Prediction of Gastric Cancer-Related Genes Based on the Graph Transformer Network

Yan Chen, Xuan Sun and Jiaxing Yang*

Department of Gastrointestinal Surgery, The First Hospital of Jilin University, Changchun, China

Gastric cancer is a complex multifactorial and multistage process that involves a large number of tumor-related gene structural changes and abnormal expression. Therefore, knowing the related genes of gastric cancer can further understand the pathogenesis of gastric cancer and provide guidance for the development of targeted drugs. Traditional methods to discover gastric cancer-related genes based on biological experiments are time-consuming and expensive. In recent years, a large number of computational methods have been developed to identify gastric cancer-related genes. In addition, a large number of experiments show that establishing a biological network to identify disease-related genes has higher accuracy than ordinary methods. However, most of the current computing methods focus on the processing of homogeneous networks, and do not have the ability to encode heterogeneous networks. In this paper, we built a heterogeneous network using a disease similarity network and a gene interaction network. We implemented the graph transformer network (GTN) to encode this heterogeneous network. Meanwhile, the deep belief network (DBN) was applied to reduce the dimension of features. We call this method "DBN-GTN", and it performed best among four traditional methods and five similar methods.

Keywords: gastric cancer, susceptibility gene, graph transformer network, deep belief network, heterogeneous network

INTRODUCTION

Gastric cancer is a malignant tumor originated from gastric mucosal epithelial cells (1). At present, due to the increase of work pressure, the change of diet structure, and *Helicobacter pylori* infection, gastric cancer is gradually showing a younger trend (2, 3). Patients with early gastric cancer often have no obvious symptoms, or only nonspecific symptoms such as abdominal discomfort and flatulence (4). These symptoms are often similar to chronic gastric symptoms such as dyspepsia, gastritis, and gastric ulcer (5). Most patients with early-stage cancer find their condition through gastroscopy. Under reasonable medical measures, the 5-year survival rate of patients with early-stage gastric cancer can reach 90% (6). However, most patients with gastric cancer are in the middle and late stage of gastric cancer when they are diagnosed. The tumor has invaded the outside of the stomach and is complicated by lymph node metastasis; thus, the odds of being cured is low. Screening related genes closely related to gastric cancer can be used as molecular targets for diagnosis (7). Different gene combinations can reflect the early diagnosis, incidence, effectiveness of

treatment, and prognosis of gastric cancer. The early stage of gastric cancer generally only contains a few gene changes. These changes are potential molecular targets for early diagnosis (8). If these changes can be detected, gastric cancer can be detected as soon as possible, which can greatly improve the cure rate of gastric cancer. The typing of gastric cancer susceptibility genes and related genes can also provide some information for the prediction of the disease (9), so as to take preventive measures as soon as possible to prevent the deterioration of the disease. With the continuous in-depth post-genome studies, more genotypes will be found to be related to the occurrence, development, and prognosis of gastric cancer. The final conclusion can provide a new theoretical basis for the discussion of the molecular mechanism of gastric cancer (10, 11).

Gastric cancer is a complex and multifactorial disease. Environmental and genetic factors play an important role in the occurrence of gastric cancer (12, 13). MiRNA precisely regulates the occurrence of gastric cancer by participating in a network system composed of a series of important biological processes such as cell proliferation, apoptosis, and differentiation (14). A large number of studies have shown that according to the difference in expression level, specific miRNAs have become a potential biomarker of malignant cancer and have an impact similar to carcinogenic or tumor suppressor genes. For example, the expression of miR-21 and miR-155 is usually increased in gastric cancer, which can promote cell proliferation and induce the occurrence of malignant cancer (15), and the expression of mir-449 is usually reduced. It can inhibit cell proliferation and inhibit the further development of gastric cancer (16). To a large extent, miRNA is almost involved in the whole process of gastric cancer pathogenesis. Therefore, with the deepening of research, it can enrich the biological function of miRNA, show a new vision for the in-depth study of the molecular mechanism of the occurrence and development of gastric cancer, and show a broader platform for the medical field. The application of gene chips can further extend the research on gastric cancer into the gene regulation network, making it possible to explore the gene expression profile of gastric cancer in different pathological stages. Gene chips have become a powerful tool to study the molecular regulation mechanism and pathway of gastric cancer progress, and they have been widely used in the field of gastric cancer research. In recent years, tumor genomics and proteomics have been widely used in biomedical and clinical research. Since the rise of gene chips and microarray technology, people have used these technologies to find new disease subclasses (17, 18), identify new tumor markers (19, 20), distinguish tumor grades (21), and predict the prognosis of the disease. For example, Wang et al. found that the increased expression of INHBA was related to the low survival rate of patients with gastric cancer through gene enrichment analysis (22). Liu et al. confirmed that extracellular matrix receptors and cell cycle signaling pathways may play an important role in gastric cancer (23). Wnt signaling pathway may lead to carcinogenesis by stimulating the migration and invasion of gastric cancer cells (24). β-Catenin is frequently mutated in gastric cancer (25). Fze3 is overexpressed in 75% of gastric cancer tissues and hsfrp is downregulated in 16% of

