JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

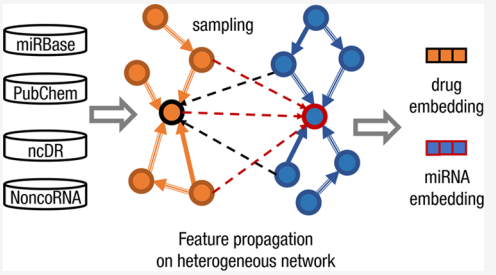
Dual-Channel Heterogeneous Graph Neural Network for Predicting microRNA-Mediated Drug Sensitivity

Lei Deng, Ziyu Fan, Xiaojun Xiao, Hui Liu,* and Jiaxuan Zhang

Cite This: J. Chem. Inf. Model. 2022, 62, 5929−5937 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Many studies have confirmed that microRNAs (miRNAs) are mediated in the sensitivity of tumor cells to anticancer drugs. MiRNAs are emerging as a type of promising therapeutic targets to overcome drug resistance. However, there is limited attention paid to the computational prediction of the associations between miRNAs and drug sensitivity. In this work, we proposed a heterogeneous network-based representation learning method to predict miRNA-drug sensitivity associations (DGNNMDA). An miRNA-drug heterogeneous network was constructed by integrating miRNA similarity network, drug similarity network, and experimentally validated miRNA-drug sensitivity associations. Next, we developed a dual-channel heterogeneous graph neural network model to perform feature propagation among the homogeneous and heterogeneous nodes so that our method can learn expressive representations for miRNA and drug nodes. On two benchmark datasets, our method outperformed other seven competitive methods. We also verified the effectiveness of the feature propagations on homogeneous and heterogeneous nodes. Moreover, we have conducted two case studies to verify the reliability of our methods and tried to reveal the regulatory mechanism of miRNAs mediated in drug sensitivity. The source code and datasets are freely available at https://github.com/19990915fzy/DGNNMDA.



INTRODUCTION

The sensitivity of cancer cells to drugs is a crucial factor for effective treatments.[1] Resistance to anticancer drugs arises from a wide range of causes, such as genetic mutations and/or epigenetic changes that allow cancer cells to avoid programmed cell death, activity variation of transport proteins that decrease the intracellular drug concentration, and other cellular and molecular mechanisms.[2] Systematic understanding of the drug response mechanism would greatly facilitate the development of novel therapeutic strategies and lead to better clinical outcomes.[3]

MicroRNAs (miRNAs), a type of small noncoding RNAs with a length of about 19−25nt,[4,5] participate in various cellular processes such as cell formation, differentiation, proliferation, and apoptosis.[6,7] Most miRNAs interact with the 3′ untranslated region (3′ UTR) of target mRNAs to induce their degradation and translational repression.[8−14] In recent years, many studies have shown that miRNAs play a non-negligible role in the regulation of drug response among different cancers.[4,15,16] Especially, some miRNAs are potential therapeutic targets to enhance drug sensitivity.[17−19] For example, inhibition of miR-21 and miR-200b can increase sensitivity to Gemcitabine in human cholangiocarcinoma cells.[20] Upregulation of miR-155-5p enhanced the sensitivity of liver carcinoma cells to Adriamycin and promoted apoptosis through the inhibition of autophagy in vitro.[21] MiR-21, miR-146a, miR-148a, miR-34a, and miR-27a have been shown to play an essential role in mediating tamoxifen sensitivity in breast cancer and is a potential therapeutic target for tamoxifen-resistant breast cancer.[18] Low expression of miR-17 and miR-20B decreased the sensitivity to Paclitaxel in breast cancer through upregulation of nuclear receptor coactivator 3 (NCOA3) levels.[22]

With the development of biochemical techniques, large-scale experiments have confirmed that many miRNAs are closely related to the sensitivity of anticancer drugs. Two databases, NoncoRNA[23] and ncDR,[24] have integrated the miRNA-drug sensitivity associations from the biomedical literature. NoncoRNA[23] is a manually curated database of experimentally supported noncoding RNAs (ncRNAs) and drug target associations that aims to provide a high-quality data resource for exploring drug sensitivity/resistance-related ncRNAs in various human cancers. The ncDR database[24] collected both experimentally validated and predicted drug resistance-associated microRNAs and long noncoding RNAs through manual curation and computational analysis. The data sources enable us to develop computational methods to predict the miRNA-mediated drug efficacy effectively.

Many computational methods formulated the association between miRNA and drug resistance, but few with regard to

Received: August 19, 2022
Published: November 22, 2022

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

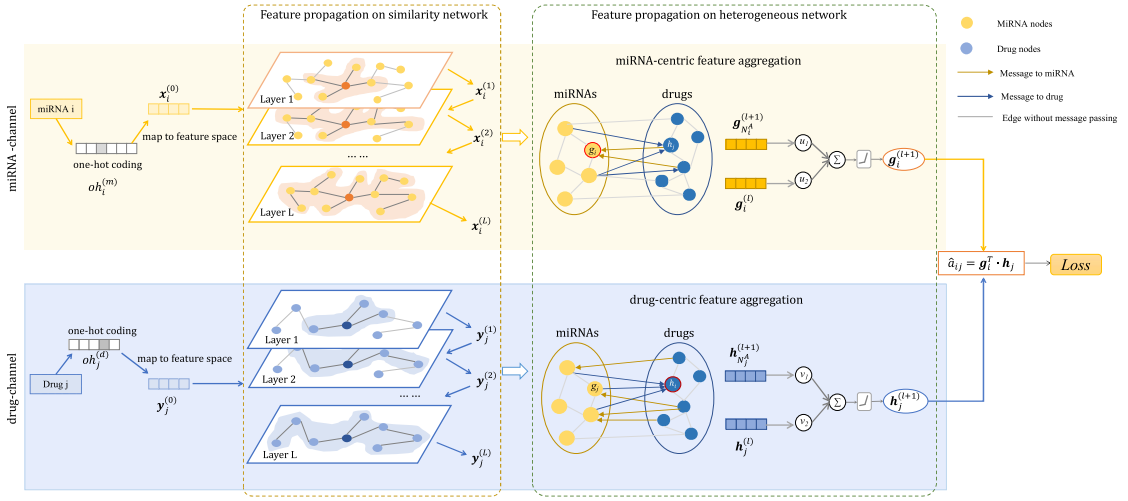


Figure 1. Illustrative flowchart of the proposed DGNNMDA model for miRNA and drug representation learning on heterogeneous graph neural network. The miRNAs and the drugs were encoded and mapped to the same feature space. Next, feature propagations on homogeneous and heterogeneous networks were successively performed to learn the representations of miRNAs and drugs. The learned embeddings were used to infer the associations between miRNAs and drugs.

drug sensitivity. For example, GCMDR[25] made use of graph convolution to build a three-layer latent factor model to predict the associations between miRNA and drug resistance. AMMGC[26] is an attentive multimodal graph convolution network method to predict miRNA-drug resistance associations, by learning the latent representations of drugs and miRNAs from four graph convolution subnetworks with distinctive combinations of features. Xu et al.[27] predicted the miRNA-drug efficacy associations using the Bi-Random walk (BiRW) algorithm on the miRNA-drug heterogeneous network. LRGCPND[28] leveraged the verified associations of ncRNAs and drug resistance to construct a bipartite graph and then developed a linear residual graph convolution approach for predicting ncRNA-associated drug resistance. To date, only one computational method has been proposed to predict the associations between miRNA and drug sensitivity. LGCMDS[29] employed LightGCN, which only retains the neighborhood aggregation component of graph convolutional network (GCN), to predict miRNA and drug sensitivity associations. From its experimental results, there is still much room for improvement of the prediction accuracy. Moreover, with the release of new data of experimentally validated miRNA-drug sensitivity associations, it is necessary to develop a reliable and accurate model to predict miRNA-drug sensitivity associations.

