

RESEARCH

Open Access



Inferring circRNA-drug sensitivity associations via dual hierarchical attention networks and multiple kernel fusion

Shanghai Lu^{1,2}, Yong Liang^{1,3*}, Le Li¹, Shuilin Liao¹, Yongfu Zou², Chengjun Yang⁴ and Dong Ouyang¹

Abstract

Increasing evidence has shown that the expression of circular RNAs (circRNAs) can affect the drug sensitivity of cells and significantly influence drug efficacy. Therefore, research into the relationships between circRNAs and drugs can be of great significance in increasing the comprehension of circRNAs function, as well as contributing to the discovery of new drugs and the repurposing of existing drugs. However, it is time-consuming and costly to validate the function of circRNA with traditional medical research methods. Therefore, the development of efficient and accurate computational models that can assist in discovering the potential interactions between circRNAs and drugs is urgently needed. In this study, a novel method is proposed, called DHANMKF, that aims to predict potential circRNA-drug sensitivity interactions for further biomedical screening and validation. Firstly, multimodal networks were constructed by DHANMKF using multiple sources of information on circRNAs and drugs. Secondly, comprehensive intra-type and inter-type node representations were learned using bi-typed multi-relational heterogeneous graphs, which are attention-based encoders utilizing a hierarchical process. Thirdly, the multi-kernel fusion method was used to fuse intra-type embedding and inter-type embedding. Finally, the Dual Laplacian Regularized Least Squares method (DLapRLS) was used to predict the potential circRNA-drug sensitivity associations using the combined kernel in circRNA and drug spaces. Compared with the other methods, DHANMKF obtained the highest AUC value on two datasets. Code is available at <https://github.com/cuntjx/DHANMKF>.

Keywords circRNA-drug sensitivity associations, Multimodal networks, Bi-typed multi-relational heterogeneous graphs, Attention mechanism, Multi-kernel fusion

Introduction

Circular RNA (circRNA) is a unique type of RNA that differs from other RNAs in that it forms a covalently closed loop and is typically considered non-coding. With the advancement of high-throughput genomics technology, circRNA has become a hot topic in RNA biology research [1]. Since the discovery of the first circRNA in RNA viruses in the 1970s [2], the advancement of biomedical technology has resulted in the discovery of an increasing amount of circRNAs. However, research into circRNA function has progressed very slowly over several decades, until 2013, when Memczak et al. and Hansen et al. proved that the circular RNA of human cerebellar

*Correspondence:

Yong Liang
yongliangresearch@gmail.com

¹ Faculty of Innovation Enginee, Macau University of Science and Technology, Avenida Wai Long, Taipa, 999078 Macao, Macao Special Administrative Region of China, China

² School of Mathematics and Physics, Hechi University, No.42, Longjiang, 546300 Guangxi, China

³ Peng Cheng Laboratory, Shenzhen, 518055 Guangdong, China

⁴ School of Artificial Intelligence and Manufacturing, Hechi University, No.42, Longjiang, 546300 Guangxi, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

degeneration-related protein has an important function in neural development [3, 4]. This discovery led to a great increase in the study of circRNA function. The most notable function of circRNAs is that they act as miRNA sponges, which regulates target gene expression by inhibiting miRNA activity. One circRNA can regulate one or multiple miRNAs through multiple miRNA binding sites in a circular sequence [5]. Previous studies have found that circRNA can regulate alternative splicing or transcription [6, 7], as well as parental gene expression [8, 9]. The results of these studies have also indicated that circRNA plays an important role in physiological and pathological processes, and that the dysregulation of circRNA is closely related to many human diseases [10]. Over the past two decades, several verified biological function experiments have shown that circRNA has potential as a new clinical diagnostic marker.

Over the years, an increasing number of studies have demonstrated that circRNA can significantly affect the drug sensitivity of cells. For example, Gao et al. [11] screened 18 circRNAs from 3093 circRNAs and then verified them in real-time by quantitative reverse transcription PCR. Finally, hsa_circ_0006528 was found to play an important role in chemotherapy resistance in breast cancer patients. Peng et al. [12] first used next-generation sequencing (NGS) technology to identify the comprehensive circRNA expression profile of multi-drug-resistant osteosarcoma(OS) cell lines and found that hsa_circ_0004674 was significantly elevated in OS-resistant cells and tissues, and was associated with poor prognosis. This was then verified by quantitative real-time PCR (qRT-PCR). A study by Wu et al. [13] found that hsa_circ_0001546 is decreased in gastric cancer, which is associated with poor prognosis and also inhibits drug resistance via the ATM/Chk2/p53-dependent pathway. Ruan et al. [14] used four identification algorithms to describe the expression profile of circRNA in approximately 1000 human cancer cell lines and observed a strong correlation between circRNA expression and drug response. That study systematically demonstrated the effect of circRNAs on drug sensitivity. However, research into the relationship between circRNA and drug sensitivity is a newly emerging field that has developed rapidly over the past decade, so our understanding of this relationship is still in its early stages.

The process of validating the relationships between circRNA and drug sensitivity using traditional biomedical methods is time-consuming and costly. Therefore, some researchers have developed computational models that can help to reveal the potential relationships between circRNA and drugs. For example, Deng et al. [15] proposed a computational model called GATECDA for predicting the association between circRNA and drug

sensitivity. GATECDA is based on the Graph Attention Auto-encoder(GATE) [16]. First, sequence information data for circRNAs, structural data for drugs, and circRNA-drug sensitivity association data were collected. Then the similarity between circRNAs and drugs were each calculated and these data as well as circRNA-drug sensitivity association data were input into the GATE, in order to generate low-dimensional vector representations of circRNA and drug nodes. Finally, the low-dimensional vector representations generated by the GATE were input into a fully connected neural network for circRNA-drug sensitivity association prediction. Later, Yang et al. [17] proposed a model called MNGACDA. The model constructs a multimodal network based on multiple information sources on circRNAs and drugs. Then, a node-level attention Graph Auto-Encoder was used to obtain low-dimensional embeddings of circRNAs and drugs from the multimodal network. Finally, the low-dimensional embeddings of circRNAs and drugs were input into an inner product decoder to score the association between circRNAs and drug sensitivity. To our knowledge, these were the first models to apply computational methods to predict the potential association between circRNAs and diseases. Thus far, no other new models have been applied in this field, and considerable advancement in the creation of new and improved models for this field of research is much needed.

Since Multiple Kernel Learning (MKL) [18] was proposed, it has been widely applied to bipartite biological networks for the improvement of model performance. Specifically, the information contained in the samples were used by MKL to compute the multiple kernel matrix, and then the optimal kernel matrix was obtained by fusing multiple kernel matrices. For example, MKGCN, which is based on MKL and GCN [19], was proposed by Yang et al. to infer novel microbe-drug associations. Yan et al. [20] proposed a computational methods called MKLC-BiRW, based on MKL and Bi-random walk algorithm, to predict potential drug-target interactions by integrating diverse drug-related and target-related heterogeneous information.

In this study, we propose a novel method, called DHANMKF, that aims to predict potential circRNA-drug sensitivity associations for further biomedical screening and validation. Firstly, multimodal networks were constructed by DHANMKF using multiple sources of information on circRNAs and drugs. Secondly, comprehensive intra-type and inter-type node representations were learned using multi-relational heterogeneous graphs, which are attention-based encoders under a hierarchical process. Thirdly, a multi-kernel fusion method was used to fuse intra-type embedding and inter-type embedding. Fourthly, the Dual Laplacian Regularized Least Squares

(DLapRLS) method was used to predict the potential circRNA-drug sensitivity associations by the combined kernel in circRNA and drug spaces. In order to evaluate the effectiveness of DHANMKF, it was compared with six state-of-the-art methods on a benchmark data set under 5-fold cross-validations (5-CV). Compared with the other methods, DHANMKF obtained the highest AUC. Furthermore, an ablation study was performed to compare the experimental results from different perspectives. Finally, case studies were conducted to demonstrate that the DHANMKF model can be a useful tool for helping with the study of circRNA-drug sensitivity associations in real situations. To the best of our knowledge, DHANMKF is the first algorithm to use dual hierarchical attention networks for the prediction of circRNA-drug sensitivity associations. Our main contributions, differing from previous approaches, are summarized as follows: (1) We classify nodes into two types, i.e., head nodes and tail nodes, based on the degree of the nodes, and then define the types of edges based on the associations between different kinds of nodes. (2) Based on the differences in types of edges, we use dual hierarchical attention networks to extract the information on circRNAs and drugs, use the multi-kernel fusion method to fuse this information, and then use the dual graph regularized least squares method to predict potential circRNA-drug associations. (3) We tested DHANMKF on two datasets, and the results show that multi-relational dual hierarchical attention networks perform better than the other methods in predicting potential circRNA-drug associations. These results can provide new insights for further research on circRNA-drug associations.