gastric cancer tissues, indicating that the expression of fze3 and hsfrp in this pathological tissue is often changed (26). Highly recombinant Shh induces the migration and invasion of gastric cancer cells by regulating tissue growth factor (TGF), which plays a role in the alk5–smad3 pathway (27). LOXL2 can promote tumor invasion through the Src/FAK signaling pathway, and its expression in gastric cancer is significantly increased (28). The loss of embryonic liver cell lining protein (ELF) can destroy the TGF-mediated signal pathway by interfering with the localization of Smad3 and Smad4 and lead to gastric cancer (29). The increase of BMP-2 concentration can significantly improve the motility and invasiveness of gastric cancer cells (30). The upregulation of cycox-61 may lead to the progression of gastric cancer. Interleukin-6 induces the invasion of gastric cancer cell line AGS cells through the activation of the c-Src/RhoA/ROCK signaling pathway (31).

Although the cost of large-scale sequencing data is decreasing and the speed is increasing, the number of clear gastric cancer-related genes remains small. A large number of multi omics data of gastric cancer have been accumulated. It is an important means to fully understand the genetic mechanism of gastric cancer to preliminarily screen potential genes through large-scale data mining algorithms and then verify them one by one through biological experiments. Systems biology aims to study the interaction of various molecules with different structures and functions at the overall level of organisms, and then add computational methods to describe and predict biological functions (32, 33), phenotypes, and behaviors. Most of these methods are based on networks (34, 35). These computational methods have been widely used in the discovery of disease-related genes (33, 35–39), genetic mechanism (40, 41), gene expression (37, 40), protein function (42, 43), metabolic association (44, 45), and drug target (46, 47). Therefore, in this paper, we developed a novel method named "DBN-GTN" to identify gastric cancer-related genes in a large scale. This method is based on the thought of systems biology. It used multiple features of genes and gastric cancer to identify the patterns of gastric cancer-related genes, which can be used to find more gastric cancer-related genes.

METHOD

Workflow

We firstly constructed a disease similarity network and a gene interaction network. We connected the two networks together based on the known relationship between diseases and genes. For example, the public databases have shown that the EGFR gene has a relationship with gastric cancer. Then, the node "gastric cancer" can be connected to the node "EGFR". Finally, we can obtain a heterogeneous network. Then, we should extract the features of diseases and genes, respectively. We used the relationship between miRNAs and both diseases and genes as the features. Therefore, gene feature is the regulatory relationship between the gene and all miRNAs. Disease feature is the known relationship between the disease and all miRNAs. Then, the deep

belief network (DBN) was applied to reduce the dimension of features. Finally, the graph transformer network was implemented to train the model and predict gastric cancer-related genes. The workflow is shown in **Figure 1**.

Construction of Heterogeneous Network

Firstly, we need to calculate the similarity of diseases. Disease ontology (DO) was applied to explore the relationship between different diseases. Every disease term in DO is related to some molecular components (such as genes, proteins, small molecules, and drugs), which are usually called annotation entities of diseases. The similarity between the two diseases is also related to their common ancestors. The similarity of two diseases from the same ancestor node is usually greater than that of two diseases that do not belong to the same ancestor node. Therefore, the similarity of two diseases can be calculated by calculating the amount of information of two disease ancestor nodes. Similarly, in DO, each disease is related to its annotation entity. The similarity between the two diseases can also be calculated by calculating the relationship between their annotation entities.

