


Article

GATCDA: Predicting circRNA-Disease Associations Based on Graph Attention Network

Chen Bian ¹, Xiu-Juan Lei ^{1,*} and Fang-Xiang Wu ^{2,*} ¹ School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; bianchen@snnu.edu.cn² Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

* Correspondence: xjlei@snnu.edu.cn (X.-J.L.); faw341@mail.usask.ca (F.-X.W.);

Tel.: +86-029-8531-0161 (X.-J.L.); +1-306-966-5280 (F.-X.W.)

Simple Summary: CircRNAs (circular RNAs), a novel kind of non-coding RNAs, play a regulatory role in cellular processes. A growing number of biological experiments has proved that circRNAs can be used as biomarkers and therapeutic targets of some cancers. As the time and financial costs of biological experiments are high, computational methods have become a better way to predict the associations between circRNAs and diseases. Graph attention network was first applied to predict circRNA-disease associations with multiple similarities of data in this study. The circRNA-miRNA interactions and disease-mRNA interactions were adopted to construct features. The computational method proposed in this study has improved the prediction performance.

Abstract: CircRNAs (circular RNAs) are a class of non-coding RNA molecules with a closed circular structure. CircRNAs are closely related to the occurrence and development of diseases. Due to the time-consuming nature of biological experiments, computational methods have become a better way to predict the interactions between circRNAs and diseases. In this study, we developed a novel computational method called GATCDA utilizing a graph attention network (GAT) to predict circRNA-disease associations with disease symptom similarity, network similarity, and information entropy similarity for both circRNAs and diseases. GAT learns representations for nodes on a graph by an attention mechanism, which assigns different weights to different nodes in a neighborhood. Considering that the circRNA-miRNA-mRNA axis plays an important role in the generation and development of diseases, circRNA-miRNA interactions and disease-mRNA interactions were adopted to construct features, in which mRNAs were related to 88% of miRNAs. As demonstrated by five-fold cross-validation, GATCDA yielded an AUC value of 0.9011. In addition, case studies showed that GATCDA can predict unknown circRNA-disease associations. In conclusion, GATCDA is a useful method for exploring associations between circRNAs and diseases.

Keywords: circRNA-disease association; graph attention network; circRNA-miRNA-mRNA axis



Citation: Bian, C.; Lei, X.-J.; Wu, F.-X. GATCDA: Predicting circRNA-Disease Associations Based on Graph Attention Network. *Cancers* **2021**, *13*, 2595. <https://doi.org/10.3390/cancers13112595>

Academic Editor: Christos K. Kontos

Received: 28 March 2021

Accepted: 22 May 2021

Published: 25 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

CircRNAs (circular RNAs) are a class of non-coding RNA molecules with a closed circular structure, without a 5'-end cap and a 3'-end ployA tail. They are mainly located in the cytoplasm or stored in exosomes, and are not affected by RNA exonuclease [1]. Although circRNAs are non-coding RNAs, some circRNAs can encode polypeptides. Currently, biological functions of circRNAs are well-recognized as follows [2]: miRNA sponges, regulatory protein binding, regulation of gene transcription, and coding functions. CircRNA expression is more stable and not easily degradable, and has been proved to exist widely in a variety of eukaryotes [1]. Most circRNAs are formed by exon loops, and some circRNAs are lariat structures formed by intron loops. Because circRNAs contain a number of miRNA response elements (MREs), they can form the catalytic core of the RNA-induced silencing complex (RISC) with AGO proteins, which eventually leads to the degradation

of circRNAs [3]. According to their sources, circRNAs can be roughly divided into four categories [4]: full-exon circRNAs, exon-introns circRNA (EIcircRNAs), intron-composed lariat circRNAs, and circRNAs produced by cyclization of viral RNA genomes (tRNA, rRNA, snRNA, etc.). Twenty years ago, scientists found circRNAs from plant viroids, yeast mitochondria, and hepatitis B viruses (HBV) as byproducts of abnormal splicing that have no regulatory function [5]. In 2013, Hansen et al. proposed and confirmed for the first time that circRNA is the regulatory mechanism of the miRNA sponge [6], providing a new field for circRNA research. With the rapid development in RNA sequencing technology and bioinformatics analysis, 14,807 candidate circRNAs have been identified in the human tissue transcriptome, and many exons have been found to form circRNAs by nonlinear reverse splicing or gene rearrangement in cells of other species.

In recent years, many studies [7–9] have shown that circRNAs are closely related to the occurrence and development of diseases, and predicted circRNAs' application prospects in aspects of diagnostic markers of diseases. For example, after brain/spinal cord injury, circRNAs can activate several biological, molecular, and cellular activities. Therefore, interventions centered on the regulation of circRNAs may be promising for traumatic brain injury and spinal cord injury [10]. Chen et al. demonstrated that circRNA circCTNNA1 promoted colorectal cancer progression by sponging miR-149-5p and regulating FOXM1 expression [11]. Wang et al. found that circCNST promoted the tumorigenesis of osteosarcoma cells by sponging miR-421 and targeting SLC25A3, providing a potential biomarker for patients with osteosarcoma [12]. Wu et al. discovered that circ_0009582, circ_0037120, and circ_0140117 may serve as potential biomarkers for predicting the occurrence of hepatocellular carcinoma in patients with HBV infection [13]. Li et al. revealed that circRNA circITGA7 may play a regulatory role in thyroid cancer and may be a potential marker for thyroid cancer diagnosis or progression [14].

At present, the number of known interactions between circRNAs and diseases obtained through biological experiments is increasing. Some relevant databases have appeared [15–18], including the interactions between circRNAs and diseases verified by biological experiments. Due to the significant time and financial costs associated with biological experiments, in recent years, it has become a hot topic to predict associations between circRNAs and diseases using computational methods with these databases. The current calculation methods can be divided into five categories [19]. The first category is a network propagating method. For example, the computational model BRWSP applies biased random walk to search paths on a multiple heterogeneous network to discover circRNA-disease associations [20]. A method called KATZHCDA uses KATZ measures for human circRNA-disease association prediction [21]. The second category is a recommendation system method. For example, Lei et al. proposed a computational method named ICFCDA based on collaboration filtering recommendation system, handling the “cold start” problem to predict potential circRNA-disease associations [22]. The third category is the matrix completion methods. For example, a computational method called iCircDA-MF was developed based on matrix factorization by Wei et al. [23]. Zhang et al. [24] utilized metapath2vec++ and matrix factorization to discover circRNA-disease associations. The fourth category is the classical machine learning methods. For example, based on a gradient boosting decision tree, a model named GBDTCDA was proposed by Lei et al. in 2019 [25]. A computational model called RWLR uses logistic regression to predict circRNA-disease associations [26]. The fifth category is the deep learning methods. For example, Wang's method [27] applies a convolutional neural network to discover unknown circRNA-disease associations. In 2020, GCNCDA was proposed based on a graph convolutional network [28].