In this paper, we focused on the prediction of miRNA-mediated drug sensitivity. We proposed a dual-channel graph neural network on the heterogeneous network, DGNNMDA, to predict the miRNA-drug sensitivity associations. By integrating miRNA similarity network, drug similarity network, and known miRNA-drug sensitivity associations, we constructed an miRNA-drug heterogeneous network. Next, we proposed feature propagation on the homogeneous and heterogeneous networks to learn the features of miRNA and drug nodes. The multilayer heterogeneous graph aggregation operations effectively obtained the information from high-order neighbor miRNAs and drugs to update node embedding. In addition, we used a random sampling strategy to increase the generalization ability during information aggregation. To evaluate the performance of our method, we built a benchmark dataset from NoncoRNA and ncDR databases, and a manually curated independent dataset from the PubMed literature. On both datasets, we verified that our method achieved better performance than seven competitive methods. The model ablation experiments verified the effectiveness of the feature propagation on homogeneous and heterogeneous networks. We also showed that our method was robust to the data imbalances. Finally, two case studies on two cytotoxic drugs, Paclitaxel and Cisplatin, were conducted. Apart from the publications supporting the predicted miRNA-drug associations, we also constructed the regulatory network consisting of drugs, miRNAs, and target mRNAs to explore the regulatory mechanism of miRNA mediated in the drug sensitivity.

MATERIALS AND METHODS

Data Source. The miRNA-drug sensitivity associations were obtained from the NoncoRNA[23] and ncDR[24] databases. The ncDR database includes both experimentally validated and predicted associations between 140 drugs and 1039 ncRNAs. The NoncoRNA database collects experimentally supported associations between 5568 ncRNAs and 154 drugs from 134 types of cancers. We chose only the experimentally validated associations between miRNAs and drug sensitivity in our study, namely, the associations involved drug resistance or non-miRNAs were excluded. In total, 2049 miRNA-drug sensitivity associations between 431 miRNAs and 140 drugs were obtained. The dataset was used as the benchmark dataset to evaluate the performance of our method.

To further verify the reliability of our method, another manually curated dataset was built by collecting validated miRNA-drug sensitivity associations from PubMed publications. We used the keywords "miR-/miRNA/microRNA", "sensitive/sensitivity" and individual drug names to retrieve publications, and manually checked whether the miRNAs mediated in drug sensitivity or not. As a result, we collected 350 miRNA-drug sensitivity associations that were not included in ncRNAs/NoncoRNA databases. As far as our knowledge, this is the largest manually curated dataset of miRNA-drug sensitivity associations.

Construction of miRNA-Drug Heterogeneous Network. *miRNA Similarity Network.* We constructed the

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

miRNA similarity network by calculating the sequence similarity between miRNAs. The miRNA sequences were obtained from the miRBase database.[30] The miRNA similarity was computed as the Levenshtein distance,[31] which is defined as the number of deletions, insertions, or substitutions required to transform one sequence to the other one (for more details, see Section S1.1 in the Supporting Information).

For the similarities of miRNA pairs, we found that its distribution nearly followed the normal distribution. We chose the top 5% miRNA pairs with the highest similarities and obtained 7812 pairs. We counted the number of edges of these miRNAs, and found that the degree of most miRNAs was less than 25. Thereby, we decided that the number of connecting edges of each miRNA should not exceed 25. In this way, we constructed the miRNA similarity network, and the representation of each miRNA node aggregated the information from its neighbor miRNA nodes (for more details, see Section S1.2 in the Supporting Information).

Drug Similarity Network. For each drug, we obtained the canonical SMILES of the drug from the PubChem database.[32] Next, we extracted the MACCS fingerprints and calculate the similarity for each drug pair. To construct the drug similarity network, we chose the top 5% drugs with the highest similarity and obtained 1333 drug pairs. As a result, most of drugs had less than 10 neighbors, and we determined that the edge number of each drug should not exceed 10 (for more details, see Section S1.2 in the Supporting Information).

miRNA-Drug Heterogeneous Network. We integrated the miRNA similarity network, drug similarity network, and known miRNA-drug sensitivity associations to construct an miRNA-drug heterogeneous network. Formally, denoted by B the heterogeneous network, we have

$$B = \begin{bmatrix} S_{\mathrm{m}} & A \\ A^{\mathrm{T}} & S_{\mathrm{d}} \end{bmatrix} \tag{1}$$

in which S_{m} and S_{d} are the miRNA and drug similarity matrices, respectively, and A is the miRNA-drug association network with $A_{ij} = 1$ if miRNA i is experimentally validated to associate with drug j, and 0 otherwise.

Model Framework. Figure 1 shows the architecture of our model. For each miRNA-drug input sample, the model outputted a prediction score representing the possibility of miRNA mediating in drug sensitivity. Overall, the model consisted of three steps: node encoding and mapping to feature space, representation learning on the heterogeneous graph, and association prediction. First, the miRNAs and the drugs were encoded and mapped to the same feature space. Next, the node representation learning was run on miRNA-drug heterogeneous graph to incorporate the similarity information between homogeneous nodes and the association information between heterogeneous nodes. Finally, miRNA and drug representations were used to predict the association between miRNA and drug sensitivity.

Node Encoding and Mapping to Feature Space. We first performed one-hot encoding based on the numerical order of the miRNAs and the drugs separately. Because the encodings of miRNAs and drugs were in different feature spaces, we mapped the initial encodings of these two types of nodes into the same dimension so that the features of heterogeneous nodes could be propagated and aggregated. Formally, for miRNA i and drug j, we used principal component analysis (PCA) to reduce the dimension of one-hot coding, and projected them to a few principal components to obtain low-dimensional vectors, denoted by $\mathbf{x}_i^{(0)}$ and $\mathbf{y}_j^{(0)}$.

Dual-Channel Heterogeneous Graph Representation Learning. In each GNN layer, the feature of each node was updated by aggregating neighbor node features.[33] However, existing GNN models neglect the heterogeneity when aggregating neighbor information. To fully learn the complex local structure and semantic associations of heterogeneous networks, we proposed the dual-channel heterogeneous graph learning to learn node representations.

