

# A graph based deep learning technology application in degenerative polyarthritis associated genes prediction

Zhenggen Qu<sup>1,2</sup>, Danying Niu<sup>3</sup>

1 College of Mathematics and Computer Application, Shangluo University, Shangluo 726000, China

2 Engineering Research Center of Qinling Health Welfare Big Data, Universities of Shaanxi Province, Shangluo 726000, China

3 Shangluo Central Hospital trauma of edics, Shangluo 726000, China

## Abstract

Degenerative polyarthritis is the most common joint disease and affects millions of people worldwide. However, there is currently no cure for degenerative polyarthritis and no effective methods to prevent or slow down its progression. Gene regulatory relationships are vital for understanding disease mechanisms and developing treatment and novel drugs. Gene regulatory networks can be obtained from the RNA sequencing. Although various single-cell and bulk RNA sequencing data are available, an effective method to integrate the data for molecular diagnosis and treatment of degenerative polyarthritis has not yet been carried out. Here, we propose a novel deep learning-based method to efficiently capture the gene regulatory features of degenerative polyarthritis. First, we integrate single-cell RNA sequencing data-based gene regulatory network to model the gene regulatory relationships between genes and transcription factors as node feature aggregation. Second, we propose a graph convolutional model named dpTF-GCN on gene regulatory graph to transmit and update the node feature for potential associated genes predicting. According to the results, dpTF-GCN achieved the best performance among represented network-based methods. Furthermore, case studies suggest that dpTF-GCN can identify potential associated genes accurately. Our research not only provides theoretical and methodological support for the study of degenerative polyarthritis, but also provides a research case for the application of graph neural network-based identification of associated genes in other diseases.

 OPEN ACCESS

**Published:** 23/06/2023

**Accepted:** 16/06/2023

**DOI:**  
10.23967/j.rimni.2023.06.004

## 1 Introduction

Degenerative polyarthritis is a non-inflammatory disease that affects the joints by causing damage to the cartilage and tissues surrounding them [1]. It is also known as osteoarthritis. It is the most common joint disease and affects millions of people worldwide [2]. The risk of developing degenerative polyarthritis increases with age and with occupations or activities that put high stress on the joint, such as heavy labor or sports [3]. Therefore, the prevalence and burden of this disorder is expected to rise rapidly in the aging population and in the modern society. Degenerative polyarthritis has no cure and no effective methods to prevent or slow down its progression. The current treatments mainly aim to relieve pain and improve function, but they have limited efficacy and side effects [2]. Knee arthroplasty, a surgical procedure to replace the damaged joint with an artificial one, can provide some relief of symptoms, but it cannot restore the normal function and activity level of the patient [4]. Degenerative polyarthritis is a complex disease that involves various pathophysiological processes that affect the whole joint structure and function [5]. To understand the molecular mechanisms underlying this disease and to identify new biomarkers and therapeutic targets, it is necessary to use deep learning-based system biology approaches that can integrate and analyze large-scale data from different sources [6].

In recent years, deep learning methods have been applied in various fields of medicine and bioinformatics, such as prediction for associations between miRNA and disease [7-9], associations between gene and disease and associations between

metabolite and disease [10-12]. These methods have promoted the development of computational models for identifying complex disease uncovered associated genes. Deep learning techniques are based on artificial neural networks (ANN), also known as representation learning techniques, which can identify hidden patterns of the data without requiring an explicit feature extraction [13]. In other words, deep learning-based architectures make automatic feature extraction possible. For disease association genes prediction, different deep learning architectures have been successfully applied, such as deep-neural network (DNN) has been adopted for aging-related diseases associated genes prediction [14], deep-autoencoder (DA) is used for detection of Parkinson's disease association genes [15] and convolutional-neural network (CNN) is also adopted for predicting lung tumor [16]. However, the molecular mechanism of disease is complex and cannot be ignored. Graph provides a natural framework for disease mechanism prediction, which are widely used to capture interactions between individual elements represented as nodes in graph. In disease associated genes prediction, specifically, nodes can represent genes or other functional regulatory elements, while the graph edges incorporate associations between genes or genes and regulatory elements in an intuitive manner. The graph neural network (GNN) is a deep learning model that focuses on graph data and has been used to perform many bioinformatics tasks [17-19]. Graph convolutional network (GCN) is a representative GNN models, aims to learn node embeddings by implementing the convolution operation on a graph based on the attributes of neighborhood nodes [11]. GCN has achieved satisfactory results in the construction of many disease associated genes prediction model, such as Alzheimer's

disease [20], Parkinson's disease [21], and some rare diseases [22]. As far as we know, the GCN-based deep learning model has not yet been developed for the prediction of degenerative polyarthritis associated genes.