In this paper, we propose a novel computational model named GATCDA to predict circRNA-disease associations with graph attention network (GAT). First, we construct a circRNA-disease association network, a circRNA-miRNA association network and a disease-mRNA association network. Second, we calculate disease symptom similarity, network similarity and information entropy similarity for both circRNAs and diseases.

Third, these similarities are integrated to create the features of circRNAs and diseases. Fourth, the circRNA-disease association network and the features of circRNAs and diseases are fed into GAT, and the output is the prediction score of associations between circRNAs and diseases.

2. Materials and Methods

The whole flowchart of GATCDA is shown in Figure 1.

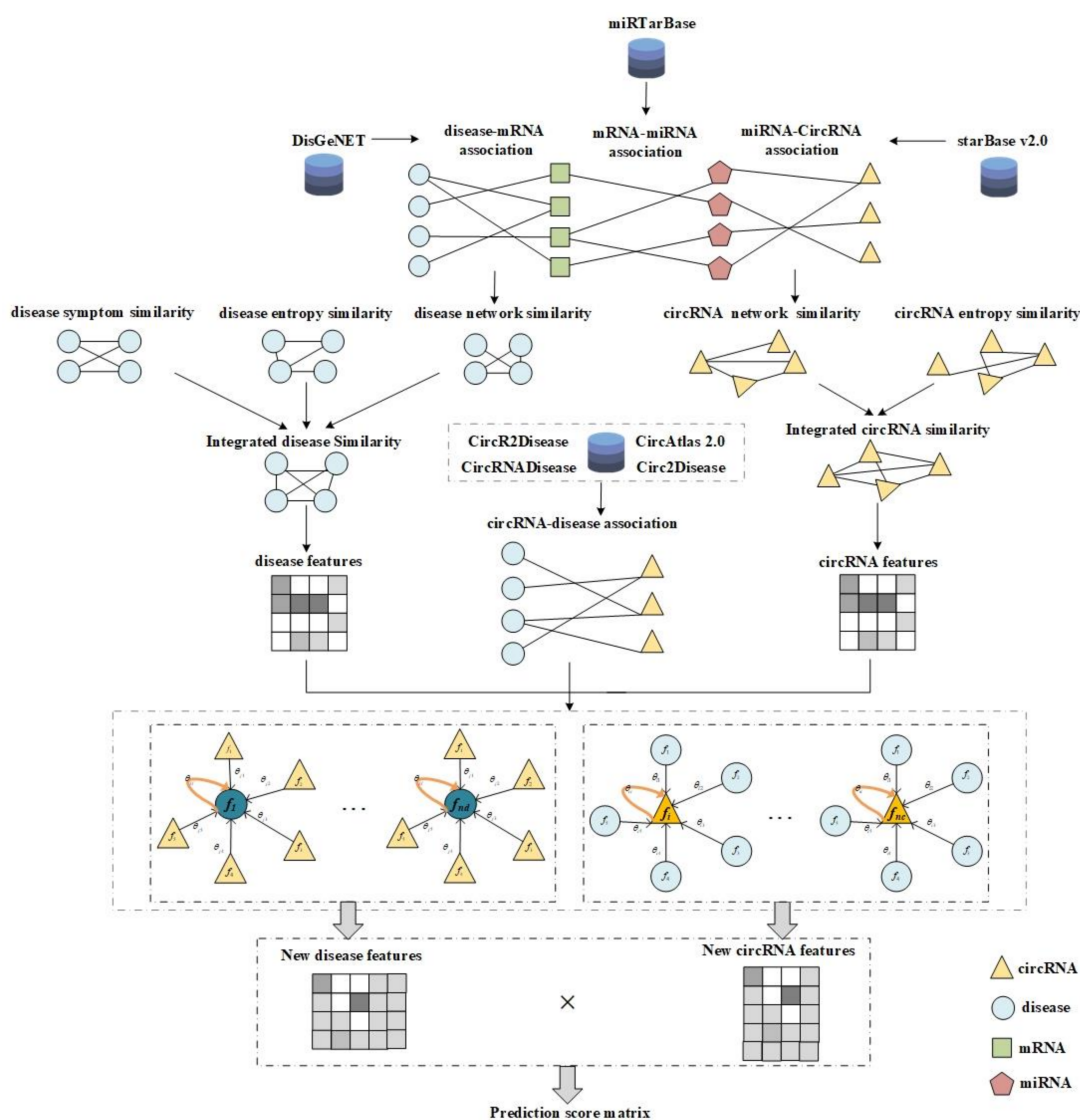


Figure 1. The flowchart of the computational method GATCDA.

2.1. Dataset Curation

2.1.1. CircRNA-Disease Association

The circRNA-disease associations were downloaded from CircR2Disease [15], CircAtlas 2.0 [16], Circ2Disease [17], and CircRNADisease [18], in which the number of circRNA-disease associations are 739, 927, 273, and 354, respectively. After integration, we obtained 768 circRNA-disease associations, including 624 circRNAs and 102 diseases.

2.1.2. CircRNA-MiRNA Association and Disease-mRNA Association

An initial circRNA-miRNA association dataset was downloaded from the starBase v2.0 [29] including 130,000+ circRNA-miRNA interactions, with 276 miRNA entries and

7018 circRNA entries. We selected 82 circRNAs common to circRNA-disease interactions and circRNA-miRNA interactions. There were only 142 miRNAs related to these 82 circRNAs in the initial circRNA-miRNA interactions. Finally, we constructed a new circRNA-miRNA association network including 509 circRNA-miRNA associations among 142 miRNAs and 82 circRNAs.

An initial disease-mRNA association dataset was downloaded from DisGeNET [30] including 60,000+ disease-mRNA interactions. We selected 37 diseases common to circRNA-disease interactions and disease-mRNA interactions. There were 820 mRNAs related to these 37 diseases in the initial disease-mRNA interactions. Finally, we constructed a new disease-mRNA association network including 1239 disease-mRNA associations among 37 diseases and 820 mRNAs.

CircRNAs act as miRNA sponges in cells and increase the expression level of target genes. The circRNA-miRNA-mRNA axis plays an important regulatory role in diseases [2,31,32]. In the new circRNA-miRNA associations, we found all the miRNAs related with diseases from circRNA-disease associations or mRNAs from disease-mRNA associations. Furthermore, in the new disease-mRNA associations, we found all the mRNAs related to the 125 of 142 miRNAs mentioned above.

2.1.3. Construction of the Interaction Network

For convenience, we formulate circRNA-disease associations as a binary matrix $Y \in R^{624 \times 102}$. If there exists an experimentally verified interaction between circRNA c_i and disease d_j , $Y(i, j) = 1$; otherwise, $Y(i, j) = 0$. At the same time, a circRNA-miRNA interaction matrix and a disease-mRNA interaction matrix are constructed in the same way based on circRNA-miRNA associations and disease-mRNA associations, respectively.

