Wei *et al. BMC Biology* (2025) 23:119
https://doi.org/10.1186/s12915-025-02221-y

BMC Biology

METHODOLOGY

Open Access

iPiDA-LGE: a local and global graph ensemble learning framework for identifying piRNA-disease associations

Hang Wei[1*], Jialu Hou[3], Yumeng Liu[4], Alexey K. Shaytan[5,6], Bin Liu[2,3,7*] and Hao Wu[3*]

Abstract

Background Exploring piRNA-disease associations can help discover candidate diagnostic or prognostic biomarkers and therapeutic targets. Several computational methods have been presented for identifying associations between piRNAs and diseases. However, the existing methods encounter challenges such as over-smoothing in feature learning and overlooking specific local proximity relationships, resulting in limited representation of piRNA-disease pairs and insufficient detection of association patterns.

Results In this study, we propose a novel computational method called iPiDA-LGE for piRNA-disease association identification. iPiDA-LGE comprises two graph convolutional neural network modules based on local and global piRNA-disease graphs, aimed at capturing specific and general features of piRNA-disease pairs. Additionally, it integrates their refined and macroscopic inferences to derive the final prediction result.

Conclusions The experimental results show that iPiDA-LGE effectively leverages the advantages of both local and global graph learning, thereby achieving more discriminative pair representation and superior predictive performance.

Keywords piRNA-disease association identification, Graph ensemble learning, Local context graph

*Correspondence:
Hang Wei
weihang@xidian.edu.cn
Bin Liu
bliu@bliulab.net
Hao Wu
hwu@bliulab.net
[1] School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi 710126, China
[2] SMBU-MSU-BIT Joint Laboratory On Bioinformatics and Engineering Biology, Shenzhen MSU-BIT University, Shenzhen, Guangdong 518172, China
[3] School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[4] College of Big Data and Internet, Shenzhen Technology University, Shenzhen, Guangdong 518118, China
[5] Department of Biology, Lomonosov Moscow State University, Moscow 119234, Russia
[6] International Laboratory of Bioinformatics, AI and Digital Sciences Institute, Faculty of Computer Science, HSE University, Moscow 109028, Russia
[7] Zhongguancun Academy, Beijing 100094, China

Background

PIWI-interacting RNA (piRNA) is a category of small non-coding RNA molecules with high conservation, species specificity, and abundant expression [1]. It is involved in multiple biological functions like transposable element silencing, gene expression regulation, embryonic development, and epigenetic modification to maintain genome stability and reproductive process by forming complexes with members of the PIWI protein family [2–4].

As piRNAs are the critical regulatory factors in biological processes, their function abnormality may promote occurrence and development of diseases [5–8]. Several studies have been conducted to explore piRNA functions [9, 10]. For example, piRNA-eQTL is the first robust eQTL database that systematically uncovers the effects of genetic variants on piRNA expression across various cancer types [11]. PiRSNP identified human and mouse piRNA-related SNPs and evaluated their impacts on piRNA-mRNA binding [7]. PiRTarBase offers predicted and experimentally identified functional piRNA targeting sites [12].

Recent studies have shown that piRNAs can serve as promising diagnostic or prognostic biomarkers and therapeutic targets for various cancers [13–15]. Therefore, numerous computational approaches have been suggested for detecting potential piRNA-disease associations, laying the groundwork for subsequent biological investigations. These approaches can primarily be categorized into two groups: classical machine-learning-based and graph-learning-based approaches [16, 17]. Most classical machine-learning-based methods manually splice the attribute information of piRNAs and diseases to construct pair features. To overcome the instability of a single classifier, iPiDA-PUL identified associations based on an ensemble of various classical classifiers [18], and iPiDA-LTR integrated different feature components through a supervised ranking framework [19]. iPiDA-sHN was proposed to improve the quality of negative pairs based on two-step positive-unlabeled learning [20]. With the successful application of graph neural networks in bioinformatics tasks [21–23], several graph-learning-based methods have been proposed to explore the hidden structural features and association patterns within piRNA-disease association networks. For example, iPiDA-GCN designed two principal sub-modules for iteratively learning the proximity features of homogeneous similarity networks and heterogeneous piRNA-disease bipartite network [24]. Considering the varying importance among bio-entity nodes, ETGPDA and PUTransGCN constructed heterogeneous networks and employed different attention-aware graph neural networks to identify potential piRNA-disease associations [25, 26]. The known piRNA-disease associations are relatively limited, posing a challenge due to the highly sparse nature of the association network. To enhance network structural semantics, iPiDA-SWGCN incorporated a supplemental weighted strategy [27], CLPiDA integrated LightGCN with a data augmentation technique [28], and iSG-PDA fused multi-source genetic information [29].

Despite the significant advancements made by the aforementioned graph-learning-based methods in predicting piRNA-disease associations, several issues still need to be addressed: (i) Current methods integrate proximity information from the global piRNA-disease network during the learning process, resulting in more informative node features. However, it may introduce irrelevant noise interference by considering the entire graph structure in each layer, causing over-smoothing of node features. (ii) Existing methods overlook local proximity relationships, which are crucial for piRNA-disease association identification task. PiRNAs may exhibit diverse functional mechanisms across different diseases [9]. However, global graph learning can only extract general and globally invariant node features, making it difficult to detect discriminative association patterns [30, 31]. Overall, fully exploring latent biological semantics while alleviating noise interference and enhancing the discriminative ability of piRNA-disease pair representation remains a challenge.

To overcome the aforementioned limitations, we propose a new method named iPiDA-LGE for identifying piRNA-disease associations based on local and global graph ensemble learning. iPiDA-LGE designs two primary graph convolutional neural network modules to capture local-specific features and general features, thereby enhancing the expressiveness of piRNA-disease pairs. Subsequently, it integrates prediction results derived from refined and macroscopic inferences based on local and global piRNA-disease association networks, respectively. Experimental results indicate that iPiDA-LGE can simultaneously maintain the advantages of local and global graph learning, facilitating a more precise and comprehensive identification of piRNA-disease associations.

Results and discussion

Parameter analysis

iPiDA-LGE comprises two primary modules including global-level graph learning and local-level graph learning. Given the parameters in global-level graph learning have been discussed in our previous study [27], we focus on analyzing parameters in local-level graph learning in this study. The impact of four important parameters including neighbor order, epoch, learning rate, and GCN layer is discussed.

As shown in Fig. 1, the results reveal the following observations: (i) The number of neighbor order influences the size and contextual scope of each local piRNA-disease graph; there is limited contextual semantic information when selecting 1-hop neighborhood to extract local piRNA-disease graph. (ii) The performance of local-level graph learning module initially improves and then decreases with the increment of epochs attributed to overfitting, and a large learning rate may cause divergence or oscillated around the optimal solution. (iii) Different from global-level graph learning, the local-level graph learning module only captures local topology structure and is relatively less sensitive with GCN layer. Ultimately, the neighbor order, epoch, learning rate, and GCN layer are set to 2, 20, 0.001, and 2 during local-level graph learning with considering runtime and predictive performance.

Global and local graph ensemble learning can improve the predictive performance

To study if ensemble learning of local and global piRNA-disease association graphs could improve the predictive performance, two baseline predictors, iPiDA-L and iPiDA-G, which are learned from local and global graph levels are compared with iPiDA-LGE. The comparison results are depicted in Fig. 2, enabling us to draw the following conclusions: (i) The fusion coefficient is relativity sensitive to the overall performance of iPiDA-LGE.