Bulk expression data and single-cell transcriptomic data are two common types of gene expression profiles. Bulk expression data have large sample sizes but they mix the expression levels of all cells in a sample or tissue, hiding important differences and signals between cells. Single-cell transcriptomic data can reveal more information about each cell's gene expression changes during differentiation, cell type features and other biological aspects [23-25]. Moreover, single-cell transcriptomic data can help to build gene regulatory network (GRN) at the single-cell level and understand how expression regulation differs across cell types [26]. For instance, many studies recently found that transcription factors (TFs) could control different genes in different subtypes of human pancreatic islets and that GRN dynamics was a key factor for pancreatic cell expression diversity [27]. Also, GRN reprogramming could affect melanoma progression and resistance to therapy [28] and single-cell GRN analysis revealed increased expression of cell-type-specific TFs in the bronchoalveolar immune cells of COVID-19 patients, indicating a highly inflammatory macrophage environment in the lungs of server COVID-19 patients [29]. Thus, GRN and the expression patterns of TFs and their target genes are crucial to fully grasp the mechanisms of disease pathogenesis.

In this study, we propose a model based on the graph convolutional neural network, named dpTF-GCN, to predict potential genes that are associated with degenerative polyarthritis, a common form of arthritis that affects the joints. The dpTF-GCN makes full use of topological information of heterogeneous networks that consist of TFs and target genes, as well as the data of similarities among TFs and target genes. We constructed a heterogeneous network composed of nodes representing genes and TFs. The nodes were connected by edges based on their regulatory relationships and regulatory similarity. We adopted a GCN-based approach to learn the feature representations of genes and TFs from the network structure and node attributes. We also designed an end-to-end framework to automatically optimize model parameters using gradient descent. We evaluated the performance of dpTF-GCN using a ten-fold cross-validation and compared it with other state-of-the-art prediction methods for degenerative polyarthritis associated genes. The results showed that dpTF-GCN achieved significantly higher accuracy and recall than other methods. Furthermore, we conducted case studies to demonstrate that dpTF-GCN can successfully infer potential disease-associated candidate genes that are supported by existing literature or biological databases.

## 2 Materials and Methods

### 2.1 Data collection

As degenerative polyarthritis is a common disease that affects the joints and bones of human, we obtained gene regulatory network (GRN) constructed by single-cell sequencing data of human bone marrow from GRNdb [30] to investigate the molecular mechanisms of this disease. This database is a valuable resource that provides gene regulatory network datasets for various human tissues and cell types. In these datasets, the nodes represent transcription factors (TFs) and target genes, and the edges represent the regulatory relationships between them. To identify the genes that are associated with degenerative polyarthritis, we used DisGeNET as a reference database [31]. This database contains a large collection of genes that have been linked to various human diseases based on different types of evidence. These genes have

been widely used in different studies to explore the genetic basis of diseases [32-34].

### 2.2 Network construction

Here, denoting that  $G = \{G_1, G_2, G_3...G_g\}$  is a set of  $g$  genes,  $T = \{T_1, T_2, T_3...T_t\}$  is a set of  $t$  TFs. To clarity, we denote the gene regulatory network as a matrix  $M_{GT} \in R^{g \times t}$ , the gene-gene similarity network is a matrix  $M_{GG} \in R^{g \times g}$  and the TF-TF similarity network is a matrix  $M_{TT} \in R^{t \times t}$ . Mathematically, the specific content and calculation process of these matrices are as follows:

$$M_{GT} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,t} \\ \vdots & \ddots & \vdots \\ w_{g,1} & \cdots & w_{g,t} \end{bmatrix} \#(1)$$

$$M_{GG} = \begin{bmatrix} s_{1,1} & \cdots & s_{1,t} \\ \vdots & \ddots & \vdots \\ s_{g,1} & \cdots & s_{g,t} \end{bmatrix} \#(2)$$

$$M_{TT} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,t} \\ \vdots & \ddots & \vdots \\ x_{g,1} & \cdots & x_{g,t} \end{bmatrix} \#(3)$$

where  $w_{i,j}$  ( $i \in g, j \in t$ ) is the weight of the links directed from TFs to target genes, the higher weights correspond to more likely regulatory links,  $s_{i,j}$  ( $i \in g, j \in g$ ) and  $x_{i,j}$  ( $i \in t, j \in t$ ) is the similarity between gene pairs and TF pairs, respectively.