Then, we need to obtain gene interaction information. We downloaded gene interaction information from HumannetV2.0 (48). The genes that can interact with each other can be connected in the gene network.

Finally, we need to connect these two networks based on the known relationship between diseases and genes. The DisGeNet database (49) was used to obtain the associations between diseases and genes. Based on the information reported by DisGeNet, we can build a heterogeneous network of diseases and genes.

Feature of Diseases and Genes

DIANA-TarBase v8 (50) collected decade-long experimentally supported miRNA–gene interactions. Using this database, we obtained the relationship between genes known to be related to disease and miRNAs. Each miRNA is one dimension of a gene feature. If a gene is reported to be regulated by the miRNA, the feature value of this gene in this characteristic dimension is 1.

Mir2disease (51) contains 349 miRNAs, 163 diseases, 3,273 miRNAs, and the association information between diseases. Using this database, we obtained the relationship between miRNAs and diseases similar to gastric cancer. Each miRNA is one dimension of a disease feature. If a disease is reported to be related to the miRNA, the feature value of this disease in this characteristic dimension is 1.

Dimensionality Reduction by Deep Belief Network

In order to reduce the feature dimension of miRNA, we constructed a DBN network architecture based on Restricted Boltzmann Machine (RBM) for miRNA feature encoding. Each RBM is a layer in the DBN network architecture, and the DBN-based miRNA feature encoding method contains a total of 3 layers of RBMs.

First, the variables in RBM are divided into hidden variables and observable variables. Among them, the observable variables are the features of miRNAs. The observable and hidden

variables are represented by the observable layer and the hidden layer, respectively. The nodes in the RBM layer are not connected, and all the nodes in the adjacent RBM layers are connected to each other. This connection method is consistent with the fully connected neural network.

Unsupervised learning is difficult because the distribution of input miRNA features is unknown. Based on the conclusions of statistical mechanics, we describe the probability distribution with an energy-based model. An RBM is composed of miRNA features and latent variables, whose energy function is defined as:

$$E(v, h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{ij} h_j =$$

$$-a^T v - b^T v - v^T W h \qquad (1)$$

Where the feature of genes can be represented $v = [v_1, v_2, \dots v_m]^T$; h is the random vector $h = [h_1, h_2, \dots h_n]^T$ W is the matrix of weight. Both a and b are bias.

With the energy function, the joint probability between the original feature of a gene and the feature after dimensionality reduction can be defined, and the conversion from the visualized node to the hidden node can be realized. Denote the joint probability distribution as $p(v, h)$, which is calculated as follows:

$$p(v, h) = \frac{1}{Z} \exp(-E(v, h))$$

$$= \frac{1}{Z} \exp(a^T v) \exp(b^T h) \exp(v^T W h) \qquad (2)$$

Where $Z = \sum_{v,h} \exp(-E(v, h))$

is the partition function and can also be called normalization coefficient.

Prediction of Gastric Cancer-Related Genes by the Graph Transformer Network

Since our network is a heterogeneous network of diseases and genes, there are multiple types of meta-paths in it. The first step is to select edge types from the adjacency matrix A. Then, we need to do matrix multiplication of two selected adjacency matrices to learn a novel meta-path network $A^{(1)}$. This new adjacency matrix can be calculated as the sum of candidate adjacency matrices based on weight. The addition process is based on 1*1 convolution with the activation function softmax.