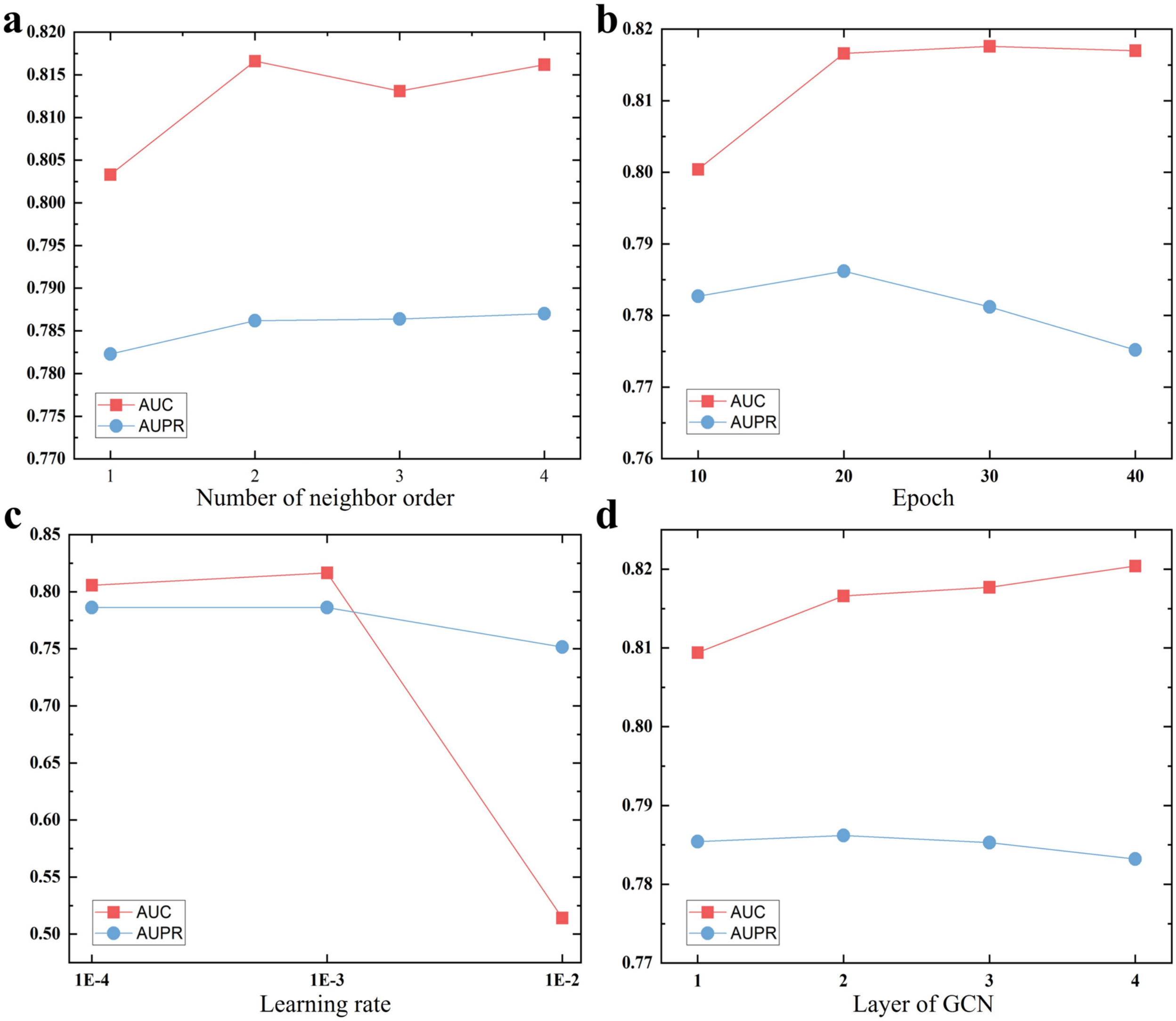


Fig. 1 Parameter analysis of local-level graph learning in iPiDA-LGE. **a**, **b**, **c**, and **d** illustrate the AUC and AUPR obtained by local-level graph learning module with different neighbor orders, epochs, learning rates, and GCN layers on $\mathbb{S}_{validation}$, respectively

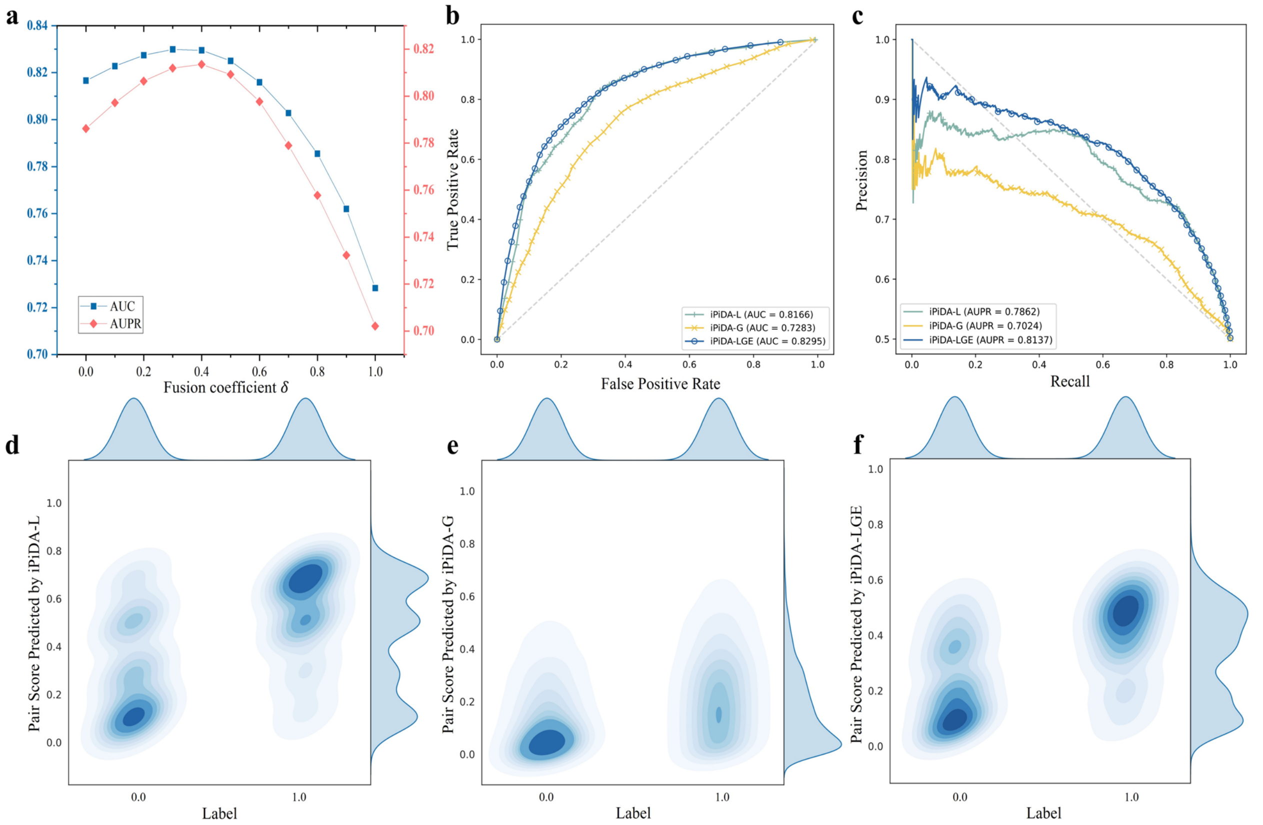


Fig. 2 Comparison between local/global graph learning and their combination. **a** shows the AUC and AUPR achieved by iPiDA-LGE with different fusion coefficients on $\mathbb{S}_{\text{validation}}$. **b** and **c** show the ROC and PR curves obtained by different predictors on $\mathbb{S}_{\text{validation}}$, respectively. **d**–**f** show the binary distribution of true labels and association scores predicted by iPiDA-L, iPiDA-G, and iPiDA-LGE on $\mathbb{S}_{\text{validation}}$

Compared to iPiDA-G, iPiDA-L based on local graph learning plays a more important role in iPiDA-LGE and achieves higher predictive performance, attributing to its ability to capture specific contextual semantics for each target piRNA-disease pair and preventing from noise interference. (ii) iPiDA-LGE is superior to iPiDA-L and iPiDA-G regarding AUC and AUPR metrics, illustrating ensemble learning of local and global piRNA-disease association graphs contributes to performance improvement. (iii) iPiDA-L may predict some false positive associations, while iPiDA-G tends to predict potential piRNA-disease associations as negatives. iPiDA-LGE can obtain more discriminative association scores by integrating iPiDA-L and iPiDA-G.

Table 1 Performance comparison of iPiDA-LGE with baseline methods on $\mathbb{S}_{\text{independent}}$

Method	AUC	AUPR
iPiDA-A	0.6321	0.5838
iPiDA-L	0.8402	0.8231
iPiDA-G	0.8178	0.8151
iPiDA-LGE	0.8537	0.8497

Feature analysis of local and global graph representation

There are different types of piRNA-disease pair features including initial attribute features, local and global graph structural features. To analyze their contributions for identifying piRNA-disease associations, four predictors based on various features are constructed and compared. Table 1 lists the performance results obtained by iPiDA-LGE along with iPiDA-A, iPiDA-L, and iPiDA-G on $\mathbb{S}_{\text{independent}}$, where iPiDA-A is a random forest-based predictor trained by concatenating piRNA and disease attribute features. It is not surprising that the AUC and AUPR obtained by iPiDA-A is significantly lower than other predictors based on graph structural features. In addition, consistent with above observation on the benchmark dataset, iPiDA-LGE achieves superior performance owing to its local and global graph ensemble learning.