Since the GRNdb only provides gene regulatory network, we use the Jaccard index as  $s_{i,j}$  and  $x_{i,j}$  for network constructing [35,36]. The  $s_{i,j}$  and  $x_{i,j}$  can be calculated by:

$$s_{i,j} = \frac{T_i^g \cap T_j^g}{T_i^g \cup T_j^g} \#(4)$$

$$x_{i,j} = \frac{G_i^t \cap G_j^t}{G_i^t \cup G_j^t} \#(5)$$

where  $T_i^g$  is the set of TFs linked to  $G_i$  ( $i \in g$ ),  $T_j^g$  is the set of TFs linked to  $G_j$  ( $j \in g$ );  $G_i^t$  is the set of target genes linked to  $T_i$  ( $i \in t$ ),  $G_j^t$  is the set of target genes linked to  $T_j$  ( $j \in t$ ).

### 2.3 Model architecture

To fully capture the imperceptible information of gene regulation relationships, we adopted graph structure data and GCN for feature representation. GCN is a multilayer connected neural network architecture for information aggregation and learning low-dimensional representations of nodes [37], which can be an effective method for extracting useful information from intricate gene regulatory networks. Specifically, the input of GCN is a graph  $G_{G-T} = (v, e)$  with  $v = (G, T)$  representing  $G$  gene nodes and  $T$  TF nodes, and  $e$  is a set of edges between each node. As some nodes  $v$  in  $G_{G-T}$  are known to be associated to degenerative polyarthritis, the aims of model to classify whether one node is associated to degenerative polyarthritis.

Here, denoting the  $G_{G-T}$  is represented by an adjacency matrix  $A_{G-T} \in R^{(G+T) \times (G+T)}$ . The weight of edges as the feature of each node and the node feature matrix can be represented as  $B_{G-T} \in R^{(G+T) \times (G+T)}$ . The graph convolution is defined on graph as the product of the input signal after the filter  $g_\theta$  in the Fourier domain. According to the definition of GCN original model architecture, the symmetric normalized Laplacian matrix

of  $A_{G-T}$  is  $L_{G-T} = U_{G-T} \Lambda_{G-T} U_{G-T}^t$ , where  $\Lambda_{G-T} = \text{diag}(\mu_1, \mu_2, \mu_3, \dots, \mu_{G+T})$  is the diagonal matrix of eigenvalues. As the feature matrix  $B_{G-T}$  needs to undergo the Fourier transform, this can be represented as  $U_{G-T}^t B_{G-T}$ . The Chebyshev polynomials  $T_K(x) = 2xT_{K-1}(x) - T_{K-2}(x)$  was used to reduce the computational complexity<sup>38</sup>. The filter  $g_\theta$  can be mathematically represented as:

$$g_\theta(\Lambda_{G-T}) = \sum_{K=0}^K \theta_K T_K(\tilde{\Lambda}_{G-T}) \# (6)$$

$$g_\theta * B_{G-T} = \sum_{K=0}^K \theta_K T_K(\tilde{L}_{G-T}) B_{G-T} \# (7)$$

where  $\theta \in R^K$  is a vector of Chebyshev coefficients,  $\tilde{\Lambda}_{G-T} = \frac{2\Lambda_{G-T}}{\mu_{max}} - I_N$ ,  $\tilde{L}_{G-T} = \frac{2L_{G-T}}{\mu_{max}} - I_N$ ,  $I_N$  is the identity matrix and  $K$  is the  $K^{\text{th}}$ -order neighborhood.

The formulation can simplify by limiting  $K=1$  as the Chebyshev polynomials is recursively [38,39]. After activation functions were introduced in each layer ( $l > 0$ ), the graph convolution operation can be represented as follows:

$$g_\theta * B_{G-T} = \text{ReLU}\left(\theta \left(D_{G-T}^{-\frac{1}{2}} (I_N + A_{G-T}) D_{G-T}^{-\frac{1}{2}}\right)\right) \# (8)$$

$$\begin{bmatrix} H_G \\ H_T \end{bmatrix} = g_\theta * B_{G-T} \# (9)$$

where  $D_{G-T}$  is the diagonal matrix with diagonal entry  $[D_{G-T}]_{i,j} = \sum_j [A_{G-T}]_{i,j}$ ,  $H_G$  is the embedding of genes and  $H_T$  is the embedding of TFs.

To make the model as an end-to-end binary classifier, we use the embedding vectors from GCN as the input of multi-layer perception (MLP). The category scores were computed by the sigmoid function that follows the output of the last hidden layer, as follows:

$$S = \text{Sigmoid}\left(W_{out} \cdot \begin{bmatrix} H_G \\ H_T \end{bmatrix} + b_{out}\right) \# (10)$$

where  $S$  is the scores of a gene associated to degenerative polyarthritis,  $W_{out}$  and  $b_{out}$  are the weight matrix and bias vector.

