

# GATNNCDA

Subjects: Cell Biology

Contributor: Cunmei Ji

Circular RNAs (circRNAs) are a new class of endogenous non-coding RNAs with covalent closed loop structure. Researchers have revealed that circRNAs play an important role in human diseases. As experimental identification of interactions between circRNA and disease is time-consuming and expensive, effective computational methods are an urgent need for predicting potential circRNA–disease associations. In this study, we proposed a novel computational method named GATNNCDA, which combines Graph Attention Network (GAT) and multi-layer neural network (NN) to infer disease-related circRNAs. Specially, GATNNCDA first integrates disease semantic similarity, circRNA functional similarity and the respective Gaussian Interaction Profile (GIP) kernel similarities. The integrated similarities are used as initial node features, and then GAT is applied for further feature extraction in the heterogeneous circRNA–disease graph. Finally, the NN-based classifier is introduced for prediction. The results of fivefold cross validation demonstrated that GATNNCDA achieved an average AUC of 0.9613 and AUPR of 0.9433 on the CircR2Disease dataset, and outperformed other state-of-the-art methods.

Keywords: circRNA–disease associations ; graph attention network ; multi-layer neural network

---

## 1. Introduction

Circular RNAs (circRNAs) are a new class of endogenous non-coding RNA lacking a 5 cap and a 3 polyadenylated tail [1]. Since circRNAs were first discovered, in the 1970s, they have been considered as splicing errors [3][4]. In the past decade, with the development of high-throughput sequencing technology, a large number of circRNAs have been identified in mammalian cells [5][6]. Researchers have found that circRNAs are widely expressed in human tissues, and have stable structure and tissue-specificity. The mechanism of circRNA expression remains unknown, and how the biogenesis of circRNA affects its unique regulatory pattern remains limited [7]. Studies have revealed that many circRNAs perform their biological functions by acting as sponges of microRNA or RNA-binding proteins, by regulating protein function or by being translated themselves [8][9][10].

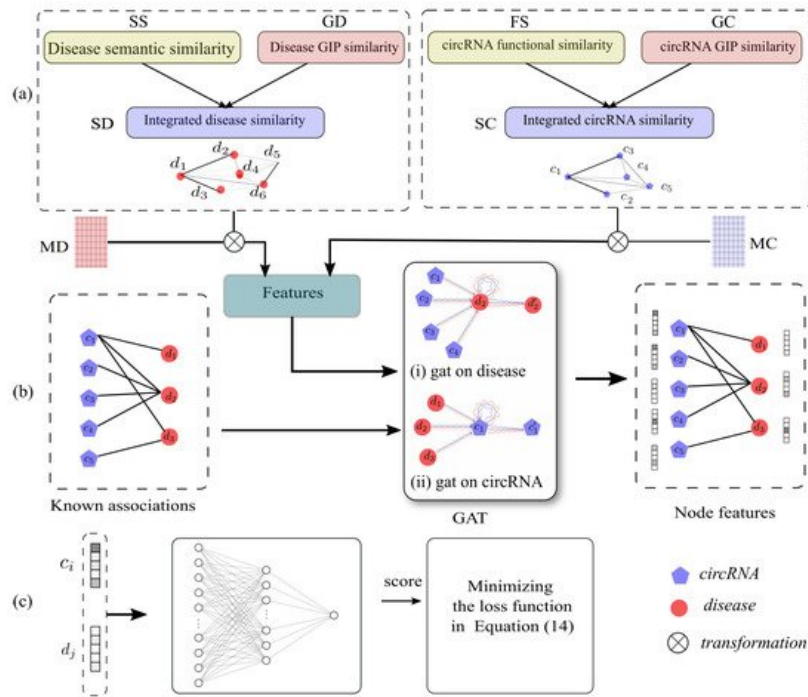
Cumulative evidence has indicated that many circRNAs are involved in human diseases, especially cancers [11]. For example, circHIPK3 has been found significantly up-regulated in colorectal cancer (CRC) tissues by sponging miR-7 to inhibit miR-7 activity [12]. Hsa\_circ\_0000190 was down-regulated in gastric cancer (GC) tissues and plasma from patients with GC. Compared with common biomarkers such as CEA and CA19-9, it has better sensitivity and specificity, and can be used as a novel biomarker for diagnosis of gastric cancer [13]. Researchers have identified that the expression of hsa\_circ\_0005075 is significantly different between hepatocellular carcinoma (HCC) and normal tissues [14]. The expression of Hsa\_circ\_0001649 was significantly different between HCC and normal liver tissues [15]. Moreover, circRNAs have also been related to other human diseases. CircANRIL is related to atherosclerotic disease by binding to pescadillo homolog 1 (PES1), which then impairs pre-rRAN processing and ribosomal biogenesis, results in the activation of p53, and thereby induces apoptosis and inhibits proliferation [14]. Recent studies have shown that the circRNA level in the brain is associated with Alzheimer's disease (AD) [16]. Compared with the control group, Li et al. have found that 112 circRNAs were up-regulated and 51 circRNAs were down-regulated in AD patients [17], which also were enriched in AD-related pathways, and the clinical guidance of circ-AXL, circ-GPHN and circ-PCCA in disease management of AD patients was identified.