The different features extracted by iPiDA-A, iPiDA-G, and iPiDA-L are further investigated and visualized in Fig. 3, revealing the following observations: (i) Compared to the spliced attribute features, the graph structural features can capture hidden association pattern between piRNAs and diseases and perform stronger discriminative ability. (ii) The pair features extracted by iPiDA-L are more discriminative and expressive compared to those by iPiDA-G. The global graph representation in iPiDA-G learns general features for each piRNA or disease node, which are then used to construct pair representations. Therefore, it is limited in its ability to differentiate pairs that share the same target piRNA or disease. In contrast, local graph representation in iPiDA-L captures pair features from their specific local-contextual graphs. It cannot only detect specific patterns across different piRNA-disease pair types, but also extract refined contextual semantics for target piRNA-disease pairs. For example, (piR-hsa-24486, Renal cell carcinoma) and (piR-hsa-24486, Parkinson disease) are two positive associations, showing distinct heatmap patterns compared to other two negative pairs. Furthermore, the local graphs constructed for the positive associations exhibit subtle differences, conforming to the discovery that piR-hsa-24486 shows differentially downregulated and upregulated expression in the two target diseases [32, 33], respectively.

Performance comparison of various methods

Six cutting-edge methods include iPiDi-PUL [18], iPiDi-GCN [24], CLPiDA [28], iPiDi-SWGCN [27], ETGPDA [25], and PUTransGCN [26]. The web servers or source codes for these methods are readily available, facilitating an unbiased performance comparison. iPiDi-PUL identifies piRNA-disease associations using classical machine-learning algorithms with manual feature engineering

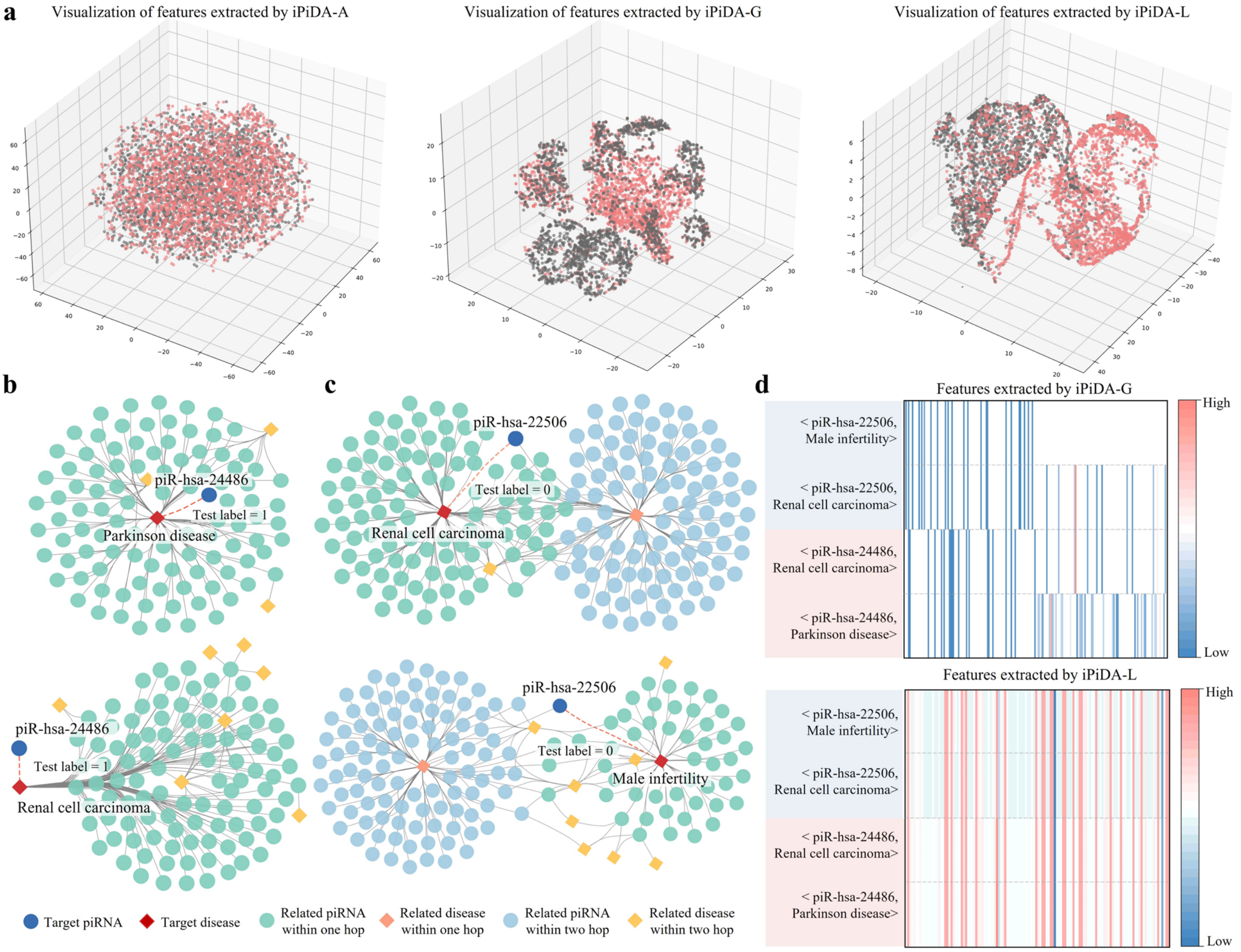


Fig. 3 Analysis of features extracted by local and global graph learning. **a** shows the t-SNE visualization of features extracted by iPiDA-A, iPiDA-G, and iPiDA-L, respectively. **b** and **c** show local context graphs for two example positive and two example negative piRNA-disease pairs in $\mathbb{S}_{\text{independent}}$, respectively. **d** shows the heatmap of features extracted by iPiDA-G and iPiDA-L for four example piRNA-disease pairs

process. The other five predictors adaptively learn pair features from the global piRNA-disease association graph using various graph neural networks.

To improve the reliability of method comparisons, we randomly partition the $\mathbb{S}_{\text{all}}^{+}$ 100 times according to the strategy described in section "Datasets," creating 100 independent test sets. The prediction results obtained from different methods across these 100 independent test sets are then compared. The average comparison results are listed in Table 2, while the statistical test results are shown in Fig. 4. The findings indicate that the local and global graph ensemble learning framework used in iPiDA-LGE demonstrates significantly superior performance across most metrics.

Leave-one-disease-out cross-validation is also conducted to evaluate the generalization ability of iPiDA-LGE. Different from focusing on uncovering missing associations within known diseases as previously discussed. In this scenario, all known associations with each target disease in $\mathbb{S}_{\text{all}}^{+}$ are firstly removed. The training set consists of known associations along with an equal number of unknown associations related to other 18 diseases.

Table 2 Performance comparison of different methods on 100 partitioned independent test sets

Methods	AUC	AUPR	F1	ACC	PRE	SPE	SEN
iPiDi-PUL	0.6749	0.6434	0.6771	0.5848	0.5556	0.2991	0.8705
iPiDA-GCN	0.7236	0.706	0.7023	0.6473	0.6086	0.4627	0.8319
CLPiDA	0.5555	0.5145	0.6771	0.5507	0.5285	0.1593	0.9422
iPiDA-SWGCN	0.8122	0.8211	0.7566	0.7476	0.7313	0.7106	0.7845
ETGPDA	0.8255	0.8232	0.7845	0.7707	0.7467	0.7100	0.8314
PUTransGCN	0.7990	0.8349	0.7821	0.7337	0.6818	0.5284	0.9389
iPiDA-LGE	0.8613	0.8637	0.8006	0.8066	0.8270	0.8370	0.7763

