MiRGraph: A transformer-based feature learning approach to identify microRNA-target interactions by integrating heterogeneous graph network and sequence information*

Pei Liu[1], Ying Liu[1], Jiawei Luo[1(✉)], and Yue Li[2(✉)]

[1] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410083, China
luojiawei@hnu.edu.cn

[2] School of Computer Science, McGill University, Montreal, QC H3A 0C6
yueli@cs.mcgill.ca

Abstract. MicroRNAs (miRNAs) play a crucial role in the regulation of gene expression by targeting specific mRNAs. They can function as both tumor suppressors and oncogenes depending on the specific miRNA and its target genes. Detecting miRNA-target interactions (MTIs) is critical for unraveling the complex mechanisms of gene regulation and identifying therapeutic targets and diagnostic markers. There is currently a lack of MTIs prediction method that simultaneously performs feature learning on heterogeneous graph network and sequence information. To improve the prediction performance of MTIs, we present a novel transformer-based multi-view feature learning method, named MiRGraph. It consists of two main modules for learning the sequence and heterogeneous graph network, respectively. For learning the sequence-based feaature embedding, we utilize the mature miRNA sequence and the complete 3'UTR sequence of the target mRNAs to encode sequence features. Specifically, a transformer-based CNN (TransCNN) module is designed for miRNAs and genes respectively to extract their personalized sequence features. For learning the network-based feature embedding, we utilize a heterogeneous graph transformer (HGT) module to extract the relational and structural information in a heterogeneous graph consisting of miRNA-miRNA, gene-gene and miRNA-target interactions. We learn the TransCNN and HGT modules end-to-end by utilizing a feedforward network, which takes the combined embedded features of the miRNA-gene pair to predict MTIs. Comparisons with other existing MTIs prediction methods illustrates the superiority of MiRGraph under standard criteria. In a case study on breast cancer, we identified plausible target genes of an oncomir *hsa-MiR-122-5p* and plausible miRNAs that regulate the oncogene *BRCA1*.

Keywords: MiRNA-Target Interaction, Heterogeneous Graph Transformer, Deep learning, Transformer, CNN.

* Supported by the National Natural Science Foundation of China under Grant 62032007, 62372165 and 62202152.

1 Introduction

MicroRNAs (miRNAs) are endogenous, non-coding, single-stranded small RNA molecules composed of approximately 22 nucleotides, which widely participate in all kinds of biological processes such as immune response, transcription, proliferation and differentiation in human cells [Bartel, 2004; Gebert and MacRae, 2019; Griffiths-Jones et al., 2006]. MiRNAs mainly play a key regulator at the post-transcriptional regulation by binding to the 3' UTRs of their target mRNAs, thereby inducing mRNAs degradation or inhibiting mRNA translation [Lai, 2002]. Their dysregulation can affect gene regulatory programs ultimately leading to diseases including cancers [Calin and Croce, 2006; Peng and Croce, 2016]. Therefore, accurate prediction of miRNA-target interactions (MTIs) is of great significance for revealing the complex diseases regulatory mechanism.

While traditional experimental methods by were developed to detect MIT [Androvic et al., 2017; Chi et al., 2009b; Hafner et al., 2010; Khan et al., 2009], they are limited to a specific miRNA. High-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation (HITS-CLIP) [Chi et al., 2009a] or crosslinking, ligation, and sequencing of hybrids (CLASH) [Helwak et al., 2013] are costly and biased towards strongly bound MTIs [Wen et al., 2011]. Therefore, computational MTI prediction methods are cost-effective way to prioritize candidate MTIs for experimental validation.

The computational algorithms of MTI predictions can be divided into two categories [Li and Zhang, 2015]: rule-based methods and machine learning (ML)-based methods. The former utilize hand-crafted features. For instance, PITA [Kertesz et al., 2007] leveraged site accessibility to calculate thermodynamic stability-based scores. miRanda [Enright et al., 2003] utilized the sequence complementarity to assess the likelihood of MTI. TargetScan [Agarwal et al., 2015] prioritized the miRNA-candidate target site (CTS) using seed region matching rule. Although these rule-based methods provide putative MTIs, they depended on the design of hand-crafted feature engineering strategies, resulting in inconsistent predictions and high false positive rates.

ML-based methods train a supervised classification method to predict MTI using a set of features. For instance, mirSVR [Betel et al., 2010] is a support vector regression (SVR) method that uses sequence information, thermodynamic stability and structural accessibility to predict MTI score. DeepTarget [Lee et al., 2016] adopted four types of sequence-based seed-matching features along with learned sequence-based features from LSTM-based autoencoder to predict CTSs. miRAW [Pla et al., 2018] applied multi-layer perceptron (MLP) to predict binding site accessibility energies. TargetNet [Min et al., 2022] utitlized a ResNet-based deep learning framework with a miRNA–CTS sequence encoding scheme to incorporate alignment of extended seed-region and mRNA sequences. Although these methods utilized CTS to reduce the search space, they rely only on the RNA sequence information and are unable to utilize the miRNA-regulatory network information. Meanwhile, with the accumulation of the relevant miRNA-mRNA interaction databases such as mirRTarBase [Hsu et al., 2011; Huang et al., 2022], TargetScan [McGeary et al., 2019], StarBase [Li et al., 2014], it has become possible to use graph neural networks to extract relational information between the two RNA species. MRMTI [Luo et al., 2022] utilized the multi-relational graph convolutional module and the Bi-LSTM module to incorporate both network topology and sequential information. SRG-Vote [Xie et al., 2022] leveraged Doc2Vec and Graph Convolutional Network (GCN) to separately generate sequential and geometrical embeddings from molecular sequences and relational network. Both methods operate on k-mers embeddings (i.e., k = 3 nucleotides). Specifically, they split a gene sequence into k-mer segments treated as "words" and then then maps them to real-valued embeddings via a pre-trained word2vec-based model. In addition to the above shortcomings, there is currently a lack of MTI prediction methods for extracting node features from heterogeneous networks by modeling structural and relational data. Moreover, while considering edge and node information simultaneously in heterogeneous graph networks has proven to be beneficial [Hu et al., 2020; Ma et al., 2023; Schlichtkrull et al., 2018], their applications on MTI prediction remains unexplored.

In this study, we present a novel transformer-based deep learning framework named MiRGraph, which considers both network and sequential information for MTIs prediction (Fig. 1). Our contributions are two-fold. As the first contribution, our method utilizes the entire raw sequences of mature miRNA and the 3'UTRs of target mRNAs to initialize the sequence features. This is in contrast to to most existing methods that rely on short k-mer to represent sequences. To address the different sequence lengths of miRNAs and mRNAs, we design a transformer-based CNN (TransCNN) module for miRNA and mRNA to extract sequence features from long sequences. Furthermore, to fully utilize the existing biological relationship data, we construct a

heterogeneous graph network consisting of miRNA-miRNA, gene-gene and miRNA-gene interactions. As the second contribution, we utilize a heterogeneous graph transformer (HGT) module to learn the relational and topological information from the heterogeneous network. The concatenated sequence-based and graph-based feature embeddings are then fed to an MLP layer followed by a bilinear transformation function to compute the MTI prediction scores. We compared MiRGraph with several existing MTIs prediction methods on two different datasets under standard metrics. The results demonstrate that our approach outperforms the state-of-the-art methods in MTIs prediction task. In a case study, we use MiRGraph to prioritize *de novo* target genes of the oncomir *hsa-miR-122-5p* and *de novo* miRNAs associated with the oncogene *BRCA1* in breast cancer and find hits that are supported by the literature.

2 MiRGraph framework

The proposed framework MiRGraph mainly consists of three parts (Fig. 1): a. Data Preprocessing; b. Feature Encoder; c. MTI Predictor. Briefly, we first built a heterogeneous information network (HIN), and obtain the initial network and sequence one-hot encoding of miRNAs and genes (Fig. 1a). They are fed into feature encoder including two modules (i.e. HGT and TransCNN) to learn network and sequence representations separately (Fig. 1b). Finally, we feed the concatenated representation into the MTI predictor (Fig. 1c). Details of each step are described below.