As researchers have realized that circRNAs are abundant in mammalian cells, evolutionarily conserved and stable, and could serve as better biomarkers [18], databases of rich circRNA information, such as circBase [19], circ2traits [20], CircFunBase [21] have been built for study. Furthermore, researchers have also manually curated evidence from published literature, established databases such as circRNADisease [19], CircR2Disease [22], Circ2Disease [23], and circAtlas [24]. While experimental verification is expensive and time-consuming, computational methods have gradually introduced inferring potential circRNA–disease associations. Lei et al. first proposed a path weighted method to predict disease-related circRNAs. They calculated disease semantic similarity, disease functional similarity and integrated with the

Gaussian Interaction Profile (GIP) kernel similarities. Then, they constructed a heterogeneous network and adopted the depth-first search (DFS) to traverse nodes in the network and calculate the predictive score [25]. Yan et al. developed the DWNN-RLS method based on Regularized Least Squares of Kronecker product kernel for predicting circRNA–disease associations, and obtained AUC values of 0.8854, 0.9205 and 0.9701 in fivefold, 10-fold and leave-one-out cross validation, respectively [26]. Another graph-based method KATZHCDA achieved the best AUC values of 0.7936 and 0.8469 in fivefold CV and LOOCV, respectively [27]. Xiao et al. developed a weighted low-rank approximation optimization method with dual-manifold regulations to infer potential circRNA–disease associations [28].

Deep learning algorithms have also been introduced in this field. Deepthi et al. proposed an ensemble method named AE-RF, which extracted features via deep autoencoder, and then used random forest for prediction. As a result, this method achieved 0.9486 and 0.9552 in fivefold and 10 fold CV, respectively [29]. Li et al. used DeepWalk to extract node features in the circRNA–disease network, and used a network consistency projection algorithm for circRNA–disease interactions prediction [30]. Wang et al. designed GCNCDA using FastGCN to extract high-level features, and by applying Forest PA classifier for prediction [31]. As a result, it achieved an AUC value of 0.909 in fivefold CV based on circR2Disease dataset. Bian et al. developed GATCDA method based on graph attention network to obtain representation of circRNAs and diseases, calculated the probability score by dot production [32], and yielded an AUC value of 0.9011.

In this study, we proposed a novel computational method named GATNNCDA to predict potential circRNA–disease associations, based on graph attention network and multi-layer neural network. To be specific, GATNNCDA first integrates circRNA functional similarity, disease semantic similarity and the GIP similarities. Secondly, GATNNCDA utilizes linear transformation to project the integrated similarity matrices into the same space, and applies a graph attention network to extract dense representations of nodes in the heterogeneous circRNA–disease graph. Furthermore, a multi-layer neural network is constructed to infer the associations between circRNAs and diseases. The framework of GATNNCDA is shown in **Figure 1**. In summary, our contributions are listed as follows:



**Figure 1.** The framework of GATNNCDA. It consists of three steps: (a) similarity integration for circRNA and disease, (b) GAT-based feature extraction, and (c) NN-based classification.