Materials and methods

Datasets

Two datasets, data271 and data251, were used in this study. Data271 is from Deng et al. [15] and data251 is from Deng et al. [15] and Peng et al. [21]. CircRNA-drug sensitivity associations were collected from the circRc database [14] by Deng et al. [15], where drug sensitivity data were obtained from the GDSC database [22]. After Wilcoxon tests with a false discovery rate < 0.05 , these significant circRNA-drug sensitivity associations were extracted as the data271 dataset, which contains $N_c = 271$ circRNAs, $N_d = 218$ drugs and 4134 circRNA-drug sensitivity associations. Integrating with the dataset of Peng et al. [21], we removed circRNAs with host-gene interaction scores ≤ 0.5 and nodes with a degree of 0. This resulted in the data251 dataset, containing $N_c = 251$ circRNAs, $N_d = 217$ drugs, and 3635 circRNA-drug sensitivity associations. Additional information on these two datasets can be found in the [Supplementary file](#). In our experiment, circRNAs and

drugs were represented as two different types of nodes in the network. The node set of N_c circRNAs was defined as $\mathbf{C} = \{c_1, \dots, c_{N_c}\}$. Similarly, the node set of N_d drugs was described as $\mathbf{D} = \{d_1, \dots, d_{N_d}\}$. An adjacency matrix $\mathbf{Y} \in R^{N_c \times N_d}$ was created for the storage of circRNA-drug associations. In this matrix, N_c rows represent the number of circRNAs and N_d columns represent the number of drugs. If circRNA $c_i (1 \leq i \leq N_c)$ is associated with drug $d_j (1 \leq j \leq N_d)$, $\mathbf{Y}_{ij} = 1$, otherwise $\mathbf{Y}_{ij} = 0$. During the training phase, all the $\mathbf{Y}_{ij} = 1$ are treated as positive samples and the others are treated as negative samples. We randomly masked some positive samples from \mathbf{Y} to get $\mathbf{Y}_{\text{train}}$. In order to calculate the similarity of circRNAs and drugs, the host gene sequences of circRNAs were downloaded from the National Center for Biotechnology Information (NCBI) gene database [23] and the drug structure data were downloaded from NCBI's PubChem database [24].

Sequence similarity of host genes of circRNAs

Applying methods similar to those of Deng et al. [15] and Yang et al. [17], we treated the sequence similarity between host genes of circRNAs as the similarity between circRNAs. In this way, the similarity calculation between circRNAs became the sequence similarity calculation between host genes of circRNAs. The sequence similarity between host genes of circRNAs was calculated based on the sequence Levenshtein distance, which was obtained using the ratio function of Python's Levenshtein package. A similarity matrix $\mathbf{CSS} \in R^{N_c \times N_c}$ was created for storing the circRNA sequence similarity.

Structural similarity of drugs

The structure of drugs has a great impact on their function. Therefore, it has become a common practice to measure the similarity of drugs based on their structure. As in previous studies [25, 26], RDKit [27] toolkit and the Tanimoto method were used to calculate the structural similarities between drugs. The specific process was as follows: first, the structural data on several drugs were obtained from the PubChem database. Then, RDKit was used to calculate the topological fingerprint of each drug. After that, the structural similarity between drugs was calculated using the Tanimoto method. Finally, the drugs structural similarity matrix $\mathbf{DSS} \in R^{N_d \times N_d}$ was derived.

Gaussian interaction profile kernel similarity for circRNAs and drugs

The Gaussian Interaction Profile (GIP) kernel similarity [28] algorithm is a collaborative filtering algorithm that has been widely used in previous studies for similarity calculation [29, 30], and it helps to obtain topological information on circRNAs and drugs in relational

graphs. Therefore, we calculated the GIP kernel similarity for circRNAs and drugs using the circRNA-drug association network. Firstly, based on the assumption that similar circRNAs are more likely to be associated with similar drugs, we utilized a binary vector $\mathbf{BI}(c_i)$, which is the i th row of the $\mathbf{Y}_{\text{train}}$ matrix, representing the associations between circRNAs c_i and all drugs in the training matrix of \mathbf{Y} . Then, the GIP kernel similarity for circRNAs $\mathbf{CGS}(c_i, c_j)$ between circRNA c_i and c_j was calculated as below:

$$\mathbf{CGS}(c_i, c_j) = \exp\left(-\gamma_c \|\mathbf{BI}(c_i) - \mathbf{BI}(c_j)\|^2\right), \quad (1)$$

$$\gamma_c = \alpha_c / \left(\frac{1}{N_c} \sum_{i=1}^{N_c} \|\mathbf{BI}(c_i)\|^2 \right). \quad (2)$$

Here, α_c has been set to 1 referring to [28]’s studies. And similarly, we calculated the GIP of drug $\mathbf{DGS}(d_i, d_j)$ between drugs d_i and d_j as follows:

$$\mathbf{DGS}(d_i, d_j) = \exp\left(-\gamma_d \|\mathbf{BI}(d_i) - \mathbf{BI}(d_j)\|^2\right), \quad (3)$$

$$\gamma_d = \alpha_d / \left(\frac{1}{N_d} \sum_{i=1}^{N_d} \|\mathbf{BI}(d_i)\|^2 \right). \quad (4)$$

Here, the binary vector $\mathbf{BI}(d_i)$ is the i th column of the $\mathbf{Y}_{\text{train}}$ matrix, representing the associations between drugs d_i and all circRNAs in the training matrix of \mathbf{Y} . α_d has been set to 1 referring to [28] studies.

Integrated similarity for circRNAs and drugs

Inspired by the study of Wang et al. [31], we used a non-linear fusion method to integrate the circRNA similarity and the drug similarity. With circRNA similarity, for example, we first normalized the sequence similarity of host genes of circRNAs using the following formula:

$$\mathbf{CSS}'(i, j) = \frac{\mathbf{CSS}(i, j)}{\sum_j \mathbf{CSS}(i, j)}. \quad (5)$$

Then, the K Nearest Neighbors (KNN) algorithm was used to measure \mathbf{CSS} ’s local affinity as follows:

$$\mathbf{CKNN1} = \begin{cases} \frac{\mathbf{CSS}(i, j)}{\sum_{k \in N'_i} \mathbf{CSS}(i, k)}, & j \in N'_i \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

N'_i in Eq. (6) is the set of KNN of c_i , including c_i in \mathbf{CSS} . This operation is based on the assumption that the higher the local similarity, the more reliable it is. Therefore, the near-end similarity is high while the far-end similarity

is set to 0. Similarly, we repeated the process for \mathbf{CGS} and then we obtained \mathbf{CGS}' and $\mathbf{CKNN2}$. After that, we updated the similarity matrix for each kind of data as follows:

$$\mathbf{CSS}'^{(t+1)} = \mathbf{CKNN1} \times \mathbf{CGS}'^{(t)} \times (\mathbf{CKNN1})^T. \quad (7)$$

$$\mathbf{CGS}'^{(t+1)} = \mathbf{CKNN2} \times \mathbf{CSS}'^{(t)} \times (\mathbf{CKNN2})^T. \quad (8)$$

After each iteration, $\mathbf{CSS}'^{(t+1)}$ is normalized by formula Eq. (5). Similarly, $\mathbf{CGS}'^{(t+1)}$ performs the same normalization. The iteration does not stop until the convergence condition is met, and the convergence condition is met when the relative change in $\frac{\|\mathbf{CSS}'^{(t+1)} - \mathbf{CSS}'^{(t)}\|}{\|\mathbf{CSS}'^{(t)}\|}$ and $\frac{\|\mathbf{CGS}'^{(t+1)} - \mathbf{CGS}'^{(t)}\|}{\|\mathbf{CGS}'^{(t)}\|}$ is less than 10^{-6} . Assuming that the process involves t iterations, the overall comprehensive similarity matrix of circRNA can be obtained by Eq. (9) when the iteration ends.

$$\mathbf{S}_c = \frac{\mathbf{CSS}'^{(t)} + \mathbf{CGS}'^{(t)}}{2}. \quad (9)$$

Based on these rules, the similarity matrix \mathbf{S}_c is an asymmetry matrix. Therefore, we calculated the $\mathbf{S}_c = \frac{\mathbf{S}_c + \mathbf{S}_c^T}{2}$ as the circRNA comprehensive similarity matrix. For drugs, we applied the same rules to \mathbf{DSS} and \mathbf{DGS} , then we obtained the comprehensive drug similarity matrix \mathbf{S}_d .

DHANMKF

Dual Hierarchical Attention Networks (DHAN) were proposed by Zhao et al. [32] in 2022. Comprehensive node representations are learned with intra-type and inter-type attention-based encoders using a hierarchical process based on the bi-typed multi-relational heterogeneous graphs in DHAN. Specifically, DHAN uses two encoders, one to aggregate information on nodes of the same type and the other to aggregate node representations of different type neighbors. Then, the complex structure of the bi-typed multi-relational heterogeneous graph is captured by the model by a hierarchical process and dual-level attention operation. It is worth noting that the association matrix \mathbf{Y} of circRNA-drug is a bi-typed single-relation heterogeneous graph. Therefore, in order to fully utilize the extraction ability of DHAN for node embedding, it is necessary to classify the relationships between nodes.