2.2. Similarity Calculation

2.2.1. Network Similarity

Zhou et al. [33] demonstrated the usefulness of network similarity. For a given miRNA mi_k , we denote the set of its interacting circRNAs by $C(mi_k)$. For a given mRNA m_k , we denote the set of its interacting diseases by $D(m_k)$. The network contribution of miRNA mi_k in the circRNA-miRNA interaction network can be calculated as follows

$$nc(mi_k) = -\ln \left(|C(mi_k)| / \sum_{k=1}^y |C(mi_k)| \right), \tag{1}$$

where $nc(mi_k)$ is the network contribution of miRNA mi_k in the circRNA-miRNA interaction network, and y is the number of miRNAs. The network contribution of mRNA m_k in the disease-mRNA interaction network can be calculated as follows

$$nc(m_k) = -\ln \left(|D(m_k)| / \sum_{k=1}^z |D(m_k)| \right), \tag{2}$$

where $nc(m_k)$ is the network contribution of mRNA m_k in the disease-mRNA interaction network, and z is the number of mRNAs. We also denote the set of miRNAs that interact with a given circRNA c_u by $Mi(c_u)$, and the set of mRNAs that interact with a given disease d_u by $M(d_u)$. The network similarity between circRNA c_u and circRNA c_v can be defined as

$$CNS(c_u, c_v) = \sum_{mi_k \in Mi(c_u) \cap Mi(c_v)} nc(mi_k), \tag{3}$$

where $CNS(c_u, c_v)$ is the network similarity between circRNA c_u and circRNA c_v . Similarly, given two diseases, d_u and d_v , the network similarity between disease d_u and disease d_v can be defined as

$$DNS(d_u, d_v) = \sum_{m_k \in M(d_u) \cap M(d_v)} nc(m_k), \tag{4}$$

where $DNS(d_u, d_v)$ is the network similarity between disease d_u and disease d_v .

2.2.2. Information Entropy Similarity

Information entropy is also used to measure topological similarities of circRNAs and diseases. For a given circRNA c_u , we denote the set of its interacting diseases by $T_m^{c_u}$. For a given circRNA c_v , we denote the set of its interacting circRNAs by $T_m^{c_v}$. Next, the information entropy of $T_m^{c_u}$ can be calculated as

$$\begin{cases} H(T_m^{c_u}) = -\sum_{i=1}^{nd} p(T_m^{c_u}(i)) \log_2(p(T_m^{c_u}(i))) \\ p(T_m^{c_u}(i)) = n(T_m^{c_u}(i))/N_{cd} \end{cases}, \tag{5}$$

where nd is the number of diseases related with circRNA c_u , N_{cd} is the total number of known circRNA-disease interactions, $n(T_m^{c_u}(i))$ is the number of interactions between the i th disease in the related disease set of circRNA c_u and all circRNAs, and $p(T_m^{c_u}(i))$ is the rate of the i th disease in the related disease set of circRNA c_u with the known circRNA-disease interactions. The information entropy similarity between circRNA c_u and circRNA c_v can be calculated as

$$CES(c_u, c_v) = \frac{2 * H(T_m^{c_u} \cap T_m^{c_v})}{H(T_m^{c_u}) + H(T_m^{c_v})}, \tag{6}$$

where $H(T_m^{c_u} \cap T_m^{c_v})$ is the information entropy of the intersection of $T_m^{c_u}$ and $T_m^{c_v}$, and $CES(c_u, c_v)$ is the information entropy similarity of circRNA c_u and circRNA c_v .

Similarly, the information entropy similarity between disease d_u and disease d_v can be calculated as follows

$$DES(d_u, d_v) = \frac{2 * H(T_n^{d_u} \cap T_n^{d_v})}{H(T_n^{d_u}) + H(T_n^{d_v})}, \tag{7}$$

where $T_n^{d_u}$ is the set of disease d_u s interacting circRNAs, $T_n^{d_v}$ is the set of disease d_v s interacting circRNAs, $H(T_n^{d_u} \cap T_n^{d_v})$ is the information entropy of the intersection of $T_n^{d_u}$ and $T_n^{d_v}$, and $DES(d_u, d_v)$ is the information entropy similarity of disease d_u and disease d_v .

2.2.3. Disease Symptom Similarity

According to the co-occurrence of diseases and symptom terms recorded in the PubMed bibliography, and the work of Zhou et al. [34], the disease similarity can be measured and a symptom-based human disease network can be constructed. Here, the symptom-based disease similarity matrix DSS was obtained from the symptom profiles of diseases.

2.2.4. Integration of Similarities

The integrated circRNA similarities and integrated disease similarities are regarded as circRNA features and disease features, respectively. The integrated circRNA similarities can be calculated as follows:

$$ICS(c_u, c_v) = \begin{cases} \beta \times CNS(c_u, c_v) + (1 - \beta) \times CES(c_u, c_v) \\ CES(c_u, c_v) \end{cases}, \tag{8}$$

where $CNS(c_u, c_v)$ is the circRNA network similarity between circRNA c_u and circRNA c_v , $CES(c_u, c_v)$ is the circRNA information entropy similarity between circRNA c_u and circRNA c_v , and $ICS(c_u, c_v)$ is the integrated circRNA similarity between circRNA c_u and circRNA c_v . β is an adjusting parameter. The dimensions of the ICS matrix are 624×624 .

The integrated disease similarities can be calculated as follows

$$IDS(d_u, d_v) = \begin{cases} \alpha \times (DNS(d_u, d_v) + DSS(d_u, d_v)) + (1 - \alpha) \times DES(d_u, d_v) \\ DES(d_u, d_v) \end{cases}, \tag{9}$$

where $DNS(d_u, d_v)$ is the disease network similarity between disease d_u and disease d_v , $DES(d_u, d_v)$ is the disease information entropy similarity between disease d_u and disease d_v , $DSS(d_u, d_v)$ is the disease symptom similarity between disease d_u and disease d_v , and $IDS(d_u, d_v)$ is the integrated disease similarity between disease d_u and disease d_v . α is an adjusting parameter. The dimensions of the IDS matrix are 102×102 .

2.3. Graph Attention Network

GAT [35] combines a weighted sum of the adjacent node features with the attention mechanism. The weight of the adjacent node features is completely dependent on the node features and independent of graph structure. GAT aims to construct a hidden self-attention layer and to learn representations for nodes on a graph by assigning different weights to different nodes in a neighborhood.

The input of graph attention layer is

$$f = \{f_1, f_2, \dots, f_N\}, f_i \in R^F, \quad (10)$$

where N is number of nodes (all circRNAs and all diseases), F is the length of features, and matrix $f \in R^{N \times F}$ denotes the features of all nodes.