Feature Propagation on Homogeneous Network. We first performed feature aggregation among homogeneous nodes, namely, message propagation run only on miRNA and drug similarity networks separately. Formally, for the miRNA channel, denoted by $\mathbf{x}_i^{(l)}$ the embedding of miRNA i obtained at the l-th layer, the output value of $(l+1)$-th layer is obtained by aggregating the features of its neighbor miRNAs as below

$$\mathbf{x}_{\mathcal{N}_i^S}^{(l+1)} = \mathrm{AGG}(\mathbf{x}_j^{(l)} | j \in \mathcal{N}_i^S(m)) \tag{2}$$

where $\mathbf{x}_{\mathcal{N}_i^S}^{(l+1)}$ is the aggregated features from the homogeneous neighbor nodes of miRNA i; $\mathcal{N}_i^S(m)$ represents the neighbors of miRNA i in the miRNA similarity network; AGG is the aggregation function that could be max pooling, average pooling or similarity-weighted average pooling. In our study, we used average pooling to aggregate the neighborhood features. Next, we applied a nonlinear transform as below

$$\mathbf{x}_i^{(l+1)} = f(\mathbf{w}_m \cdot [\mathbf{x}_i^{(l)}, \mathbf{x}_{\mathcal{N}_i^S}^{(l+1)}]) \tag{3}$$

where $f()$ is the nonlinear Leaky ReLU function and w_{m} is a learnable vector parameter used to convey the importance of features of miRNA i itself and its neighbor nodes.

For the drug channel, we ran similar feature propagations on the drug similarity network as below

$$\mathbf{y}_{\mathcal{N}_j^S}^{(l+1)} = \mathrm{AGG}(\mathbf{y}_k^{(l)} | k \in \mathcal{N}_j^S(d)) \tag{4}$$

$$\mathbf{y}_j^{(l+1)} = f(\mathbf{w}_d \cdot [\mathbf{y}_j^{(l)}, \mathbf{y}_{\mathcal{N}_j^S}^{(l+1)}]) \tag{5}$$

where $\mathbf{y}_j^{(l)}$ is the output feature of drug j at l-th layer, $\mathcal{N}_j^S(d)$ represents the neighbors of drug j in the drug similarity network, and w_{d} is a learnable vector parameter used to convey the importance of features of drug j itself and its neighbor nodes.

The initial features $\mathbf{x}_i$ and $\mathbf{y}_j$ of miRNA i and drug j were set to the mapped value from their one-hot encoding processed by PCA, respectively. After L-layer message propagations, we aggregated the features of miRNAs and drugs on their similarity network, respectively. L is a hyperparameter that could be tuned to optimize performance.

Feature Propagation on Heterogeneous Network. Based on the aggregated features from the homogeneous network, we performed feature propagation among heterogeneous nodes based on the aggregated features on the similarity network. Denote by $\mathbf{g}^{(l)}$ and $\mathbf{h}^{(l)}$ the features of miRNA i and drug j aggregated from heterogeneous neighbor nodes, respectively. Their initial values were set as $\mathbf{g}^{(0)} = \mathbf{x}_i^{(L)}$ and $\mathbf{h}^{(0)} = \mathbf{y}_j^{(L)}$. Let us first consider the miRNA-centric feature propagation. For miRNA i, we computed the aggregated feature on the heterogeneous network as below

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

$$\mathbf{g}_{\mathcal{N}_i^A}^{(l+1)} = \frac{1}{2|\mathcal{N}_i^A|} \sum_{j \in \mathcal{N}_i^A} \left(\mathbf{h}_j^{(l)} + \frac{1}{|\mathcal{N}_j^A|} \sum_{k \in \mathcal{N}_j^A} \mathbf{g}_k^{(l)} \otimes \mathbf{h}_j^{(l)} \right) \quad (6)$$

where $\mathcal{N}_i^A$ represents the neighbor drug nodes associated with miRNA i, $\mathcal{N}_j^A$ represents the neighbor miRNA nodes associated with drug j, and $\otimes$ is the element-wise multiplication operation. From the perspective of miRNA, this formulation aggregated the feature of the first-order heterogeneous neighbors (drugs), but also the fused second-order heterogeneous neighbors (miRNAs). In particular, when the known miRNA-drug associations are sparse, the term $\mathbf{g}_k^{(l)} \otimes \mathbf{h}_j^{(l)}$ actually supplemented association information regarding second-order neighbors.

Next, we integrated the feature of miRNA i itself and conducted nonlinear transform as below

$$\mathbf{g}_i^{(l+1)} = f(\mathbf{u}_m \cdot [\mathbf{g}_i^{(l)}, \mathbf{g}_{\mathcal{N}_i^A}^{(l+1)}]) \quad (7)$$

where $f()$ is the nonlinear Leaky ReLU function and $\mathbf{u}_m$ represents the learnable weight vector used to reflect the importance of features of miRNA i itself and its neighbor nodes.

For drug-centric feature aggregation, we adopted a similar formulation as below

$$\mathbf{h}_{\mathcal{N}_i^A}^{(l+1)} = \frac{1}{2|\mathcal{N}_i^A|} \sum_{j \in \mathcal{N}_i^A} \left(\mathbf{g}_j^{(l)} + \frac{1}{|\mathcal{N}_j^A|} \sum_{k \in \mathcal{N}_j^A} \mathbf{h}_k^{(l)} \otimes \mathbf{g}_j^{(l)} \right) \quad (8)$$

$$\mathbf{h}_i^{(l+1)} = f(\mathbf{u}_d \cdot [\mathbf{h}_i^{(l)}, \mathbf{h}_{\mathcal{N}_i^A}^{(l+1)}]) \quad (9)$$

where $\mathcal{N}_i^A$ represents the neighbor miRNA nodes associated with drug i, $\mathcal{N}_j^A$ is the neighbor drug nodes associated with miRNA j, and $\mathbf{u}_d$ represents the learnable weight vector used to reflect the importance of features of drug i itself and its neighbor nodes.

It is worth noting that we sampled only a portion of neighbor nodes from which the information was aggregated. More precisely, for the aggregation in the similarity networks, when the neighbor nodes are more than 10, we randomly selected 10 neighbor nodes to aggregate their information. Random sampling has a couple of advantages: (1) Sampling in the training stage increased randomness, thus promoting the generalization ability of our model. (2) It effectively alleviates the over-smoothing of node embedding. (3) Sampling makes it possible for our method to scale to large-scale data.

Association Prediction. Following the representation learning, the learned embeddings of miRNAs and drugs can be used to predict the associations. We adopted the simple inner product to score miRNA-drug sensitivity associations. Suppose $\hat{a}_{ij}$ be the predictive score between miRNA i and drug j, we defined

$$\hat{a}_{ij} = \mathbf{g}_i^T \cdot \mathbf{h}_j \quad (10)$$

and adopted the pair-wise ranking-based loss function

$$\mathcal{L} = -\sum_{i=1}^{M} (\ln \sigma(\hat{a}_{ij} - \hat{a}_{ik}) + \lambda \|\Theta\|), (i, j) \in A \cup (i, k) \in A^-$$

where σ is the sigmoid function and Θ represents all of the parameters involved in the model. When minimizing the loss function, the miRNA-drug pairs with known associations tend to have similar representations, while those without known associations tend to have dissimilar representations. For all trainable parameters, we initialized them with a Gaussian distribution with mean 0 and standard deviation 0.01.