It is well-known that the adjacency matrix describing different objects in the biomedical field is sparse. This means that there are many nodes with small degrees. Histograms of the degree distributions of circRNAs and drugs can be found in the [Supplementary file](#). It can be seen that most of the nodes have small degrees regardless

of whether they are circRNA nodes or drug nodes. Intuitively, the biomedical significance of a drug being associated with only a few circRNAs or a drug being associated with many circRNAs are different. Inspired by Liu et al. [33], we categorized the nodes into head nodes and tail nodes according to the value of their degrees. That is, for every node $v \in V$, where V is the set of nodes in a graph. N_v is denoted by the set of neighboring nodes of v , and the number of elements in set N_v is defined as the degree of v . Here we let V_h and V_t denote the set of head and tail nodes, respectively. For some threshold K , we define tail nodes as nodes with a degree not exceeding K , i.e., $V_t = \{v : |N_v| \leq K\}$, whereas head nodes are the complement of tail nodes, that is, $V_h = \{v : |N_v| > K\}$. K is treated as a hyperparameter in our study. In this way, the association of circRNAs with drugs in dataset data271 changes from being one type to being the following four types.

1. Association between the head node of circRNA and the head node of the drug.
2. Association between the head node of circRNA and the tail node of the drug.
3. Association between the tail node of circRNA and the head node of the drug.
4. Association between the tail node of circRNA and the tail node of the drug.

Because the node types in the circRNA similarity network and the drug similarity network are the same, that is, they are either all circRNA or all drugs, the following three types of associations will be in these two similarity networks.

1. Associations between the head nodes.
2. Associations between head node and tail node.
3. Associations between tail nodes.

In summary, we represent intra-type relationships and inter-type relationships as $R_{intra} = \{1, 2, 3\}$ and $R_{inter} = \{1, 2, 3, 4\}$, respectively. Whereas in the data251 dataset, circRNAs were split into two types depending on whether the host gene of the circRNA was associated with a disease or not, which is analogous to splitting circRNAs into head nodes and tail nodes. Thus the same number of edge types can also be obtained, and the definition is more biologically meaningful in this way.

Intra-type attention-based encoder

After the computational process above, the associated network of circRNA and drug becomes a bi-type multi-relationship heterogeneous network, given a

node pair $(n_i, n_j) \in C$ that are connected via node intra-type relationship $\Phi_k \in R_{intra}^{(c)} = \{1, 2, 3\}$. Firstly, we initialized the representation matrix of circRNA to $H_0^c = S_c Y_{train} W_{intra}^c = [h_1^T, \dots, h_{N_c}^T]^T$. Where $W_{intra}^c \in R^{N_d \times d'}$ is a learnable parameter $H_0^c \in R^{N_c \times d'}$, and $h_i \in R^{d'} (1 \leq i \leq N_c)$ is the feature vector of the node n_i . Secondly, self-attention was performed on the circRNA nodes to formulate the importance $e_{ij}^{\Phi_k}$ of a specific-relation based node pair (n_i, n_j) as follows:

$$e_{ij}^{\Phi_k} = att_{local}(h_i, h_j; \Phi_k) = LeakyRelu(a_{\Phi_k^T} \cdot [h_i || h_j]). \tag{10}$$

Where $||$ denotes the concatenate operation, and $a_{\Phi_k^T} \in R^{2d' \times 1}$ denotes the shared node-level attention weight vector under relation Φ_k . LeakyRelu is the non-linearity activation function, which is widely used in attention-based neural networks. In the third step, $e_{ij}^{\Phi_k}$ is standardized using the Eq. (11) to facilitate comparison of importance between different nodes.

$$\alpha_{ij}^{\Phi_k} = softmax_j(e_{ij}^{\Phi_k}) = \frac{exp(e_{ij}^{\Phi_k})}{\sum_{n_p \in N_{intra}^{\Phi_k}(n_i)} exp(e_{ij}^{\Phi_k})} \tag{11}$$

Where $N_{intra}^{\Phi_k}(n_i)$ denotes specific relation-based neighbors of n_i , the embedding $h_i^{\Phi_k}$ of node n_i under given relation Φ_k is obtained as follows:

$$h_i^{\Phi_k} = LeakyRelu \left(Norm_{\Phi_k} \left(\sum_{n_j \in N_{intra}^{\Phi_k}(n_i)} \alpha_{ij}^{\Phi_k} \cdot h_j \right) \right). \tag{12}$$

Where $Norm_{\Phi_k}$ denotes the relation-specific layer normalization operation; $h_i^{\Phi_k}$ is semantic-specific. Therefore, by using Eq. (12) to fuse the aggregated information of nodes with different specific relations, more comprehensive node embeddings can be obtained as follows:

$$g_i^{\Phi_k} = \mathbf{q}^T (h_i || h_i^{\Phi_k}). \tag{13}$$

Where $\mathbf{q} \in R^{2d' \times 1}$ is a trainable parameter. Similar to Eq. (11), we standardize $g_i^{\Phi_k}$ by using the softmax function as follows:

$$\beta_{ij}^{\Phi_k} = softmax_k(g_i^{\Phi_k}) = \frac{exp(g_i^{\Phi_k})}{\sum_{\Phi_l \in N_{intra}^{(c)}} exp(g_i^{\Phi_l})}. \tag{14}$$

Here $\beta_{ij}^{\Phi_k}$ is used to measure the local importance of intra-relation Φ_k . Finally, the intra-type attention-based representation of circRNA node n_i can be obtained as follows:

$$\mathbf{z}_i^c = \sum_{\Phi_l \in N_{intra}^{(c)}} \left(t\beta_G^{\Phi_l} + (1-t)\beta_i^{\Phi_l} \right) \cdot \mathbf{h}_i^{\Phi_l}. \tag{15}$$

Where $\mathbf{z}_i^c \in R^{d' \times 1}$, $\beta_G^{\Phi_l}$ denotes how important intra-type Φ_l is for all circRNA nodes and can be regarded as a global importance parameter. The global and local importance of the intra-type relationship Φ_l is smoothed by the parameter t . Both $\beta_G^{\Phi_l}$ and t can be learned from training. The aggregated information for node n_i under intra-type relation Φ_l is represented by $\mathbf{h}_i^{\Phi_l}$. Initialize the representation matrix of drug to $\mathbf{H}_0^d = \mathbf{S}_d \mathbf{Y}_{train}^T \mathbf{W}_{intra}^d$ where $\mathbf{W}_{intra}^d \in R^{N_c \times d'}$ is a learnable parameter and $\mathbf{H}_0^d \in R^{N_c \times d}$. Using the same process above, we can get the intra-type attention-based representation of drug node n_i , which can be represented as $\mathbf{z}_i^d \in R^{d' \times 1}$. Let $\mathbf{Z}_1^c = [(\mathbf{z}_1^c)^T, \dots, (\mathbf{z}_{N_c}^c)^T]^T$ and $\mathbf{Z}_1^d = [(\mathbf{z}_1^d)^T, \dots, (\mathbf{z}_{N_d}^d)^T]^T$ respectively represent the first layer output of the intra-type attention-based encoder, that is, the node embedding matrix of circRNAs and drugs. Assuming that the intra-type attention-based encoder has t layers, the output of the previous layer is taken as the input of the next layer. Repeating this process can obtain t node embedding matrices about circRNA and drugs as follows: $\mathbf{Z}_1^c, \dots, \mathbf{Z}_t^c, \mathbf{Z}_1^d, \dots, \mathbf{Z}_t^d$.

Inter-type attention-based encoder

The purpose of the intra-type attention-based encoder is to learn node embeddings by aggregating the node information of the same type neighbors, while the purpose of the inter-type attention-based encoder is to handle interactions between different types of nodes. Let $n_i \in C$ and $n_j \in D$, respectively. \mathbf{z}_i^c and \mathbf{z}_j^d are the learned representations of the circRNA node n_i and drug node n_j by intra-type attention networks, respectively. The node-level importance $c_{ij}^{\Phi_m}$ can be calculated by Eq. (16) and normalized by Eq. (17) as follows:

$$c_{ij}^{\Phi_m} = att_{node}(\mathbf{z}_i^c, \mathbf{z}_j^d; \Phi_m) = LeakyRelu\left(\mathbf{a}_{\Phi_m}^T \cdot \left[\mathbf{W}_{inter}^c \mathbf{z}_i^c \parallel \mathbf{W}_{inter}^d \mathbf{z}_j^d \right]\right), \tag{16}$$

$$\gamma_{ij}^{\Phi_m} = softmax_j(c_{ij}^{\Phi_m}) = \frac{exp(c_{ij}^{\Phi_m})}{\sum_{n_k \in N_{inter}^{(\Phi_m)}(n_i)} exp(c_{ij}^{\Phi_m})}. \tag{17}$$

$N_{inter}^{\Phi_m}(n_i)$ denotes the neighbors of node n_i under specific inter-relation Φ_m . \mathbf{W}_{inter}^c and $\mathbf{W}_{inter}^d \in R^{d' \times d}$ are two type-specific matrices that map their features \mathbf{z}_i^c and \mathbf{z}_i^d into a common space. $\mathbf{a}_{\Phi_m} \in R^{2d'}$ is a learnable weight vector. The relationship embedding of circRNA node n_i can be aggregated from the embeddings of its neighbors

of different types(that is, the nodes of a drug), with corresponding coefficients as follows:

$$\mathbf{z}_i^{\Phi_m} = LeakyRelu\left(Norm_{\Phi_m}\left(\sum_{n_j \in N_{inter}^{\Phi_m}(n_i)} \gamma_{ij}^{\Phi_m} \mathbf{W}_{inter}^d \mathbf{z}_j^d\right)\right). \tag{18}$$