2.1 Data Preprocessing

HIN Construction The predicted miRNA-gene interactions network was obtained from TargetScan [Agarwal et al., 2015; McGeary et al., 2019]. We used the experimentally validated MTI from miRTarbase [Hsu et al., 2011] to label the TargetScan-predicted edges as positive edges and the rest as negative edges. The above procedure resulted in 22,221 positive edges and 59,897 negative edges between 279 miRNAs and 14,676 genes. Considering that the negative edges of the network built on this basis are too large, we retained only the negative edges for each gene g, which involves the top 10 miRNAs with sequence most similar to the miRNA involved in the positive MTI. This reduces the negative samples and improves the specificity of our model to discriminate similar but false MTIs. We obtained the gene-gene network from the protein-protein interaction network with 1,043,837 edges between 14,676 genes from the STRING database [Szklarczyk et al., 2019]. The miRNA-miRNA interaction network with 209 edges between 279 miRNAs was built with miRNA family information [Liu et al., 2020]. The resulting HIN $A_{(m+g)\times(m+g)} \in [0,1]^{(279+14676)\times(279+14676)}$ consists of three networks, i.e., the miRNA-gene ($A_{m\times g} \in [0,1]^{279\times 14676}$), gene-gene ($A_{g\times g} \in [0,1]^{14676\times 14676}$), and miRNA-miRNA ($A_{m\times m} \in [0,1]^{279\times 279}$).

Feature Initialization We converted the miRNA sequences into one-hot representations with the maximum length set to $k_m = 25$. We processed the target gene sequences in the same way with the maximum length set to $k_g = 35,526$. Because the maximum length of the complete 3'UTR sequences of all genes is too long, each gene sequence was divided into 15 fixed-length shorter sequences. As a result, for all miRNAs and genes, we obtained their initial sequence features $X_m^{seq} \in R^{279\times 4\times 25}$ and $X_g^{seq} \in R^{14676\times 4\times 15\times 2500}$. This is equivalent to dividing a long 2D image into 15 segments each having height of 4 and width of 2500.

To obtain the initial miRNA and gene network node features, we first constructed the diagonal matrix based on miRNAs and genes identities using one-hot encoding i.e. $X_m^{one-hot} \in [0,1]^{279\times 279}$ and $X_g^{one-hot} \in [0,1]^{14676\times 14676}$. To enrich node features, we integrated the one-encoded node feature with the corresponding network matrices via element-wise addition: $X_m^{net} = X_m^{one-hot} + A_{m\times m}$ and $X_g^{one-hot} = X_g^{net} + A_{g\times g}$ (Fig. 1a).

2.2 Feature Encoder

Sequence Feature Encoding Module In order to fully capture the MTI-specific sequence elements in miRNA and gene sequences, we designed a novel sequence learning module TransCNN by combining the convolutional neural network (CNN) and the Transformer encoder (Fig. 2). Specifically, for gene $x_g^{seq} \in R^{14676\times 4\times 15\times 2500}$, we devised a three-layer encoder namely $TransCNN_gene$ (Fig. 2a):

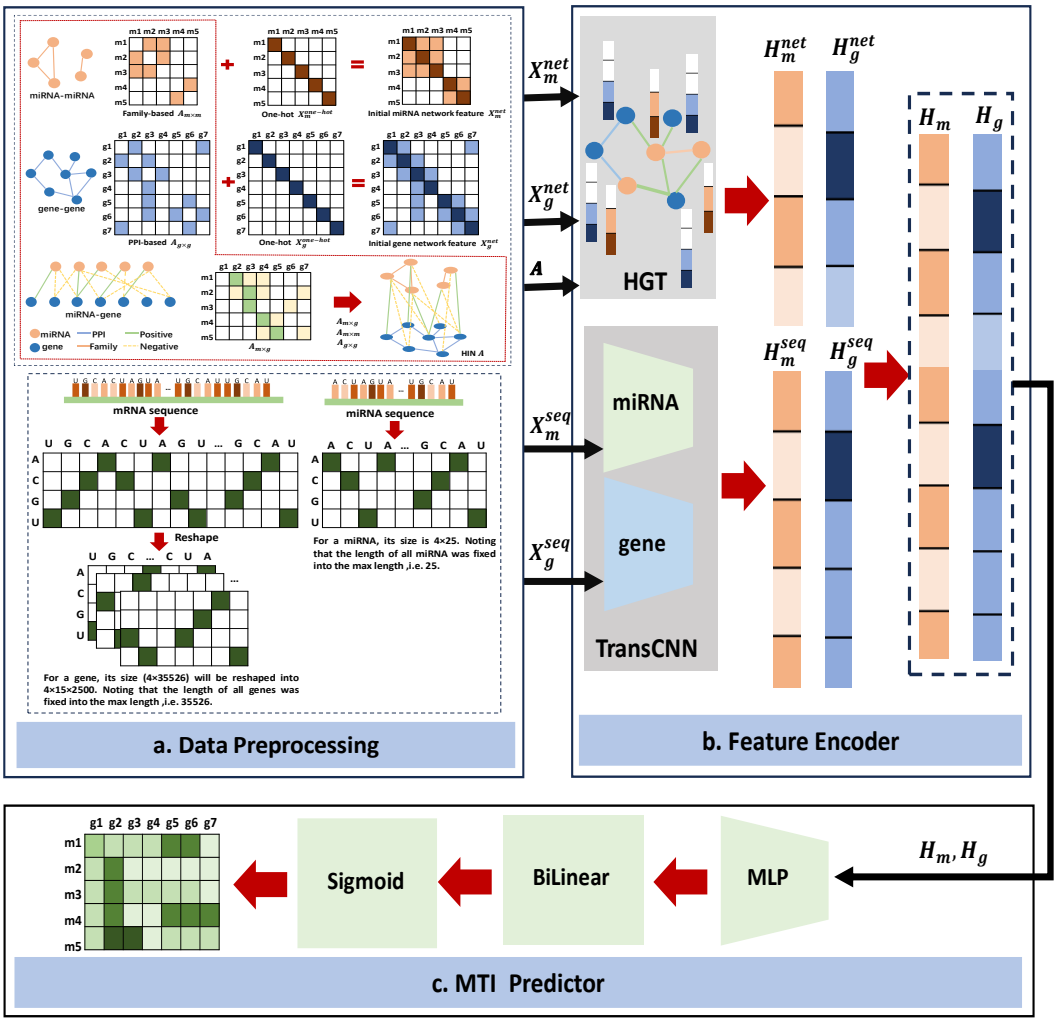

Fig. 1. Overall framework of MiRGraph. **a. Data Preprocessing.** We built a heterogeneous information network (HIN) A that comprises miRNA-miRNA, gene-gene and miRNA-gene interaction networks ($A_{m\times m}$, $A_{g\times g}$ and $A_{m\times g}$). And then we obtained the initial miRNA and gene network features (X_m^{net} and X_g^{net}) by combining their one-hot features ($X_m^{one-hot}$ and $X_g^{one-hot}$) and adjacency matrices of networks ($A_{m\times m}$ and $A_{g\times g}$). Also, we obtained one-hot encodings of the miRNA and gene sequences (X_m^{seq} and X_g^{seq}). **b. Feature Encoder.** We first learn from the HIN and sequences feature embeddings of miRNAs (H_m^{net} and H_m^{seq}) and genes (H_g^{net} and H_g^{seq}), using 'HGT' and 'TransCNN' modules separately. We then concatenate them into the final feature embeddings of miRNAs and genes (H_m and H_m). **c. MTI predictor.** We feed the learned features of miRNAs and genes (H_m and H_m) into MLP. Finally, we use the 'BiLinear' and 'Sigmoid' functions to calculate their interaction scores.

– **2D convolutional layer.** Two layers of 2D convolutional layer $conv2d(.)$ are used to integrate the 15 sequence segments of each gene. This is followed by a max-pooling layer $maxpool2d(.)$ to reduce the sequence length. To effectively learn features, we treat the 15 sub-sequences as 15 channels and introduce a channel attention mechanism layer (ECA-Net) [Wang et al., 2020] between the two $conv2d(.)$ layers.