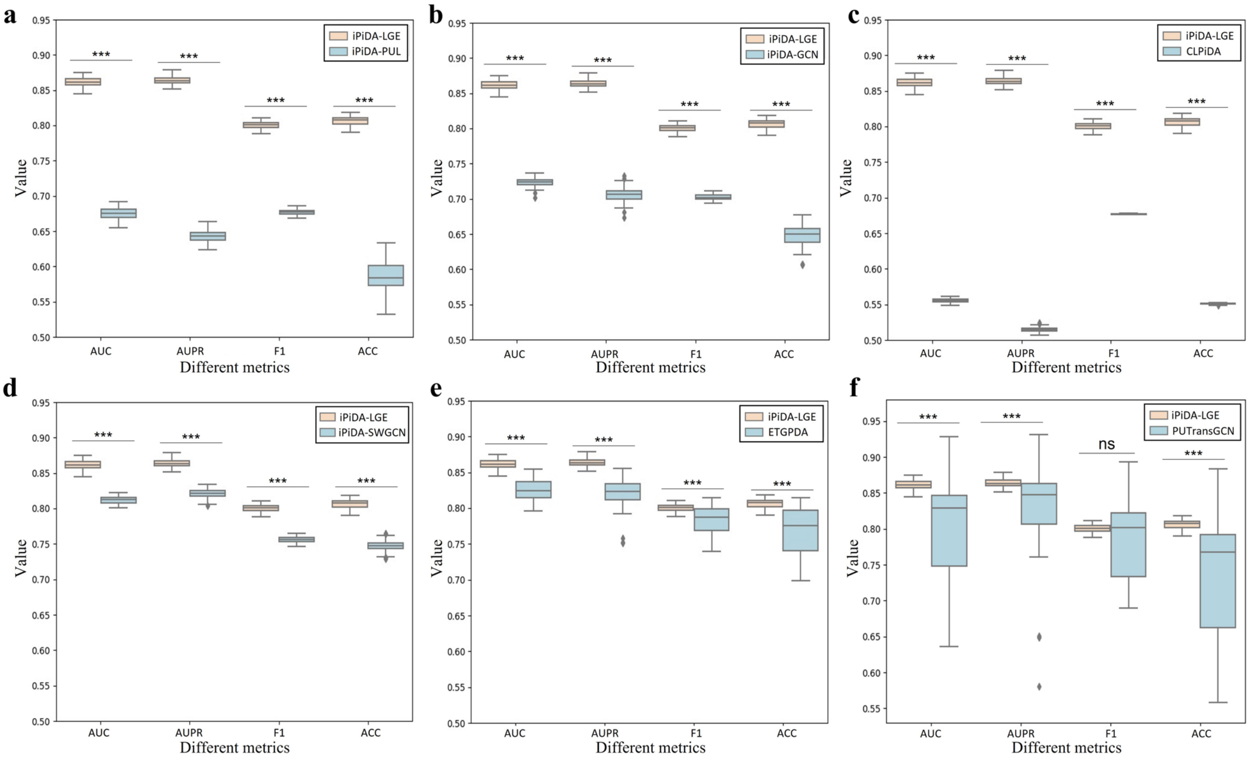


Fig. 4 Comparison of different methods across four comprehensive metrics on 100 partitioned independent test sets. Wilcoxon rank-sum test is used to calculate the statistical difference between two groups of results. Comparisons with *p* values < 0.05 are marked with *, *p* values < 0.01 with **, *p* values < 0.001 with ***, and "ns" indicates no significant difference

The trained model is then used to predict the positive and negative associations for each target disease. We compared the predicted results of different methods using leave-one-disease-out cross-validation. The average comparison results are shown in Additional file 1: Table S1, with the statistical test results shown in Additional file 1: Fig. S1. Compared to the scenario of identifying missing associations for known diseases, the performance of most methods in leave-one-disease-out cross-validation scenario is relatively unstable and shows decline, likely due to significant differences in the distribution of associations between different diseases and the incomplete network structure of the target disease. Nevertheless, iPiDA-LGE consistently outperforms other methods in comprehensive metrics such as AUC and AUPR and achieves superior or at least comparable performance in F1 score and accuracy.

Case study

Case studies are implemented to further investigate the prediction quality of disease-associated piRNAs identified by iPiDA-LGE. We choose five significant diseases, specifically "Parkinson's disease," "cardiovascular disease," "renal cell carcinoma," "Alzheimer's disease," and "male infertility." The top five predicted piRNAs associated with each target disease in the independent test dataset are provided in Table 3, indicating that all of them have biological literature supporting. The predicted disease-associated piRNAs show significantly differential expression between different sample groups. For instance, piR-hsa-28421 exhibits significantly elevated expression levels in cell samples derived from patients afflicted with Parkinson's disease [32]. PiR-hsa-4946 shows abnormally downregulated in cardiovascular disease, and its expression in cardiomyocytes is more than 4 times higher than that in cardiomyocyte derived cells [34]. PiR-hsa-26731 and piR-hsa-20294 have been shown to exhibit differential expression levels in metastatic renal cell tumors compared to non-metastatic tumors, with piR-hsa-26731 being over fourfold lower and piR-hsa-20294 being over 40-fold higher [33]. In brain tissues of Alzheimer's disease patients, piR-hsa-25783 exhibits an approximately eight-fold increase in expression compared to that of normal individuals [35]. PiR-hsa-32046 is significantly downregulated in the seminal plasma of infertile patients against the fertile control groups [36]. In addition, multiple single nucleotide polymorphism (SNP) variations have been identified on most predicted disease-associated piRNAs, such as rs891709032, rs532580755, and rs998079688 have been identified on piR-hsa-4946 and rs149511548, rs898504240, rs938072950, and rs555921708 have been identified on piR-hsa-26731. These mutations may potentially trigger abnormal expression and affect their biological function [7].

Table 3 The top five piRNAs associated with different diseases predicted by iPiDA-LGE

Disease	Rank	piRNA[a]	Expression	Evidence[b]
Parkinson's disease	1	piR-hsa-28421	Downregulated	29986767
	2	**piR-hsa-6841**	Downregulated	29986767
	3	piR-hsa-183	Downregulated	29986767
	4	piR-hsa-16796	Downregulated	29986767
	5	**piR-hsa-859**	Downregulated	29986767
Cardiovascular disease	1	**piR-hsa-4946**	Downregulated	27131603
	2	**piR-hsa-28185**	Downregulated	27131603
	3	piR-hsa-27715	Downregulated	27131603
	4	piR-hsa-2219	Downregulated	27131603
	5	**piR-hsa-27854**	Downregulated	27131603
Renal cell carcinoma	1	**piR-hsa-26731**	Downregulated	26071182
	2	**piR-hsa-27910**	Upregulated	26071182
	3	**piR-hsa-15980**	Upregulated	26071182
	4	**piR-hsa-31522**	Downregulated	26071182
	5	piR-hsa-20294	Upregulated	26071182
Alzheimer disease	1	**piR-hsa-25783**	Upregulated	28127595
	2	piR-hsa-27438	Upregulated	28654860
	3	**piR-hsa-27622**	Upregulated	28127595
	4	**piR-hsa-1100**	Upregulated	26934981
	5	**piR-hsa-16724**	Downregulated	28654860
Male infertility	1	**piR-hsa-32046**	Downregulated	27068805
	2	piR-hsa-7144	Downregulated	27068805
	3	**piR-hsa-7264**	Downregulated	27068805
	4	**piR-hsa-18608**	Downregulated	27068805
	5	**piR-hsa-18727**	Downregulated	27068805

[a] The piRNAs with multiple SNPs are highlighted in bold

[b] The PMIDs of supporting literature in PubMed have been provided

To evaluate the scalability of the proposed graph ensemble learning framework, it is applied to other two bio-entity association prediction tasks: miRNA-disease and circRNA-disease association prediction. For the miRNA-disease prediction task, HMDD v3.2 data [37] including 12,446 known associations between 853 miRNAs and 591 diseases is used to construct and evaluate iPiDA-LGE model. For the circRNA-disease prediction task, CircR2Disease v2.0 data [38] including 1820 experimentally validated associations covering 1429 circRNAs and 122 diseases is used to construct and evaluate proposed framework. In the both bio-entity association prediction tasks, we randomly extracted 20% of the known associations as test positive samples, while the remaining 80% were used as training positive samples. Then, we randomly selected an equal number of unknown pairs as negative samples for both the test and training sets. To ensure the reliability of the model performance evaluation, this dataset construction strategy was repeated 100 times, and the average performance metrics obtained by our proposed framework for both bio-entity association prediction tasks are listed in Additional file 1: Table S2. Specifically, the proposed graph ensemble learning framework achieved a high AUC of 0.9435 and AUPR of 0.9437 in the miRNA-disease association prediction task, and competitive performance with an AUC of 0.8623 and AUPR of 0.8621 in the circRNA-disease association prediction task. These results demonstrate the scalability of our proposed graph ensemble learning framework and its potential for further performance optimization in domain-specific tasks.

To facilitate researchers' experimental validation and downstream analysis, we further investigated the top-ranked disease-associated miRNAs and circRNAs predicted by iPiDA-LGE. For the miRNA-disease prediction task, a graph ensemble predictor is constructed using all 12,446 known associations in HMDD v3.2 [37]. The trained predictor identified miRNAs associated with five significant diseases: "Alzheimer's disease," "breast neoplasms," "colorectal neoplasms," "leukemia," and "liver neoplasms." Additional file 1: Table S3 lists the top ten predicted miRNAs for each disease, along with their corresponding validation information. Notably, 42 of these predicted associations can be corroborated by HMDD v4.0 [39]. In contrast to miRNAs, circRNAs currently have fewer known associations with diseases. For the circRNA-disease association prediction task, ten important diseases are selected: "colorectal cancer," "breast carcinoma," "bladder carcinoma," "lung adenocarcinoma," "Alzheimer's disease," "esophageal squamous cell carcinoma," "pancreatic cancer," "glioblastoma," "hepatoblastoma," and "ovarian cancer." During predictor construction, we implemented strict leave-one-disease-out cross-validation by excluding all known circRNA-disease associations for each target disease from the CircR2Disease v2.0 data. Additional file 1: Table S4 lists the top five predicted circRNAs for each target disease, along with relevant evidence. A total of 36 predictions were validated by CircR2Disease v2.0.

In general, iPiDA-LGE can predict novel piRNA-disease associations and provide candidate piRNAs for biological experiments to study pathological mechanisms. Moreover, the proposed local and global graph ensemble learning framework provides innovative computational insights and can be optimized for miRNA-disease and circRNA-disease association prediction tasks.