$Norm_{\Phi_m}$ denotes the layer normalization operation related to the inter-type relation. Then, the importance of relation embedding $\mathbf{z}_i^{\Phi_m}$ related to node n_i are obtained by fusing all relational representations by Eq. (19), and it is normalized by Eq. (20) for making relation importance comparable within inter-type relations.

$$f_i^{\Phi_m} = \tilde{\mathbf{q}}^T(\mathbf{z}_i^c \parallel \mathbf{z}_i^{\Phi_m}), \tag{19}$$

$$\epsilon_i^{\Phi_m} = softmax_m(f_i^{\Phi_m}) = \frac{exp(f_i^{\Phi_m})}{\sum_{\Phi_n \in N_{inter}} exp(f_i^{\Phi_n})}. \tag{20}$$

Finally, the representation \mathbf{u}_i of circRNA node n_i is obtained by fusing these relation-specific representations as follows:

$$\mathbf{u}_i^c = \sum_{\Phi_m \in N_{inter}} \epsilon_i^{\Phi_m} \cdot \mathbf{z}_i^{\Phi_m}. \tag{21}$$

Similarly, we can get the inter-type attention-based representation of drug node n_j , which can be represented as \mathbf{u}_j^d . Let $\mathbf{U}_1^c = [(\mathbf{u}_1^c)^T, \dots, (\mathbf{u}_{N_c}^c)^T]^T$ and $\mathbf{U}_1^d = [(\mathbf{u}_1^d)^T, \dots, (\mathbf{u}_{N_d}^d)^T]^T$ respectively represent the first layer output of the inter-type attention-based encoder, that is, the node embedding matrix of circRNAs and drugs. Assuming the inter-type attention-based encoder has M layers, the output of the previous layer is taken as the input of the next layer. Repeating this process can obtain M node embedding matrices about circRNA and drugs as follows: $\mathbf{U}_1^c, \dots, \mathbf{U}_M^c, \mathbf{U}_1^d, \dots, \mathbf{U}_M^d$.

Multi-kernel fusion

We can extract multiple embeddings from the intra-type attention-based encoder and the inter-type attention-based encoder that represent the information on circRNA nodes and drug nodes of different types and different relationships. For all the embeddings of circRNA and drug, we used the GIP kernel similarity function to calculate the circRNA and drug kernel matrices in each layer as follows:

$$\mathbf{IC}_l(i, j) = exp\left(-\gamma_l \|\mathbf{H}_c^{(l)}(i) - \mathbf{H}_c^{(l)}(j)\|^2\right), \tag{22}$$

$$\mathbf{ID}_l(i, j) = exp\left(-\gamma_l \|\mathbf{H}_d^{(l)}(i) - \mathbf{H}_d^{(l)}(j)\|^2\right). \tag{23}$$

Where $\mathbf{IC}_l \in R^{N_c \times N_c}$, $\mathbf{ID}_l \in R^{N_d \times N_d}$, $\mathbf{H}_c^{(l)}(i)$ and $\mathbf{H}_d^{(l)}(i)$ represent the i -th row of matrix $\mathbf{H}_c^{(l)}$ and $\mathbf{H}_d^{(l)}$ respectively; $\mathbf{H}_c^{(l)}$ and $\mathbf{H}_d^{(l)}$ are the l -th element of the circRNA embedding matrix set $\mathbf{H}_c = \{\mathbf{H}_0^c, \mathbf{Z}_1^c, \dots, \mathbf{Z}_t^c, \mathbf{U}_1^c, \dots, \mathbf{U}_M^c\}$ and the drug embedding matrix set $\mathbf{H}_d = \{\mathbf{H}_0^d, \mathbf{Z}_1^d, \dots, \mathbf{Z}_t^d, \mathbf{U}_1^d, \dots, \mathbf{U}_M^d\}$, respectively. γ_l denotes the corresponding bandwidth, we set $\gamma_l = \gamma, l = 1, \dots, K + 1$, and γ is a hyperparameter, and $K + 1$ is the number of elements in the circRNA embedding matrix set \mathbf{H}_c and the drug embedding matrix set \mathbf{H}_d .

We integrated all the kernels above with multiple kernel fusion in order to fully utilize the information and improve the performance of predicting circRNA-drug associations, then the final kernel matrices of circRNA and drug were obtained as follows:

$$\mathbf{IC} = \sum_{i=0}^K \omega_i^c \mathbf{IC}_i, \tag{24}$$

$$\mathbf{ID} = \sum_{i=0}^K \omega_i^d \mathbf{ID}_i. \tag{25}$$

Where $\mathbf{IC} \in R^{N_c \times N_c}$, $\mathbf{ID} \in R^{N_d \times N_d}$, $\omega_i^m = \frac{1}{K+1}$, and $\omega_i^d = \frac{1}{K+1}$ are the corresponding weight of circRNA kernels and drug kernels, respectively.

Dual Laplacian regularized least squares model

Inspired by previous studies [34] and [35], the Dual Laplacian Regularized Least Squares (DLapRLS) method was adopted by us to predict circRNA-drug associations. Overfitting was avoided by adding graph regularization with DLapRLS. Thus, the loss function can be defined as follows:

$$\min J = \|\mathbf{IC}\alpha_c + (\mathbf{ID}\alpha_d)^T - 2\mathbf{Y}_{train}\|_F^2 + \phi_c \text{tr}(\alpha_c^T \mathbf{L}_c \alpha_c) + \phi_d \text{tr}(\alpha_d^T \mathbf{L}_d \alpha_d). \tag{26}$$

Where $\|\cdot\|_F$ is the Frobenius norm, α_c and $\alpha_d^T \in R^{N_c \times N_d}$ are learnable matrices, ϕ_c and ϕ_d are regularization parameters; $\mathbf{L}_c \in R^{N_c \times N_c}$ and $\mathbf{L}_d \in R^{N_d \times N_d}$ are the normalized Laplacian matrices, as follows:

$$\mathbf{L}_c = \mathbf{V}_c^{-1/2} \mathbf{\Delta}_c \mathbf{V}_c^{-1/2}, \mathbf{\Delta}_c = \mathbf{V}_c - \mathbf{IC}, \tag{27}$$

$$\mathbf{L}_d = \mathbf{V}_d^{-1/2} \mathbf{\Delta}_d \mathbf{V}_d^{-1/2}, \mathbf{\Delta}_d = \mathbf{V}_d - \mathbf{ID}. \tag{28}$$

Where $\mathbf{V}_c = \sum_{i=1}^{N_c} \mathbf{IC}$ and $\mathbf{V}_d = \sum_{i=1}^{N_d} \mathbf{ID}$ are diagonal degree matrix. Finally, the prediction $\hat{\mathbf{F}}$ for circRNA-drug associations from \mathbf{IC} and \mathbf{ID} is obtained as follows:

$$\hat{\mathbf{F}} = \frac{\mathbf{IC}\alpha_c + (\mathbf{ID}\alpha_d)^T}{2}. \tag{29}$$

Training

Except for parameters α_c and α_d , the parameters of our model are updated by Adam [36]. The parameters of α_c and α_d are updated by calculating the partial derivatives for the parameters of DLapRLS. The specific calculation process is as follows: we first assume that α_d is a constant matrix when α_d is optimized. Thus, the partial derivative of the loss function Eq. (26) with respect to α_c can be calculated as follows:

$$\frac{\partial J}{\partial \alpha_c} = 2\mathbf{IC}(\mathbf{IC}\alpha_c + (\mathbf{ID}\alpha_d)^T - 2\mathbf{Y}_{train}) + 2\phi_c \mathbf{L}_c \alpha_c \tag{30}$$

Let $\frac{\partial J}{\partial \alpha_c} = 0$, then α_c can be obtained as follows:

$$\begin{aligned} (\mathbf{IC} + \phi_c \mathbf{L}_c) \alpha_c &= \mathbf{IC} [2\mathbf{Y}_{train} - \alpha_d^T \mathbf{ID}^T], \\ \alpha_c &= (\mathbf{IC} + \phi_c \mathbf{L}_c)^{-1} \mathbf{IC} [2\mathbf{Y}_{train} - \alpha_d^T \mathbf{ID}^T]. \end{aligned} \tag{31}$$

Similarly, the partial derivative of the loss function Eq. (26) with respect to α_d can be calculated as follows:

$$\frac{\partial J}{\partial \alpha_d} = 2\mathbf{ID}(\mathbf{ID}\alpha_d + (\mathbf{IC}\alpha_c)^T - 2\mathbf{Y}_{train}^T) + 2\phi_d \mathbf{L}_d \alpha_d. \tag{32}$$

Same as above, we let $\frac{\partial J}{\partial \alpha_d} = 0$, and then α_d can be obtain as follows:

$$\begin{aligned} (\mathbf{ID} + \phi_d \mathbf{L}_d) \alpha_d &= \mathbf{ID} [2\mathbf{Y}_{train}^T - \alpha_c^T \mathbf{IC}^T], \\ \alpha_d &= (\mathbf{ID} + \phi_d \mathbf{L}_d)^{-1} \mathbf{ID} [2\mathbf{Y}_{train}^T - \alpha_c^T \mathbf{IC}^T]. \end{aligned} \tag{33}$$

α_c and α_d were randomly initialized at the beginning of our model training, and then they were calculated by Eqs. (31) and (33) directly in each iteration, while other parameters were optimized by Adam. The flowchart of our model is shown in Fig. 1. All experimentally verified circRNA-drug associations were treated as positive samples, and the unknown circRNA-drug associations were treated as negative samples, similar to the work of Deng et al. [15] and Yang et al. [17]. Then, the same number of negative samples were randomly selected from all the the unknown circRNA-drug associations. Finally, the same number of positive and negative samples were selected for training.

