

Diagnostic Value of Glycosylated Extracellular Vesicle microRNAs in Gastric Cancer

Shunda Wang^{1,*}, Cuidie Ma^{2,*}, Zhihua Ren^{3,*}, Yufei Zhang², Kun Hao⁴, Chengxiu Liu⁴, Lida Xu⁴, Shun He⁵, Jianwei Zhang¹

¹Department of Pancreatic and Gastric Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, People's Republic of China; ²College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, People's Republic of China; ³Department of General Surgery, Qilu Hospital of Shandong University, Jinan, 250012, People's Republic of China; ⁴Beijing Hotgen Biotech Co., Ltd, Beijing, 102600, People's Republic of China; ⁵Department of Endoscopy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, People's Republic of China

*These authors contributed equally to this work

Correspondence: Shun He, Department of Endoscopy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, People's Republic of China, Email heshuns@126.com; Jianwei Zhang, Department of Pancreatic and Gastric Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, People's Republic of China, Email panchutong@163.com

Introduction: Early diagnosis is crucial for improving the prognosis of patients with gastric cancer (GC). However, the currently used biomarkers for diagnosing GC have limited sensitivity and specificity. This study aimed to develop a novel diagnostic model based on miRNAs from glycosylated extracellular vesicles and evaluate its effectiveness in diagnosing gastric cancer.

Methods: GlyExo-capture technology was used to isolate glycosylated extracellular vesicles from serum samples. The signatures were screened in a discovery cohort of GC patients (n=55) and non-disease controls (n=46) using an integrated process, including high-throughput sequencing technology, screening using a complete bioinformatics algorithm, validation using RT-qPCR, and evaluation by constructing a diagnostic model. The diagnostic model was evaluated in an independent validation cohort (n=139).

Results: We developed a diagnostic model for GC based on five miRNA pairs. This diagnostic model demonstrated high sensitivity, specificity, and stable performance in distinguishing GC patients from non-cancer controls with AUC of 0.930 in the independent validation cohort, particularly in differentiating early-stage GC from benign patients. The markers also showed excellent performance in indicating perineural invasion status and lymph node metastasis in the testing cohort.

Discussion: The model demonstrated high sensitivity and specificity in diagnosing patients with GC, especially in differentiating early-stage GC from benign patients. The five miRNA pairs could also aid in making treatment decisions. Thus, miRNAs derived from glycosylated exosomes are promising biomarkers for cancer diagnosis.

Keywords: biomarker, diagnosis, gastric cancer, extracellular vesicles, microRNAs

Introduction

Gastric cancer (GC) is a prevalent malignancy worldwide, with increasing annual incidence rates. According to the latest data, GC is the fifth leading cause of cancer-related deaths worldwide.¹ Despite the growing popularity of gastroscopy, many patients with GC are diagnosed with locally advanced disease, which leads to poor prognosis and limited treatment options. The five-year survival rate for patients with advanced GC is estimated to be 10%, while the five-year survival rate for early GC is 90%.² Therefore, early diagnosis is crucial for improving postoperative survival rates in patients with gastrointestinal cancer.

Currently used tumor markers for diagnosing GC, including carcinoembryonic antigen (CEA), CA19-9, CA125, CA72-4, and pepsinogen (PG I/II),³ have limited sensitivity and specificity, and none of them are unique to GC. Therefore, there is a need to identify early diagnostic markers for gastric cancer to improve treatment outcomes. Extracellular vesicles (EV) are generated and actively secreted by various cells (lymphocytes, dendritic cells, macrophages, and tumor cells) and are widely present in various body fluids.⁴ EV contain exosomes and exosomes are a part of

EV. They contain various bioactive substances such as RNA, DNA, and proteins,⁵ making them an abundant source of biomarkers for diseases. Extracellular vesicles have also been shown to participate in multiple steps of gastrointestinal tumorigenesis, including tumor proliferation, invasion, angiogenesis, immune escape, and chemoresistance,⁶ making them potential indicators of cancer prognosis. Besides, extracellular vesicles secreted by tumor have typical abnormal glycan characteristics, and glycosylation of them play certain role in tumor development.⁷