The output of the graph attention layer is

$$f' = \{f'_1, f'_2, \dots, f'_N\}, f'_i \in R^{F'}, \quad (11)$$

where F' denotes the dimension of new features and matrix $f' \in R^{N \times F'}$ denotes the new features of all nodes.

The first step is to learn the importance of the neighbors for a given node. GAT implements the self-attention mechanism for every node. The attention coefficient e_{ij} for an association pair between circRNA c_i and disease d_j is formulated as follows

$$e_{ij}(c_i, d_j) = att(Wf_i, Wf_j), \quad (12)$$

where att denotes a single-layer feed-forward neural network that transforms input features into high-level features for circRNAs and diseases, and $W \in R^{F' \times F}$ is a weight matrix.

To make the attention coefficient comparable across different nodes, GAT further normalizes the attention coefficient e_{ij} as follows

$$\theta_{ij} = softmax(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{t \in N_i} \exp(e_{it})}, \quad (13)$$

where N_i is the set of neighbor nodes of circRNA c_i , and θ_{ij} is the normalized attention coefficient indicating the importance of disease d_j for circRNA c_i in the process of information propagation.

By combining Formulas (12) and (13), the complete attention mechanism can be obtained as follows

$$\theta_{ij} = \frac{\exp(leakyReLU(a^T [Wf_i || Wf_j]))}{\sum_{t \in N_i} \exp(leakyReLU(a^T [Wf_i || Wf_t]))}, \quad (14)$$

where $leakyReLU$ is a nonlinearity activation function assigning all negative values a non-zero slope, T denotes transposition, $||$ is the concatenation operation, and $a \in R^{2F'}$ is the weight coefficient matrix of the graph attention layer.

The second step is to fuse the representations of the neighbors for a given node according to their attention coefficients. The embedding of a given node can be fused by the projected node features of neighbors with different weights as follows

$$f'_i = \sigma(\sum_{t \in N_i} \theta_{it} Wf_t), \quad (15)$$

where σ is a nonlinear activation function.

GAT applies a multi-head attention mechanism to increase the stability of the learning process of self-attention. Multi-head attention is the combination of multiple self-attention structures. Each head learns the features in different representation spaces, and the focus of attention learned by multiple heads may be slightly different, which increases the capacity of the model. Specifically, K -independent attention mechanisms are integrated to achieve embedding as follows:

$$f'_i = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{t \in N_i} \theta_{it}^k \cdot W^k f_t\right), \tag{16}$$

where K is the number of attention mechanisms and W^k is the weight matrix for the k th attention mechanism.

Ultimately, the probability score matrix S can be calculated as follows

$$S = U \times V^T, \tag{17}$$

where $U \in R^{nc \times F'}$ is the final representation matrix of the circRNAs, in which nc is the number of circRNAs; and $V \in R^{nd \times F'}$ is the final representation matrix of diseases, in which nd is the number of diseases. The dimension of probability score matrix S is $nc \times nd$.

The detailed procedures of using GAT to predict the associations between circRNAs and diseases are shown in Figure 2. As shown in Figure 2, the circRNA-disease association network is fed into a GAT in which the final node representation is obtained through feature propagation and attention fusion. Finally, the prediction score is calculated according to node representation. In the case of disease d_3 and circRNA c_2 , the dark blue row represents d_3 , the dark yellow row represents c_2 , and the red grid represents the predicted score of the association between d_3 and c_2 .

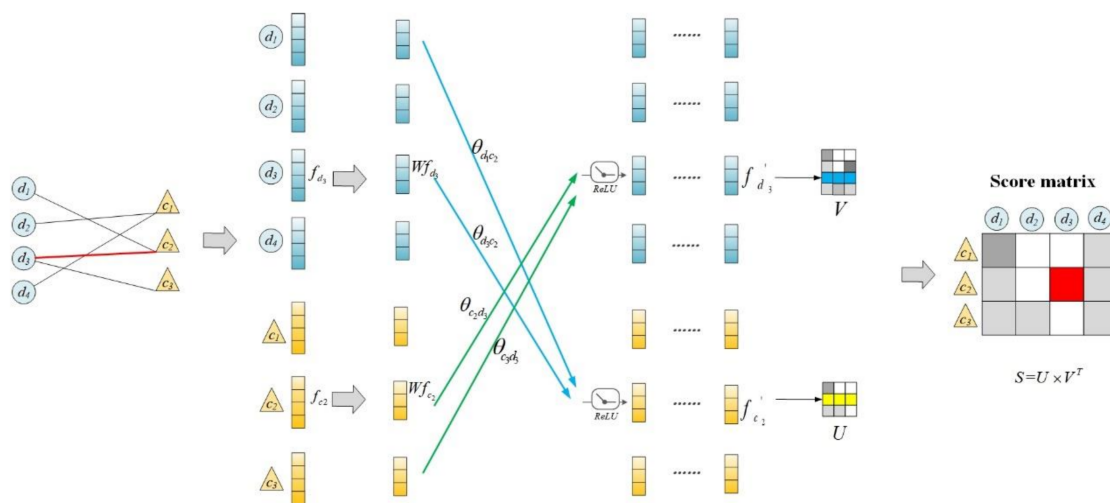


Figure 2. The detailed procedures of using GAT to predict the associations between circRNAs and diseases.

3. Results

3.1. Performance Evaluation

The five-fold cross-validation (5CV) technique was used to evaluate the prediction performance of our model. The 5CV technique randomly divides the positive samples into five equal parts, and takes out one part of them as a testing sample while the rest of samples are regarded as training samples. Next, the predicted scores are sorted in descending order. We drew the receiver operating characteristics (ROC) curve via plotting the true positive rate (TPR) versus the false positive rate (FPR) at different score thresholds. TPR (FPR) refers to the percentage of positive (negative) cases that are correctly identified. Generally, the area under the ROC curve (AUC) is calculated and employed to evaluate the prediction performance. Specifically, the closer the AUC value is to one, the better

the prediction performance. As a result, in 5CV, GATCDA achieved an AUC of 0.9011. In addition, GATCDA yielded an accuracy of 0.8710, with a precision of 0.9013.

3.2. Adjustment of Parameters

The GATCDA model involves two parameters, α and β , which adjust the influence of similarity data when calculating integrated similarities. We let α and β both range between 0.1 and 0.9. As a result, GATCDA ($\alpha = 0.1$, $\beta = 0.1$) gained the highest AUCs of 0.9011 in 5CV as shown in Figure 3.