Conclusions

Detecting piRNA-disease associations holds significant importance in understanding disease mechanisms and biomarker discovery. In this study, we propose a new computational method named iPiDA-LGE to detect piRNA-disease associations. In contrast to competing methods, it predominantly possesses the following advantages: (i) The global graph learning module incorporates side information like piRNA sequence and disease ontology and learns various basic predictors to construct supplementary heterogeneous association network. Therefore, the apparent sparsity issue of original association can be alleviated by enriching biological semantics. (ii) Diverging from general pair features obtained by global graph learning, the local graph learning module in iPiDA-LGE considers the specific functional mechanism of piRNAs in different diseases and encodes each target piRNA-disease pair as a local graph. As a result, it can learn a more discriminative summary representation by capturing specific contextual information. (iii) Lastly, iPiDA-LGE integrates the local and global graph representation learning, which can simultaneously achieve refined inferences based on local graphs and overarching judgments derived from global graphs, leading to an enhancement in predictive performance.

While iPiDA-LGE achieves superior capability in identifying reliable disease-related piRNAs, it is important to note that further validation through biological experiments is needed for some of the predicted associations that have not been confirmed. In addition, iPiDA-LGE may encounter challenges in elucidating comprehensive insights into how piRNAs influence disease development. Therefore, future enhancements will focus on providing more fine-grained guidance for understanding disease pathology at the piRNA level. Additional phenotypic and genotypic information, such as SNPs, expression profiles, piRNA targets, and multiple bio-entity associations, can be incorporated to construct a more comprehensive heterogeneous biological network. Moreover, the introduction of denoising techniques, causal inference, and interpretable mechanisms will improve the model robustness and biological implication.

Methods

Datasets

To construct a heterogeneous piRNA-disease graph, piRNA sequence and disease ontology information are downloaded from piRBase [40] and Disease Ontology [41] databases, respectively. In addition, the known associations between piRNAs and diseases are retrieved from MNDR v3.0 [42]. Following the removal of duplicate entries, a total of 11,981 experimentally validated associations involving 10,149 piRNAs and 19 diseases are compiled. The dataset can be denoted by:

$$\begin{cases} \mathbb{S}_{\text{all}} = \mathbb{S}_{\text{all}}^{+} \bigcup \mathbb{S}_{\text{all}}^{-} \\ \mathbb{S}_{\text{all}}^{+} = \mathbb{S}_{\text{independent}}^{+} \cup \mathbb{S}_{\text{benchmark}}^{+} \\ \mathbb{S}_{\text{independent}} = \mathbb{S}_{\text{independent}}^{+} \bigcup \mathbb{S}_{\text{independent}}^{-} \end{cases} \tag{1}$$

where $\mathbb{S}_{\text{all}}^{+}$ represents the positive set encompassing 10,149 known piRNA-disease associations, while $\mathbb{S}_{\text{all}}^{-}$ is the negative set composed of all possible unknown piRNA-disease pairs between 10,149 piRNAs and 19 diseases. Independent test set $\mathbb{S}_{\text{independent}}$ is constructed for testing different computational methods. Sequentially partition known associations related to each specific disease in $\mathbb{S}_{\text{all}}^{+}$ into five folds. One fold from all 19 diseases is integrated to form the independent positive set $\mathbb{S}_{\text{independent}}^{+}$, which is then combined with a randomly selected equal number of negative pairs from $\mathbb{S}_{\text{all}}^{-}$ to constitute the independent negative set $\mathbb{S}_{\text{independent}}^{-}$. The remaining four folds of positive associations, along with an equal number of negative pairs, are used to create $\mathbb{S}_{\text{benchmark}}$. To train predictor and optimize hyper-parameters, $\mathbb{S}_{\text{benchmark}}$ is further divided and can be denoted as:

$$\begin{cases} \mathbb{S}_{\text{benchmark}} = \mathbb{S}_{\text{train}} \cup \mathbb{S}_{\text{validation}} \\ \mathbb{S}_{\text{train}} = \mathbb{S}_{\text{train}}^{+} \cup \mathbb{S}_{\text{train}}^{-} \\ \mathbb{S}_{\text{validation}} = \mathbb{S}_{\text{validation}}^{+} \cup \mathbb{S}_{\text{validation}}^{-} \end{cases} \tag{2}$$

where the positive set in $\mathbb{S}_{\text{benchmark}}$ is divided into five folds, four folds constitute $\mathbb{S}_{\text{train}}^{+}$, and $\mathbb{S}_{\text{validation}}^{+}$ is composed of remaining portion. $\mathbb{S}_{\text{train}}^{-}$ and $\mathbb{S}_{\text{validation}}^{-}$ are comprised of randomly selected negative pairs, equating in number to the corresponding positive set. The datasets can be downloaded from http://bliulab.net/iPiDA-LGE/download/.

Method overview

As illustrated in Fig. 5, the iPiDA-LGE framework encompasses three primary steps: graph construction, graph representation, and association prediction. Detailed elaboration of each step will be provided in subsequent sections.

Heterogeneous graph construction

To enhance the biological semantics of original piRNA-disease bipartite graph, the heterogeneous association graph is constructed and supplemented followed by our previous study [27]. Firstly, piRNA sequence similarity

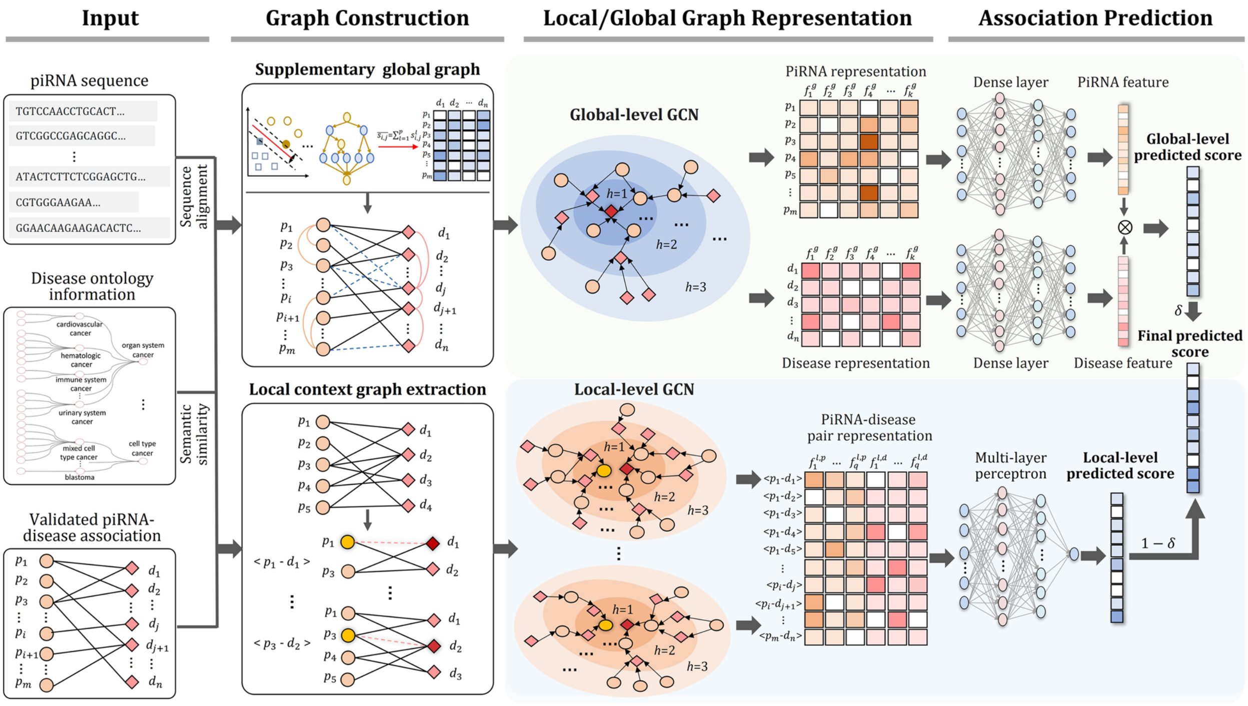
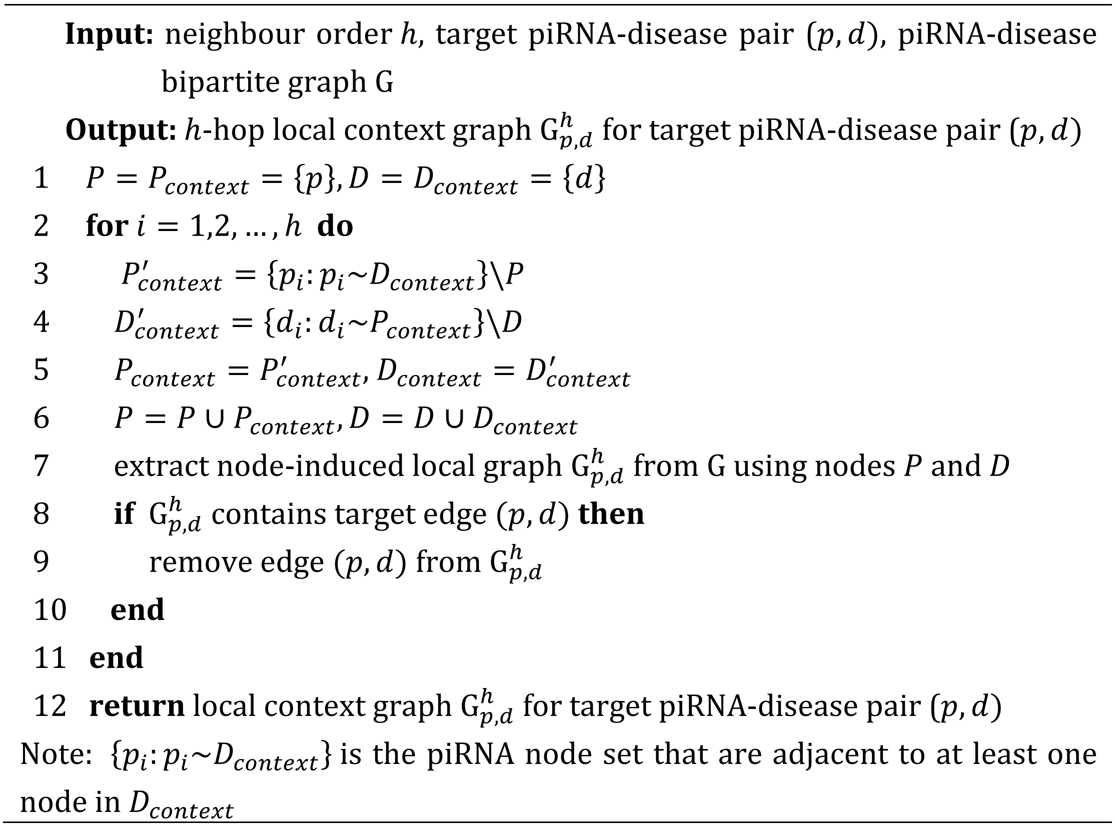