Negative Sample Generation Strategy. In most studies for association prediction, negative samples are randomly generated by sampling pairs of nodes. However, randomly generated negative samples often include noisy labels that may lead to biased decision boundary. In fact, the number of validated miRNA-drug associations is far less than the true ones, as the potential miRNA-drug associations are huge. As a result, randomly generated negatives may include positive samples, which probably biased the decision boundary in the model training process. We built highly credible negative samples using the strategy proposed by our group.[34] The main idea is that miRNA is unlikely to interact with a drug if it is not similar to all miRNAs associated with the drug, and vice versa. Based on this rationale, we ranked the calculated miRNA-miRNA similarities from low to high, and selected the miRNAs with the lowest similarity to those with associations to drugs. Meanwhile, we ranked the calculated drug−drug similarities from low to high and selected the drugs with the lowest similarity to those with associations to any miRNA. Finally, we randomly paired the selected miRNAs and drugs to generate highly credible negative samples.

RESULTS AND DISCUSSION

Experimental Setup. We implemented the proposed method using Tensorflow framework. The learning rate was set to 1×10^{-6}. The maximum training epoch was 100. Mini-batch was set to the default value 2048. The model was trained on a workstation with 3 NVIDIA GeForce RTX 3090 graphic cards of 24G memory.

Performance Evaluation on Benchmark Dataset. To verify the performance of our method, we first evaluated it on the benchmark dataset by fivefold cross-validations. The training samples were randomly split into five subsets of roughly equal size, each subset was taken in turn as a test set and the remaining four subsets were used to train the model, whose prediction accuracy on the test set was then evaluated. To avoid random bias, the cross-validation process was repeated 10 times, and the performance metrics were averaged as final results.[35] As shown in Figure 2 and Table S1, our method achieved average ROC-AUC, AUPR, ACC, F1-score, Precision and Recall values of 0.9255, 0.9224, 0.8523, 0.8726, 0.8261, and 0.8481, respectively. The experimental result indicated that our method performed well in predicting the associations between miRNA and drug sensitivity.

Hyperparameter Optimization. We next explored the influence of two main hyperparameters, the number of GNN layers and the dimension of embedding vector. The number of GNN layers determined the scope of feature propagation of each node. A larger number of GNN layers means feature aggregations from higher-order neighbor nodes but loss of local structure. We adopted the grid search to tune the two parameters. The number of GNN layers increased from one to four by step 1. The dimension of embedding was selected from the list {8, 16, 32, 64, 128, 256}. The experimental results are shown in Figure 3 and Table S2. When the layer number was 2

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

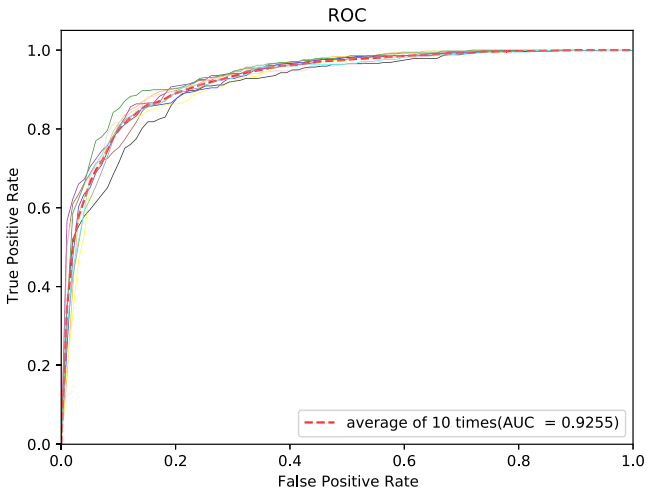


Figure 2. Receiver operator characteristic (ROC) curves of our method evaluated on benchmark dataset by fivefold cross-validations.

and the embedding dimension was 16, our method achieved the best performance.

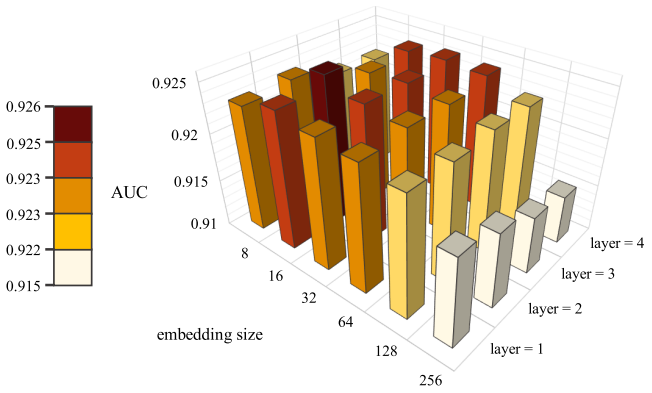


Figure 3. Impact on the performance of two hyperparameters: the number of GNN layers and embedding dimension.

Model Selection and Ablation Experiments. To explore the effect of feature propagation on homogeneous and heterogeneous nodes, we conducted model ablation experiments. We have tested four variants: (1) Feature propagation was only performed on homogeneous nodes, and the number of propagation layers was set to 1, 2, or 3; (2) Feature propagation was only performed on heterogeneous nodes, and the number of propagation layers was set to 1, 2 or 3; (3) The feature propagation on homogeneous and heterogeneous nodes was performed alternately, and the feature propagation repeated for one, two, or three times; (4) The feature propagation on homogeneous and heterogeneous nodes was performed sequentially. As shown in Table 1, when only homogeneous feature propagation was performed, two GNN layers achieved the best performance. If only heterogeneous feature propagation was performed, we also found that two GNN layers performed best. When both homogeneous and homogeneous feature propagations were run, the performance is better than the variants that run homogeneous or homogeneous feature propagations alone. More importantly, we found that the sequential mode outperformed the alternative mode. In particular, the performance of two-layer homogeneous feature propagations followed by one-layer heterogeneous feature propagation achieved the highest performance among all model variants, whose area under the curve (AUC) and AUPR reached 0.9255 and 0.9224, respectively.

Moreover, it can be seen from Table 1 that the AUC, AUPR, and accuracy scores of "Homo2 + Heter1" are higher than other combinations, but the precision, recall, and F1-score are relatively low. This would be explained from the calculation of AUC and F1-score. Generally speaking, the larger the AUC, the greater difference between true positives and true negatives. Instead, F1-score is a comprehensive consideration of precision and recall. The F1-score measure drives the model to classify positive and negative samples as much as possible, while AUC tends to reduce false positives. Our study focused on the highly reliable predictions of miRNA-drug sensitivity associations for downstream biochemical experiments. Due to the manpower and cost of wet-lab experiments, low-confidence predictions should be excluded as much as possible. Therefore, we tuned our model to achieve a high AUC so as to reduce false positives as much as possible, which leads to relatively low precision, recall, and F1-score.

Performance Comparison with Other Methods. To verify the performance of our method, we compared it with seven state-of-the-art methods for link prediction. We concisely introduced the seven methods as below:

- **GANLDA**[36] is a computational model based on the graph attention network to infer the association between lncRNAs and diseases.
- **LGCMDS**[29] used the LightGCN and retained only the neighborhood aggregation of GCN to predict miRNA-drug sensitivity associations.