	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$	$\beta = 0.8$	$\beta = 0.9$
$\alpha = 0.1$	0.9011	0.8802	0.8745	0.8742	0.8719	0.8632	0.8622	0.8615	0.8611
$\alpha = 0.2$	0.8608	0.8594	0.8584	0.8584	0.8572	0.8571	0.8554	0.8526	0.8519
$\alpha = 0.3$	0.8517	0.8511	0.8508	0.8499	0.8494	0.8485	0.8471	0.8451	0.8438
$\alpha = 0.4$	0.8411	0.8406	0.8399	0.8392	0.8390	0.8383	0.8371	0.8359	0.8352
$\alpha = 0.5$	0.8349	0.8347	0.8330	0.8323	0.8319	0.8311	0.8301	0.8300	0.8285
$\alpha = 0.6$	0.8281	0.8277	0.8275	0.8260	0.8245	0.8241	0.8237	0.8236	0.8229
$\alpha = 0.7$	0.8228	0.8211	0.8208	0.8180	0.8171	0.8171	0.8151	0.8139	0.8139
$\alpha = 0.8$	0.8125	0.8122	0.8096	0.8084	0.8071	0.8060	0.8050	0.8022	0.7982
$\alpha = 0.9$	0.7943	0.7928	0.7912	0.7842	0.7690	0.7647	0.7529	0.7479	0.7378

Figure 3. Heatmap of AUC results of adjustment parameters.

3.3. Compared with Other Methods

To analyze the performance of GATCDA in predicting circRNA-disease associations, we compared GATCDA with other four methods: DWNN-RLS [36], KATZHCDA [21], bi-random walks (BiRWR) [37], and DeepWalk [38]. DWNN-RLS uses the regularized least squares of the Kronecker product kernel to predict circRNA-disease associations. KATZHCDA uses KATZ measures for human circRNA-disease association prediction. BiRWR predicts circRNA-disease associations by walking in a circRNA subnetwork and disease subnetwork. DeepWalk is a way to learn the potential representation of nodes in a graph structure. The ROC curve and AUC value of each method using 5CV are shown in Figure 4. The precision-recall curve and the area under the precision-recall curves (AUPR) value of each method with 5CV are shown in Figure 5. Through comparison, it can be seen that the extraction of circRNA and disease features by GAT can achieve better prediction performance compared with DeepWalk. In addition, as a deep learning method, GAT also shows better prediction performance compared with the other two link-based prediction methods (KATZHCDA and BiRWR).

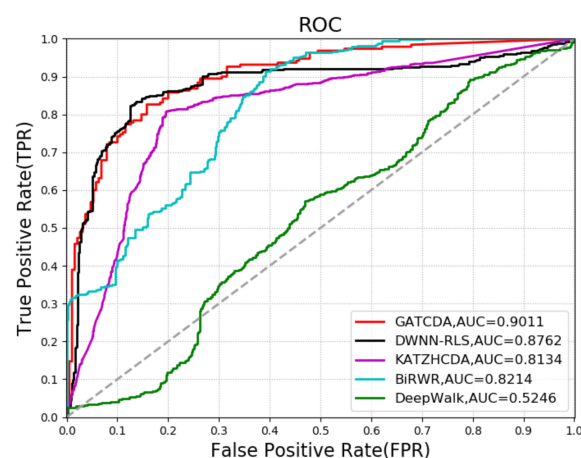


Figure 4. The ROC curves and AUCs of five methods using 5CV.

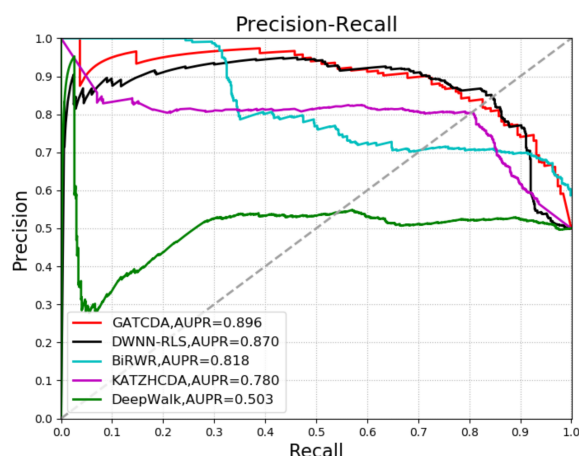


Figure 5. Comparison of five methods in PR curves and AUPRs (5CV).

3.4. Case Study

To further evaluate the prediction performance of GATCDA, we also carried out case studies on three common diseases, i.e., bladder cancer, diabetes retinopathy, and rheumatoid arthritis.

Bladder cancer is the most frequent cancer affecting the urinary tract [39], and has a high rate of recurrence [40]. Diabetes retinopathy is a common chronic metabolic disorder, increasing with an ageing population and the growing number of cases of diabetes [41]. Rheumatoid arthritis is the most common chronic inflammatory arthritis, which can lead to cartilage and bone damage and disability [42]. There is increasing evidence that circRNAs can be used as effective biomarkers for the diagnosis of bladder cancer, diabetes retinopathy, and rheumatoid arthritis. Therefore, we selected bladder cancer, diabetes retinopathy, and rheumatoid arthritis to verify the predictive ability of GATCDA.

In this work, all known associations between the investigated disease and circRNAs were assumed to be unknown. Through the calculation of GATCDA, the circRNAs with the top 10 scores were selected among all the predicted associations between the investigated disease and circRNAs. Then, through searching the related literature, some circRNAs were confirmed to be related to the investigated disease.

The results of the case studies of the three diseases (bladder cancer, diabetes retinopathy, and rheumatoid arthritis) are shown in Table 1. For bladder cancer, we can see that 8 of the top 10 candidates with the highest prediction scores are confirmed by the relevant literature. Notably, the seventh circRNA (hsa_circ_0075828) predicted by GATCDA is related to bladder cancer. For diabetes retinopathy and rheumatoid arthritis, 7 of the top 10 candidates with the highest prediction scores are confirmed by the relevant literature. For example, Li et al. found that hsa_circ_0001859 regulates ATF2 expression by functioning as an MiR-204/211 sponge in human rheumatoid arthritis [43]. Zhang et al. revealed that hsa_circ_0005015 acts as an miR-519d-3p sponge to inhibit miR-519d-3p activity, leading to increasing MMP-2, XIAP, and STAT3 expression in diabetes retinopathy [44].

In order to verify the prediction performance of GATCDA, we compared it with other models in the case studies of these three kinds of same diseases, as shown in Table 2. From Table 2, 8, 7, and 7 out of these top 10 circRNAs predicted by GATCDA were verified to be associated with bladder cancer, diabetes retinopathy, and rheumatoid arthritis, respectively, which are the highest among competing methods. Therefore, GATCDA also outperforms the competing prediction models in terms of the hit rate in case studies.

Table 1. Candidate circRNAs identified by GATCDA for bladder cancer, diabetes retinopathy and rheumatoid arthritis.