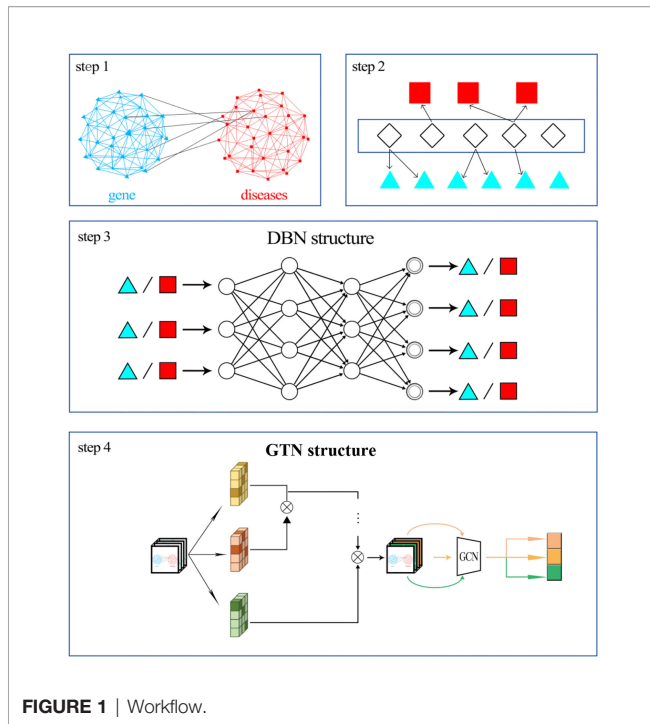$$Q = F(A, W) = \sigma(A, \qquad \text{softmax}(W)) \qquad (3)$$

$\sigma()$ represents a convolutional layer and W is the weight of it.

In this way, GTN can generate new meta-path adjacency matrices (52). Then, we can implement graph convolutional network (GCN) on these adjacency matrices. Each GCN layer can be calculated as:

$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^l W^l) \qquad (4)$$

Finally, each node in GTN can be encoded as:

$$Z = \| \sigma(D^{-1} A X W) \qquad (5)$$

FIGURE 1 | Workflow.

RESULTS

Compare With Traditional Methods

We obtained a total of 435 genes that are reported to be related to gastric cancer. We randomly selected 435 genes as the positive samples and selected part of the remaining genes as negative samples to train the model. We compared DBN-GTN with several traditional methods, which include support vector machine (SVM), back-propagation artificial neural networks (BP-ANN), naive Bayes, and random forest. Since these methods do not have the ability to encode a network, we simply combined the features of genes and diseases to construct a disease–gene pair. We input these disease–gene pairs into these traditional methods and build models to predict gastric cancer-related genes. The performance of these methods is shown in **Table 1**.

As we can see in **Table 1**, DBN-GTN performed best among these methods. The main reason why the accuracy of our analysis of DBN-GTN is significantly higher than other methods is that it considers the association between diseases and the interaction between genes, while other traditional methods are limited by their own shortcomings and cannot incorporate this information into the models.

TABLE 1 | AUC and AUPR of traditional methods and DBN-GTN.

Method	AUC	AUPR
DBN-GTN	0.93	0.86
SVM	0.78	0.68
BP-ANN	0.80	0.73
Naive Bayes	0.72	0.63
Random Forest	0.75	0.69

Compare With Similar Methods

Two methods make up the DBN-GTN, and we try to replace the two methods with similar methods to test whether the accuracy of the method is the highest. DBN mainly plays the function of dimensionality reduction, and principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) have a similar function. Therefore, we try to use these two methods to replace DBN and test the performance. In addition, GCN can be used to encode a homogeneous network. To compare the difference between encoding a heterogeneous network and encoding two homogeneous networks separately, we used GCN to replace GTN. GCN was implemented to encode a gene interaction network and a disease similarity network, respectively. Then, the features of genes and diseases are combined together to train the GCN model. The experimental results are shown in **Figure 2**.

As we can see from **Figure 2**, DBN-GTN performed best among these methods and t-SNE-GCN performed worst. From the impact of dimensionality reduction on accuracy, DBN outperforms t-SNE and PCA, and t-SNE has the worst accuracy. This is because PCA can manually select the amount of information contained after dimensionality reduction, while t-SNE can only reduce the data to 2 to 3 dimensions. From the perspective of the influence of the coding network method on accuracy, the performance of GTN is better than that of GCN. This is because GTN can encode heterogeneous networks and obtain more information than two homogeneous networks by GCN.