Results

Implementation details and performance evaluation

The model used in this study was implemented based on PyTorch and PyG, and we evaluated the predictive

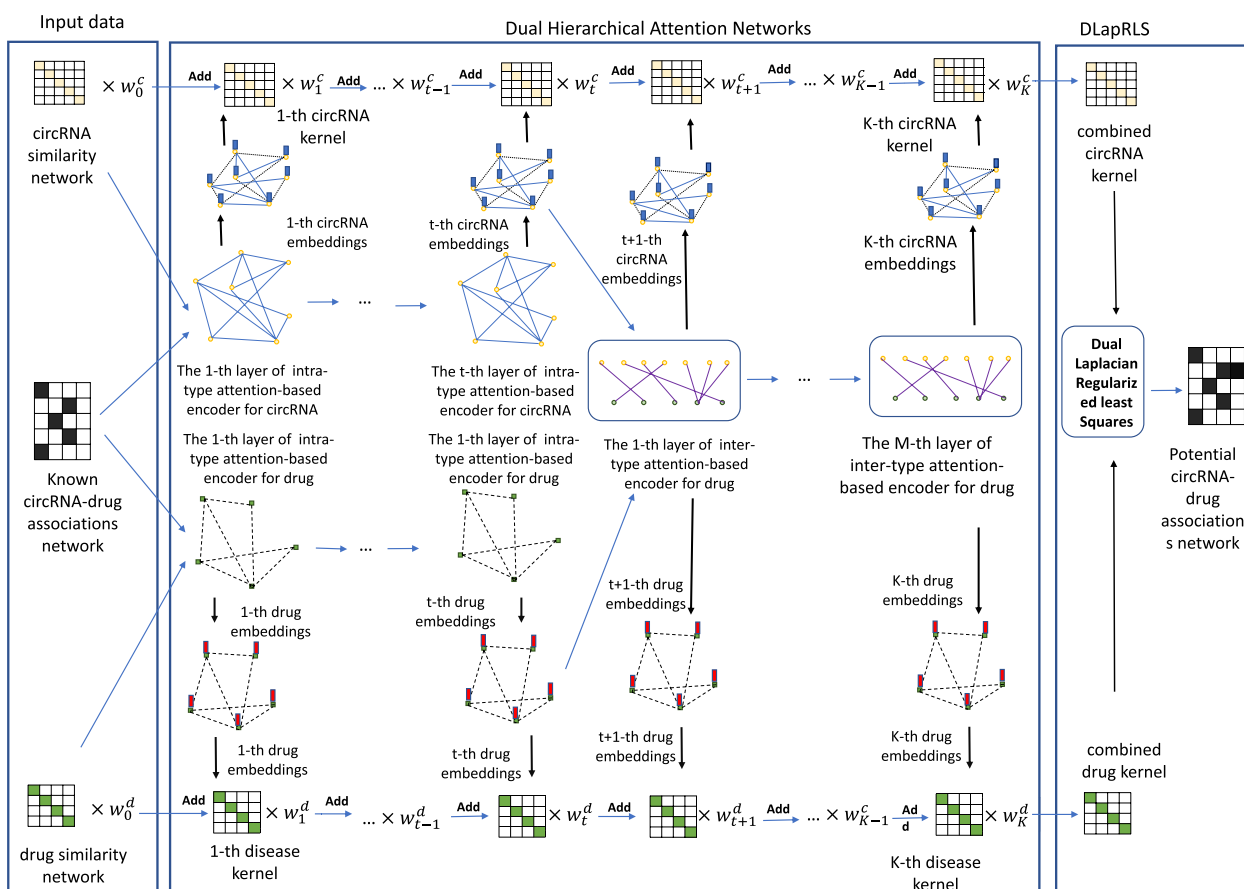


Fig. 1 The overview of our proposed method

performance of our model using 5-fold cross-validation (5CV). The training epochs were set to 40, the learning rate to 0.05 and the weight decay to 0.01. The number of layers for both the intra-type attention-based encoder and the inter-type attention-based encoder were set to 1 and the output dimensions were both set to 16. The thresholds for distinguishing the head and tail nodes of circRNAs and drugs were set at 27 and 39, respectively. Multi-headed attention was set to 5, and the remaining hyperparameters were set as follows: $\phi_c = \phi_d = \frac{1}{120}$, $\gamma_c = \gamma_d = \gamma = \frac{1}{75}$.

During evaluation, we randomly divided all the samples into 5 folds. Four of these folds were used as a training set while the remaining fold was treated as a test set. Seven metrics are used to compare model performance: AUC, AUPR, Accuracy, Precision, Recall, F1-Score, and Specificity. It is well-established that improved model performance is reflected by higher AUC and AUPR values. F1-Score is the average of accuracy and recall, while specificity measures the ability of the classifier to correctly identify negative cases.

Performance comparison with other methods under 5-CV

The current computational methods for predicting circRNA-drug sensitivity associations are restricted. We found that GATECDA [15] and MNGACDA [17] are specifically designed for predicting circRNA-drug sensitivity associations. Thus, like Ref. [15] and Ref. [17], we compared our model with seven state-of-the-art models from different domains, namely MNGACDA [17], GATECDA [15], MINIMDA [37], LAGCN [38], MMGCN [39], and GANLDA [40]. Brief descriptions of these models are provided below:

- MNGACDA [17] : a computational framework for predicting circRNA-drug sensitivity associations. This model uses multimodal networks to learn the embedded representations of circRNAs and drugs, then captures the internal information between nodes in the networks with node-level attention Graph Auto-Encoder.
- GATECDA [15] : a computational model based on Graph Attention Auto-encoder (GATE) for predicting circRNA-drug sensitivity associations.

- MKGCN [35] : a computational model based on GCN and MKL for predicting microbe-drug associations.
- MINIMDA [37] : a method of predicting miRNA-disease associations by constructing integrated similarity networks and using multimodal networks to obtain embedding representations of miRNAs and diseases. These representations are then fed into a multilayer perceptron for prediction.
- LAGCN [38] : LAGCN integrates various associations into a heterogeneous network, learns embeddings of drugs and diseases by Graph Convolution operations, and then combines multiple layers by using an attention function.
- MMGCN [39] : MMGCN differs from simple multi-source integration in that it uses a GCN encoder to obtain miRNA and disease features in different similarity views and enhances the learned representations for association prediction by using multichannel attention that adaptively learns the importance of different features.
- GANLDA [40] : this method combines heterogeneous data of lncRNA and disease as original features and reduces noise by using Principal Component Analysis (PCA). Then the Graph Attention Network is used to extract information from the features. Finally, a multi-layer perceptron is used to predict lncRNA-disease associations.

The prediction performance of each method was evaluated by a 5CV experiment using the same settings and

optimal parameters recommended in their respective studies. From Table 1, it can be seen that DHANMKF achieved the highest AUC and AUPR values. This indicates that DHANMKF performed better overall compared to the other models.

Evaluation of parameters

The prediction performance of DHANMKF is affected by various parameter values. The parameters of DHANMKF can be divided into four parts: the parameters in the inter-type attention-based encoder and the intra-type attention-based encoder, bandwidth parameter γ in MKE, regularization parameters (ϕ_c and ϕ_d) in DLapRL, and degree threshold parameters (K_c and K_d) for distinguishing circRNA and drug nodes as head and tail nodes. Here, the process of parameter evaluation is demonstrated using data271 as the baseline dataset. The parameter settings of DHANMKF on the data251 dataset have been put into the [Supplementary file](#).

Optimizable parameters in the intra-type attention-based encoder and the inter-type attention-based encoder

- *Learning rate and its weight decay*. Learning rate and its weight decay are the same in the intra-type attention-based encoder and the inter-type attention-based encoder. Based on the research conducted by Zhao et al. [32], we set them as 0.05 and 0.01 respectively.