Disease	Rank	CircRNA	Source
Bladder cancer	1	hsa_circ_0091017	[45]
	2	hsa_circ_0002495	[46]
	3	hsa_circ_0071410	- ¹
	4	hsa_circ_0001141	[47]
	5	hsa_circ_0007915	-
	6	hsa_circ_0041103	[48]
	7	hsa_circ_0075828	[49]
	8	hsa_circ_0061265	[48]
	9	hsa_circ_0002768	[50]
	10	hsa_circ_0082582	[48]
Diabetes retinopathy	1	hsa_circ_0098964	-
	2	hsa_circ_0057093	[44]
	3	hsa_circ_0051172	-
	4	hsa_circ_0087215	[44]
	5	hsa_circ_0081162	[44]
	6	hsa_circ_0066922	[44]
	7	hsa_circ_0026388	[44]
	8	hsa_circ_0005525	-
	9	hsa_circ_0000615	[51]
	10	hsa_circ_0005015	[44]
Rheumatoid arthritis	1	hsa_circ_0083964	[52]
	2	hsa_circ_0064996	[52]
	3	hsa_circ_0004712	[52]
	4	hsa_circ_0061893	-
	5	hsa_circ_0052012	[52]
	6	hsa_circ_0032683	[52]
	7	hsa_circ_0001859	[43]
	8	hsa_circ_0088036	[53]
	9	hsa_circ_0003028	-
	10	hsa_circ_0010090	-

¹—means has no source.

Table 2. The number of circRNAs confirmed by evidence in the top 10 potential disease-related circRNAs predicted by GATCDA and other models in case studies of the three kinds of diseases such as bladder cancer, diabetes retinopathy, and rheumatoid arthritis.

Model	Bladder Cancer	Diabetes Retinopathy	Rheumatoid Arthritis
GATCDA	8	7	7
DWNN-RLS	7	5	4
KATZHCDA	5	4	4
BiRWR	5	3	4
DeepWalk	3	1	2

4. Discussion

With the rapid development in RNA sequencing technology and bioinformatics analysis, various studies have shown that circRNAs are closely related to the occurrence and development of disease, and circRNAs act a potential biomarkers for patients with certain cancers. Therefore, discovering associations between circRNAs and diseases is significative for disease diagnosis and treatment. Nevertheless, biological experiments are very costly in terms of time and money. It has become a hot topic to predict associations between circRNAs and diseases using computational methods in recent years. The lack of data on the interactions between circRNA and disease limits the predictive power of most computational methods.

In this study, we proposed a new computational model named GATCDA to identify underlying circRNA-disease associations. We performed 5CV experiments to assess the predictive performance of GATCDA. Our method yielded an AUC value of 0.9011 and an AUPR value of 0.896, which are higher than those of DWNN-RLS, KATZHCDA, BiRWR, and DeepWalk. In addition, the predicted top 10 circRNA-disease interactions in the case studies of three diseases (bladder cancer, diabetes retinopathy, and rheumatoid arthritis) have been confirmed in the relevant literature, which suggests that GATCDA can be an effective tool for predicting circRNA-disease associations.

The accurate predictive performance of GATCDA is attributed to the following factors: First, in order to identify more interactions between circRNAs and diseases, circRNA-disease interactions were integrated from four databases, i.e., CircR2Disease, CircAtlas 2.0, Circ2Disease, and CircRNADisease. Therefore, the number of positive samples input to GAT algorithm is higher. Second, as the circRNA-miRNA-mRNA axis plays an important role in the generation and development of diseases, the circRNA-miRNA interactions and disease-mRNA interactions were adopted to construct features, in which mRNAs are related to 88% of miRNAs. CircRNAs have several distinct modes of action. From the functional perspective of circRNAs as miRNA sponges, an interaction network can be constructed for circRNA network similarity calculations. Other functions of circRNA are difficult to quantify. Third, more similarities are involved in GATCDA, i.e., disease symptom similarity, disease network similarity, disease information entropy similarity, circRNA network, and circRNA information entropy similarity, which are integrated effectively. Fourth, GAT has obvious advantages, including learning representations for nodes on a graph using an attention mechanism. Therefore, it can assign different weights to different nodes in a neighborhood.

GATCDA also has limitations. Compared with other non-coding RNAs, the interactions between circRNAs and diseases are still insufficient. Therefore, the circRNA-disease association matrix is sparse, which has an impact on prediction performance. In the future, we will collect more data on the associations between circRNAs and diseases.

5. Conclusions

In this study, we proposed a new computational model named GATCDA to identify underlying circRNA-disease associations. Specifically, GAT was used to predict circRNA-disease associations based on multiple similarities of circRNA and disease. This work has two highlights: First, as the circRNA-miRNA-mRNA axis plays an important role in the generation and development of diseases, circRNA-miRNA interactions and disease-mRNA interactions are adopted to construct features, in which mRNAs are related to 88% of miRNAs. Second, GAT is used to predict the interactions between circRNAs and diseases. GAT can assign different learning weights to different neighbors, and the correlation between vertex features can be better integrated into the model. In terms of predictive performance, GATCDA achieves an AUC of 0.9011 in 5CV, and in case studies of three diseases, 70% of experimentally validated relationships were predicted. In summary, GATCDA is a powerful tool for predicting circRNA-disease associations.