  ECA-Net improves the expressiveness of the feature representation by a set of globally adaptive weights of the feature map of each channel. Specifically, it first dynamically determines the kernel size k after aggregating convolution features using global average pooling (GAP) without dimensionality reduction, and then performs 1D convolution followed by Sigmoid transformation to learn the channel attention. Finally, it calculates the element-wise product between channel attention and original feature to obtain the enhanced feature (Fig. 2b).

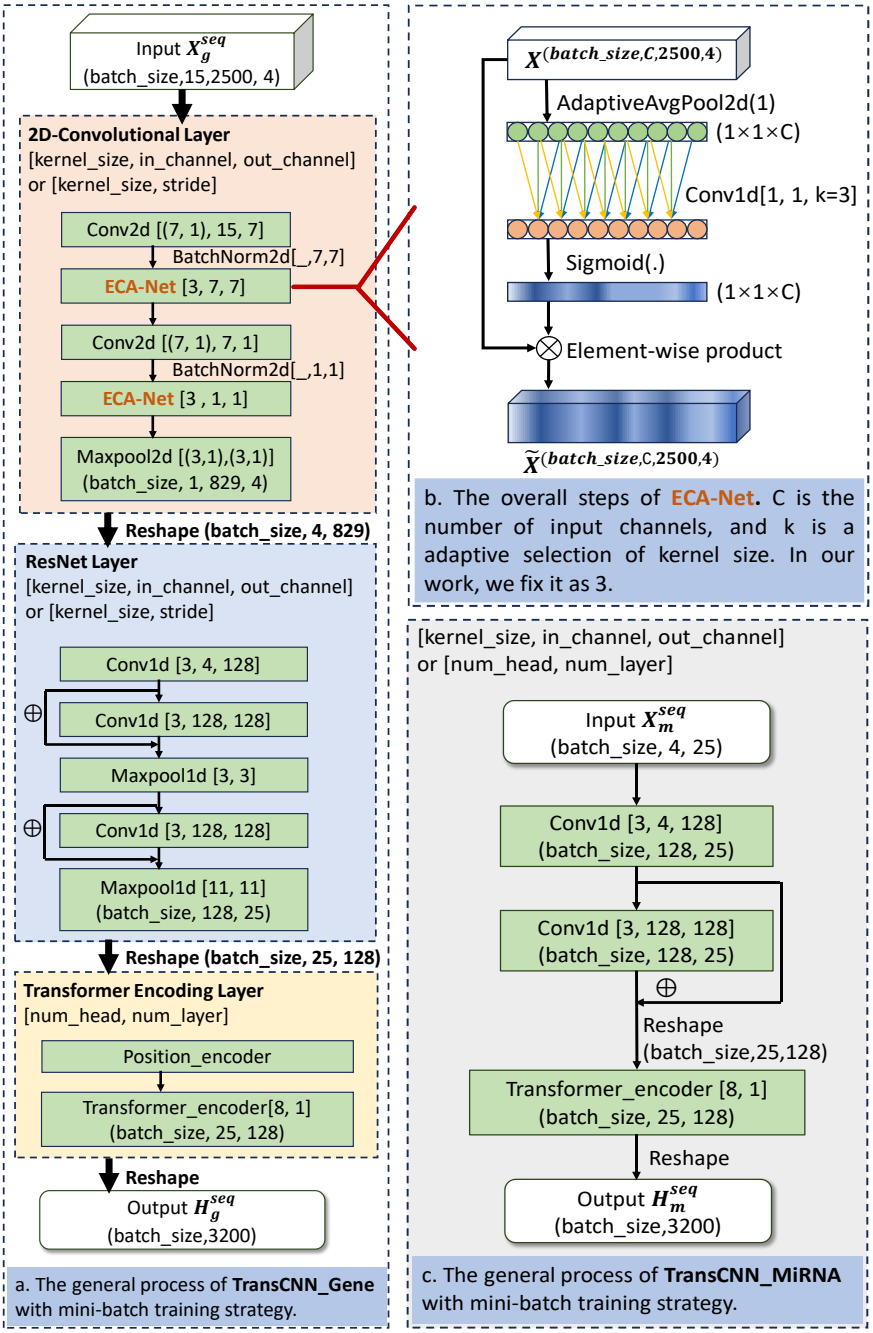

Fig. 2. The TransCNN module. The diagram depicts the sequence encoder module as the part of the MiRGraph encoder (Fig. 1b). It is composed of two sub-modules namely TransCNN_Gene and TransCNN_MiRNA. **a. TransCNN_Gene**. This sub-module consists of three layer types: '2D Convolutional layer' used to learned cross-channel interaction information, 'ResNet layer' used to learned intra-channel specific information, and 'Transformer encoding layer' used to capture more fine-grained information. **b. Detailed view of ECA-net**. ECA-net is used to model local cross-channel attention values. **c. TransCNN_MiRNA**. Due to short sequences of miRNA, this sub-module only uses residual connection after the 1D convolution layer $conv1d(.)$, followed by transformer encoder.

– **Residual connection layer.** Residual connection is adopted after the 1D convolution layer $conv1d(.)$. A max-pooling layer $maxpool1d(.)$ is applied after each residual connection to further reduce the sequence length.

– **Transformer layer**. One layer of transformer encoder $transformer_encoder(.)$ is adopted to learn global contextual information from the sequence feature map.

For the input of miRNAs $X_m^{seq} \in R^{279\times4\times25}$, we utilize similar operations in steps 2 and 3 in $TransCNN_gene$ to learn sequence features of miRNAs (Fig. 2c). In order to integrate the two types of sequence information, we designed the convolution filters to match the dimensions between the final sequence feature embedding of miRNAs and genes, i.e., $H_m^{seq} \in R^{279\times3200}$ and $H_g^{seq} \in R^{14676\times3200}$.

Network Feature Encoding Module Our HIN graph $A = (V, E, B, D)$ comprises node $v \in V$, edge $e \in E$, and the corresponding node and edge types B and D, which can be mapped via functions $\tau(v) : V \to B$ and $\theta(e) : E \to D$. For any edge $e = (s, t)$ from source node s to target node t, its meta-relation is expressed as $\langle\tau(s), \theta(e), \tau(t)\rangle$. We adapted the Heterogeneous Graph Transformer (HGT) model [Hu et al., 2020], which uses the meta-relations of the heterogeneous graphs to parameterize its weights for computing the heterogeneous mutual attention, messages, and propagation steps. HGT expresses different types of vertices and edges with unique representation with dedicated type-specific parameters. It combines information from different types of higher-order neighbors using inter-layer message passing to enable interactions between different types of vertices. Therefore, HGT can automatically learn and extract the salient 'meta-paths' for different downstream tasks, even if the input to HGT is only one-hop edge without manually designed meta-paths. Specifically, the HGT is composed of three key parts as follows.

Heterogeneous Mutual Attention. Given a target node t and a source node $s \in N(t)$, where $N(t)$ is the set of neighbor nodes of t, their mutual attention ($Attention(s, e, t)$) grounded by their meta relations $\langle\tau(s), \theta(e), \tau(t)\rangle$ triplets is defined as the following equations:

$$
\begin{aligned}
&Attention(s,e,t) = \underset{s\in N(t)}{Softmax}\left(\underset{i\in[1,h]}{\|}(ATT_head^i(s,e,t))\right)\\
&ATT_head^i(s,e,t) = \left(K^i(s)W_{\theta(e)}^{ATT}Q^i(t)^T\right)\cdot\frac{\mu_{\langle\tau(s),\theta(e),\tau(t)\rangle}}{\sqrt{d}}\\
&K^i(s) = K_Linear^i_{\tau(s)}\left(H^{l-1}[s]\right)\\
&Q^i(t) = Q_Linear^i_{\tau(t)}\left(H^{l-1}[t]\right)
\end{aligned}
\tag{1}
$$

where $\|$ (.) is the concatenation function. $ATT_head^i(s, e, t)$ is the output of the ith attention head between s and t. $W_{\theta(e)}^{ATT} \in \mathbb{R}^{\frac{d}{h}\times\frac{d}{h}}$ is a edge-based matrix to learn different meta-relation features even between the same node type pairs, where h is the number of attention heads and $\frac{d}{h}$ is the vector dimension per head. $(.)^T$ is the transposal function. $\mu_{<\tau(s),\theta(e),\tau(t)>}$ is a prior tensor, serving as an adaptive scaling of attention to denote the significance for each meta-relation. $K^i(s)$ is the ith Key vector of source node s and $Q^i(t)$ is the ith Query vector of target node t, and they are separately mapped by the linear projection functions $K_Linear^i_{\tau(s)} : R^d \to R^{\frac{d}{h}}$ and $Q_Linear^i_{\tau(s)} : R^d \to R^{\frac{d}{h}}$. $H^{l-1}[t]$ and $H^{l-1}[s]$ are node embedding of nodes t and s at the $l-1^{th}$ layer, respectively.