- We proposed an end-to-end framework for inferring disease-related circRNAs, which can effectively and accurately infer the potential associations between circRNAs and diseases.
- We made use of GAT to extract low-dimensional dense representations of circRNAs and diseases, and these presentations had rich structural and semantic information of the heterogeneous circRNA–disease graph.
- We proposed a NN-based classifier, and applied a sampling strategy to construct balanced samples. In addition, we designed cross-entropy loss with L2 regularization to make the training process fast and robust.
- We demonstrated the predictive performance of our method by extensive experiments via fivefold cross validation and case studies, and achieved competitive results on CircR2Disease and circRNADisease datasets.

## 2. Case Studies

To further evaluate the prediction ability of our proposed method, we performed two case studies in this section. We trained GATNNCDA on CircR2Disease dataset [22], and then verified the candidates on circRNADisease [19] and circAtlas v2.0 [24] datasets. The first case study was conducted on breast cancer, which is one of the most common cancers in women. In particular, we constructed the positive samples with all known associations between circRNAs and diseases in the CircR2Disease. Meanwhile, we randomly chose the same number of negative samples from the unknown associations. Based on these training samples, we built the GATNNCDA and calculated the scores between breast cancer and each circRNA. Finally, we selected the top 20 related circRNAs for analysis. As shown in

**Table 1.** Top 20 predicted circRNAs related to Breast cancer based on circR2Disease dataset.

Rank	circRNA	Evidence	Rank	circRNA	Evidence
1	hsa_circ_0007534	//	11	hsa_circ_0068033	I; //
2	hsa_circ_0011946	//	12	circamotl1hsa_circ_0004214	I; //
3	hsa_circ_0093859	//	13	hsa_circ_0006528	I; //
4	circrna-000911	//	14	hsa_circ_0002874	I; //>
5	circrna-001283	PMID:29431182	15	hsa_circ_0001667	I; //
6	circrna-001175	//	16	hsa_circ_0085495	I; //
7	circrna-100438	PMID:29431182	17	hsa_circ_0086241	I; //
8	hsa_circ_0001982	I; //	18	hsa_circ_0092276	I; //
9	hsa_circ_0001785	I	19	hsa_circ_0003838	I; //
10	hsa_circ_0108942	I; //	20	circvrk1	I; //

I, // denote circRNADisease, circAtlas v2.0.

The second case study is performed on hepatocellular carcinoma. It is the most common form of liver cancer, with a higher incidence in patients with long-term liver diseases [33]. We utilized GATNNCDA to calculate the correlation score with circRNAs and then sorted by descending order. The top 20 hepatocellular carcinoma related circRNAs are listed in **Table 2**. We can see that 10 of the top 20 are verified by the validation datasets, and the other eight candidates have been conformed in relevant literature, e.g., hsa\_circ\_0000520 is one of the three circRNAs that showed significantly different expression levels in HCC tissues [14]. Therefore, the unknown associations with high scores are likely to be correlated.

**Table 2.** Top 20 predicted circRNAs related to hepatocellular carcinoma based on circR2Disease dataset.

Rank	circRNA	Evidence
1	circ3p1	//
2	hsa_circ_0067531	//
3	circarsp91hsa_circ_0085154	//
4	circmto1hsa_circrna_0007874hsa_circrna_104135	//
5	hsa_circ_0005986	I; //
6	hsa_circrna_100338circsnx27	PMID:28710406
7	hsa_circrna_104075	I; //
8	hsa_circrna_102049	PMID:28710406
9	circrna_000839	//
10	circzkscan1hsa_circ_0001727	I; //
11	hsa_circ_0004018	I; //
12	hsa_circ_0005075	//

Rank	circRNA	Evidence
13	hsa_circrna_100571	PMID: 29609527
14	hsa_circrna_400031	PMID:29609527
15	hsa_circrna_102032	PMID: 29609527
16	hsa_circrna_103096	PMID:29609527
17	hsa_circrna_102347	PMID:29609527
18	hsa_circrna_000167hsa_circ_0000518	unknown
19	hsa_circ_0000520	PMID:27258521
20	hsa_circ_0000172	unknown

I, // denote circRNADisease, circAtlas v2.0.