Fig. 5 The framework of iPiDA-LGE. There exist three primary processes: (i) Graph construction. The global graph is constructed and supplemented with piRNA sequences, disease ontology knowledge, and validated piRNA-disease relationships. In addition, the local context graph for each target pair is extracted from original bipartite graph. (ii) Local/global graph representation. The representations of piRNA and disease nodes are captured by global-level GCN, while the pair representations are obtained by local-level GCN. (iii) Association prediction. The dense layer and multi-layer perceptron are applied to reduce feature dimensionality and calculate association scores. Finally, the global-level and local-level association scores are integrated with different weight coefficients to predict the relationships between piRNAs and diseases

is obtained based on Smith-Waterman local alignment algorithm [43–45], and disease semantic similarity is calculated utilizing the DOSE package along with disease ontology knowledge [46]. Then, the global piRNA-disease association graph can be constructed by integrating three different types of edges: piRNA similarities, disease similarities, and piRNA-disease relationships. For alleviating limited neighbor information aggregation caused by high-sparse known association, the supplementarily weighted global graph is constructed. Fifteen predictors based on different machine-leaning algorithms and training samples are learned to score unknown piRNA-disease pairs, and the supplementary edges weighted with average scores predicted by fifteen basic predictors can be incorporated into original piRNA-disease association graph. The details of constructing a supplementary global piRNA-disease association graph can be referred to our previous study [27].

A local context graph for each target piRNA-disease pair is constructed for capturing local-specific structural features. Given the original piRNA-disease bipartite graph G, the local context graphs are constructed based on the enclosing subgraph extraction strategy [47]. Specifically, the piRNA and disease in a target pair (p, d) are considered as core nodes, and their h-hop contextual neighbors are extracted by Breadth-First Search procedure. Then, the node-induced local graph $G^{h}_{p,d}$ can be extracted from G using corresponding core and context nodes. Particularly, if $G^{h}_{p,d}$ contains target piRNA-disease association (p, d), we should remove it to prevent information leakage. The main steps for extracting local context piRNA-disease graph are described in Algorithm 1.

Algorithm 1. Local context graph extraction



```
Input: neighbour order $h$, target piRNA-disease pair $(p, d)$, piRNA-disease
        bipartite graph G
Output: $h$-hop local context graph $G^{h}_{p,d}$ for target piRNA-disease pair $(p, d)$
$P = P_{context} = \{p\}, D = D_{context} = \{d\}$
for $i = 1,2, \ldots, h$ do
    $P'_{context} = \{p_i : p_i \sim D_{context}\} \backslash P$
    $D'_{context} = \{d_i : d_i \sim P_{context}\} \backslash D$
    $P_{context} = P'_{context}, D_{context} = D'_{context}$
    $P = P \cup P_{context}, D = D \cup D_{context}$
    extract node-induced local graph $G^{h}_{p,d}$ from G using nodes $P$ and $D$
    if $G^{h}_{p,d}$ contains target edge $(p, d)$ then
        remove edge $(p, d)$ from $G^{h}_{p,d}$
    end
end
return local context graph $G^{h}_{p,d}$ for target piRNA-disease pair $(p, d)$
Note: $\{p_i : p_i \sim D_{context}\}$ is the piRNA node set that are adjacent to at least one
node in $D_{context}$
```

The extracted local context piRNA-disease graphs are independent of global graph; a node labeling strategy is further applied to differentiate the node types and their hop stages [47]. The core nodes representing piRNAs are designated as 0, while those for diseases are labeled as 1. Other contextual nodes are assigned labels based on their node types and hop stages. A piRNA node at the ith hop can be labeled as $2i$, while a disease node at the ith hop can be labeled as $2i+1$. Therefore, the nodes in local context graphs have distinguishable labels for following local-level graph representation learning.

Global-level graph representation

Graph convolutional network (GCN) exhibits powerful capability to excavate intricate topological properties across diverse networks [48] and has been effectively applied to different bioinformatics tasks [49–52]. To capture global neighborhood features for piRNAs and diseases, GCN is adopted on the supplementary global association graph. At the lth layer, the node feature matrix $\mathbf{H}^l \in \mathrm{R}^{(m+n)\times k}$ obtained by GCN can be represented as follows:

$$\mathbf{H}^l = \sigma\left(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}_g}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{l-1}\mathbf{W}^{l-1}\right) \tag{3}$$

and

$$\widetilde{\mathbf{A}_g} = \mathbf{I} + \mathbf{A}_g \tag{4}$$

$$\widetilde{\mathbf{D}}(i,i) = \sum_j \widetilde{\mathbf{A}_g}(i,j) \tag{5}$$

$$\mathbf{A}_g = \begin{bmatrix} \mathbf{S}_\mathrm{p} & \mathbf{A}_\mathrm{s} \\ \mathbf{A}_\mathrm{s}^\mathbf{T} & \mathbf{S}_\mathrm{d} \end{bmatrix} \tag{6}$$

where $\mathbf{A}_g \in \mathrm{R}^{(m+n)\times(m+n)}$ represents the piRNA-disease adjacency matrix for the global association graph, and $\mathbf{I}$ is the identity matrix. $\mathbf{S}_\mathrm{p} \in \mathrm{R}^{m\times m}$ and $\mathbf{S}_\mathrm{d} \in \mathrm{R}^{n\times n}$ denote matrices of piRNA and disease similarity, respectively, and $\mathbf{A}_\mathrm{s} \in \mathrm{R}^{m\times n}$ denotes the adjacency matrix for the supplementary piRNA-disease bipartite graph. σ is defined as ReLU activation function. To acquire the attribute features of piRNA and disease nodes, we utilize the random walk with restart [53] algorithm on the matrices $\mathbf{S}_\mathrm{p}$ and $\mathbf{S}_\mathrm{d}$, and then dense layers are structured to uncover their hidden features. Thus, $\mathbf{H}^l$ is initialized by merging the attribute features of piRNAs and diseases, both with identical dimension. More details of global-level graph representation learning can be referred to our previous study [27].

Local-level graph representation

Different from global-level graph representation, local-level graph representation learning aims to capture local contextual features for each piRNA-disease pair. Given a local context graph $\mathbf{G}_{p,d}$ for target piRNA-disease pair (p,d), each node feature vector x_i^l learned at the lth layer can be calculated by [47]:

$$\mathrm{x}_i^l = \sigma\left(\mathbf{W}_1^{l-1}\mathrm{x}_i^{l-1} + \sum_{j\in\mathbb{N}_i}\mathrm{x}_j^{l-1}\mathbf{W}_2^{l-1}\right) \tag{7}$$

where $\mathbf{W}_1^{l-1}$ and $\mathbf{W}_2^{l-1}$ are learnable parameter matrices, $\mathbb{N}_i$ represents the set of neighbors of node i, while σ is defined as tanh activation function. For each node, its final representation x_i can be obtained by concatenating L message passing layers:

$$\mathrm{x}_i = \mathrm{concat}(\mathrm{x}_i^1, \mathrm{x}_i^2, \ldots, \mathrm{x}_i^L) \tag{8}$$

Then, considering the greater importance of core nodes compared to other context nodes for the target piRNA-disease pair, the concatenate pooling layer is set to capture the local contextual features for $\mathbf{G}_{p,d}$:

$$\mathrm{x}_{p,d} = \mathrm{concat}(\mathrm{x}_p, \mathrm{x}_d) \tag{9}$$

where x_p and x_d are the final representations of target piRNA p and disease d, respectively.