Heterogeneous Message Passing. To pass information from source nodes to target nodes, the multi-head message for a pair of nodes $e = (s, t)$ is calculated by:

$$
\begin{aligned}
&Message(s,e,t) = \underset{i\in[1,h]}{\|}\left(MSG_head^i(s,e,t)\right)\\
&MSG_head^i(s,e,t) = V^i(s)W_{\theta(e)}^{MSG}\\
&V^i(s) = V_Linear^i_{\tau(s)}\left(H^{l-1}[s]\right)
\end{aligned}
\tag{2}
$$

where $MSG_head^i(s, e, t)$ represents the ith head message between s and t. $V_Linear^i_{\tau(s)} : R^d \to R^{\frac{d}{h}}$ is a linear projection function to map each source node s in the ith head into a message vector $V^i(s)$. $W_{\theta(e)}^{MSG} \in R^{\frac{d}{h}\times\frac{d}{h}}$ is a edge-based matrix similar to $W_{\theta(e)}^{ATT} \in R^{\frac{d}{h}\times\frac{d}{h}}$.

Target-Specific Aggregation. To get the output $H^l[t]$ of the target node t after l layers, we first aggregate the heterogeneous multi-head attentions and messages from all source nodes $s \in N(t)$ to target

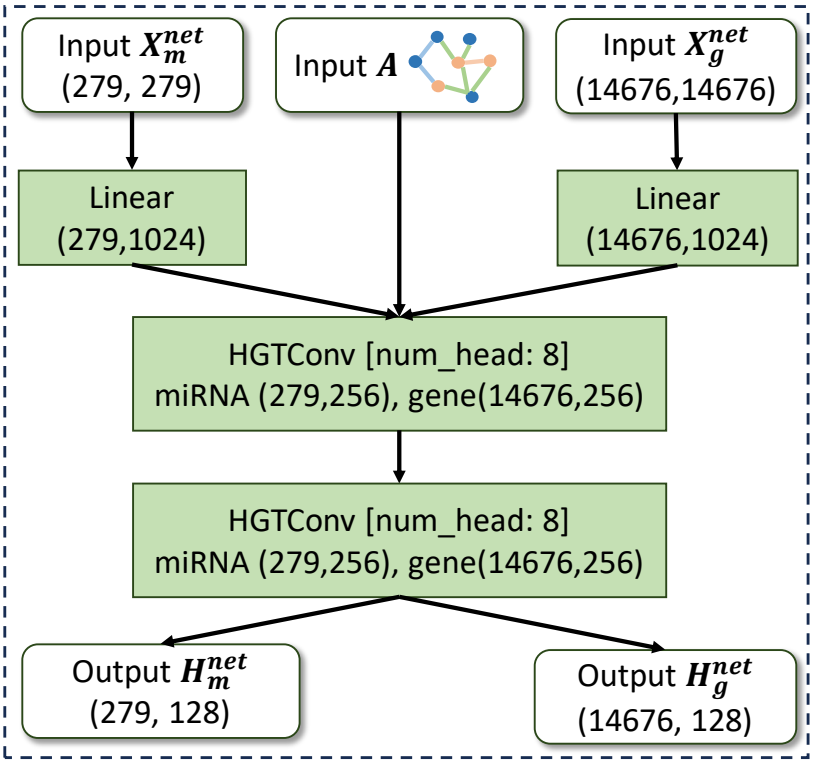

Fig. 3. The HGT module. The diagram depicts the network feature encoder module as part of the MiRGraph encoder (Fig. 1b). Given a heterogeneous graph A with two node types $B = \{miRNA, gene\}$ and three edge types $D = \{'miRNA - miRNA', 'miRNA - gene', 'gene - gene'\}$, we first feed the initial network features of miRNAs and genes $X_m^{net} \in R^{279\times279}$ and $X_g^{net} \in R^{14676\times14676}$ to two separate linear layers $Linear(.)$. The resulting node features along with the three adjacency matrices (i.e., $A_{m\times g} \in [0,1]^{279\times14676}$, $A_{g\times g} \in [0,1]^{14676\times14676}$ and $A_{m\times m} \in [0,1]^{279\times279}$) are fed into the $HGTConv(.)$ layer to learn the nuanced features of different types of nodes. After two layers of $HGTConv(.)$, we obtain the final network feature representations of miRNAs and genes.

node t, and then use a linear projection function $A_Linear_{\tau(t)}$ to update the embedding $\tilde{H}^l[t]$ followed by a residual connection:

$$\tilde{H}^l[t] = \bigoplus_{s\in N(t)} (Attention(s,e,t) \cdot Message(s,e,t))$$
$$H^l[t] = A_Linear_{\tau(t)}\left(\sigma(H^l[t])\right) + H^{l+1}[t] \tag{3}$$

where $\sigma(.)$ is an activation function. Through the above three steps, HGT can generate a highly contextualized representation H for each node in the heterogeneous graph. In summary, Fig. 3 depicts our HGT module, which outputs the network features of miRNAs and genes i.e., $H_m^{net} \in R^{279\times128}$ and $H_g^{net} \in R^{14676\times128}$.

Feature Integration Module We first concatenate the two types of node embeddings, and then apply the nonlinear activation function $GELU(.)$ to obtain the final representation of miRNAs and genes, i.e. $H_m \in R^{279\times3328}$ and $H_g \in R^{14676\times3328}$ (i.e., $3200 + 128 = 3328$ (Fig. 1b):

$$H_m = GELU\left(\| \left([H_m^{seq}, H_m^{net}]\right)\right)$$
$$H_g = GELU\left(\| \left([H_g^{seq}, H_g^{net}]\right)\right) \tag{4}$$

2.3 MTI Predictor

Predicting Score Given the input of H_m and H_g, we first use $MLP(.)$ to separately map them onto the same low-dimensional space, and then use a bilinear layer followed by sigmoid function $Sigmoid(.)$ to

compute the MTI probability:

$$p_{m,g} = Sigmoid\left(MLP(H_m)^T \times W \times MLP(H_g)\right) \tag{5}$$

Loss Function To account for the imbalance of small number of positive and large number of negative examples (i.e., true and false MTI edges), we use the focal loss function [Lin et al., 2017] to train the model:

$$FL(p) = \begin{cases} -\alpha(1-p)^{\gamma} log(p), & y = 1 \\ -(1-\alpha)p^{\gamma} log(1-p), & y = 0 \end{cases} \tag{6}$$

where y is the true label, p is the predicting score, $\alpha > 0$ is a hyperparameter to balance the importance of positive/negative examples, and $\gamma > 0$ is a hyperparameter. By default, we set $\alpha = 0.65$ and $\gamma = 2$.