Table 1. Performance of the Ablated Model by Running Feature Propagations on Homogeneous and/or Heterogeneous

model variant	AUC	AUPR	accuracy	precision	recall	F1-score
Homo × 1	0.9009	0.8684	0.8395	0.8500	0.8258	0.8372
Homo × 2	0.9013	0.8879	0.8316	0.8473	0.8091	0.8274
Homo × 3	0.8925	0.8762	0.8105	0.8094	0.8144	0.8114
Heter × 1	0.9058	0.8795	0.8471	**0.8748**	0.8115	0.8411
Heter × 2	0.9198	0.9148	0.8481	0.8680	0.8215	0.8440
Heter × 3	0.9134	0.9064	0.8409	0.8303	**0.8584**	0.8436
(Homo + Heter) × 1	0.9077	0.8797	0.8318	0.8951	0.7545	0.8174
(Homo + Heter) × 2	0.9210	0.9158	0.8495	0.8553	0.8421	**0.8483**
(Homo + Heter) × 3	0.9133	0.9102	0.8323	0.8688	0.7871	0.8237
Homo × 2 + Heter × 1	**0.9255**	**0.9224**	**0.8523**	0.8726	0.8261	0.8481
Homo × 2 + Heter × 2	0.9205	0.9154	0.8486	0.8665	0.8244	0.8448

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

- **LAGCN**[37] constructed heterogeneous networks of diseases and drugs, and predicted drug-disease associations by embedding different convolutional layers and attentional mechanisms.
- **ABHMDA**[38] employed adaptive boosting to reveal the associations between diseases and microbes by calculating the relation probability of disease−microbe pair using a robust classifier.
- **SDLDA**[39] combined singular value decomposition and deep learning techniques to predict lncRNA-disease associations.
- **DMFCDA**[40] is a circRNA-disease association prediction model that utilized a projection layer to learn representations and multilayer neural networks to capture nonlinear associations.
- **KATZMDA**[41] calculated the miRNA similarity and disease similarity based on the KATZ model and predicted miRNA-disease associations.

We evaluated these competitive methods on the benchmark dataset (for parameter settings of these methods, see Section S3 in the Supporting Information). Figure 4 shows the ROC curves and AUC values of our method DGNNMDA and seven competitive methods, from which we found that our method achieved the highest AUC value 0.9255, followed by GANLDA. Moreover, Table 2 lists other performance metrics and verified that our method outperformed other methods in the metrics AUC, AUPR, Accuracy, Recall, and F1-score except for Precision metric. Because the associations between miRNA and drug sensitivity are relatively sparse, the competitive methods failed to fit the data well or suffered from overfitting, thereby performing inferior to our method.

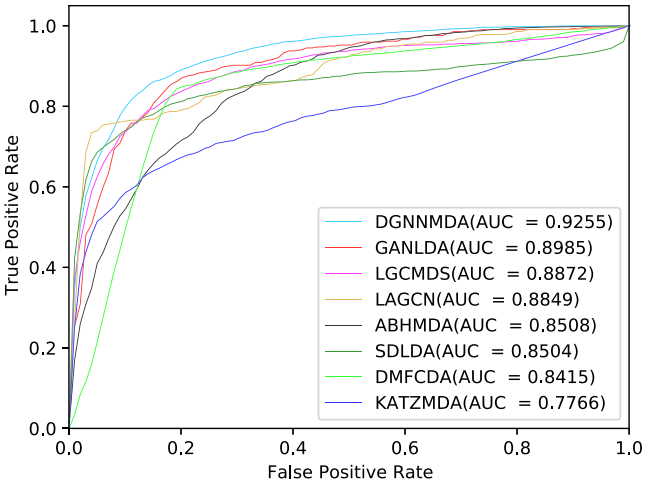


Figure 4. ROC curves and AUC values of DGNNMDA and seven competitive methods on the benchmark set by fivefold cross-validations.

Performance Evaluation on Independent Dataset. For objective evaluation, we trained our model on the benchmark dataset and then tested it on the manually curated independent set. We conducted similar experiments for other seven competitive methods to evaluate their performance. Figure 5 shows the ROC curves and AUC values of our method and other competitive methods, from which we can see that our model obtained the best performance on the independent test set by AUC value 0.82, while the other competitive methods did not reach 0.8 AUC value. These experimental results strongly demonstrated that our method has obtained better generalization ability than other existing methods.

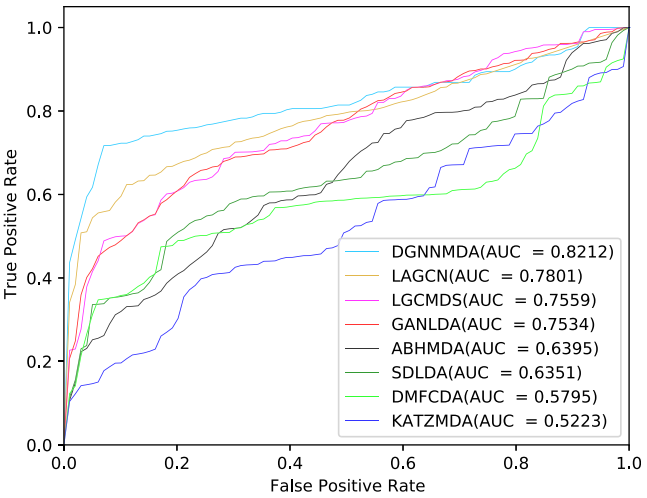


Figure 5. ROC curves and AUC values of DGNNMDA and seven competitive methods on independent test set.

Impact of Negative Samples. We were also interested in the impact of negative samples on predictive performance. For this purpose, we compared the randomly generated negative samples and the selected negative samples. Also, we tested different ratios of the number between positive and negative samples. The experimental results are shown in Table 3. We found when the ratio of positive and negative samples was equal to 1:3, our model achieved the highest performance. The reliable negative samples selection strategy improved the performance moderately, compared to random negative sample generation. The results showed that our method was robust to the selection of negative samples.

Table 2. Performance Comparison of DGNNMDA and Seven Competitive Methods on Benchmark Dataset

method	AUC	AUPR	accuracy	precision	recall	F1
DGNNMDA	**0.9255**	**0.9224**	**0.8523**	0.8726	**0.8261**	**0.8481**
GANLDA	0.8985	0.8999	0.8194	**0.8761**	0.7440	0.8047
LGCMDS	0.8872	0.9026	0.8240	0.8370	0.8049	0.8204
LAGCN	0.8849	0.9070	0.8218	0.8625	0.7656	0.8112
ABHMDA	0.8508	0.8384	0.7686	0.7456	0.8173	0.7794
SDLDA	0.8504	0.8932	0.8168	0.8697	0.7461	0.8026
DMFCDA	0.8415	0.7226	0.8180	0.8158	0.8240	0.8183
KATZMDA	0.7766	0.8226	0.7381	0.7809	0.6619	0.7163

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

Table 3. Performance Evaluation of Different Negative Sample Generation Strategies and Ratios

	negative strategy		positive vs negative ratio			
	selected	random	1:1	1:2	1:3	1:5
AUC	**0.9255**	0.9212	0.9248	0.9249	**0.9255**	0.9247
AUPR	**0.9224**	0.9143	0.9223	**0.9225**	0.9224	0.9221

Case Studies. To check the reliability of our method, we conducted case studies about two antitumor drugs Paclitaxel and Cisplatin. Paclitaxel is a cytotoxic agent proven to be effective to multiple types of tumors, especially ovarian and breast cancers.[42,43] We removed all Paclitaxel-involved miRNA associations from the training set, and then fed the remaining candidate associations into the trained model. We list the top 10 miRNAs with the highest predicted association scores in Table 4. Through PubMed literature retrieval, we found 7 miRNAs have been proven to mediate in the regulation of Paclitaxel sensitivity in various cancers.