### 3. Conclusions

Cumulative evidence has shown that circRNAs play an important role in progression of human diseases, and are suitable as promising disease biomarkers for prevention, diagnosis and treatment. As traditional biological identification is very costly and time-consuming, more and more computational methods have been introduced in this field. In this study, we proposed a novel computational method called GATNNCDA for predicting potential circRNA–disease associations. GATNNCDA achieved a better performance than other state-of-the-art methods by combining similarity integration, graph attention network and multi-layer neural network. In particular, we performed fivefold CV for evaluation, and obtained the best performance of AUC of 0.9742, AUPR of 0.9707. The average values of AUC and AUPR for under 50 experiments were 0.9613 and 0.9452. Furthermore, case studies on breast cancer and hepatocellular carcinoma have also demonstrated that GATNNCDA can be a useful tool for predicting potential disease-related circRNAs.

However, GATNNCDA still has some limitations. The initial node features may not be perfect. Recall that similarity integration as initial node representations would affect the final performance. Nonetheless, known interactions between circRNA–disease associations are insufficient. In addition, circRNA functional similarity and GIP similarity may be inaccurate. Therefore, more biological information such as circRNA–miRNA association or circRNA sequence will be used for further study to construct more accurate node features, especially for some unseen circRNAs. Furthermore, the NN-based classifier of GATNNCDA requires negative samples for training, which are rarely reported in the literature. Randomly sampling from the unknown associations in a CircR2Disease dataset would introduce bias. In the future, we will seek a better negative sampling strategy to promote the performance of GATNNCDA.

### References

- Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S.D.; Gregersen, L.H.; Munschauer, M.; et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013, 495, 333–338.
- Meng, S.; Zhou, H.; Feng, Z.; Xu, Z.; Tang, Y.; Li, P.; Wu, M. CircRNA: Functions and properties of a novel potential biomarker for cancer. *Mol. Cancer* 2017, 16, 1–8.
- Sanger, H.L.; Klotz, G.; Riesner, D.; Gross, H.J.; Kleinschmidt, A.K. Viroids are single stranded covalently closed circular RNA molecules existing as highly base paired rod like structures. *Proc. Natl. Acad. Sci. USA* 1976, 73, 3852–3856.
- Coca-Prados, M.; Hsu, M.T. Electron microscopic evidence for circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 1979, 280, 339–340.
- Salzman, J.; Gawad, C.; Wang, P.L.; Lacayo, N.; Brown, P.O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* 2012, 7, e30733.
- Jeck, W.R.; Sharpless, N.E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 2014, 32, 453–461.
- Chen, L.L. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 2020, 21, 475–490.