Author Contributions: Conceptualization, C.B., X.-J.L. and F.-X.W.; methodology, C.B. and X.-J.L.; data curation, C.B.; writing and drafting of the manuscript, C.B., X.-J.L. and F.-X.W.; critical revision of the manuscript for important intellectual content, X.-J.L. and F.-X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 61972451 and 61902230; and the Fundamental Research Funds for the Central Universities of China, grant number GK201901010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, Y.; Zheng, Q.; Bao, C.; Li, S.; Guo, W.; Zhao, J.; Chen, D.; Gu, J.; He, X.; Huang, S. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis. *Cell Res.* **2015**, *25*, 981–984. [[CrossRef](#)] [[PubMed](#)]
2. Rong, D.; Sun, H.; Li, Z.; Liu, S.; Dong, C.; Fu, K.; Tang, W.; Cao, H. An emerging function of circRNA-miRNAs-mRNA axis in human diseases. *Oncotarget* **2017**, *8*, 73271–73281. [[CrossRef](#)]
3. Chen, Y.; Wei, S.; Wang, X.; Zhu, X.; Han, S. Progress in research on the role of circular RNAs in lung cancer. *World J. Surg. Oncol.* **2018**, *16*, 215. [[CrossRef](#)]
4. Zhang, L.; Hou, C.; Chen, C.; Guo, Y.; Yuan, W.; Yin, D.; Liu, J.; Sun, Z. The role of N6-methyladenosine (m6A) modification in the regulation of circRNAs. *Mol. Cancer* **2020**, *19*, 105. [[CrossRef](#)]
5. Patop, I.L.; Wüst, S.; Kadener, S. Past, present, and future of circRNAs. *EMBO J.* **2019**, *38*, e100836. [[CrossRef](#)]
6. Hansen, T.B.; Jensen, T.I.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nat. Cell Biol.* **2013**, *495*, 384–388. [[CrossRef](#)] [[PubMed](#)]
7. Han, B.; Chao, J.; Yao, H. Circular RNA and its mechanisms in disease: From the bench to the clinic. *Pharmacol. Ther.* **2018**, *187*, 31–44. [[CrossRef](#)]
8. Zhu, L.-P.; He, Y.-J.; Hou, J.-C.; Chen, X.; Zhou, S.-Y.; Yang, S.-J.; Li, J.; Zhang, H.-D.; Hu, J.-H.; Zhong, S.-L.; et al. The role of circRNAs in cancers. *Biosci. Rep.* **2017**, *37*. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, Z.; Yang, T.; Xiao, J. Circular RNAs: Promising Biomarkers for Human Diseases. *EBioMedicine* **2018**, *34*, 267–274. [[CrossRef](#)] [[PubMed](#)]
10. Yuan, J.; Botchway, B.O.A.; Zhang, Y.; Wang, X.; Liu, X. Role of Circular Ribonucleic Acids in the Treatment of Traumatic Brain and Spinal Cord Injury. *Mol. Neurobiol.* **2020**, *57*, 4296–4304. [[CrossRef](#)] [[PubMed](#)]
11. Chen, P.; Yao, Y.; Yang, N.; Gong, L.; Kong, Y.; Wu, A. Circular RNA circCTNNA1 promotes colorectal cancer progression by sponging miR-149-5p and regulating FOXM1 expression. *Cell Death Dis.* **2020**, *11*, 557. [[CrossRef](#)] [[PubMed](#)]
12. Wang, J.-H.; Wu, X.-J.; Duan, Y.-Z.; Li, F. Circular RNA_CNST Promotes the Tumorigenesis of Osteosarcoma Cells by Sponging miR-421. *Cell Transplant.* **2020**, *29*. [[CrossRef](#)]
13. Wu, C.; Deng, L.; Zhuo, H.; Chen, X.; Tan, Z.; Han, S.; Tang, J.; Qian, X.; Yao, A. Circulating circRNA predicting the occurrence of hepatocellular carcinoma in patients with HBV infection. *J. Cell. Mol. Med.* **2020**, *24*, 10216–10222. [[CrossRef](#)] [[PubMed](#)]
14. Li, S.; Yang, J.; Liu, X.; Guo, R.; Zhang, R. circITGA7 Functions as an Oncogene by Sponging miR-198 and Upregulating FGFR1 Expression in Thyroid Cancer. *BioMed Res. Int.* **2020**, *2020*, 8084028. [[CrossRef](#)]
15. Fan, C.; Lei, X.; Fang, Z.; Jiang, Q.; Wu, F.-X. CircR2Disease: A manually curated database for experimentally supported circular RNAs associated with various diseases. *Database* **2018**, *2018*, bay044. [[CrossRef](#)]
16. Ji, P.; Wu, W.; Chen, S.; Zheng, Y.; Zhou, L.; Zhang, J.; Cheng, H.; Yan, J.; Zhang, S.; Yang, P.; et al. Expanded Expression Landscape and Prioritization of Circular RNAs in Mammals. *Cell Rep.* **2019**, *26*, 3444–3460.e5. [[CrossRef](#)] [[PubMed](#)]
17. Yao, D.; Zhang, L.; Zheng, M.; Sun, X.; Lu, Y.; Liu, P. Circ2Disease: A manually curated database of experimentally validated circRNAs in human disease. *Sci. Rep.* **2018**, *8*, 11018. [[CrossRef](#)]
18. Zhao, Z.; Wang, K.; Wu, F.; Wang, W.; Zhang, K.; Hu, H.; Liu, Y.; Jiang, T. circRNA disease: A manually curated database of experimentally supported circRNA-disease associations. *Cell Death Dis.* **2018**, *9*, 475. [[CrossRef](#)] [[PubMed](#)]
19. Lei, X.; Mudiyansele, T.B.; Zhang, Y.; Bian, C.; Lan, W.; Yu, N.; Pan, Y. A comprehensive survey on computational methods of non-coding RNA and disease association prediction. *Brief. Bioinform.* **2020**. [[CrossRef](#)]
20. Lei, X.; Zhang, W. BRWSP: Predicting circRNA-Disease Associations Based on Biased Random Walk to Search Paths on a Multiple Heterogeneous Network. *Complexity* **2019**, *2019*, 5938035. [[CrossRef](#)]
21. Fan, C.; Lei, X.; Wu, F.-X. Prediction of CircRNA-Disease Associations Using KATZ Model Based on Heterogeneous Networks. *Int. J. Biol. Sci.* **2018**, *14*, 1950–1959. [[CrossRef](#)] [[PubMed](#)]
22. Lei, X.; Fang, Z.; Guo, L. Predicting circRNA-Disease Associations Based on Improved Collaboration Filtering Recommendation System with Multiple Data. *Front. Genet.* **2019**, *10*, 897. [[CrossRef](#)] [[PubMed](#)]
23. Hang, W.; Bin, L. iCircDA-MF: Identification of circRNA-disease associations based on matrix factorization. *Brief. Bioinform.* **2019**, *21*, 1356–1367.
24. Zhang, Y.; Lei, X.; Fang, Z.; Pan, Y. CircRNA-disease associations prediction based on metapath2vec++ and matrix factorization. *Big Data Min. Anal.* **2020**, *3*, 280–291. [[CrossRef](#)]
25. Lei, X.; Fang, Z. GBDTCDA: Predicting circRNA-disease Associations Based on Gradient Boosting Decision Tree with Multiple Biological Data Fusion. *Int. J. Biol. Sci.* **2019**, *15*, 2911–2924. [[CrossRef](#)]
26. Ding, Y.; Chen, B.; Lei, X.; Liao, B.; Wu, F.-X. Predicting novel CircRNA-disease associations based on random walk and logistic regression model. *Comput. Biol. Chem.* **2020**, *87*, 107287. [[CrossRef](#)] [[PubMed](#)]