3 Experiments

3.1 MTI gold-standard data

We evaluated the proposed method with two experimentally validated MTI datasets: **miRTarBase MTIs**. This was the main labelled dataset we used to train and evaluate our method. The details were described in the *HIN Construction* section above. We split the positive (miRTarbase+TargetScan) and negative edges (TargetScan only) into 70%, 20%, and 10% for training, validation and testing, respectively. **miRAW benchmark MTIs**. To further demonstrate the utility of MiRGraph, we used another benchmark dataset compiled by [Pla et al., 2018]. This dataset was originally derived from DIANA-TarBase [Karagkouni et al., 2018] and miRTarBase [Hsu et al., 2011]. It was used by the TargetNet authors [Min et al., 2022]. After removing several genes without complete 3'UTR sequence information, we constructed a miRAW-based miRNA-gene network with 7322 genes and 823 miRNAs. Also, we constructed a miRNA-miRNA network with 1210 edges and a gene-gene network with 341,760 edges. The training set contains 22,104 positive and 467 negative edges, a validation set contains 1892 positive and 46 negative edges, and the 10 testing sets contains a total of 5333 positive edges and 543 negative edges. Among them, the number of positive and negative samples in each testing set is roughly the same. For TargetNet, the training set contains 26,711 positive and 27,209 negative miRNA–CTS pairs, a validation set contains 2182 positive and 2136 negative miRNA–CTS pairs, and the positive and negative pairs in each of testing set are the same as MiRGraph.

3.2 Implementation details of MiRGraph

We implemented MiRGraph in Python (v 3.9.6) using Pytorch (v 1.10.0) and PyTorch Geometric (v 2.1.0) as backend. We used NVIDIA A100-SXM4-40GB for the GPU job and AMD EPYC 7413 24-Core for the CPU job with the memory of 256G. The experiments were performed on the Compute Canada Narval server. We leveraged Kaiming [He et al., 2015] to initialize the model parameters and AdamW as the optimizer for model training. The learning rates of HGT, TransCNN and the prediction MLP layer were set as 0.001, 0.0001 and 0.0001 respectively. Dropout rate and weight decay of them were set as 0.3 and 0.005.

Given the large graph of HIN interactions (i.e., 14,676+279 by 14,676+279), we designed two efficient training strategies:

- **End-to-end training strategy.** We first pre-trained TransCNN and HGT (i.e., separately training TransCNN+MLP and HGT+MLP) (Figure 1b). We then transferred the parameters of the trained HGT and TransCNN to MiRGraph (i.e., [TransCNN, HGT] + MLP) for an end-to-end fine-tuning (Fig. 1c).
- **Step-by-step training strategy.** We first trained TransCNN and HGT separately. We fixed the HGT and TransCNN and only trained the prediction layer using the combined feature representations of miRNAs and genes from the two modules for final MTI predictions.

Table 1. Details and differences about MiRGraph and comparison methods.

	sequence information	network information	predictor	loss function	gene sequence segmentation
RGCN	——	✓	MLP+BiLinear	Focal Loss	——
HGT_BiLSTM	✓	✓	MLP+BiLinear	Focal Loss	3-mer
MRMTI	✓	✓	MLP+BiLinear	Focal Loss	3-mer
TargetNet	——	✓	Linear	Binary Cross Entropy	binding site sequence in 3'UTR
HGT	✓	——	BiLinear	Focal Loss	——
TransCNN	——	✓	MLP+BiLinear	Focal Loss	complete 3'UTR
MiRGraph	✓	✓	MLP+BiLinear	Focal Loss	complete 3'UTR

We observed that the second strategy takes much less time than the first one (on average, 25 seconds versus 3 minutes per epoch) while conferring similar performances. Therefore, we only demonstrated the results of the second strategy. Specifically, training the HGT alone took 45 seconds per epoch on the CPU. It took approximately 300 epochs for the model to converge (about 3.5 hours). Note that we used CPU to train HGT because of the large memory requirement for training the large heterogeneous graph. Training TransCNN on the GPU took on average 42 seconds per epoch, and it took approximately 300 epochs for the model to converge (about 3.5 hours). The final training of the MLP prediction layer of MiRGraph was performed on CPU and approximately 1000 epochs to converge (about 7 hours).

3.3 Implementations of baseline methods

We compared MiRGraph with several state-of-the-art methods, namely Relational Graph Convolution Network (RCGN) [Schlichtkrull et al., 2018], MRMTI [Luo et al., 2022], HGT [Hu et al., 2020], HGT_BiLSTM and TargetNet [Pla et al., 2018]. Except for TargetNet (`https://github.com/mswzeus/TargetNet`), we implemented all of these methods below ourselves based on the description in the original paper using the PyTorch Geometric API.

Specifically, RGCN and HGT are the extensions of the GCN framework to model heterogeneous graph network. To apply them to MTI prediction task, we adopted the operator *RGCNConv* and *HGTConv*, implemented in *torch_geometric.nn* package[3],[4]. MRMTI is an MTI prediction model using RCGN and Bi-LSTM to learn network and sequence information separately. HGT_BiLSTM can be seen as a variant of MRMTI, which combines HGT and Bi-LSTM. Integrating RGCN and HGT with Bi-LSTM, we implemented MRMTI and HGT_BiLSTM. Bi-LSTM was implemented by the operator *LSTM* in *torch.nn*.

For training the miRTarBase MTIs data, we set the learning rate of RGCN, HGT, HGT_BiLSTM and TargetNet as 0.001, and the learning rate of MRMTI as 0.0001. All of them were trained 500 epochs in a full-batch way with AdamW optimizer, and the weight_decay is set as 0.005. The dropout rate of MRMTI and HGT_BiLSTM is 0.1. Besides, these methods were trained 200 epochs on the miRAW MTIs data using the same hyperparameters setting as above, except for TargetNet using the same hyperparameters setting as the original paper.

All implementations are available at `https://github.com/Liangyushi/MiRGraph/tree/main`. Table 1 summarizes ours and the existing methods.

3.4 Evaluation metrics

We used standard supervised learning metrics to evaluate the performance of our proposed method. These include Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves, the Area Under the ROC (AUROC) and PR curves (AUPR) can quantify the predicting accuracy. In addition, we also utilize

[3] `https://pytorch-geometric.readthedocs.io/en/latest/_modules/torch_geometric/nn/conv/rgcn_conv.html`

[4] `https://pytorch-geometric.readthedocs.io/en/latest/_modules/torch_geometric/nn/conv/hgt_conv.html`

the following evaluation metrics to obtain a more comprehensive assessment:

$$
\begin{aligned}
&Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad FPR = \frac{FP}{TN+FP} \\
&Recall = \frac{TP}{TP+FN}, \quad Precision = \frac{TP}{TP+FP} \\
&F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad Specificity = \frac{TN}{TN+FP} \\
&Negative_Precision(NP) = \frac{TN}{TN+FN}
\end{aligned} \tag{7}
$$

where TP and TN are the number of true positives and true negatives, FP and FN are false positives and false negatives, respectively.

4 Results

4.1 Benchmark of miRTarBase MTI predictions

We compared RCGN [Schlichtkrull et al., 2018], MRMTI [Luo et al., 2022], HGT [Hu et al., 2020], HGT_BiLSTM, TransCNN and MiRGraph (i.e., HGT+TranCNN) in terms of the miRTarBase MTI predictions using common evaluation metrics (Eq. 7 in Supplementary Material). Specifically, MiRGraph conferred the best prediction performance with the highest AUROC and AUPR (Fig. 4). MiRGraph not only conferred AUROC and AUPR that are 3.38 and 4.38 percent higher than those from the runner-up method but also achieved the best performance on most other metrics (Table 2). We also observed a large overlap between the three methods of HGT, TranCNN and MiRGraph in terms of the predicted MTIs, especially in true positive MTIs

Table 2. Comparison of MTIs prediction performance on the miRTarBase MTIs dataset. The bold font represents the method or component proposed in this paper, the bold number represents the best performance, and the underlined number represents the sub-optimal performance.