The model optimizer is cross-entropy loss  $L$ :

$$L = \sum_{i,j \in Y \cup Y^-} (y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log (1 - \hat{y}_{ij})) \# (11)$$

where  $y_{ij}$  represents the true label of the nodes, which will be 1 or 0,  $Y$  and  $Y^-$  denote the set of all nodes contained in the positive nodes set and negative nodes set, respectively. The model was trained by back propagation algorithm in an end-to-end architecture (Figure 1).

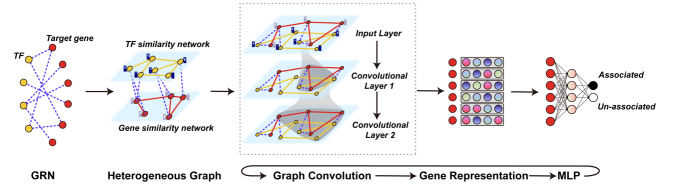


Figure 1: The workflow of dpTF-GCN

## 2.4 Experimental setting and hyperparameters

We used ten-fold cross-validation (10-CV) to test the performance of dpTF-GCN. Because of the limited number of true labeled genes, we randomly selected 10 genes as the validation dataset for case study and excluded them from the training and testing process. The remaining known associated genes were randomly divided into ten equal-sized subsets. We repeated the cross-validation process ten times and used each subset as the testing dataset once while using the other nine subsets as the training dataset. We chose AUC and AUPR as the main evaluation metrics because they can measure the performance of model without a specific threshold. We also computed some threshold-based metrics such as precision (PRE), recall (REC), accuracy (ACC) and F1-score (F1).

The hyperparameters in dpTF-GCN are the number of layers  $y \in \{2, 3\}$ , learning rate of optimizer  $\gamma \in \{0.00001, 0.00005, 0.0001, 0.0002, 0.0005\}$  and the total training epochs  $\alpha \in \{1000, 2000, 4000, 6000, 8000\}$ . By adjusting the parameters empirically, we set  $y = 3$ ,  $\gamma = 0.0001$  and  $\alpha = 4000$  for dpTF-GCN in the following experiments.

## 3 Results

### 3.1 Graph construction and model overall performance

We obtained a gene regulatory network (GRN) from GRNdb database [30], which contains information about how transcription factors (TFs) regulate the expression of target genes. The GRN was derived from human single-cell RNA sequencing data of adult bone marrow cells. This data set consisted of 1,834 cells that were analyzed by the SCENIC pipeline [26] to infer the regulatory interactions between 107 TFs and 4,009 target genes. The resulting GRN had 17,318 TF-target pairs. We searched the DisGeNET database [31] for genes that are associated with degenerative polyarthritis, a common joint disease. We found 498 genes in our GRN that were marked as degenerative polyarthritis related genes and had supporting evidence from biological research. We then constructed a gene-gene similarity network based on the number of shared TFs between each pair of genes. We used the Jaccard Index (JI) to measure the similarity score of each gene pair. Similarly, a TF-TF similarity network was also built based on the number of shared target genes between each pair of TFs. We also used JI to quantify the similarity score of each TF pair.

To better identify the potential associated genes of degenerative polyarthritis, we deployed a graph convolutional neural network (GCN) algorithm to construct an end-to-end framework for predicting. Our framework, which we named dpTF-GCN, integrates the gene regulatory network and the gene-gene similarity network to capture the complex interactions between transcription factors and target genes. By using graph theory and GCN, our framework can extract the features of gene regulation that are relevant to degenerative polyarthritis. We evaluated the performance of dpTF-GCN on six metrics: area under the receiver operating characteristic curve

(AUC), area under the precision-recall curve (AUPRC), precision (PRE), recall (REC), accuracy (ACC) and F1-score (F1). All these metrics were higher than 0.8 (Figure 2), indicating that dpTF-GCN can effectively identify potential degenerative polyarthritis associated genes. Moreover, we observed that AUC and AUPR, which are two metrics that do not depend on a specific threshold for classification, were also high (Figure 3). This suggests that dpTF-GCN has a good balance between sensitivity and specificity in predicting degenerative polyarthritis associated genes.