Table 1 Performance comparison based on five-fold cross-validation

Method	F1-Score	Accuracy	Recall	Specificity	Precision	AUC	AUPR
Dataset: data271							
DHANMKF(our)	0.8500	0.8520	0.8648	0.8636	0.8618	0.9178	0.9262
MNGACDA [17]	0.8472	0.8424	0.8723	0.8155	0.8247	0.9139	0.9209
GATECDA [15]	0.8224	0.8186	0.8404	0.7966	0.8054	0.8873	0.8915
MKGCN [35]	0.8230	0.8208	0.8221	0.8193	0.8350	0.8768	0.8984
MMGCN [39]	0.8190	0.8183	0.8231	0.8135	0.8156	0.8766	0.8664
MINIMDA [37]	0.7988	0.7901	0.8331	0.7472	0.7684	0.8562	0.8534
LAGCN [38]	0.7900	0.7786	0.8338	0.7233	0.7516	0.8505	0.8478
GANLDA [40]	0.7936	0.7822	0.8384	0.7259	0.7542	0.8517	0.8468
Dataset: data251							
DHANMKF(our)	0.8597	0.8588	0.8655	0.8521	0.8552	0.9263	0.9286
MNGACDA [17]	0.8444	0.8429	0.8528	0.8330	0.8362	0.9136	0.9211
MKGCN [35]	0.8253	0.8247	0.8285	0.8208	0.8222	0.8941	0.9090
MMGCN [39]	0.8065	0.8120	0.7836	0.8404	0.8308	0.8842	0.9049
GATECDA [15]	0.8336	0.8287	0.8225	0.8354	0.8451	0.8802	0.9034
MINIMDA [37]	0.8053	0.7950	0.8476	0.7424	0.7670	0.8575	0.8502
LAGCN [38]	0.7880	0.7857	0.7966	0.7748	0.7796	0.8540	0.8615
GANLDA [40]	0.7883	0.7956	0.8089	0.7823	0.7880	0.8539	0.8551

- *Dropout and number of model training epochs.* We selected the dropout to be $\{0.02, 0.021, \dots, 0.03\}$. When the value of the dropout is 0.026, the model performance reaches its optimum, and with an increase in the value of the dropout, the model performance gradually declines. The loss of DHANMKF started converging at 40 training epochs, so the number of epochs for our model was set to 40.
- *The number of attention heads.* To have a more powerful representation learning capacity, the multi-head attention mechanism was incorporated into the model. This parameters was tuned using 5CV. As shown in Fig. 2A, when the number of attention heads is equal to 5, the model performance reaches its optimum.
- *The output dimensions.* We analyzed the output dimensions of the intra-type attention-based encoder and inter-type attention-based encoder, as shown in Fig. 2B. When the output dimension was 16, the AUC performance was best.
- *The number of layers of the intra-type attention-based encoder and the inter-type attention-based encoder.* As shown in Fig. 3A, when the number of layers of the intra-type attention-based encoder and the inter-type attention-based encoder are both 1, the AUC of DHANMKF reaches its optimal value.

Optimizable parameters in MKF and DLapRL

- The bandwidth parameter γ in MKF is actually the $\frac{1}{2\sigma^2}$ of the Gaussian kernel function, that is, $\gamma = \frac{1}{2\sigma^2}$. Parameter σ determines the smoothness of the Gaussian filter. The larger σ is, the smoother

it is. Therefore, by adjusting γ , a compromise can be reached between over-smoothing and under-smoothing. As shown in Fig. 3B, when $\gamma = \frac{1}{75}$, the AUC of DHANMKF reaches its optimal value.

- The parameters ϕ_c and ϕ_d play a regulating role in DLapRL, and they can be adjusted to balance under-fitting and overfitting. From Fig. 4A, we can see that the AUC of the model reaches its maximum when ϕ_c and ϕ_d are both $\frac{1}{120}$.

Optimization of head and tail node thresholds

The threshold of the head and tail nodes can adjust the number of head and tail nodes and the number of associated types, thus affecting the embedding of corresponding nodes. As shown in Fig. 4B, the maximum AUC value of the model is achieved when the thresholds of the circRNA node and drug node are 27 and 39, respectively.

Ablation tests

Ablation experiments were conducted from two perspectives: 1. Analyzing the importance of the intra-type attention-based encoder and the inter-type attention-based encoder; 2. Analyzing the effects of multiple relationships. Therefore, we constructed three ablation experiments. The first one is called DHANMKF-intra, which means that DHANMKF removes the embedding produced by the intra-type attention-based encoder when doing Multi-Kernel Fusion. The second one is called DHANMKF-inter, which means that DHANMKF removes the embedding produced by the inter-type attention-based encoder when doing multi-core fusion.