	AUROC	AUPR	Precision	Recall	F1	Accuracy	FPR	Specificity	NP
RGCN	0.6666	0.4659	0.4686	0.3900	0.4257	0.7152	0.1641	0.8359	0.7869
HGT_BiLSTM	0.8367	0.7400	0.6425	0.6631	0.6526	0.8090	0.1369	0.8631	0.8735
MRMTI	0.8412	0.7754	0.7337	0.6716	0.7013	0.8451	0.0905	0.9095	0.8818
HGT	0.8549	0.7671	0.6748	0.6955	0.6850	0.8269	0.1244	0.8756	0.8857
TransCNN	0.8678	0.8072	0.6752	**0.7220**	0.6978	0.8308	0.1289	0.8711	**0.8941**
MiRGraph(end-to-end)	0.8626	0.7954	0.7141	0.7090	0.7115	0.8444	0.1053	0.8947	0.8923
MiRGraph(step-by-stey)	**0.8706**	**0.8109**	**0.7937**	0.6784	**0.7315**	**0.8652**	**0.0654**	**0.9346**	0.8867

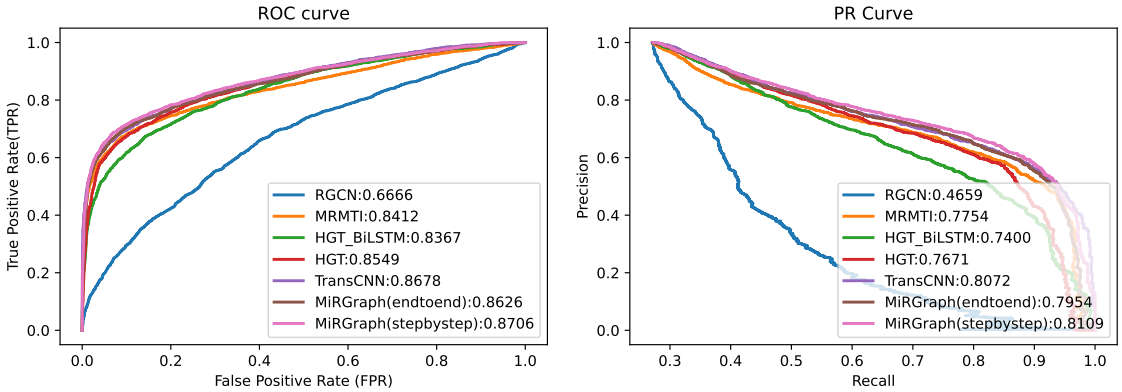

Fig. 4. The ROC and PR curves of MTI predictions on the test miRTarBase MTI data. The figure legend displays the method names and the AUROC and AUPR, respectively.

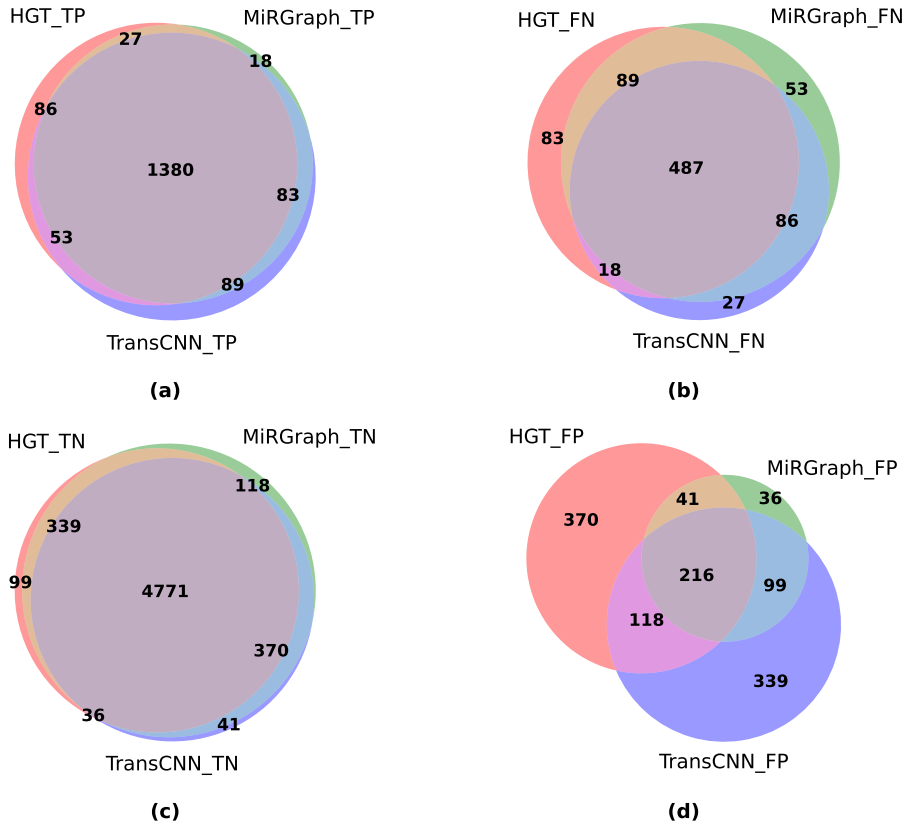

Fig. 5. Venn diagram comparing HGT, TranCNN and MiRGraph predicted MTI. TP, TN, FP and FN stand for True positive, True Negative, False Positive, and False Negative, respectively. HGT_TP, MiRGraph_TP and TransCNN_TP indicate the number of TP predicted by HGT, MiRGraph and TransCNN respectively. Panel a-d display the 3-way overlap of TP, FN, TN, and FP among HGT, MiRGraph, and TransCNN, respectively.

and true negative MTIs (Fig. 5). Nonetheless, we observed that the number of false positive MTIs predicted by MiRGraph was much smaller than that of HGT and TranCNN (Fig. 5d). The improved specificity is likely attributed to the fusion of heterogeneous network and sequence features.

In addition, the performance of RGCN is significantly worse than HGT, which demonstrates the advantage of the transformer layer in learning the attention weights between different relationships to better capture the heterogeneous information in the graph network. Interestingly, comparing HGT, RGCN, HGT_BiLSTM and MRMTI, we found that the prediction performance of MRMTI (fusing graph and sequence information with RGCN and BiLSTM) is better than RGCN. On the other hand, HGT_BiLSTM (i.e., using HGT and BiLSTM to learn graph and sequence information, respectively) is worse than HGT. This implies the limitation of the BiLSTM module compared to our TransCNN module given the same capability of graph feature representation. The ablation study also shows that the sequence features learned by the $Trans_CNN$ (AUROC 0.8678 and AUPR 0.8072) is more informative than the graph features learned by the HGT (AUROC 0.8549, AUPR 0.7671). Nonethless, combining the features learned by the two modules led to the highest performance, implying MiRGraph's ability to learn complementary information from both sequences and HIN.

4.2 Benchmark of miRAW MTI predictions

Similar to the miRTarBase benchmark, we evaluated the performance of seven methods including RGCN, HGT, HGT_BiLSTM, TransCNN, MiRGraph and TargetNet [Min et al., 2022] on predicting miRAW-defined MTIs. Table 3 shows that MiRGraph achieves optimal performance on AUROC, AUPR, Recall and Nagetive

Table 3. Comparison of MTIs prediction performance on the miRAW MTIs dataset.

	AUROC	AUPR	Precision	Recall	F1	Accuracy	FPR	Specificity	NP
RGCN	0.7094	0.6588	0.5821	0.8799	0.7007	0.6242	0.6317	0.3683	0.7548
HGT	0.7547	0.6915	0.6821	0.8932	0.7735	0.7385	0.4162	0.5838	0.8456
TargetNet	0.8005	0.7644	0.6570	0.9417	0.7740	0.7250	0.4917	0.5083	0.8973
HGT_BiLSTM	0.8089	0.7539	0.6521	0.9731	0.7809	0.7269	0.5193	0.4807	0.9471
MRMTI	0.8203	0.7946	0.6785	0.9639	0.7964	0.7536	0.4567	0.5433	0.9378
TransCNN	0.8979	0.8591	**0.7659**	0.9578	**0.8512**	**0.8325**	**0.2928**	**0.7072**	0.9438
MiRGraph(end-to-end)	**0.9030**	**0.8686**	0.6743	**0.9914**	0.8027	0.7563	0.4788	0.5212	**0.9837**

Precision (NP) metrics, and our sequence-based module namely TransCNN confers the best performance on other metrics. This is perhaps because the miRAW data are largely based on physical binding sites measured by CLIP and CLASH, which may not be necessarily functional [Pla et al., 2018]. Besides, TargetNet with resnet structure is obviously inferior to TransCNN, and HGT_BiLSTM and MRMTI able to learn both sequence and network information are not better than TransCNN yet. This further demonstrates that combining transformer with convolutional network can effectively learn sequence information.