CONCLUSION

Biologists discovered some genes related to gastric cancer through large-scale transcriptome and genome sequencing. These results suffer from sample heterogeneity and insufficient sample size. At the same time, these experiments also cost a lot of time and money. Therefore, from the perspective of systems biology, this paper mines the association patterns between diseases and genes, and establishes a model through deep learning algorithms to identify large-scale gastric cancer-related genes. Although a large number of previous studies have used computational methods to identify gastric cancer-related genes, most of these methods focus on extracting information from homogeneous networks and cannot fully incorporate the association between diseases and genes into the model. In this paper, we established a disease similarity network and a gene interaction network, and connected the two networks through the known correlation between the two to form a disease–gene heterogeneous network. At the same time, we extracted the features of the disease and gene based on their relationship with miRNAs. In other words, a bridge between diseases and genes is established through miRNAs. We employ deep belief networks for feature dimensionality reduction and GTN for heterogeneous network encoding. We call this method DBN-GTN. We compare the accuracy of this method with four traditional methods and five similar methods. Experimental results show that DBN-GTN outperforms our chosen
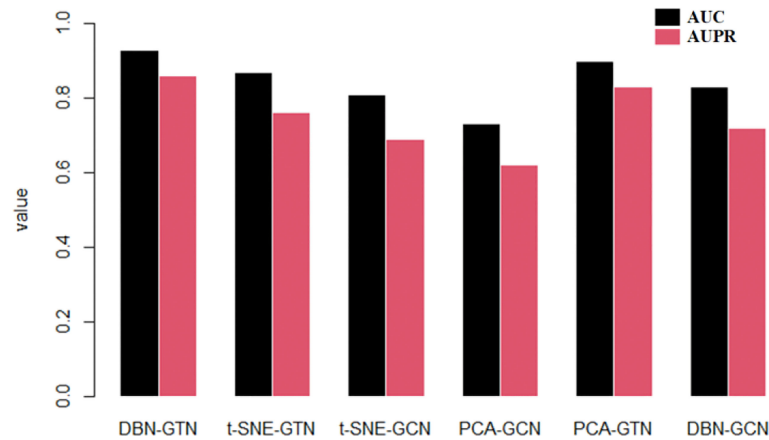
FIGURE 2 | AUC and AUPR of DBN-GTN and similar methods.

traditional method and similar methods, which shows that DBN-GTN is superior in the task of large-scale identification of gastric cancer genes. This paper provides support to further explain the genetic risk, susceptibility, and drug screening of gastric cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Experimental design: YC, XS, and JY. Data analysis: YC and XS. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.902616/full#supplementary-material

REFERENCES

1. Thrift AP, El-Serag HB. Burden of Gastric Cancer. *Clin Gastroenterol Hepatol* (2020) 18:534–42. doi: 10.1016/j.cgh.2019.07.045
2. Pyo JH, Lee H, Min YW, Min B-H, Lee JH, Kim K-M, et al. Young Age and Risk of Lymph Node Metastasis in Differentiated Type Early Gastric Cancer. *Ann Surg Oncol* (2018) 25:2713–9. doi: 10.1245/s10434-018-6659-3
3. Sonkar C, Doharey PK, Rathore AS, Singh V, Kashyap D, Sahoo AK, et al. Repurposing of Gastric Cancer Drugs Against COVID-19. *Comput Biol Med* (2021) 137:104826. doi: 10.1016/j.compbiomed.2021.104826
4. Gao F, Li M, Xiang R, Zhou X, Zhu L, Zhai Y. Expression of CLDN6 in Tissues of Gastric Cancer Patients: Association With Clinical Pathology and Prognosis. *Oncol Lett* (2019) 17:4621–5. doi: 10.3892/ol.2019.10129
5. Miyamoto R, Inagawa S, Sano N, Tadano S, Adachi S, Yamamoto M. The Neutrophil-to-Lymphocyte Ratio (NLR) Predicts Short-Term and Long-Term Outcomes in Gastric Cancer Patients. *Eur J Surg Oncol* (2018) 44:607–12. doi: 10.1016/j.ejso.2018.02.003
6. Shafabakhsh R, Yousefi B, Asemi Z, Nikfar B, Mansournia MA, Hallajzadeh J. Chitosan: A Compound for Drug Delivery System in Gastric Cancer-a Review. *Carbohydr Polymer* (2020) 242:116403. doi: 10.1016/j.carbpol.2020.116403