8. Zheng, Q.; Bao, C.; Guo, W.; Li, S.; Chen, J.; Chen, B.; Luo, Y.; Lyu, D.; Li, Y.; Shi, G.; et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 2016, 7, 1–13.
9. Abdelmohsen, K.; Panda, A.C.; Munk, R.; Grammatikakis, I.; Dudekula, D.B.; De, S.; Kim, J.; Noh, J.H.; Kim, K.M.; Martindale, J.L.; et al. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol.* 2017, 14, 361–369.
10. Kristensen, L.S.; Andersen, M.S.; Stagsted, L.V.; Ebbesen, K.K.; Hansen, T.B.; Kjems, J. The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 2019, 20, 675–691.
11. Vo, J.N.; Cieslik, M.; Zhang, Y.; Shukla, S.; Xiao, L.; Wu, Y.M.; Dhanasekaran, S.M.; Engelke, C.G.; Cao, X.; Dan, R.; et al. The landscape of circular RNA in cancer. *Cell* 2020, 176, 869–881.e13.
12. Zeng, K.; Chen, X.; Xu, M.; Liu, X.; Hu, X.; Xu, T.; Sun, H.; Pan, Y.; He, B.; Wang, S. CircHIPK3 promotes colorectal cancer growth and metastasis by sponging miR-7 article. *Cell Death Dis.* 2018, 9.
13. Chen, S.; Li, T.; Zhao, Q.; Xiao, B.; Guo, J. Using circular RNA hsa\_circ\_0000190 as a new biomarker in the diagnosis of gastric cancer. *Clin. Chim. Acta* 2017, 466, 167–171.
14. Shang, X.; Li, G.; Liu, H.; Li, T.; Liu, J.; Zhao, Q.; Wang, C. Comprehensive circular RNA profiling reveals that hsa-circ-0005075, a new circular RNA biomarker, is involved in hepatocellular carcinoma development. *Medicine* 2016, 95, e3811.
15. Qin, M.; Liu, G.; Huo, X.; Tao, X.; Sun, X.; Ge, Z.; Yang, J.; Fan, J.; Liu, L.; Qin, W. Hsa-circ-0001649: A circular RNA and potential novel biomarker for hepatocellular carcinoma. *Cancer Biomark.* 2016, 16, 161–169.
16. Lukiw, W.J. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front. Genet.* 2013, 4, 1–2.
17. Li, Y.; Fan, H.; Sun, J.; Ni, M.; Zhang, L.; Chen, C.; Hong, X.; Fang, F.; Zhang, W.; Ma, P. Circular RNA expression profile of Alzheimer's disease and its clinical significance as biomarkers for the disease risk and progression. *Int. J. Biochem. Cell Biol.* 2020, 123, 105747.
18. Chen, L.L. The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 2016, 17, 205–211.
19. Glažar, P.; Papavasileiou, P.; Rajewsky, N. CircBase: A database for circular RNAs. *RNA* 2014, 20, 1666–1670.
20. Ghosal, S.; Das, S.; Sen, R.; Basak, P.; Chakrabarti, J. Circ2Traits: A comprehensive database for circular RNA potentially associated with disease and traits. *Front. Genet.* 2013, 4, 1–9.
21. Meng, X.; Hu, D.; Zhang, P.; Chen, Q.; Chen, M. CircFunBase: A database for functional circular RNAs. *Database* 2019, 2019, baz003.
22. 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.
23. 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, 1–6.
24. Wu, W.; Ji, P.; Zhao, F. CircAtlas: An integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* 2020, 21, 1–14.
25. Lei, X.; Fang, Z.; Chen, L.; Wu, F.X. Pwcd: Path weighted method for predicting circrna-disease associations. *Int. J. Mol. Sci.* 2018, 19, 3410.
26. Yan, C.; Wang, J.; Wu, F.X. DWNN-RLS: Regularized least squares method for predicting circRNA–disease associations. *BMC Bioinform.* 2018, 19, 520.
27. 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.
28. Xiao, Q.; Luo, J.; Dai, J. Computational Prediction of Human Disease- Associated circRNAs Based on Manifold Regularization Learning Framework. *IEEE J. Biomed. Health Inform.* 2019, 23, 2661–2669.
29. Deepthi, K.; Jereesh, A.S. Inferring Potential circRNA–disease Associations via Deep Autoencoder-Based Classification. *Mol. Diagn. Ther.* 2021, 25, 87–97.
30. Li, G.; Luo, J.; Wang, D.; Liang, C.; Xiao, Q.; Ding, P.; Chen, H. Potential circRNA–disease association prediction using DeepWalk and network consistency projection. *J. Biomed. Inform.* 2020, 112, 103624.
31. 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, e7568.
32. Bian, C.; Lei, X.J.; Wu, F.X. GATCDA: Predicting circRNA–disease associations based on graph attention network. *Cancers* 2021, 13, 2595.

33. Llovet, J.M.; Kelley, R.K.; Villanueva, A.; Singal, A.G.; Pikarsky, E.; Roayaie, S.; Lencioni, R.; Koike, K.; Zucman-Rossi, J.; Finn, R.S. Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* 2021, 7.
- 

Retrieved from <https://encyclopedia.pub/entry/history/show/31450>