Noting that due to the imbalance of positive and negative samples in the training and validation set of miRAW-defined MTIs data (i.e., positive samples are about 40 times of negative samples), and the adoption of mini-batch strategy in training TransCNN, in order to prevent the situation that there is only one label in a batch, we over-sampled the negative edges. Specifically, we replicated the negative samples in the training and validation set, and made the final number of negatives edges 20 times of the original.

4.3 Case study in cancer

We used MiRGraph trained on the miRTarBase dataset to investigate the target genes of an oncomir and miRNA regulators of an oncogene in cancer. In particular, we chose to investigate *hsa-miR-122-5p*, which is associated with cancer development and metastasis by targeting specific oncogenes [Duan et al., 2018]. Table 4 lists the top five MiRGraph-predicted target genes. Among them, *SOCS1* is supported by the existing literature as it is modulated by *hsa-miR-122-5p* in regulation of Type I interferon expression in the hepatitis C virus (HCV)-infected patients [Li et al., 2013]. The other four novel candidate genes deserve further investigation.

We then investigated the predicted miRNAs that bind to a specific oncogene *BRCA1*, which is a well-known hereditary gene associated with the risk of breast and ovarian cancers. Table 4 shows the top five predicted miRNAs and their supports. Interaction between *BRCA1* and *hsa-miR-181a-5p* has been found from a high-confidence high-throughput experimental assay called CLASH [Helwak et al., 2013]. Furthermore, we used TAM [Lu et al., 2010] to conduct functional analysis of the top-10 candidate miRNAs of *BRCA1*. We found that most of the top-10 miRNAs are under the 'Cluster', 'Family', 'Function' and 'HMDD' categories, which illustrates that these miRNAs have the same functional mechanism (Table 5 in Supplementary Material). Interestingly, the predicted *BRCA1* miRNA regulator *hsa-miR-181a-5p* is associated with devel-

Table 4. Top five candidate genes/miRNAs predicted by MiRGraph for *hsa-miR-122-5p* and *BRCA1*.

miRNA	gene	evidence
hsa-miR-122-5p	SOCS1	PMID: 23348614
	PPP1R11	–
	PDIA6	–
	ANKRA2	–
	DEFB108B	–
hsa-miR-302d-3p	BRCA1	–
hsa-miR-302a-3p		–
hsa-miR-373-3p		–
hsa-miR-181a-5p		PMID: 23622248
hsa-miR-520a-3p		–

Table 5. Functional analysis results of potential biological significance of top-10 candidate miRNAs of *BRCA1* using TAM version 2.0. Noting that the top-10 candidates includes *hsa-miR-302d-3p, hsa-miR-302a-3p, hsa-miR-373-3p, hsa-miR-181a-5p, hsa-miR-520a-3p, hsa-miR-302e,hsa-miR-372-3p, hsa-miR-302b-3p, hsa-miR-19b-3p and hsa-miR-520c-3p.*

Category	Term	Count	P-value	Bonferroni	FDR	miRNAs
Cluster	hsa-mir-1283 cluster	4	3.91E-03	3.13E-01	1.32E-01	hsa-mir-520a, hsa-mir-520c, hsa-mir-372, hsa-mir-373
Cluster	hsa-mir-302a cluster	3	4.48E-05	3.59E-03	4.04E-03	hsa-mir-302d, hsa-mir-302b, hsa-mir-3023
Family	mir-302 family	4	8.80E-07	7.01E-05	4.74E-04	hsa-mir-302e, hsa-mir-302d, hsa-mir-302b, hsa-mir-3023
Function	cell fate determination	4	7.28E-04	5.83E-02	3.03E-02	hsa-mir-302e, hsa-mir-302d, hsa-mir-302b, hsa-mir-302a
Function	onco-miRNAs	3	1.56E-02	1.00E+00	4.43E-01	hsa-mir-373, hsa-mir-372, hsa-mir-19b
Function	Human embryonic stem cell (hESC) regulation	6	1.58E-03	1.26E-01	6.11E-02	**hsa-mir-181a**, hsa-mir-302d, hsa-mir-302b, hsa-mir-302a, hsa-mir-373,hsa-mir-372
HMDD	Azoospermia	4	3.92E-05	3.14E-03	4.24E-03	hsa-mir-373, hsa-mir-372, hsa-mir-302a, hsa-mir-19b
HMDD	Breast Neoplasms	5	3.24E-02	1.00E+00	7.97E-01	hsa-mir-520a, hsa-mir-302d, hsa-mir-302b, hsa-mire302a, hsa-mir-19b
HMDD	Carcinoma	6	2.57E-05	2.05E-03	4.62E-03	hsa-mir-302e, hsa-mir-302d, hsa-mir-302b, hsa-mir-302a, hsa-mir-373, hsa-mir-372
HMDD	Glioma	5	3.02E-05	2.41E-03	4.08E-03	hsa-mir-373, hsa-mir-302d, **hsa-mir-181a**, hsa-mir-302a, hsa-mir-302b
HMDD	Heart Failure	8	4.73E-05	3.79E-03	3.66E-03	hsa-mir-520a, hsa-mir-520c, **hsa-mir-181a**, hsa-mir302b, hsa-mir-302a,hsa-mir-19b, hsa-mir-373, hsa-mir-372
HMDD	Neoplasms	7	2.59E-04	2.07E-02	1.56E-02	**hsa-mir-181a**, hsa-mir-302d, hsa-mir-302b, hsa-mir-302a, hsa-mir-19b,hsa-mir-373, hsa-mir-372
HMDD	Pre-Eclampsia	4	9.72E-05	7.77E-03	6.57E-03	hsa-mir-302e, hsa-mir-302d, hsa-mir-302b, hsa-mir-302a
HMDD	Skin Neoplasms	3	1.81E-05	1.45E-03	4.90E-03	hsa-mir-302d, hsa-mir-302b, hsa-mir-302a

opment and diseases, including 'Human embryonic stem cell (hESC) regulation', 'Glioma', 'Heart Failure' and 'Neoplasms'.

5 Discussion

The investigation on the miRNA-target interactions can help reveal the complex gene regulation and disease treatment. Currently, few method can integrate multi-relationship network information and sequence information to identify functional MTIs. In this study, we develop MiRGraph, a multi-view deep learning framework to improve MTI prediction. To represent sequence features, we design two feature encoders namely TransCNN_miRNA and TransCNN_gene to jointly learn the MTI-specific sequence elements from miRNA and the entire 3' UTR sequence of a candidate target genes, respectively. This design of TransCNN_gene is similar to Enformer [Avsec et al., 2021], which was nonetheless designed to predict chromatin profiles from DNA sequences and was unable to leverage graph information.

To strengthen the role of multi-relational and structural information for the network feature learning of miRNAs and genes, we utilize a heterogeneous graph transformer (HGT) to effectively extract complex heterogeneous network information. Finally, we combine a MLP layer and a bilinear function to obtain the MTIs prediction scores. Our experimental results suggest that MiRGraph conferred significant performance improvements in MTIs predictions over the state-of-the-art methods. Our case studies highlight putative targets of the oncomir *hsa-miR-122-5p* in cancer development and the putative miRNA regulators of *BRCA1*.

We consider several future extensions. MiRGraph is trained to distinguish whether the TargetScan-predicted MTIs are experimentally validated in miRTarBase. One caveat is that some TargetScan-predicted edges may be true positives but do not yet have experimental supports recorded in miRTarbase. Alternatively,

we can use HITS-CLIP experimental data to identify negative edges. Another caveat is that we only use the 279 miRNAs, which have experimentally validated targets in miRTarbase. Alternatively, we can formulate a semi-supervised learning by considering the entire repertoire of human miRNAs, which consists of over a thousand of mature miRNAs. Furthermore, MiRGraph can be extended to utilize gene regulatory network and gene expression data.