7. Kahroba H, Hejazi MS, Samadi N. Exosomes: From Carcinogenesis and Metastasis to Diagnosis and Treatment of Gastric Cancer. *Cell Mol Life Sci* (2019) 76:1747–58. doi: 10.1007/s00018-019-03035-2
8. Biagioni A, Skalamera I, Peri S, Schiavone N, Cianchi F, Giommoni E, et al. Update on Gastric Cancer Treatments and Gene Therapies. *Cancer Metastasis Rev* (2019) 38:537–48. doi: 10.1007/s10555-019-09803-7
9. Yusefi AR, Lankarani KB, Bastani P, Radinmanesh M, Kavosi Z. Risk Factors for Gastric Cancer: A Systematic Review. *Asian Pacific J Cancer Prevent: APJCP* (2018) 19:591–603. doi: 10.22034/APJCP.2018.19.3.591
10. Li Z, Zhang T, Lei H, Wei L, Liu Y, Shi Y, et al. Research on Gastric Cancer's Drug-Resistant Gene Regulatory Network Model. *Curr Bioinf* (2020) 15:225–34. doi: 10.2174/1574893614666190722102557
11. Machlowska J, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci* (2020) 21:4012. doi: 10.3390/ijms21114012
12. Guggenheim DE, Shah MA. Gastric Cancer Epidemiology and Risk Factors. *J Surg Oncol* (2013) 107:230–6. doi: 10.1002/jso.23262
13. Gu Y, Gao Y, Tang X, Xia H, Shi K. Bioinformatics Analysis Identifies CPZ as a Tumor Immunology Biomarker for Gastric Cancer. *Curr Bioinf* (2021) 16:98–105. doi: 10.2174/1574893615999200707145643

14. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van De Velde CJ, Nicolson M, et al. Perioperative Chemotherapy Versus Surgery Alone for Resectable Gastroesophageal Cancer. *N Engl J Med* (2006) 355:11–20. doi: 10.1056/NEJMoa055531

15. Sasaki CT, Vageli DP. miR-21, miR-155, miR-192, and miR-375 Deregulations Related to NF-kappaB Activation in Gastroduodenal Fluid–Induced Early Preneoplastic Lesions of Laryngeal Mucosa *In Vivo*. *Neoplasia* (2016) 18:329–38. doi: 10.1016/j.neo.2016.04.007

16. Jang S-G, Yoo CW, Park SY, Kang S, Kim HK. Low Expression of miR-449 in Gynecologic Clear Cell Carcinoma. *Int J Gynecol Cancer* (2014) 24:1558–63. doi: 10.1097/IGC.0000000000000267

17. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu C-Q, Strumpf D, et al. Three-Gene Prognostic Classifier for Early-Stage non–Small-Cell Lung Cancer. *J Clin Oncol* (2007) 25:5562–9. doi: 10.1200/JCO.2007.12.0352

18. Yoshihara K, Tajima A, Komata D, Yamamoto T, Kodama S, Fujiwara H, et al. Gene Expression Profiling of Advanced-Stage Serous Ovarian Cancers Distinguishes Novel Subclasses and Implicates ZEB2 in Tumor Progression and Prognosis. *Cancer Sci* (2009) 100:1421–8. doi: 10.1111/j.1349-7006.2009.01204.x

19. Li B-S, Zhao Y-L, Guo G, Li W, Zhu E-D, Luo X, et al. Plasma microRNAs, miR-223, miR-21 and miR-218, as Novel Potential Biomarkers for Gastric Cancer Detection. (2012) 7:e41629. doi: 10.1371/journal.pone.0041629

20. Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. Tumor Origin Detection With Tissue-Specific miRNA and DNA Methylation Markers. *Bioinformatics* (2018) 34:398–406. doi: 10.1093/bioinformatics/btx622

21. Fèvre-Montange M, Champier J, Durand A, Wierinckx A, Honnorat J, Guyotat J, et al. Microarray Gene Expression Profiling in Meningiomas: Differential Expression According to Grade or Histopathological Subtype. *Int J Oncol* (2009) 35:1395–407. doi: 10.3892/ijo_00000457

22. Wang Q, Wen Y-G, Li D-P, Xia J, Zhou C-Z, Yan D-W, et al. Upregulated INHBA Expression is Associated With Poor Survival in Gastric Cancer. *Med Oncol* (2012) 29:77–83. doi: 10.1007/s12032-010-9766-y

