizing radiation. *Int J Radiat Oncol Biol Phys* 2010;**78**(3):879–87.
49. Wang Z, Song W, Aboukameel A, *et al.* Retracted: Tw-37, a small-molecule inhibitor of bcl-2, inhibits cell growth and invasion in pancreatic cancer. *Int J Cancer* 2008;**123**(4):958–66.
50. Wang H, Zhang Z, Wei X, *et al.* Small-molecule inhibitor of bcl-2 (tw-37) suppresses growth and enhances cisplatin-induced apoptosis in ovarian cancer cells. *J Ovarian Res* 2015;**8**(1):1–8.
51. Lei S, Ding Y, Yun F, *et al.* The preclinical analysis of tw-37 as a potential anti-colorectal cancer cell agent. *PLoS One* 2017;**12**(10):e0184501.
52. Ahn C-H, Lee WW, Jung YC, *et al.* Antitumor effect of tw-37, a bh3 mimetic in human oral cancer. *Lab Anim Res* 2019;**35**(1):1–8.
53. Li Y, Li P-K, Roberts MJ, *et al.* Multi-targeted therapy of cancer by niclosamide: a new application for an old drug. *Cancer Lett* 2014;**349**(1):8–14.
54. Pan J-X, Ding K, Wang C-Y. Niclosamide, an old antihelminthic agent, demonstrates antitumor activity by blocking multiple signaling pathways of cancer stem cells. *Chin J Cancer* 2012;**31**(4):178.
55. Rubens RD, Knight RK, Fentiman IS, *et al.* Controlled trial of adjuvant chemotherapy with melphalan for breast cancer. *Lancet* 1983;**321**(8329):839–43.
56. Steven M, Piver. Treatment of ovarian cancer at the crossroads: 50 years after single-agent melphalan chemotherapy. *Oncology* 2006;**20**(10):1157–7.
57. Gandhi R, Snedeker SM. Critical evaluation of dichlorvos' breast cancer risk. *Comments Toxicol* 2002;**8**(1):85–123.
58. Yang M, Brackenbury WJ. Membrane potential and cancer progression. *Front Physiol* 2013;**4**:185.
59. Sundelacruz S, Levin M, Kaplan DL. Role of membrane potential in the regulation of cell proliferation and differentiation. *Stem Cell Rev Rep* 2009;**5**(3):231–46.
60. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;**144**(5):646–74.
61. Monteith GR, Prevarskaya N, Roberts-Thomson SJ. The calcium–cancer signalling nexus. *Nat Rev Cancer* 2017;**17**(6):373–80.
62. Taddei ML, Pardella E, Pranzini E, *et al.* Role of tyrosine phosphorylation in modulating cancer cell metabolism. *Biochim Biophys Acta (BBA)-Rev Cancer* 2020;**1874**(2):188442.
63. Conrads TP, Veenstra TD. An enriched look at tyrosine phosphorylation. *Nat Biotechnol* 2005;**23**(1):36–7.
64. Julien SG, Dubé N, Hardy S, *et al.* Inside the human cancer tyrosine phosphatome. *Nat Rev Cancer* 2011;**11**(1):35–49.
65. Sebolt-Leopold JS, Herrera R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer* 2004;**4**(12):937–47.
66. Huang A, Garraway LA, Ashworth A, *et al.* Synthetic lethality as an engine for cancer drug target discovery. *Nat Rev Drug Discov* 2020;**19**(1):23–38.
67. Ashworth A, Lord CJ. Synthetic lethal therapies for cancer: what's next after PARP inhibitors? *Nat Rev Clin Oncol* 2018;**15**(9):564–76.
68. Setton J, Zinda M, Riaz N, *et al.* Synthetic lethality in cancer therapeutics: the next generation. *Cancer Discov* 2021;**11**(7):1626–35.
69. Wang J, Zhang Q, Han J, *et al.* Computational methods, databases and tools for synthetic lethality prediction. *Brief Bioinform* 2022;**23**(3):bbac106.
70. Topatana W, Juengpanich S, Li S, *et al.* Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *J Hematol Oncol* 2020;**13**(1):1–22.
71. Guo J, Liu H, Zheng J. Synlethdb: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res* 2016;**44**(D1):D1011–7.
72. Cheng F, Weiqiang L, Liu C, *et al.* A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat Commun* 2019;**10**(1):1–14.
73. Liberzon A, Subramanian A, Pinchback R, *et al.* Molecular signatures database (msigdb) 3.0. *Bioinformatics* 2011;**27**(12):1739–40.
74. Kurpios-Piec D, Grosicka-Maciąg E, Woźniak K, *et al.* Thiram activates nf-kappab and enhances icam-1 expression in human microvascular endothelial hmec-1 cells. *Pest Biochem Physiol* 2015;**118**:82–9.
75. Schmidt C, Berger T, Groettrup M, *et al.* Immunoproteasome inhibition impairs t and b cell activation by restraining erk signaling and proteostasis. *Front Immunol* 2018;**9**:2386.
76. Chen L, Wang L, Shen H, *et al.* Anthelminthic drug niclosamide sensitizes the responsiveness of cervical cancer cells to paclitaxel via oxidative stress-mediated mtor inhibition. *Biochem Biophys Res Commun* 2017;**484**(2):416–21.
77. Yan Zhu, Tingting Zhang, Shuwu Xie, *et al.* Gestrinone inhibits growth of human uterine leiomyoma may relate to activity regulation of erα, src and p38 mapk. *Biomed Pharmacother*, **66**(8):569–577, 2012.
78. Wen PY, Omuro A, Ahluwalia MS, *et al.* Phase i dose-escalation study of the pi3k/mtor inhibitor voxtalisib (sar245409, xl765) plus temozolomide with or without radiotherapy in patients with high-grade glioma. *Neuro Oncol* 2015;**17**(9):1275–83.
79. Rui Y, Yefen L, Ren Y, *et al.* Synergistic effects of tw-37 and abt-263 on renal cell carcinoma cells. *Cancer Manage Res* 2021;**13**:953.
80. Ma J, Fong SH, Luo Y, *et al.* Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nat Cancer* 2021;**2**(2):233–44.