We went further to reveal the possible functional mechanism of miRNA mediating drug sensitivity. As miRNAs induce the degradation and translational repression of their target mRNAs, we constructed a regulatory network consisting of the 10 miRNAs and their target mRNAs. From the miRDB[44] and miRWalk[45] databases, we selected the miRNA-targeted mRNAs that have been reported to affect Paclitaxel sensitivity in various tumors. The regulatory network is shown in Figure 6, in which the yellow nodes represent mRNAs, the red nodes represent the miRNAs associated with Paclitaxel sensitivity in the training set, the green nodes represent the miRNAs verified by the PubMed literature, and the blue nodes represent miRNAs without supportive evidence. Moreover, we focused on a subnetwork with 53 pairs of miRNA-mRNA associations between 9 miRNAs and 39 unique mRNAs, as shown in Figure 6B. We think that the sensitivity of Paclitaxel is affected by a number of mRNAs targeted by several miRNAs that inhibit their translation level. For example, ZEB1 inhibits the expression of sorcin by inhibiting the transcription of hsa-miR-142-5p, which is often overexpressed in tumor cells resistant to Paclitaxel.[46] The integrated stress response (ISR) plays an important role in the survival of cancer cells and is more related to the production of anticancer drug sensitivity. This mechanism involves two genes, EIF2AK3 and EIF2AK4,[47] which are regulated by hsa-miR-431-5p, hsa-miR-4443, and hsa-miR-663a.

For another drug Cisplatin, we conducted a similar study to verify most of the predicted miRNAs reported to involve in its sensitivity/resistance (for more details, see Section S4 and Table S3 in the Supporting Information).

DISCUSSION AND CONCLUSIONS

In recent years, noncoding RNAs have emerged as a new type of antiviral drugs. Especially, the nucleic acid drug has stood out as an effective treatment to fight against COVID-19.[48] However, the study that focused on the associations between drug sensitivity and miRNAs is few. Our understanding of the relationship between miRNA and drug sensitivity is far from guiding the development of miRNA-targeted drugs. Traditional biochemical experimental methods used to reveal miRNA-drug sensitivity are time-consuming, laborious, and expensive. Therefore, computational methods for the reliable prediction of miRNA-drug sensitivity can effectively promote the discovery of miRNA-mediated drug sensitivity.

We proposed a new computational method to predict miRNA-drug sensitivity association. We developed a dual-channel heterogeneous graph neural network model to learn the latent representations of miRNAs and drugs. By encoding and mapping miRNAs and drugs into the same embedding space, we ran feature aggregation on homogeneous nodes and heterogeneous neighbor nodes through multilayer graph aggregation to obtain node embeddings. Also, the integration of the miRNA similarity network and drug−drug association network into the miRNA-drug heterogeneous network can effectively alleviate the problem of sparse connections in the dataset. To evaluate the proposed method, we constructed a benchmark dataset and a manually curated independent test set. The performance comparison experiments verified that our method achieves excellent performance and outperforms six other state-of-the-art methods for link prediction. Moreover, two case studies for two drugs, Paclitaxel and Cisplatin, showed that the heterogeneous graph representation learning actually pulls close the associated miRNAs and drugs in the latent space, while pushing other miRNAs away from the drugs. This illustrated that our method obtained both expressive and interpretable features of miRNAs and drugs.

To the best of our knowledge, we are the first to highlight the importance of miRNA-mediated drug sensitivity. Also, the dataset built in this study is the largest so far.

Despite the merits of our method, it still has much room for improvement. First, for the initialization of drug latent representation, the chemical fingerprint rather than random values can be leveraged.[49,50] In addition, self-supervised learning strategy has developed rapidly. By constructing pretexts, it can learn meaningful representations from a large number of unlabeled data and improve the performance of

Table 4. Top 10 miRNAs Predicted to Mediate in Paclitaxel Sensitivity in Various Cancers

rank	miRNA	PMID	description
1	miR-508	32988253	Mir-508-3p plays a role in the anticancer effect of Paclitaxel in a variety of cancers.
2	miR-29a*	30229821	Downregulation of miR-29a* can promote the improve the sensitivity of Paclitaxel and apoptosis of colorectal cancer cells.
3	miR-663a	Unconfirmed	none
4	miR-431	Unconfirmed	none
5	miR-4443	31615089	Mir-4443 participates in the regulation of chemoresistance through drug genetic targets.
6	miR-215	26676658	Upregulation of miR-215 promotes apoptosis and increased sensitivity to the drug Paclitaxel.
7	miR-200a	25327865	After artificially upregulating mir-200a, it was found that SKOV-3 was more sensitive to Paclitaxel.
8	miR-518c-AS	Unconfirmed	none
9	miR-506-3p	34541682	Downregulation of miR-506-3p can increase the sensitivity of Paclitaxel in OC cells.
10	miR-31-5p	23552883	MiR-31 will bind to MET gene, resulting in the decrease of met so as to improve the sensitivity of Paclitaxel.

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937

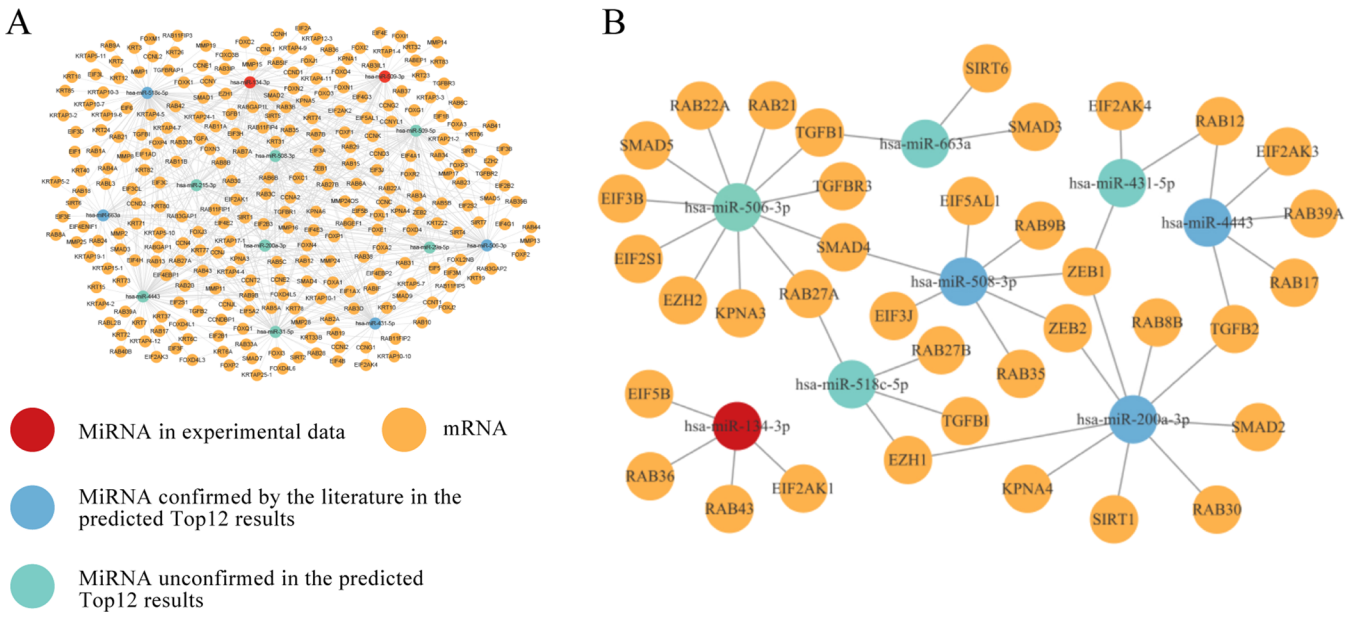


Figure 6. Regulatory network of top 10 predicted Paclitaxel-associated miRNAs and their target mRNAs.