23. Liu P, Wang X, Hu C, Hu T. Bioinformatics Analysis With Graph-Based Clustering to Detect Gastric Cancer-Related Pathways. *Genet Mol Res* (2012) 11:3497–504. doi: 10.4238/2012.September.26.5

24. Kurayoshi M, Oue N, Yamamoto H, Kishida M, Inoue A, Asahara T, et al. Expression of Wnt-5a is Correlated With Aggressiveness of Gastric Cancer by Stimulating Cell Migration and Invasion. *Cancer Res* (2006) 66:10439–48. doi: 10.1158/0008-5472.CAN-06-2359

25. Clements WM, Wang J, Sarnaik A, Kim OJ, Macdonald J, Fenoglio-Preiser C, et al. β-Catenin Mutation is a Frequent Cause of Wnt Pathway Activation in Gastric Cancer. *Cancer Res* (2002) 62:3503–6. doi: 10.1002/cncr.10589

26. To K, Chan MW, Leung W, Yu J, Tong JH, Lee T, et al. Alterations of Frizzled (FzE3) and Secreted Frizzled Related Protein (hsFRP) Expression in Gastric Cancer. *Life Sci* (2001) 70:483–9. doi: 10.1016/S0024-3205(01)01422-9

27. Yoo YA, Kang MH, Kim JS, Oh SC. Sonic Hedgehog Signaling Promotes Motility and Invasiveness of Gastric Cancer Cells Through TGF-β-Mediated Activation of the ALK5-Smad 3 Pathway. *Carcinogenesis* (2008) 29:480–90. doi: 10.1093/carcin/bgm281

28. Peng L, Ran Y-L, Hu H, Yu L, Liu Q, Zhou Z, et al. Secreted LOXL2 is a Novel Therapeutic Target That Promotes Gastric Cancer Metastasis *via* the Src/FAK Pathway. *Carcinogenesis* (2009) 30:1660–9. doi: 10.1093/carcin/bgp178

29. Kim SS, Shetty K, Katuri V, Kitisin K, Baek HJ, Tang Y, et al. TGF-β Signaling Pathway Inactivation and Cell Cycle Deregulation in the Development of Gastric Cancer: Role of the β-Spectrin, ELF. *Biochem Biophys Res Commun* (2006) 344:1216–23. doi: 10.1016/j.bbrc.2006.03.236

30. Kang MH, Kim JS, Seo JE, Oh SC, Yoo YA. BMP2 Accelerates the Motility and Invasiveness of Gastric Cancer Cells *via* Activation of the Phosphatidylinositol 3-Kinase (PI3K)/Akt Pathway. *Exp Cell Res* (2010) 316:24–37. doi: 10.1016/j.yexcr.2009.10.010

31. Lin MT, Lin BR, Chang CC, Chu CY, Su HJ, Chen ST, et al. IL-6 Induces AGS Gastric Cancer Cell Invasion *via* Activation of the C-Src/RhoA/ROCK Signaling Pathway. *Int J Cancer* (2007) 120:2600–8. doi: 10.1002/ijc.22599

32. Zhao T, Hu Y, Peng J, Cheng L. DeepLGP: A Novel Deep Learning Method for Prioritizing lncRNA Target Genes. *Bioinformatics* (2020) 36:4466–72. doi: 10.1093/bioinformatics/btaa428

33. Li D, Zhang S, Ma X. Dynamic Module Detection in Temporal Attributed Networks of Cancers. *IEEE/ACM Trans Comput Biol Bioinform* (2021) Epub ahead of print. doi: 10.1109/TCBB.2021.3069441

34. Ma X, Sun PG, Gong M. An Integrative Framework of Heterogeneous Genomic Data for Cancer Dynamic Modules Based on Matrix Decomposition. *IEEE/ACM Trans Comput Biol Bioinf* (2020).