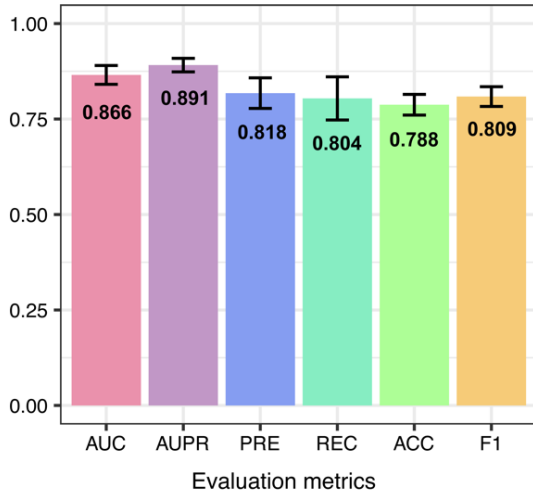


Figure 2: The six-evaluation metrics of model.

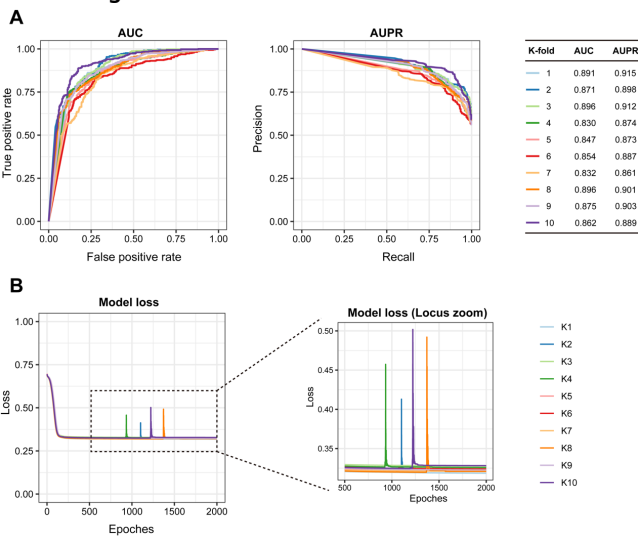


Figure 2: Performance during dpTF-GCN training. (A) The AUC and AUPR curves generated by dpTF-GCN on each fold dataset; (B) The training loss curves of each fold dataset.

### 3.2 Comparing dpTF-GCN with baseline methods

To demonstrate the superiority of dpTF-GCN especially for degenerative polyarthritis associated genes prediction, we compared it with other baseline methods by 10-fold cross-validation on gene regulatory network data. Three representative disease associated genes prediction methods are used for comparing with dpTF-GCN, including the PMFMDA which employ matrix factorization strategy [40], the MeSHHeading2vec which conduct feature extraction via graph embedding algorithms [41], and the BiRW designed by random walk strategy [42]. Each method was conducted prediction tasks on the GRN network dataset with the default optimal

parameters, and the AUC and AUPR were calculated for evaluation and comparison. According to the results (Figure 4), dpTF-GCN outperforms all comparison methods in terms of both evaluation metrics. Compared with the matrix factorization, graph embedding and random walk-based methods (PMFMDA, MeSHHeading2vec and BiRW), the GCN-based method dpTF-GCN achieves 13.3% and 26.8% improvement on average over them in terms of AUC and AUPR, respectively. The graph convolutional network-based strategy performs better than other feature extraction strategy, suggesting that GCN may lead to better aggregation of graph topological information.

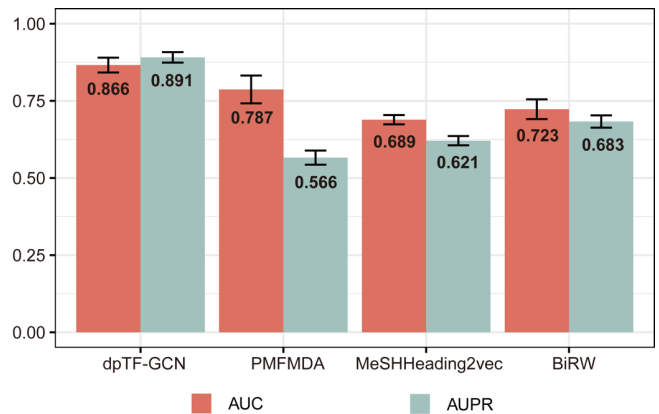


Figure 4: Comparisons with other different methods on same dataset

### 3.3 Ablation experiment of graph components

To test the effectiveness of different components in the heterogeneous graph used in dpTF-GCN, we conduct an ablation study. This involved performing leave-one-out validation on each part of the graph to determine whether all components were necessary for effective feature extraction and predicting. The heterogeneous graph used in dpTF-GCN consists of three components: a transcription factor and target-gene regulatory network (G-T), a gene similarity network (G-G), and a transcription factor similarity network (T-T). Since the goal of dpTF-GCN is to predict potential degenerative polyarthritis associated genes by aggregating gene regulatory information, the G-T network is considered an essential part of the graph for information aggregation and convolution.