6 Acknowledgments

Jiawei Luo is supported by funds from the National Natural Science Foundation of China (Grant #62032007, #62372165 and #62202152). Yue Li is supported by Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (RGPIN-2019-0621).

Bibliography

V. Agarwal, G. W. Bell, J.-W. Nam, and D. P. Bartel. Predicting effective microrna target sites in mammalian mrnas. *elife*, 4:e05005, 2015.

P. Androvic, L. Valihrach, J. Elling, R. Sjoback, and M. Kubista. Two-tailed rt-qpcr: a novel method for highly accurate mirna quantification. *Nucleic acids research*, 45(15):e144–e144, 2017.

Ž. Avsec, V. Agarwal, D. Visentin, J. R. Ledsam, A. Grabska-Barwinska, K. R. Taylor, Y. Assael, J. Jumper, P. Kohli, and D. R. Kelley. Effective gene expression prediction from sequence by integrating long-range interactions. *Nature methods*, 18(10):1196–1203, 2021.

D. P. Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *cell*, 116(2):281–297, 2004.

D. Betel, A. Koppal, P. Agius, C. Sander, and C. Leslie. Comprehensive modeling of microrna targets predicts functional non-conserved and non-canonical sites. *Genome biology*, 11:1–14, 2010.

G. A. Calin and C. M. Croce. Microrna signatures in human cancers. *Nature reviews cancer*, 6(11):857–866, 2006.

S. W. Chi, J. B. Zang, A. Mele, and R. B. Darnell. Argonaute hits-clip decodes microrna–mrna interaction maps. *Nature*, 460(7254):479–486, 2009a.

S. W. Chi, J. B. Zang, A. Mele, and R. B. Darnell. Argonaute hits-clip decodes microrna–mrna interaction maps. *Nature*, 460(7254):479–486, 2009b.

Y. Duan, Y. Dong, R. Dang, Z. Hu, Y. Yang, Y. Hu, and J. Cheng. Mir-122 inhibits epithelial mesenchymal transition by regulating p4ha1 in ovarian cancer cells. *Cell biology international*, 42(11):1564–1574, 2018.

A. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. Marks. Microrna targets in drosophila. *Genome biology*, 4:1–27, 2003.

L. F. Gebert and I. J. MacRae. Regulation of microrna function in animals. *Nature reviews Molecular cell biology*, 20(1):21–37, 2019.

S. Griffiths-Jones, R. J. Grocock, S. Van Dongen, A. Bateman, and A. J. Enright. mirbase: microrna sequences, targets and gene nomenclature. *Nucleic acids research*, 34(suppl_1):D140–D144, 2006.

M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A.-C. Jungkamp, M. Munschauer, et al. Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip. *Cell*, 141(1):129–141, 2010.

K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey. Mapping the human mirna interactome by clash reveals frequent noncanonical binding. *Cell*, 153(3):654–665, 2013.

S.-D. Hsu, F.-M. Lin, W.-Y. Wu, C. Liang, W.-C. Huang, W.-L. Chan, W.-T. Tsai, G.-Z. Chen, C.-J. Lee, C.-M. Chiu, et al. mirtarbase: a database curates experimentally validated microrna–target interactions. *Nucleic acids research*, 39(suppl_1):D163–D169, 2011.

Z. Hu, Y. Dong, K. Wang, and Y. Sun. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, pages 2704–2710, 2020.

H.-Y. Huang, Y.-C.-D. Lin, S. Cui, Y. Huang, Y. Tang, J. Xu, J. Bao, Y. Li, J. Wen, H. Zuo, et al. mirtarbase update 2022: an informative resource for experimentally validated mirna–target interactions. *Nucleic acids research*, 50(D1):D222–D230, 2022.

D. Karagkouni, M. D. Paraskevopoulou, S. Chatzopoulos, I. S. Vlachos, S. Tastsoglou, I. Kanellos, D. Papadimitriou, I. Kavakiotis, S. Maniou, G. Skoufos, et al. Diana-tarbase v8: a decade-long collection of experimentally supported mirna–gene interactions. *Nucleic acids research*, 46(D1):D239–D245, 2018.

M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, and E. Segal. The role of site accessibility in microrna target recognition. *Nature genetics*, 39(10):1278–1284, 2007.

A. A. Khan, D. Betel, M. L. Miller, C. Sander, C. S. Leslie, and D. S. Marks. Transfection of small rnas globally perturbs gene regulation by endogenous micrornas. *Nature biotechnology*, 27(6):549–555, 2009.

E. C. Lai. Micrornas are complementary to 3 utr sequence motifs that mediate negative post-transcriptional regulation. *Nature genetics*, 30(4):363–364, 2002.

B. Lee, J. Baek, S. Park, and S. Yoon. deeptarget: end-to-end learning framework for microrna target prediction using deep recurrent neural networks. In *Proceedings of the 7th ACM international conference on bioinformatics, computational biology, and health informatics*, pages 434–442, 2016.

A. Li, W. Song, J. Qian, Y. Li, J. He, Q. Zhang, W. Li, A. Zhai, W. Kao, Y. Hu, et al. Mir-122 modulates type i interferon expression through blocking suppressor of cytokine signaling 1. *The international journal of biochemistry & cell biology*, 45(4):858–865, 2013.

J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, and J.-H. Yang. starbase v2. 0: decoding mirna-cerna, mirna-ncrna and protein–rna interaction networks from large-scale clip-seq data. *Nucleic acids research*, 42(D1):D92–D97, 2014.

Y. Li and Z. Zhang. Computational biology in microrna. *Wiley Interdisciplinary Reviews: RNA*, 6(4): 435–452, 2015.

T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

P. Liu, J. Luo, and X. Chen. mircom: tensor completion integrating multi-view information to deduce the potential disease-related mirna-mirna pairs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1747–1759, 2020.

M. Lu, B. Shi, J. Wang, Q. Cao, and Q. Cui. Tam: a method for enrichment and depletion analysis of a microrna category in a list of micrornas. *BMC bioinformatics*, 11(1):1–7, 2010.

J. Luo, W. Ouyang, C. Shen, and J. Cai. Multi-relation graph embedding for predicting mirna-target gene interactions by integrating gene sequence information. *IEEE Journal of Biomedical and Health Informatics*, 26(8):4345–4353, 2022.

A. Ma, X. Wang, J. Li, C. Wang, T. Xiao, Y. Liu, H. Cheng, J. Wang, Y. Li, Y. Chang, et al. Single-cell biological network inference using a heterogeneous graph transformer. *Nature Communications*, 14(1):964, 2023.

S. E. McGeary, K. S. Lin, C. Y. Shi, T. M. Pham, N. Bisaria, G. M. Kelley, and D. P. Bartel. The biochemical basis of microrna targeting efficacy. *Science*, 366(6472):eaav1741, 2019.

S. Min, B. Lee, and S. Yoon. Targetnet: functional microrna target prediction with deep neural networks. *Bioinformatics*, 38(3):671–677, 2022.

Y. Peng and C. M. Croce. The role of micrornas in human cancer. *Signal transduction and targeted therapy*, 1(1):1–9, 2016.

A. Pla, X. Zhong, and S. Rayner. miraw: A deep learning-based approach to predict microrna targets by analyzing whole microrna transcripts. *PLoS computational biology*, 14(7):e1006185, 2018.

M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer, 2018.

D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, et al. String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1): D607–D613, 2019.

Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.

J. Wen, B. J. Parker, A. Jacobsen, and A. Krogh. Microrna transfection and ago-bound clip-seq data sets reveal distinct determinants of mirna action. *Rna*, 17(5):820–834, 2011.

W. Xie, Z. Zheng, W. Zhang, L. Huang, Q. Lin, and K.-C. Wong. Srg-vote: Predicting mirna-gene relationships via embedding and lstm ensemble. *IEEE Journal of Biomedical and Health Informatics*, 26(8): 4335–4344, 2022.