PiRNA-disease association prediction

The global-level graph representation learning can capture general node features by considering the heterogeneous association and homogeneous similarity in the supplementary global graph. The local-level graph representation learning encodes each piRNA-disease pair as a sub-graph to aggregate refined contextual semantics. For comprehensively exploring association patterns between piRNAs and diseases, we integrate the local- and global-level graph representations to predict piRNA-disease association scores.

For the node features obtained from supplementary global piRNA-disease association graph, three dense layers with 400, 200, and 100 neurons are designed to extract high-level features. Then, for a target piRNA-disease pair (p,d), its association score can be predicted by inner product $\mathbf{Y}_{p,d}^{\mathrm{global}} = \mathbf{h}_{p'}\mathbf{h}_{d'}{}^{\mathrm{T}}$, where $\mathbf{h}_{p'}$ and $\mathbf{h}_{d'}$ denote the features for target piRNA p and disease d. To avoid prediction bias problem, a variant loss function based on mean square error [54] is set:

$$\mathbf{L}^{\mathrm{global}} = \|\mathbf{A}^{'} - \mathbf{Y}^{\mathrm{global}}\|_{\mathrm{F}}^2 + \mu\|\mathbf{W}\|_2^2 \tag{10}$$

and

$$\mathbf{A}'_{p,d} = \begin{cases} 0 \ \text{if} \ (p,d) \in \mathbb{S}_{\text{independent}} \\ 0 \qquad \text{if} \ \mathbf{A}_{p,d} = 0 \\ \alpha \qquad \text{otherwise} \end{cases} \tag{11}$$

$\mathbf{A} \in \mathrm{R}^{m \times n}$ denotes the adjacency matrix for the original piRNA-disease bipartite graph. $\mathbf{A}'$ and μ denote the variant adjacency matrix and decay factor, respectively. μ aims to regulate the regularization of trainable parameter matrix $\mathbf{W}$ in global-level graph representation learning.

For the piRNA-disease pair feature extracted from local context graph, the multilayer perceptron with 128 and 1 neurons is trained to predict piRNA-disease association scores:

$$\mathbf{Y}^{\text{local}} = \mathbf{W}_2 \bullet \text{ReLU}(\mathbf{W}_1\mathbf{X} + b_1) + b_2 \tag{12}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are two learnable parameter matrices, while $\mathbf{X}$ denotes the piRNA-disease pair feature matrix. We employ the mean square error loss function to guide local context graph representation learning.

Given a test piRNA-disease pair (p,d), its final association score is computed by:

$$\mathbf{Y}_{p,d} = \delta \mathbf{Y}^{\text{global}}_{p,d} + (1-\delta)\mathbf{Y}^{\text{local}}_{p,d} \tag{13}$$

where δ is a parameter controlling the balance between global- and local-level predicted association scores.

Performance evaluation

Different indicators, specifically the area under the receiver operating characteristics curve (AUC), area under the precision-recall curve (AUPR), F1 score, accuracy (ACC), precision (PRE), specificity (SPE), and sensitivity (SEN), are utilized to comprehensively estimate predictive performance [55–57]. AUC and AUPR respectively emphasize the trade-offs between sensitivity and specificity, and precision and recall across different threshold settings [58, 59].

Abbreviations

piRNA	PIWI-interacting RNA
GCN	Graph convolutional network
AUC	Area under the receiver operating characteristics curve
AUPR	Area under the precision-recall curve
ACC	Accuracy
PRE	Precision
SPE	Specificity
SEN	Sensitivity

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02221-y.

Additional file 1. Supplementary information. The performance comparison of different methods across metrics using leave-one-disease-out cross-validation is listed in Table S1. The comparison of different methods across four metrics using leave-one-disease-out cross-validation is shown in Fig. S1. The performance metrics obtained by iPiDA-LGE for other two bio-entity association prediction tasks is listed in Table S2. The top ten miRNAs associated with different diseases predicted by iPiDA-LGE is listed in Table S3. The top five circRNA associated with different diseases predicted by iPiDA-LGE is listed in Table S4.

Acknowledgements

We are very much indebted to the three anonymous reviewers, whose constructive comments are very helpful for enhancing the quality of this paper.

Authors' contributions

H.W was involved in the implementation, manuscript writing, and correcting, J.L.H was involved in the implementation and programming, Y.M.L was involved in the manuscript correcting. AKS was involved in the manuscript correcting. B.L and H.W conceived the project and were involved in the manuscript writing and correcting. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 62372041, 62302359, 62302316), Zhongguancun Academy (Project No. 20240101), the Entrepreneurship Talent Project of Qin Chuang Yuan under Grant (No. QCYRCXM2023-115), and Shenzhen Science and Technology Program (No. RCBS20221008093227027). AKS is supported by HSE basic research program.

Data availability

The iPiDA-LGE webserver is accessible at http://bliulab.net/iPiDA-LGE. The code and datasets used in this study can be found in online repositories. The name of the repository and accession number for the data reported in this paper is zenodo, https://doi.org/10.5281/zenodo.15080670 [60]. All data generated or analysed during this study are included in this published article, its supplementary information files, and publicly available repositories.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 May 2024 Accepted: 23 April 2025
Published online: 09 May 2025

References

1. Seto AG, Kingston RE, Lau NC. The coming of age for Piwi proteins. Mol Cell. 2007;26(5):603–9.
2. Wang X, Ramat A, Simonelig M, Liu MF. Emerging roles and functional mechanisms of PIWI-interacting RNAs. Nat Rev Mol Cell Biol. 2023;24(2):123–41.
3. Guo X, Huang Z, Ju F, Zhao C, Yu L. Highly accurate estimation of cell type abundance in bulk tissues based on single-cell reference and domain adaptive matching. Adv Sci. 2024;11(7):2306329.
4. Zuo Y, Zou Q, Lin J, Jiang M, Liu X. 2lpiRNApred: a two-layered integrated algorithm for identifying piRNAs and their functions based on LFE-GM feature selection. RNA Biol. 2020;17(6):892–902.
5. Li YH, Xu JY, Tao L, Li XF, Li S, Zeng X, Chen SY, Zhang P, Qin C, Zhang C. SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. PLoS ONE. 2016;11(8): e0155290.
6. Jiang L, Guo F, Tang J, Leng S, Ness S, Ye F, Kang H, Samuels DC, Guo Y. Global autozygosity is associated with cancer risk, mutational signature and prognosis. Cancers. 2020;12(12):3646.