Here, we denote three types of experiments: MASK-GG, MASK-TT and MASK-GGTT. The MASK-GG experiment represents graph used in our model without G-G network. The MASK-TT experiment denotes an experiment without T-T network, and MASK-GGTT represents our model only conduct on a graph of G-T network. The results of these experiments showed that removing either the G-G or T-T networks resulted in a decrease in both AUC and AUPR values by more than 15% (Table 1). This suggests that all three components are important for effective and accurate predicting.

Table 1: The ablation experiment of dpTF-GCN on the heterogeneous graph

Evaluation metrics	MASK-GG	MASK-TT	MASK-GGTT
AUC	0.697±0.021	0.712±0.032	0.611±0.019
AUPR	0.711±0.030	0.739±0.026	0.663±0.022
PRE	0.578±0.041	0.581±0.024	0.489±0.033
REC	0.523±0.022	0.492±0.031	0.552±0.017



ACC	0.512±0.030	0.522±0.014	0.537±0.021
F1	0.549±0.027	0.530±0.025	0.519±0.023

Notably, when both G-G and G-T networks are removed, the AUC of dpTF-GCN drops sharply to 55% and the AUPR is only 60%. The results demonstrate that topology graph modeling is necessary for fully feature extraction, and the combination can improve the prediction performance significantly.

### 3.4 Case studies

In this section, we use dpTF-GCN to make predictions on the validation dataset. This dataset contains 10 labeled genes that have been supported by literature and biological databases but were not used in the training or testing steps. This means that the validation dataset has never been seen by dpTF-GCN before and can effectively demonstrate the model's generalization. The results of case study show that dpTF-GCN was able to successfully classify all the genes in the validation dataset with high probability scores (Table 2). All the ten genes for validation have been supported by existing biological research. For example, the gene *P2RX7* had the highest score of 0.911 and had six pieces of literature evidence supporting its association with degenerative polyarthritis. Overall, these results suggest that dpTF-GCN is an effective and general tool for predicting new associated genes of degenerative polyarthritis.

**Table 2:** The prediction score and literature evidence of ten genes in validation dataset

Gene	dpTF-GCN score	Literature evidence
P2RX7	0.911	PMID:22447075, PMID:30317598, PMID:24934217, PMID:29511609, PMID:28343378, PMID:29845461
TBX5	0.910	PMID:31376087, PMID:25320281
IL33	0.903	PMID:29867945, PMID:26520876, PMID:21441054, PMID:29095435
RNR2	0.898	PMID:15567815
SAA2	0.877	PMID:25849372
CPB1	0.871	PMID:24449579, PMID:21804193
MGP	0.866	PMID: 21724703, PMID: 31215457, PMID: 28855172
TNNC1	0.855	PMID:21762512, PMID:28722504
TKTL1	0.843	PMID:28719557, PMID:27996342, PMID:29143404, PMID:30671597
PRPF3	0.829	PMID:31268737

### 4 Discussions

Network-based methods are widely used for predicting and analyzing the associations between biological entities, such as genes and diseases. In this study, we propose a novel GCN-based computational approach, called dpTF-GCN, to predict the genes that are associated with degenerative polyarthritis. Our method dpTF-GCN can automatically learn the low-dimensional representations of genes and transcription factors by systematically integrating the complex topology of the heterogeneous network that consists of gene regulatory relationships, the neighborhood information of genes and transcription factors, and the gene- and transcription factor-specific attributes. The learned embeddings and the associated genes classification models are jointly optimized in an end-to-end fashion. We conduct extensive experiments to evaluate the performance of our method on recovering missing associated genes from the training data, and on discovering novel potential associated genes for degenerative polyarthritis that are not present in the training data. Satisfactory results confirm

the excellent performance of dpTF-GCN.

In dpTF-GCN, we use gene-gene and TF-TF similarity networks to capture the functional relationships between genes and transcription factors. These networks are constructed by computing the Jaccard index (JI) based on the overlap of transcription factors or target-genes for each pair of genes or transcription factors. This method reflects the similarity of their regulatory roles rather than their structural or sequence features [43,44]. Moreover, our gene regulatory network is derived from single-cell RNA sequencing data, which can provide more insights into the dynamic changes of gene expression during differentiation, the characteristics of different cell types and other biological aspects. By integrating these networks into a heterogeneous graph, we can incorporate diverse and rich gene regulatory information into dpTF-GCN. This may be one of the reasons why dpTF-GCN achieves excellent performance in predicting cell type-specific transcription factors.