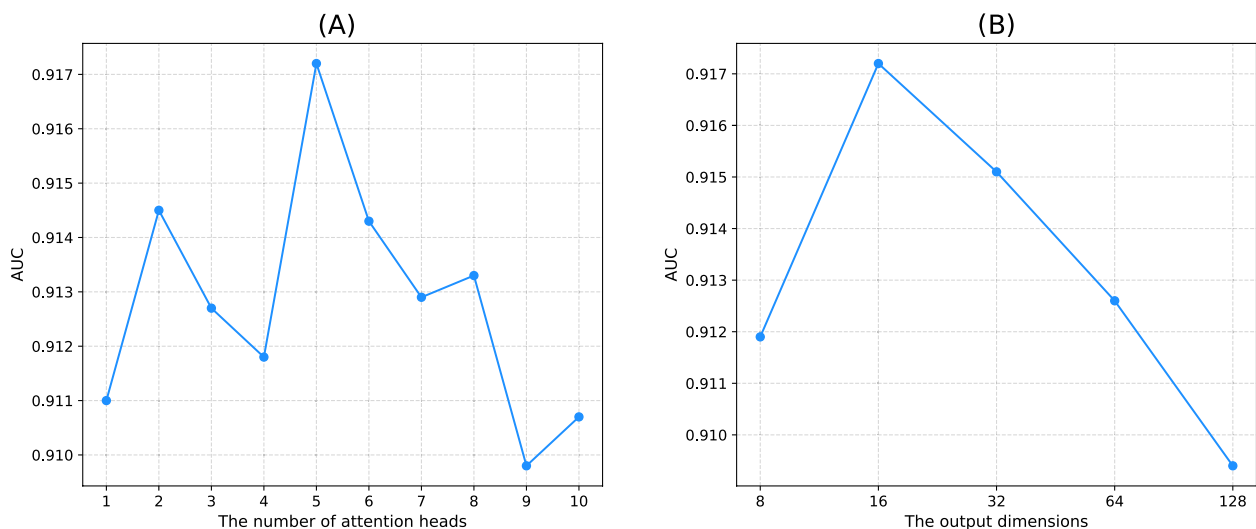


Fig. 2 DHANMKF's attention heads and output dimensions

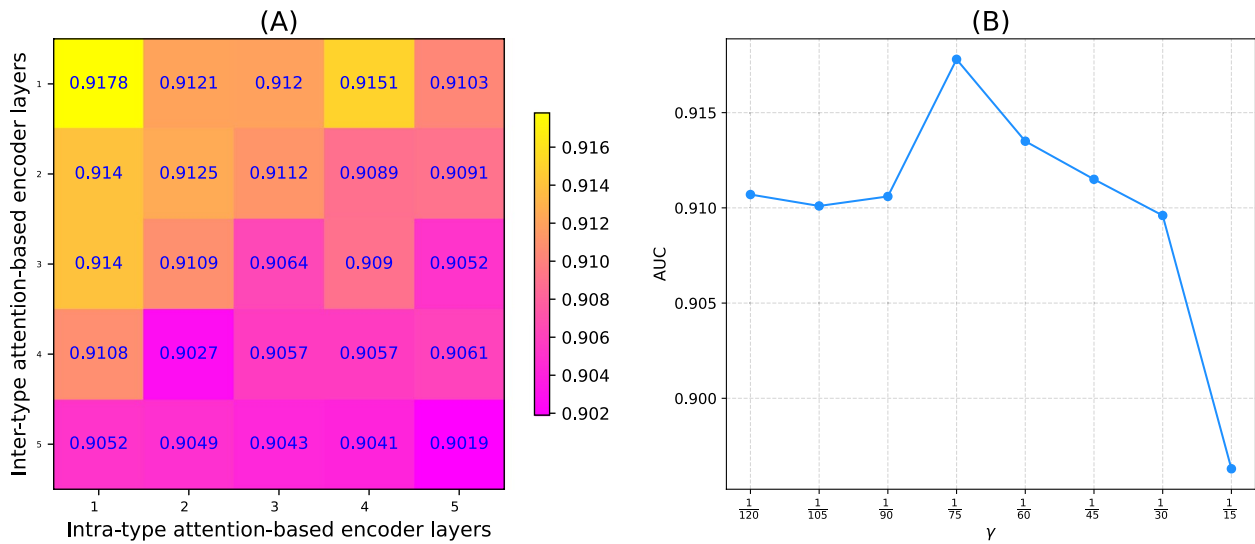


Fig. 3 DHANMKF's layers and γ

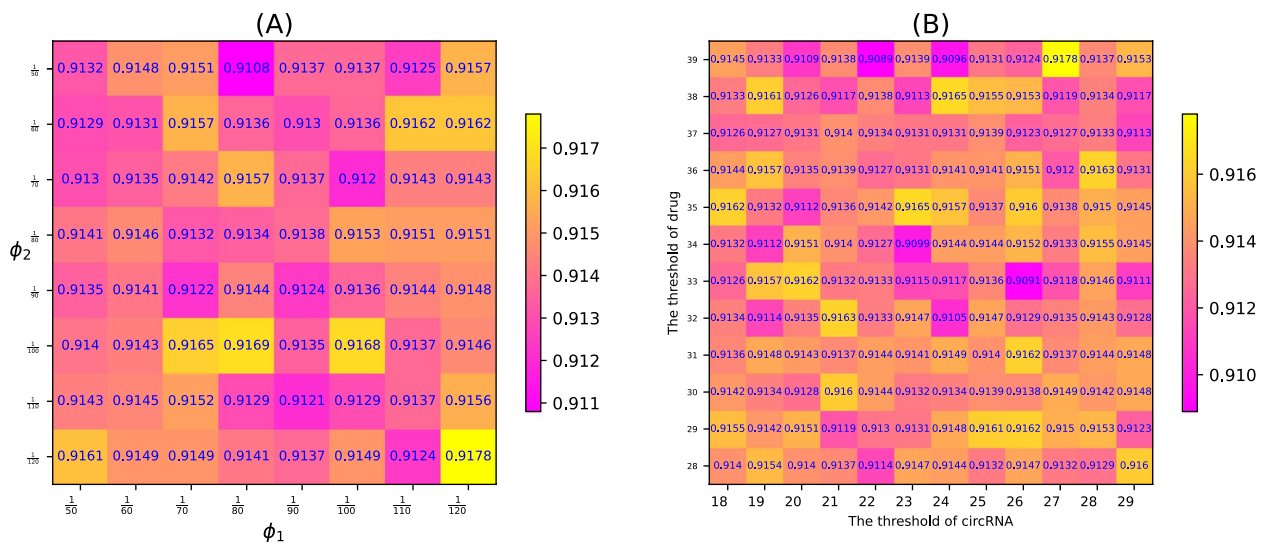


Fig. 4 ϕ_1 and ϕ_2 in DLapRL, the thresholds for the circRNA node and drug node

The third one is called DHANMKF-multi, which means that the model no longer divides the relationships between nodes into multiple categories.

Table 2 shows the comparison results of the 5CV. From Table 2 we can see that DHANMKF performs better than all other models. This shows that: (1) Compared with DHANMKF-intra and DHANMKF-inter, DHANMKF performs better, which means that the embeddings produced by the intra-type attention-based encoder and the inter-type attention-based encoder improve the performance of the model. (2) Compared with DHANMKF-multi, DHANMKF can generate node embeddings corresponding to different relationships between nodes.

Table 2 Ablation experiment

Model	Dataset: data271		Dataset: data251	
	AUC	AUPR	AUC	AUPR
DHANMKF	0.9178	0.9262	0.9263	0.9286
DHANMKF-intra	0.9143	0.9238	0.9230	0.9245
DHANMKF-inter	0.9150	0.9228	0.9238	0.9227
DHANMKF-multi	0.9131	0.9226	0.9186	0.9212

In summary, there are two main reasons why DHANMKF can outperform other models. The first reason is that our model can fully capture the complex structures

of the bi-typed multi-relational heterogeneous graphs. The second reason is that biological networks in reality are sparse, so it is reasonable to divide the nodes in biological networks into head nodes and tail nodes for analysis.

Case studies

To further evaluate the predictive performance of our model, we selected two drugs, PAC-1 and Vorinostat, for case studies. Similar to Deng et al. [15] and Yang et al. [17], we used the circRNA-drug associations in the GDSC database as the training set and those in the CTRP database as the testing set. For each drug, we chose the top 20 circRNAs with the highest predicted scores from our model's circRNA-drug association prediction outputs for validation.

PAC-1 is the first known small molecule drug that directly activates procaspase-3 to caspase-3 [41]. It not only enhances procaspase-3 activity but also induces cancer cell apoptosis. Vitro experiments have shown that PAC-1 exhibits cytotoxicity against lymphoma, multiple myeloma, and many other cancer cells [42]. Currently, PAC-1 has been used in clinical trials for the treatment of various tumors, including but not limited to lymphoma, melanoma, solid tumors, breast cancer, and lung cancer [43]. As shown in Table 3, among the top 20 circRNAs predicted by our method to be associated with PAC-1, 16 have been identified in CTRP.

Belinostat is a small-molecule hydroxamate-type inhibitor that can inhibit the activity of class I, II and IV histone deacetylase enzymes. It has been used to treat relapsed or refractory peripheral T-cell lymphoma [44]. Table 4 shows that 17 of the top 20 circRNAs predicted by our method have been confirmed in circRic.

In order to demonstrate the performance of DHANMKF in predicting the potential association between

new drugs and circRNA, we chose two drugs for *ab initio* testing, both of which had only one known circRNA-drug association. During the training phase, we removed the unique association between these two drugs and circRNA. At this point, these two drugs were not associated with any circRNAs and were treated as new drugs during training. These two drugs were Bortezomib and MS-275 (Entinostat). Bortezomib is a novel proteasome inhibitor with potent chemo/radio-sensitizing effects that can overcome the traditional resistance of tumors when used in combination with chemotherapy [45]. In addition, existing clinical applications have shown that Bortezomib can improve clinical outcomes in the treatment of hematologic malignancies [46].

MS-275, also known as Entinostat, is effective in human leukemia cells and lymphoma cells. It can reduce the level of Bcl-XL in cells, induce p21 protein expression, cause cell cycle arrest (G1 phase), and induce cell apoptosis [47]. In addition, when used in combination with other drugs, entinostat can enhance the activity of some anticancer drugs, including Rituximab, Gemcitabine, Doxorubicin, Sorafenib and Bortezomib. Currently, Entinostat is undergoing phase III clinical trials and its clinical data shows that it has great potential for treating breast cancer [48].

As shown in Table 5, 6 of the top 10 predicted circRNAs associated with Bortezomib have been confirmed in circRic, and 7 of the top 10 circRNAs related to MS-275 have been confirmed in circRic.

Conclusions

Recent research over the past twenty years has shown that circRNA plays an important role in drug sensitivity. Therefore, predicting the potential association between circRNA and drug sensitivity can be helpful in drug development and utilization, thus benefiting

Table 3 Top 20 circRNAs related to PAC-1 predicted by DHANMKF

Ranking	circRNA	Evidence	Ranking	circRNA	Evidence
1	SPARC	CTRP	11	MEF2D	CTRP
2	ARID1B	CTRP	12	PEA15	CTRP
3	LTBP4	Nonsignificant	13	PTMS	CTRP
4	VIM	CTRP	14	CRIM1	CTRP
5	POLR2A	CTRP	15	TCOF1	Nonsignificant
6	SPINT2	Nonsignificant	16	ANP32B	CTRP
7	CTTN	CTRP	17	ENO2	Nonsignificant
8	ASPH	CTRP	18	NCL	CTRP
9	COL1A2	CTRP	19	COL1A1	CTRP
10	THBS1	CTRP	20	ADPGK	CTRP

Table 4 Top 20 circRNAs related to Belinostat predicted by DHANMKF

Ranking	circRNA	Evidence	Ranking	circRNA	Evidence
1	ASPH	CTRP	11	CTSD	CTRP
2	CTTN	CTRP	12	MEF2D	CTRP
3	THBS1	CTRP	13	ILF3	CTRP
4	CRIM1	CTRP	14	FBLN1	CTRP
5	ANXA2	CTRP	15	PHF21A	CTRP
6	LTBP3	CTRP	16	MYC	CTRP
7	PTMS	Nonsignificant	17	GSE1	CTRP
8	POLR2A	Nonsignificant	18	PEA15	CTRP
9	MGAT4B	CTRP	19	COL18A1	CTRP
10	KRT19	Nonsignificant	20	NOP53	CTRP

Table 5 The Top 10 predicted circRNAs associated with the two new drugs Bortezomib and MS-275

Bortezomib			MS-275		
Ranking	circRNA	Evidence	Ranking	circRNA	Evidence
1	POLR2A	Nonsignificant	1	POLR2A	Nonsignificant
2	MEF2D	CTRP	2	CTTN	CTRP
3	VIM	Nonsignificant	3	ASPH	CTRP
4	CRIM1	CTRP	4	ANP32B	CTRP
5	THBS1	CTRP	5	SPINT2	Nonsignificant
6	ADPGK	CTRP	6	CRIM1	CTRP
7	TCOF1	Nonsignificant	7	PTMS	CTRP
8	ENO2	Nonsignificant	8	VIM	Nonsignificant
9	FBLN1	CTRP	9	THBS1	CTRP
10	PEA15	CTRP	10	FBLN1	CTRP

patients. In this study, we proposed a method, based on intra-type attention and inter-type attention called DHANMKF, for discovering potential circRNA-drug sensitivity associations. To verify the effectiveness of the model, DHANMKF was compared with six state-of-the-art methods based on 5CV on benchmark datasets. The results showed that DHANMKF achieved the best performance. In addition, to further evaluate the ability of the model to discover new drugs, a case study was conducted and the model's prediction results were validated using an independent database. The validation results clearly demonstrate that DHANMKF is an effective tool for predicting new circRNA-drug sensitivity associations.