downstream tasks. We believe that self-supervised learning can be incorporated into our learning framework to promote performance.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01060.

Levenshtein distance and details in the construction of similarity networks (Section S1); display of all experimental results (Section S2); parameter settings of the comparison methods (Section S3); case study about Cisplatin (Section S4); and more case studies (Section S5) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Hui Liu − *School of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China;* orcid.org/0000-0001-7158-913X; Email: hliu@njtech.edu.cn

Authors

Lei Deng − *School of Computer Science and Engineering, Central South University, Changsha 410083, China*

Ziyu Fan − *School of Computer Science and Engineering, Central South University, Changsha 410083, China*

Xiaojun Xiao − *Software School, Xinjiang University, Urumqi 830091, China*

Jiaxuan Zhang − *Department of Electrical and Computer Engineering, University of California, San Diego, San Diego, California 92161, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c01060

Author Contributions

Z.F. and L.D. conceived the main idea. X.L. helped to improve the idea. Z.F., X.X., and J.Z. conducted the experiments. Z.F. and H.L. wrote the manuscript. H.L. reviewed and revised the manuscript.

Notes

The authors declare no competing financial interest.

The source code and datasets are freely available at https://github.com/19990915fzy/DGNNMDA.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under grants nos. 62072058 and 61972422.

■ REFERENCES

(1) Vasan, N.; Baselga, J.; Hyman, D. M. A view on drug resistance in cancer. *Nature* **2019**, *575*, 299−309.

(2) Mansoori, B.; Ali, M.; Sadaf, D.; Solmaz, S.; Behzad, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv. Pharm. Bull.* **2017**, *7*, 339−348.

(3) Wang, X.; Zhang, H.; Chen, X.; Resistance, C. D. Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist.* **2019**, *2*, 141−160.

(4) Si, W.; Shen, J.; Zheng, H.; Fan, W. The role and mechanisms of action of microRNAs in cancer drug resistance. *Clin. Epigenet.* **2019**, *11*, 1−24.

(5) Garzon, R.; Calin, G. A.; Croce, C. M. MicroRNAs in Cancer. *Annu. Rev. Med.* **2009**, *60*, 167−179.

(6) Acunzo, M.; Romano, G.; Wernicke, D.; Croce, C. M. MicroRNA and cancer−a brief overview. *Adv. Biol. Regul.* **2015**, *57*, 1−9.

(7) Su, X.; Xing, J.; Wang, Z.; Lei, C.; Jiang, B. microRNAs and ceRNAs: RNA networks in pathogenesis of cancer. *Chin. J. Cancer Res.* **2022**, *25*, 235−239.

(8) Enerly, E.; Steinfeld, I.; Kleivi, K.; Leivonen, S. K.; Aure, M. R.; Russnes, H. G.; Ronneberg, J.; Johnsen, H.; Navon, R.; Rodland, E. miRNA-mRNA Integrated Analysis Reveals Roles for miRNAs in Primary Breast Tumors. *PLoS One* **2015**, *6*, No. e16915.

(9) Tonevitsky, A. G.; Maltseva, D. V.; Abbasi, A.; Samatov, T. R.; Sakharov, D. A.; Shkurnikov, M. U.; Lebedev, A. E.; Galatenko, V. V.; Grigoriev, A. I.; Northoff, H. Dynamically regulated miRNA-mRNA networks revealed by exercise. *BMC Physiol.* **2013**, *13*, 1−11.

(10) Wu, C.; Zhao, Y.; Liu, Y.; Yang, X.; Yan, M.; Min, Y.; Pan, Z.; Qiu, S.; Xia, S.; Yu, J.; et al. Identifying mirna-mrna regulation