27. Wang, L.; You, Z.-H.; Huang, Y.-A.; Huang, D.-S.; Chan, K.C.C. An efficient approach based on multi-sources information to predict circRNA-disease associations using deep convolutional neural network. *Bioinformatics* **2019**, *36*, 4038–4046. [[CrossRef](#)]
28. Wang, L.; You, Z.-H.; Li, Y.-M.; Zheng, K.; Huang, Y.-A. GCNCDA: A new method for predicting circRNA-disease associations based on Graph Convolutional Network Algorithm. *PLoS Comput. Biol.* **2020**, *16*, e1007568. [[CrossRef](#)]
29. Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [[CrossRef](#)] [[PubMed](#)]
30. Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **2019**, *48*, D845–D855. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, Z.; Yue, L.; Wang, Y.; Jiang, Y.; Xiang, L.; Cheng, Y.; Ju, D.; Chen, Y. A circRNA-miRNA-mRNA network plays a role in the protective effect of diosgenin on alveolar bone loss in ovariectomized rats. *BMC Complement. Med. Ther.* **2020**, *20*, 220. [[CrossRef](#)]
32. Su, Q.; Lv, X. Revealing new landscape of cardiovascular disease through circular RNA-miRNA-mRNA axis. *Genomics* **2020**, *112*, 1680–1685. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, Y.K.; Shen, Z.A.; Yu, H.; Luo, T.; Gao, Y.; Du, P.F. Predicting lncRNA-Protein Interactions with miRNAs as Mediators in a Heterogeneous Network Model. *Front. Genet.* **2019**, *10*, 1341. [[CrossRef](#)] [[PubMed](#)]
34. Zhou, X.Z.; Menche, J.R.; Barabási, A.-L.; Sharma, A. Human symptoms-disease network. *Nat. Commun.* **2014**, *5*, 4212. [[CrossRef](#)]
35. Velikovi, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2017**, arXiv:1710.10903.
36. Yan, C.; Wang, J.; Wu, F.-X. DWN-RLS: Regularized least squares method for predicting circRNA-disease associations. *BMC Bioinform.* **2018**, *19*, 520. [[CrossRef](#)]
37. Lei, X.; Tie, J. Prediction of disease-related metabolites using bi-random walks. *PLoS ONE* **2019**, *14*, e0225380. [[CrossRef](#)]
38. Chen, H.; Perozzi, B.; Al-Rfou, R.; Skiena, S. A Tutorial on Network Embeddings. *arXiv* **2018**, arXiv:1808.02590.
39. Hindy, J.-R.; Souaid, T.; Kourie, H.R.; Kattan, J. Targeted therapies in urothelial bladder cancer: A disappointing past preceding a bright future? *Future Oncol.* **2019**, *15*, 1505–1524. [[CrossRef](#)]
40. Kwan, M.L.; Garren, B.; Nielsen, M.E.; Tang, L. Lifestyle and nutritional modifiable factors in the prevention and treatment of bladder cancer. *Urol. Oncol. Semin. Orig. Investig.* **2019**, *37*, 380–386. [[CrossRef](#)]
41. Silpa-Archa, S.; Ruamviboonsuk, P. Diabetic Retinopathy: Current Treatment and Thailand Perspective. *J. Med. Assoc. Thai. Chotmaihet Thangphaet* **2017**, *100* (Suppl. S1), S136–S147.
42. Smolen, J.S.; Aletaha, D.; McInnes, I.B. Rheumatoid arthritis. *Lancet* **2016**, *388*, 2023–2038. [[CrossRef](#)]
43. Li, B.; Li, N.; Zhang, L.; Li, K.; Xie, Y.; Xue, M.; Zheng, Z. Hsa_circ_0001859 Regulates ATF2 Expression by Functioning as an MiR-204/211 Sponge in Human Rheumatoid Arthritis. *J. Immunol. Res.* **2018**, *2018*, 9412387. [[CrossRef](#)]
44. Zhang, S.-J.; Chen, X.; Li, C.-P.; Li, X.-M.; Liu, C.; Liu, B.-H.; Shan, K.; Jiang, Q.; Zhao, C.; Yan, B. Identification and Characterization of Circular RNAs as a New Class of Putative Biomarkers in Diabetes Retinopathy. *Investig. Ophthalmol. Vis. Sci.* **2017**, *58*, 6500–6509. [[CrossRef](#)]
45. Zhang, L.; Xia, H.B.; Zhao, C.Y.; Shi, L.; Ren, X.L. Cyclic RNA hsa_circ_0091017 inhibits proliferation, migration and invasiveness of bladder cancer cells by binding to microRNA-589-5p. *Eur. Rev. Med. Pharmacol. Sci.* **2020**, *24*, 86–96. [[CrossRef](#)]
46. Cai, D.; Liu, Z.; Kong, G. Molecular and Bioinformatics Analyses Identify 7 Circular RNAs Involved in Regulation of Oncogenic Transformation and Cell Proliferation in Human Bladder Cancer. *Med. Sci. Monit.* **2018**, *24*, 1654–1661. [[CrossRef](#)] [[PubMed](#)]
47. Yang, C.; Yuan, W.; Yang, X.; Li, P.; Wang, J.; Han, J.; Tao, J.; Li, P.; Yang, H.; Lv, Q.; et al. Circular RNA circ-ITCH inhibits bladder cancer progression by sponging miR-17/miR-224 and regulating p21, PTEN expression. *Mol. Cancer* **2018**, *17*, 19. [[CrossRef](#)] [[PubMed](#)]
48. Zhong, Z.; Lv, M.; Chen, J. Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci. Rep.* **2016**, *6*, 30919. [[CrossRef](#)]
49. Zhuang, C.; Huang, X.; Yu, J.; Gui, Y. Circular RNA hsa_circ_0075828 Promotes Bladder Cancer Cell Proliferation through Activation of CREB1. *BMB Rep.* **2020**, *53*, 82–87. [[CrossRef](#)]
50. Zhong, Z.; Huang, M.; Lv, M.; He, Y.; Duan, C.; Zhang, L.; Chen, J. Circular RNA MYLK as a competing endogenous RNA promotes bladder cancer progression through modulating VEGFA/VEGFR2 signaling pathway. *Cancer Lett.* **2017**, *403*, 305–317. [[CrossRef](#)] [[PubMed](#)]
51. Liu, C.; Yao, M.-D.; Li, C.-P.; Shan, K.; Yang, H.; Wang, J.-J.; Liu, B.; Li, X.-M.; Yao, J.; Jiang, Q.; et al. Silencing of Circular RNA-ZNF609 Ameliorates Vascular Endothelial Dysfunction. *Theranostics* **2017**, *7*, 2863–2877. [[CrossRef](#)] [[PubMed](#)]
52. Zheng, F.; Yu, X.; Huang, J.; Dai, Y. Circular RNA expression profiles of peripheral blood mononuclear cells in rheumatoid arthritis patients, based on microarray chip technology. *Mol. Med. Rep.* **2017**, *16*, 8029–8036. [[CrossRef](#)] [[PubMed](#)]
53. Zhong, S.; Ouyang, Q.; Zhu, D.; Huang, Q.; Zhao, J.; Fan, M.; Cai, Y.; Yang, M. Hsa_circ_0088036 promotes the proliferation and migration of fibroblast-like synoviocytes by sponging miR-140-3p and upregulating SIRT 1 expression in rheumatoid arthritis. *Mol. Immunol.* **2020**, *125*, 131–139. [[CrossRef](#)] [[PubMed](#)]