Degenerative polyarthritis is a condition that causes inflammation and pain in multiple joints and there is a need for more effective and personalized treatments for it. Our research aims to address this need by providing a novel approach for identifying potential associated genes for degenerative polyarthritis. We use a sophisticated GCN model that can capture both local and global features of the gene regulatory information. The dpTF-GCN not only offers as a new way to treat degenerative polyarthritis at the molecular level but also provides a valuable resource and method for screening drug targets. Furthermore, our research also has implications for the integration of artificial intelligence and precision medicine.

**Acknowledgement:** The authors thank to lab members for assistance.

**Funding Statement:** This study was supported by the National Social Science Fund West Project of China (21XJY015, to Zhenggeng Qu) and Shaanxi Provincial Philosophy and Social Science Research Project (2023QN0233, to Zhenggeng Qu).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

### References

- [1] Buch M.H., Eyre S. and McGonagle D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. *Nat. Rev. Rheumatol.*, 17:17–33, 2021.
- [2] Buckwalter J.A., Saltzman C. and Brown T. The impact of osteoarthritis: Implications for research. *Clin. Orthop. Relat. Res.*, 427:S6–S15, 2004.
- [3] Bijlsma J.W.J., Berenbaum F. and Lafeber F.P.J.G. Osteoarthritis: An update with relevance for clinical practice. *Lancet*, 377:2115–2126, 2011.
- [4] Cole B.J. and Harner C.D. Degenerative arthritis of the knee in active patients: Evaluation and management. *J. Am. Acad. Orthop. Surg.*, 7:389–402, 1999.
- [5] Sandell L.J. and Aigner T. Articular cartilage and changes in arthritis: Cell biology of osteoarthritis. *Arthritis Res. Ther.*, 3:107, 2001.
- [6] Wieland H.A., Michaelis M., Kirschbaum B.J. and Rudolph K.A. Osteoarthritis – an untreatable disease?. *Nat. Rev. Drug Discov.*, 4:331–344, 2005.
- [7] Chen X., Sun L.G. and Zhao Y. Ncmcmda: Mirna-disease association prediction through neighborhood constraint matrix

completion. *Brief. Bioinformatics*, 22:485-496, 2020.

[8] Chen X., Li T.H., Zhao Y., Wang C.C. and Zhu C.C. Deep-belief network for predicting potential mirna-disease associations. *Brief. Bioinformatics*, 22:bbaa186, 2021.

[9] Chen X., Zhu C.C. and Yin J. Ensemble of decision tree reveals potential mirna-disease associations. *PLoS Comput. Biol.*, 15:e1007209, 2019.

[10] Azman B., Hussain S., Azmi N., Ghani M. and Norlen N. Prediction of distant recurrence in breast cancer using a deep neural network. *Rev. Int. Metodos Numer. para Calc. Diseno Ing.*, 38:12, 2022.

[11] Ata S.K., Wu M., Fang Y., Ou-Yang L., Kwok C.K. et al. Recent advances in network-based methods for disease gene prediction. *Brief. Bioinformatics*, 22:bbaa303, 2021.

[12] Sun F., Sun J. and Zhao Q. A deep learning method for predicting metabolite-disease associations via graph neural network. *Brief. Bioinformatics*, 23:bbac266, 2022.

[13] Gautam R. and Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. *J. Med. Syst.*, 44:49, 2020.

[14] Ye J., Wang S., Yang X. and Tang X. Gene prediction of aging-related diseases based on DNN and mashup. *BMC Bioinform.*, 22:1-16, 2021.

[15] Peng J., Guan J. and Shang X. Predicting Parkinson's disease genes based on node2vec and autoencoder. *Front Genet*, 10, 2019.

[16] Alameen A. Smart lung tumor prediction using dual graph convolutional neural network. *Intell. Autom. Soft Comput.*, 36:369-383, 2023.

[17] Sun M., Zhao S., Gilvary C., Elemento O., Zhou J. et al. Graph convolutional networks for computational drug development and discovery. *Brief. Bioinformatics*, 21:919-935, 2019.

[18] Cai R., Chen X., Fang Y., Wu M. and Hao Y. Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. *Bioinformatics*, 36:4458-4465, 2019.

[19] Long Y., Wu M., Kwok C.K., Luo J. and Li X., Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics*, 36:4918-4927, 2020.

[20] Parisot S., Ktena S.I., Ferrante E., Lee M., Guerrero R., Glocker B. et al. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease. *Med. Image Anal.*, 48:117-130, 2018.