35. Huang Z, Wang Y, Ma X. Clustering of Cancer Attributed Networks by Dynamically and Jointly Factorizing Multi-Layer Graphs. *IEEE/ACM Trans Comput Biol Bioinform* (2021). Epub ahead of print. doi: 10.1109/TCBB.2021.3090586

36. Nguyen TM, Kim N, Kim DH, Le HL, Piran MJ, Um S-J, et al. Deep Learning for Human Disease Detection, Subtype Classification, and Treatment Response Prediction Using Epigenomic Data. *Biomedicines* (2021) 9:1733. doi: 10.3390/biomedicines9111733

37. Wu W, Liu Z, Ma X. jSRC: A Flexible and Accurate Joint Learning Algorithm for Clustering of Single-Cell RNA-Sequencing Data. *Brief Bioinf* (2021) 22: bbaa433. doi: 10.1093/bib/bbaa433

38. Zhao T, Lyu S, Lu G, Juan L, Zeng X, Wei Z, et al. SC2disease: A Manually Curated Database of Single-Cell Transcriptome for Human Diseases. *Nucleic Acids Res* (2021) 49:D1413–9. doi: 10.1093/nar/gkaa838

39. Ma X, Sun P, Gong M. An Integrative Framework of Heterogeneous Genomic Data for Cancer Dynamic Modules Based on Matrix Decomposition. *IEEE/ACM Trans Comput Biol Bioinform* (2022) 19:305–16. doi: 10.1109/TCBB.2020.3004808

40. Peng J, Zhao T. Reduction in TOM1 Expression Exacerbates Alzheimer's Disease. *Proc Natl Acad Sci* (2020) 117:3915–6. doi: 10.1073/pnas.1917589117

41. Zhao T, Hu Y, Zang T, Cheng L. MRTFB Regulates the Expression of NOMO1 in Colon. *Proc Natl Acad Sci* (2020) 117:7568–9. doi: 10.1073/pnas.2000499117

42. Yao S, You R, Wang S, Xiong Y, Huang X, Zhu S. NetGO 2.0: Improving Large-Scale Protein Function Prediction With Massive Sequence, Text, Domain, Family and Network Information. *Nucleic Acids Res* (2021) 49: W469–75. doi: 10.1093/nar/gkab398

43. Zhao T, Liu J, Zeng X, Wang W, Li S, Zang T, et al. Prediction and Collection of Protein–Metabolite Interactions. *Brief Bioinf* (2021) 22:bbab014. doi: 10.1093/bib/bbab014

44. Kim Y, Kim GB, Lee SY. Machine Learning Applications in Genome-Scale Metabolic Modeling. *Curr Opin Syst Biol* (2021) 25:42–9. doi: 10.1016/j.coisb.2021.03.001

45. Zhao T, Hu Y, Cheng L. Deep-DRM: A Computational Method for Identifying Disease-Related Metabolites Based on Graph Deep Learning Approaches. *Brief Bioinf* (2021) 22:bbaa212. doi: 10.1093/bib/bbaa212

46. Abbasi K, Razzaghi P, Poso A, Ghanbari-Ara S, Masoudi-Nejad A. Deep Learning in Drug Target Interaction Prediction: Current and Future Perspectives. *Curr Med Chem* (2021) 28:2100–13. doi: 10.2174/0929867327666200907141016

47. Zhao T, Hu Y, Valsdottir LR, Zang T, Peng J. Identifying Drug–Target Interactions Based on Graph Convolutional Network and Deep Neural Network. *Briefings Bioinf* (2021) 22:2141–50. doi: 10.1093/bib/bbaa044

48. Hwang S, Kim CY, Yang S, Kim E, Hart T, Marcotte EM, et al. HumanNet V2: Human Gene Networks for Disease Research. *Nucleic Acids Res* (2019) 47: D573–80. doi: 10.1093/nar/gky1126

49. Piñero J, Bravo ∥À., Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-Associated Genes and Variants. *Nucleic Acids Res* (2017) 45(D1):D833–9. doi: 10.1093/nar/gkw943

50. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase V8: A Decade-Long Collection of Experimentally Supported miRNA–Gene Interactions. *Nucleic Acids Res* (2018) 46:D239–45. doi: 10.1093/nar/gkx1141

51. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. Mir2disease: A Manually Curated Database for microRNA Deregulation in Human Disease. *Nucleic Acids Res* (2009) 37:D98–D104. doi: 10.1093/nar/gkn714

52. Yun S, Jeong M, Kim R, Kang J, Kim HJ. Graph Transformer Networks. *Adv Neural Inf Process Syst* (2019) 32:1–11. doi: 10.48550/arXiv.1911.06455