network of major depressive disorder in ovarian cancer patients. *Oncol. Lett.* **2018**, *16*, 5375−5382.
(11) Bai, Y.; Baker, S.; Exoo, K.; Dai, X.; Ding, L.; Khattak, N.; Li, H.; Liu, H.; Liu, X. MMiRNA-Viewer2, a bioinformatics tool for visualizing functional annotation for MiRNA and MRNA pairs in a network. *BMC Bioinf.* **2020**, *21*, 1−10.
(12) Wang, N.; Li, Y.; Liu, S.; Gao, L.; Liu, C.; Bao, X.; Xue, P. Analysis and Validation of Differentially Expressed MicroRNAs with their Target Genes Involved in GLP-1RA Facilitated Osteogenesis. *Curr. Bioinf.* **2021**, *16*, 928−942.
(13) Geraci, F.; Manzini, G. EZcount: An all-in-one software for microRNA expression quantification from NGS sequencing data. *Comput. Biol. Med.* **2021**, *133*, No. 104352.
(14) Li, J.; Liu, L.; Cui, Q.; Zhou, Y. Comparisons of MicroRNA Set Enrichment Analysis Tools on Cancer De-regulated miRNAs from TCGA Expression Datasets. *Curr. Bioinf.* **2021**, *15*, 1104−1112.
(15) Ma, J.; Dong, C.; Ji, C. MicroRNA and drug resistance. *Cancer Gene Ther.* **2010**, *17*, 523−531.
(16) Rajarajan, D.; Kaur, B.; Penta, D.; Natesh, J.; Meeran, S. miR-145−5p as a predictive biomarker for breast cancer stemness by computational clinical investigation. *Comput. Biol. Med.* **2021**, *135*, No. 104601.
(17) Li, M.; Gao, M.; Xie, X.; Zhang, Y.; Gu, K.; et al. MicroRNA-200c reverses drug resistance of human gastric cancer cells by targeting regulation of the NER-ERCC3/4 pathway. *Oncol. Lett.* **2019**, *18*, 145−152.
(18) Ye, P.; Fang, C.; Zeng, H.; Shi, Y.; Pan, Z.; An, N.; He, K.; Zhang, L.; Long, X. Differential microRNA expression profiles in tamoxifen-resistant human breast cancer cell lines induced by two methods. *Oncol. Lett.* **2018**, *15*, 3532−3539.
(19) Shen, L.; Liu, F.; Huang, L.; Liu, G.; Zhou, L.; Peng, L. VDA-RWLRLS: An anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Comput. Biol. Med.* **2021**, *140*, No. 105119.
(20) Meng, F.; Henson, R.; Lang, M.; Wehbe, H.; Maheshwari, S.; Mendell, J. T.; Jiang, J.; Schmittgen, T. D.; Patel, T. Involvement of Human Micro-RNA in Growth and Response to Chemotherapy in Human Cholangiocarcinoma Cell Lines. *Gastroenterology* **2006**, *130*, 2113−2129.
(21) Yu, Q.; Xu, X. P.; Yin, X. M.; Peng, X. Q. miR-155-5p increases the sensitivity of liver cancer cells to adriamycin by regulating ATG5-mediated autophagy. *Neoplasma* **2021**, *68*, 87−95.
(22) Ao, X.; Nie, P.; Wu, B.; Xu, W.; Zhang, T.; Wang, S.; Chang, H.; Zou, Z. Decreased expression of microRNA-17 and microRNA-20b promotes breast cancer resistance to taxol therapy by upregulation of NCOA3. *Cell Death Dis.* **2016**, *7*, No. e2463.
(23) Li, L.; Wu, P.; Wang, Z.; Meng, X.; Zha, C.; Li, Z.; Qi, T.; Zhang, Y.; Han, B.; Li, S.; et al. NoncoRNA: a database of experimentally supported non-coding RNAs and drug targets in cancer. *J. Hematol. Oncol.* **2020**, *13*, 1−4.
(24) Dai, E.; Yang, F.; Wang, J.; Zhou, X.; Song, Q.; An, W.; Wang, L.; Jiang, W. ncDR: a comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics* **2017**, *33*, 4010−4011.
(25) Huang, Y. A.; Hu, P.; Chan, K.; You, Z. H. Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* **2019**, *36*, 851−858.
(26) Niu, Y.; Song, C.; Gong, Y.; You, Z. H. MiRNA-Drug Resistance Association Prediction Through the Attentive Multimodal Graph Convolutional Network. *Bioinformatics* **2021**, *12*, 799108.
(27) Xu, P.; Wu, Q.; Rao, Y.; Kou, Z.; Han, H.; et al. Predicting the Influence of MicroRNAs on Drug Therapeutic Effects by Random Walking. *IEEE Access* **2020**, *8*, 117347−117353.
(28) Li, Y.; Runqi, W.; Shuo, Z.; Hanlin, X.; Lei, D. LRGCPND: Predicting Associations between ncRNA and Drug Resistance via Linear Residual Graph Convolution. *Int. J. Mol. Sci.* **2021**, *22*, 10508.
(29) Song, Y.; Hanlin, X.; Yizhan, L.; Dayun, L.; Lei, D. In *LGCMDS: Predicting miRNADrug Sensitivity based on Light Graph Convolution Network*, 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2021; pp 217−222.
(30) Griffiths-Jones, S.; Grocock, R. J.; Stijn, V.; Alex, B.; Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **2006**, *34*, 140−144.
(31) Navarro, G. A Guided Tour to Approximate String Matching. *ACM Comput. Surv.* **2001**, *33*, 31−88.
(32) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102−D1109.
(33) Yang, C.; Wang, P.; Tan, J.; Liu, Q.; Li, X. Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks. *Comput. Biol. Med.* **2021**, *139*, No. 104963.
(34) Liu, H.; Sun, J.; Guan, J.; Zheng, J.; Zhou, S. Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics* **2015**, *31*, i221−i229.
(35) Yang, Y.; Chen, L. Identification of Drug-Disease Associations by Using Multiple Drug and Disease Networks. *Curr. Bioinf.* **2022**, *17*, 48−59.
(36) Lan, W.; Wu, X.; Chen, Q.; Peng, W.; Wang, J.; Chen, Y. GANLDA: Graph attention network for lncRNA-disease associations prediction. *Neurocomputing* **2021**, *469*, 384−393.
(37) Yu, Z.; Huang, F.; Zhao, X.; Xiao, W.; Zhang, W. Predicting drug−disease associations through layer attention graph convolutional network. *Briefings Bioinf.* **2020**, *43*, No. bbaa243.
(38) Peng, L. H.; Yin, J.; Zhou, L.; Liu, M. X.; Yan, Z. Human Microbe-Disease Association Prediction Based on Adaptive Boosting. *Front. Microbiol.* **2018**, *9*, 2440.
(39) Zeng, M.; Lu, C.; Zhang, F.; Li, Y.; Li, M.; et al. SDLDA: lncRNA−disease association prediction based on singular value decomposition and deep learning. *Methods* **2020**, *179*, 73−80.
(40) Lu, C.; Zeng, M.; Zhang, F.; Wu, F.; Li, M.; Wang, J. Deep matrix factorization improves prediction of human circRNA-disease associations. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 891−899.
(41) Qu, Y.; Zhang, H.; Liang, C.; Dong, X. KATZMDA: Prediction of miRNA-disease associations based on KATZ model. *IEEE Access* **2018**, *6*, 3943−3950.
(42) Riedl, J. M.; Posch, F.; Horvath, L.; Gantschnigg, A.; Gerger, A.; et al. Gemcitabine/nab-Paclitaxel versus FOLFIRINOX for palliative first-line treatment of advanced pancreatic cancer: A propensity score analysis. *Eur. J. Cancer* **2021**, *151*, 3−13.
(43) Blomstrand, H.; Batra, A.; Cheung, W. Y.; Elander, N. O. Real-world evidence on first- and second-line palliative chemotherapy in advanced pancreatic cancer. *World J. Clin. Oncol.* **2021**, *12*, 787.
(44) Chen, Y.; Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **2020**, *48*, D127−D131.
(45) Sticht, C.; Carolina, D.; Parveen, A.; Gretz, N.; Campbell, M. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One* **2018**, *13*, No. e0206239.
(46) Zhang, J.; Guan, W.; Xu, X.; Wang, F.; Li, X.; Xu, G. A novel homeostatic loop of sorcin drives paclitaxel-resistance and malignant progression via Smad4/ZEB1/miR-142-5p in human ovarian cancer. *Oncogene* **2021**, *40*, 4906−4918.
(47) Chen, L.; He, J.; Zhou, J.; Xiao, Z.; Ding, N.; Duan, Y.; Li, W.; Sun, L. EIF2A promotes cell survival during paclitaxel treatment in vitro and in vivo. *J. Cell. Mol. Med.* **2019**, *23*, 6060−6071.
(48) Le, T. T.; Andreadakis, Z.; Kumar, A.; Román, R. G.; Tollefsen, S.; Saville, M.; Mayhew, S. The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discovery* **2020**, *19*, 305−306.
(49) Yu, C.; Yongshun, G.; Yuansheng, L.; Bosheng, S.; Quan, Z. Molecular design in drug discovery: a comprehensive review of deep generative models. *Briefings Bioinf.* **2021**, *4*, No. bbab34.
(50) Ru, X.; Ye, X.; Sakurai, T.; Zou, Q.; Xu, L.; Lin, C. Current status and future prospects of drug−target interaction prediction. *Briefings Funct. Genomics* **2021**, *20*, 312−322.

https://doi.org/10.1021/acs.jcim.2c01060
J. Chem. Inf. Model. 2022, 62, 5929−5937