[21] Zhang X., He L., Chen K., Luo Y., Zhou J. et al. Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease. *AMIA Annu. Symp. Proc.*, 2018:1147, 2018.

[22] Rao A., Vg S., Joseph T., Kotte S., Sivasadan N. et al. Phenotype-driven gene prioritization for rare diseases using graph convolution on heterogeneous networks. *BMC Medical Genom.*, 11:57, 2018.

[23] Iacono G., Massoni-Badosa R. and Heyn H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.*, 20:110, 2019.

[24] Fiers M.W.E.J., Minnoye L., Aibar S., González-Blas C.B., Atak Z.K. et al. Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics*, 17:246-254, 2018.

[25] Zhao M., He W., Tang J., Zou Q. and Guo F. A hybrid deep learning framework for gene regulatory network inference from single-cell transcriptomic data. *Brief. Bioinformatics*, 23:bbab568, 2022.

[26] Aibar S., González-Blas C.B., Moerman T., Huynh-Thu V.A., Imrichova H. et al. Scenic: Single-cell regulatory network inference and clustering. *Nat. Methods*, 14:1083-1086, 2017.

[27] Li Y., Chen J., Xu Q., Han Z., Tan F. et al. Single-cell transcriptomic analysis reveals dynamic alternative splicing and gene regulatory networks among pancreatic islets. *Science China Life Sciences*, 64:174-176, 2021.

[28] Rambow F., Rogiers A., Marin-Bejar O., Aibar S., Femel J. et al. Toward minimal residual disease-directed therapy in melanoma. *Cell*, 174:843-855.e819, 2018.

[29] Liao M., Liu Y., Yuan J., Wen Y., Xu G. et al. Single-cell landscape of bronchoalveolar immune cells in patients with covid-19. *Nat. Med.*, 26:842-844, 2020.

[30] Fang L., Li Y., Ma L., Xu Q., Tan F. et al. Grndb: Decoding the gene regulatory networks in diverse human and mouse conditions. *Nucleic Acids Res.*, 49:D97-D103, 2020.

[31] Piñero J., Saüch J., Sanz F. and Furlong L.I. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput. and Struct. Biotechnol. J.*, 19:2960-2967, 2021.

[32] Piñero J., Ramírez-Anguita J.M., Saüch J., Ronzano F., Centeno E. et al. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.*, 48:D845-D855, 2019.

[33] Joshi A., Rienks M., Theofilatos K. and Mayr M. Systems biology in cardiovascular disease: A multiomics approach. *Nat. Rev. Cardiol.*, 18:313-330, 2021.

[34] Gaudet T., Day B., Jambas A.R., Soman J., Regep C. et al. Utilizing graph machine learning within drug discovery and development. *Brief. Bioinformatics*, 22:bbab159, 2021.

[35] Yu Z., Huang F., Zhao X., Xiao W. and Zhang W. Predicting drug-disease associations through layer attention graph convolutional network. *Brief. Bioinformatics*, 22:bbaa243, 2021.

[36] Deng Y., Xu X., Qiu Y., Xia J., Zhang W. et al. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*, 36:4316-4322, 2020.

[37] Kipf T.N. and Welling M. Semi-supervised classification with graph convolutional networks. *arXiv*, 1609.02907, 2016.

[38] Defferrard M., Bresson X. and Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process Syst.*, 29, 2016.

[39] Hammond D.K., Vandergheynst P. and Gribonval R. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmon. Anal.*, 30:129-150, 2011.

[40] Xu J., Cai L., Liao B., Zhu W., Wang P. et al. Identifying potential mirnas-disease associations with probability matrix factorization. *Front. Genet.*, 10, 2019.

[41] Guo Z.H., You Z.H., Huang D.S., Yi H.C., Zheng K. et al. Meshheading2vec: A new method for representing mesh headings as vectors based on graph embedding algorithm. *Brief. Bioinformatics*, 22:2085-2095, 2020.

[42] Xie M., Xu Y., Zhang Y., Hwang T. and Kuang R. Network-based phenome-genome association prediction by bi-random walk. *PLoS ONE*, 10:e0125138, 2015.

[43] Shang J. and Sun Y. Cherry: A computational method for accurate prediction of virus-prokaryotic interactions using a graph encoder-decoder model. *Brief. Bioinformatics*, 23:bbac182, 2022.

[44] Li C.C. and Liu B. Motifcnn-fold: Protein fold recognition based on fold-specific features extracted by motif-based convolutional neural networks. *Brief. Bioinformatics*, 21:2133-2141, 2019.