7. Liu Y, Li A, Zhu Y, Pang X, Hei X, Xie G, Wu F-X. piRSNP: a database of piRNA-related SNPs and their effects on cancer-related piRNA functions. Curr Bioinform. 2023;18(6):509–16.
8. Qiao J, Jin J, Yu H, Wei L. Towards retraining-free RNA modification prediction with incremental learning. Inform Sci. 2024;660:120105.
9. Chen S, Ben S, Xin J, Li S, Zheng R, Wang H, Fan L, Du M, Zhang Z, Wang M. The biogenesis and biological function of PIWI-interacting RNA in cancer. J Hematol Oncol. 2021;14(1):93.
10. Zhang TJ, Chen L, Li RZ, Liu N, Huang XB, Wong G. PIWI-interacting RNAs in human diseases: databases and computational models. Brief Bioinform. 2022;23(4):bbac217. https://doi.org/10.1093/bib/bbac217.
11. Xin J, Du M, Jiang X, Wu Y, Ben S, Zheng R, Chu H, Li S, Zhang Z, Wang M. Systematic evaluation of the effects of genetic variants on PIWI-interacting RNA expression across 33 cancer types. Nucleic Acids Res. 2021;49(1):90–7.
12. Wu WS, Brown JS, Chen TT, Chu YH, Huang WC, Tu S, Lee HC. piRTarBase: a database of piRNA targeting sites and their roles in gene regulation. Nucleic Acids Res. 2019;47(D1):D181–7.
13. Guo B, Li D, Du L, Zhu X. piRNAs: biogenesis and their potential roles in cancer. Cancer Metastasis Rev. 2020;39(2):567–75.
14. Zhou J, Xie H, Liu J, Huang R, Xiang Y, Tian D, Bian E. PIWI-interacting RNAs: critical roles and therapeutic targets in cancer. Cancer Lett. 2023;562: 216189.
15. Tang W, Wan SX, Yang Z, Teschendorff AE, Zou Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics. 2018;34(3):398–406.
16. Wang R, Jiang Y, Jin J, Yin C, Yu H, Wang F, Feng J, Su R, Nakai K, Zou Q. DeepBIO: an automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis. Nucleic Acids Res. 2023;51(7):3017–29.
17. Ai C, Yang H, Ding Y, Tang J, Guo F. Low rank matrix factorization algorithm based on multi-graph regularization for detecting drug-disease association. Ieee Acm T Comput Bi. 2023;20(5):3033–43.
18. Wei H, Xu Y, Liu B. iPiDi-PUL: identifying Piwi-interacting RNA-disease associations based on positive unlabeled learning. Brief Bioinform. 2021;22(3):bbaa058.
19. Zhang W, Hou J, Liu B. iPiDA-LTR: identifying piwi-interacting RNA-disease associations based on learning to rank. Plos Comput Biol. 2022;18(8): e1010404.
20. Wei H, Ding Y, Liu B. iPiDA-sHN: identification of Piwi-interacting RNA-disease associations by selecting high quality negative samples. Comput Biol Chem. 2020;88: 107361.
21. Yi HC, You ZH, Huang DS, Kwoh CK. Graph representation learning in bioinformatics: trends, methods and applications. Brief Bioinform. 2022;23(1):bbab340.
22. Niu M, Wang C, Zhang Z, Zou Q. A computational model of circRNA-associated diseases based on a graph neural network: prediction and case studies for follow-up experimental validation. BMC Biol. 2024;22(1):24.
23. Yan K, Lv HW, Guo YC, Peng W, Liu B. sAMPpred-GAT: prediction of antimicrobial peptide by graph attention network and predicted peptide structure. Bioinformatics. 2023;39(1):btac715.
24. Hou J, Wei H, Liu B. iPiDA-GCN: identification of piRNA-disease associations based on graph convolutional network. Plos Comput Biol. 2022;18(10): e1010671.
25. Meng X, Shang J, Ge D, Yang Y, Zhang T, Liu JX. ETGPDA: identification of piRNA-disease associations based on embedding transformation graph convolutional network. BMC Genomics. 2023;24(1):279.
26. Chen Q, Zhang L, Liu Y, Qin Z, Zhao T. PUTransGCN: identification of piRNA–disease associations based on attention encoding graph convolutional network and positive unlabelled learning. Brief Bioinform. 2024;25(3):bbae144.
27. Hou J, Wei H, Liu B. iPiDA-SWGCN: identification of piRNA-disease associations based on supplementarily weighted graph convolutional network. Plos Comput Biol. 2023;19(6): e1011242.
28. Hu X, Zhang Y, Zhang J, Deng L. CLPiDA: a contrastive learning approach for predicting potential PiRNA-disease associations. In: International Conference on Bioinformatics and Biomedicine (BIBM). Istanbul: IEEE; 2023. p.159-164. https://doi.org/10.1109/BIBM58861.2023.10385399.
29. Wang L, Li Z-W, Hu J, Wong L, Zhao B-W, You Z-H. A PiRNA-disease association model incorporating sequence multi-source information with graph convolutional networks. Appl Soft Comput. 2024;157: 111523.
30. Liu M, Li C, Chen R, Cao D, Zeng X. Geometric deep learning for drug discovery. Expert Syst Appl. 2023;240:122498.
31. Ma T, Lin X, Song B, Philip SY, Zeng XJIToK, Engineering D. Kg-mtl: knowledge graph enhanced multi-task learning for molecular interaction. Ieee T Knowl Data En. 2023;35(7):7068–81.
32. Schulze M, Sommer A, Plötz S, Farrell M, Winner B, Grosch J, Winkler J, Riemenschneider MJ. Sporadic Parkinson's disease derived neuronal cells show disease-specific mRNA and small RNA signatures with abundant deregulation of piRNAs. Acta Neuropathol Commun. 2018;6(1):58.
33. Busch J, Ralla B, Jung M, Wotschofsky Z, Trujillo-Arribas E, Schwabe P, Kilic E, Fendler A, Jung K. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. J Exp Clin Cancer Res. 2015;34(1):61.
34. Vella S, Gallo A, Lo Nigro A, Galvagno D, Raffa GM, Pilato M, Conaldi PG. PIWI-interacting RNA (piRNA) signatures in human cardiac progenitor cells. Int J Biochem Cell Biol. 2016;76:1–11.
35. Roy J, Sarkar A, Parida S, Ghosh Z, Mallick B. Small RNA sequencing revealed dysregulated piRNAs in Alzheimer's disease and their probable role in pathogenesis. Mol Biosyst. 2017;13(3):565–76.
36. Hong YT, Wang C, Fu Z, Liang HW, Zhang SY, Lu ML, Sun W, Ye C, Zhang CY, Zen K, et al. Systematic characterization of seminal plasma piRNAs as molecular biomarkers for male infertility. Sci Rep-Uk. 2016;6:6.
37. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, Zhou Y, Cui Q. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Res. 2019;47(D1):D1013–7.
38. 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 Bioinform. 2022;20(3):435–45.
39. Cui C, Zhong B, Fan R, Cui Q. HMDD v4.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Res. 2024;52(D1):D1327–32.
40. Wang J, Shi Y, Zhou H, Zhang P, Song T, Ying Z, Yu H, Li Y, Zhao Y, Zeng X, et al. piRBase: integrating piRNA annotation in all aspects. Nucleic Acids Res. 2022;50(D1):D265–72.
41. Bello SM, Shimoyama M, Mitraka E, Laulederkind SJF, Smith CL, Eppig JT, Schriml LM. Disease Ontology: improving and unifying disease annotations across species. Dis Model Mech. 2018;11(3):dmm032839.
42. Ning L, Cui T, Zheng B, Wang N, Luo J, Yang B, Du M, Cheng J, Dou Y, Wang D. MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation. Nucleic Acids Res. 2021;49(D1):D160–4.
43. Yau SST, Zhao X, Tian K, Yu H. Sequence alignment. In: Yau SST, Zhao X, Tian K, Yu H, editiors. Mathematical principles in bioinformatics. Cham: Springer Nature Switzerland; 2023. p. 27–42.
44. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-Pro: support bio-sequence machine for proteins. Sci China Inf Sci. 2024;67(11):212106.
45. Tang FR, Chao JN, Wei YM, Yang FL, Zhai YX, Xu L, Zou Q. HAlign 3: fast multiple alignment of ultra-large numbers of similar DNA/RNA sequences. Mol Biol Evol. 2022;39(8):msac166.
46. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015;31(4):608–9.
47. Zhang M, Chen Y. Inductive matrix completion based on graph neural networks. In: International Conference on Learning Representations (ICLR). Virtual Conference; 2020:917.
48. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR). Toulon: 2017: poster.
49. Sun X, Wang B, Zhang J, Li M. Partner-specific drug repositioning approach based on graph convolutional network. Ieee J Biomed Health. 2022;26(11):5757–65.
50. Bai T, Yan K, Liu B. DAmiRLocGNet: miRNA subcellular localization prediction by combining miRNA–disease associations and graph convolutional networks. Brief Bioinform. 2023;24(4):bbad212.
51. Song Q, Su J, Zhang W. scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. Nat Commun. 2021;12(1):3826.
52. Zhu H, Hao H, Yu L. Identifying disease-related microbes based on multi-scale variational graph autoencoder embedding Wasserstein distance. BMC Biol. 2023;21(1):294.
53. Tong H, Faloutsos C, Pan JY. Fast random walk with restart and its applications. In: International Conference on Data Mining (ICDM). Hong Kong: 2006. p. 613–622. https://doi.org/10.1109/ICDM.2006.70.

54. Han P, Yang P, Zhao P, Shang S, Liu Y, Zhou J, Gao X, Kalnis P. GCN-MF: disease-gene association identification by graph convolutional networks and matrix factorization. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019. p. 705–713.
55. Zhu W, Yuan SS, Li J, Huang CB, Lin H, Liao B. A first computational frame for recognizing heparin-binding protein. Diagnostics. 2023;13(14):2465.
56. Momanyi BM, Zhou YW, Grace-Mercure BK, Temesgen SA, Basharat A, Ning L, Tang L, Gao H, Lin H, Tang H. SAGESDA: Multi-GraphSAGE networks for predicting SnoRNA-disease associations. Curr Res Struct Biol. 2024;7: 100122.
57. Alhatemi RAJ, Savaş S. A weighted ensemble approach with multiple pre-trained deep learning models for classification of stroke. Medinformatics. 2023;1(1):10–9.
58. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
59. Flach P, Kull M. Precision-recall-gain curves: PR analysis done right. In: Advances in neural information processing systems. 2015. p. 28.
60. Wei H. iPiDA-LGE: a local and global graph ensemble learning framework for identifying piRNA-disease associations. 2025. https://doi.org/10.5281/zenodo.15080670.

Publisher's Note