The results show that our model outperforms the baseline models. We believe the main reasons are the following: (1) We classify the nodes into head and tail nodes, which in turn defines the types of edges connecting these two types of nodes. This allows our model to extract node embeddings from the circRNA-drug heterogeneous graph based on different types of edges. (2) The intra-type attention-based encoder can efficiently aggregate information from nodes of the same type. (3) The inter-type attention-based encoder adequately extracts node representations from different types of nodes. (4) The MKL method fuses the multi-relational heterogeneous graph information captured by the two encoders in order to improve the overall performance of the model. In future studies, we plan to integrate more biomedical data, in order to generate more comprehensive circRNA and drug kernels and further improve model performance. Currently, there are few studies that use computational methods to predict potential associations between circRNA and drug sensitivity, so further investigation in this field is merited.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09899-w>.

Additional file 1.

Acknowledgements

Not applicable.

Authors' contributions

SHL designed the framework, conducted the experiments, and wrote the manuscript. YL, LL, SLL, YFZ, CJY, and DOY modified the manuscript. All author(s) revised and approved the manuscript.

Funding

This work was supported in part by the major key project of PCL(PCL2021A12), the Macau Science and Technology Development Funds Grands No. 0056/2020/AFJ from the Macau Special Administrative Region of the People's Republic of China, the Key Project for University of Educational Commission of Guangdong Province of China Funds (Natural, Grant No. 2019GZDXM005), Young and Middle aged Teachers Research Basic Ability Improvement Project of Guangxi Universities, China (No. 2022KY0608).

Availability of data and materials

The data sets and source codes used in this study are freely available at <https://github.com/cuntjx/DHANMKF>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 August 2023 Accepted: 12 December 2023

Published online: 21 December 2023

References

- Nisar S, Bhat AA, Singh M, Karedath T, Rizwan A, Hashem S, et al. Insights into the role of CircRNAs: biogenesis, characterization, functional, and clinical impact in human malignancies. *Frontiers Cell Dev Biol.* 2021;9:617281.
- Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci.* 1976;73(11):3852–6.
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature.* 2013;495(7441):333–8.
- Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature.* 2013;495(7441):384–8.
- Yu CY, Kuo HC. The emerging roles and functions of circular RNAs and their generation. *J Biomed Sci.* 2019;26(1):1–12.
- Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell.* 2014;56(1):55–66.
- Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol.* 2014;32(5):453–61.
- Zhang Y, Yang L, Chen LL. Life without A tail: new formats of long non-coding RNAs. *Int J Biochem Cell Biol.* 2014;54:338–49.

9. Li F, Zhang L, Li W, Deng J, Zheng J, An M, et al. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/ β -catenin pathway. *Oncotarget*. 2015;6(8):6001.
10. Fan C, Lei X, Tie J, Zhang Y, Wu FX, Pan Y. CircR2Disease v2. 0: an updated web server for experimentally validated circRNA–disease associations and its application. *Genomics Proteomics Bioinforma*. 2022;20(3):435–45.
11. Gao D, Zhang X, Liu B, Meng D, Fang K, Guo Z, et al. Screening circular RNA related to chemotherapeutic resistance in breast cancer. *Epigenomics*. 2017;9(9):1175–88.
12. Kun-Peng Z, Xiao-Long M, Lei Z, Chun-Lin Z, Jian-Ping H, Tai-Cheng Z. Screening circular RNA related to chemotherapeutic resistance in osteosarcoma by RNA sequencing. *Epigenomics*. 2018;10(10):1327–46.
13. Wu Q, Wang H, Liu L, Zhu K, Yu W, Guo J. Hsa_circ_0001546 acts as a miRNA-421 sponge to inhibit the chemoresistance of gastric cancer cells via ATM/Chk2/p53-dependent pathway. *Biochem Biophys Res Commun*. 2020;521(2):303–9.
14. Ruan H, Xiang Y, Ko J, Li S, Jing Y, Zhu X, et al. Comprehensive characterization of circular RNAs in ~ 1000 human cancer cell lines. *Genome Med*. 2019;11:1–14.
15. Deng L, Liu Z, Qian Y, Zhang J. Predicting circRNA-drug sensitivity associations via graph attention auto-encoder. *BMC Bioinformatics*. 2022;23(1):1–15.
16. Salehi A, Davulcu H. Graph attention auto-encoders. *arXiv preprint arXiv:1905.10715*. 2019.
17. Yang B, Chen H. Predicting circRNA-drug sensitivity associations by learning multimodal networks using graph auto-encoders and attention mechanism. *Brief Bioinforma*. 2023;24(1):bbac596.
18. Gönen M, Alpaydin E. Multiple kernel learning algorithms. *J Mach Learn Res*. 2011;12:2211–68.
19. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*. 2016.
20. Yan XY, Zhang SW, He CR. Prediction of drug–target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. *Comput Biol Chem*. 2019;78:460–7.
21. Peng W, Wu R, Dai W, Ning Y, Fu X, Liu L, et al. MiRNA–gene network embedding for predicting cancer driver genes. *Brief Funct Genomics*. 2023;22(4):341–50. <https://doi.org/10.1093/bfpg/ela059>.
22. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 2012;41(D1):D955–61.
23. Rangwala SH, Kuznetsov A, Ananiev V, Asztalos A, Borodin E, Evgeniev V, et al. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Res*. 2021;31(1):159–69.
24. Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, et al. Pubchem bioassay: 2017 update. *Nucleic Acids Res*. 2017;45(D1):D955–63.
25. 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*. 2022;140:105119.
26. Tian X, Shen L, Wang Z, Zhou L, Peng L. A novel lncRNA–protein interaction prediction method based on deep forest with cascade forest structure. *Sci Rep*. 2021;11(1):18881.
27. Landrum G, et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*. 2013;8:1–31.
28. Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, et al. WBSMDA: within and between score for MiRNA–disease association prediction. *Sci Rep*. 2016;6(1):1–9.
29. Van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*. 2011;27(21):3036–43.
30. Zou S, Zhang J, Zhang Z. A novel approach for predicting microbe–disease associations by bi-random walk on the heterogeneous network. *PLoS ONE*. 2017;12(9):e0184394.
31. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11(3):333–7.
32. Zhao Y, Wei S, Du H, Chen X, Li Q, Zhuang F, et al. Learning Bi-typed Multi-relational Heterogeneous Graph via Dual Hierarchical Attention Networks. *IEEE Trans Knowl Data Eng*. 2022;35(9):9054–66.
33. Liu Z, Nguyen TK, Fang Y. Tail-gnn: Tail-node graph neural networks. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. New York, United States: Association for Computing Machinery; 2021. p. 1109–19.
34. Ding Y, Tang J, Guo F. Identification of drug–target interactions via dual laplacian regularized least squares with multiple kernel fusion. *Knowl-Based Syst*. 2020;204:106254.
35. Yang H, Ding Y, Tang J, Guo F. Inferring human microbe–drug associations via multiple kernel fusion on graph neural network. *Knowl-Based Syst*. 2022;238:107888.
36. Kingma, Diederik P, and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*. 2014.
37. Lou Z, Cheng Z, Li H, Teng Z, Liu Y, Tian Z. Predicting miRNA–disease associations via learning multimodal networks and fusing mixed neighborhood information. *Brief Bioinforma*. 2022;23(5).
38. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug–disease associations through layer attention graph convolutional network. *Brief Bioinforma*. 2021;22(4):1–11.
39. Tang X, Luo J, Shen C, Lai Z. Multi-view multichannel attention graph convolutional network for miRNA–disease association prediction. *Brief Bioinforma*. 2021;22(6):bbab174.
40. Lan W, Wu X, Chen Q, Peng W, Wang J, Chen YP. GANLDA: graph attention network for lncRNA–disease associations prediction. *Neurocomputing*. 2022;469:384–93.
41. Peterson QP, Goode DR, West DC, Ramsey KN, Lee JJ, Hergenrother PJ. PAC-1 activates procaspase-3 in vitro through relief of zinc-mediated inhibition. *J Mol Biol*. 2009;388(1):144–58.
42. S Roth H, J Hergenrother P. Derivatives of procaspase-activating compound 1 (PAC-1) and their anticancer activities. *Curr Med Chem*. 2016;23(3):201–41.
43. Assempour N, lynkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, et al. 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074–82.
44. Poole RM. Belinostat: first global approval. *Drugs*. 2014;74:1543–54.
45. Adams J, Kauffman M. Development of the proteasome inhibitor Velcade™ (Bortezomib). *Cancer Investig*. 2004;22(2):304–11.
46. Mujtaba T, Dou QP. Advances in the understanding of mechanisms and therapeutic use of bortezomib. *Discov Med*. 2011;12(67):471.
47. Knipstein J, Gore L. Entinostat for treatment of solid tumors and hematologic malignancies. *Expert Opin Investig Drugs*. 2011;20(10):1455–67.
48. Connolly RM, Rudek MA, Piekarczyk R. Entinostat: a promising treatment option for patients with advanced breast cancer. *Future Oncol*. 2017;13(13):1137–48.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.