Owing to the nanoscale size of extracellular vesicles, non-coding RNAs, such as microRNAs (miRNAs), account for a large proportion of their abundance.⁸ To date, numerous circulating miRNA biomarkers with diagnostic and prognostic significance for cancers have been discovered in many researches.^{9–12} For example, several miRNAs, such as miRNA-21 and miR-376c miR-196a/b, can distinguish between GC patients and healthy controls with higher sensitivity and specificity compared to CEA or CA.19–9[¹³] Moreover, combined miRNA panels, such as a five-miRNA panel (miR-16, miR-25, miR-92a, miR-451, and miR-486-5p) and a four-miRNA panel (miR-501-3p, miR-143-3p, miR-451a, and miR-146a), were reported to be differentially expressed in the plasma of gastric patients and normal controls.^{14,15} Exosome miR-21, miR-320c, and miR-1225-5p in peritoneal lavage fluid have been proposed as prognostic markers for evaluating peritoneal recurrence of GC.¹⁶ However, variations in study methodologies, populations, and clinical factors have led to inconsistent findings regarding the selected miRNAs across different research laboratories.

In this study, we aimed to focus on the diagnostic value of miRNAs from EV in the serum of gastric cancer patients, with the goal of identifying novel early diagnostic biomarkers. Unlike traditional isolation methods such as ultracentrifugation, size-exclusion chromatography, and polymer precipitation, we adopted a novel purification method called GlyExo-capture. This technique selectively captures glycosylated extracellular vesicles from serum samples and cell supernatants with lectin specificity. The surface of secreted vesicles may carry glycosylated peptides from parent cells, and the glycosylating profile of tumor cells is specific to a certain extent, providing a theoretical basis for this technology.^{17,18} Extracellular vesicles secreted by gastric cancer tissues have typical abnormal glycan characteristics, for example, Neu5Ac α (2,6)-GalNAc α -Thr/Ser (STn antigen) can be used as biomarkers in GC.¹⁹ Moreover, they are easy to elute by destroying the affinity of lectins, which provides the advantage of keeping them intact and protecting their abundance during operation. Using this technique, we developed a novel diagnostic model for the early detection of GC based on five miRNA pairs from glycosylated extracellular vesicles using GlyExo-capture technology, which showed high accuracy, sensitivity, and specificity in distinguishing gastric cancer patients from controls.

Materials and Methods

Patient Enrollment

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital (No.NCC2022C-599). The samples in the discovery cohort were collected from the outpatient and inpatient departments between November 1, 2021, and April 30, 2022, then participants were further independently recruited from July 1 to October 31 in 2022. The diagnosis was confirmed by pathological evidence, including gastroscopy biopsy and surgical operation. The inclusion criteria for gastric cancer patients were age between 18 and 75 years, no neoadjuvant chemotherapy or radiotherapy, and no other malignant tumors. All patients with GC were treatment-naïve at the time of blood sampling. Benign disease (BN) refers to gastritis, including atrophic and non-atrophic gastritis. Healthy control (HC) includes patients with no abnormalities after gastroscopic examination. Written informed consent was obtained from all participants. Patient clinical characteristics, including baseline information such age, gender, alcohol consumption, smoke, hypertension, diabetes and family history of gastric cancer, pathological information such as tumor stage, differential grade, pathological subtypes, lymph node metastasis, and Lauren classification were collected.

Isolation and Characterization of Extracellular Vesicles

Peripheral blood samples were collected in vacuum tubes containing a separator gel (BD, Franklin Lakes, NJ, USA) and centrifuged at 1800 g for 10 min at room temperature (RT) within 1 h of collection. The resulting serum was further centrifuged at 3000 g for 10 min at RT to remove cellular debris. After aspiration, the samples were aliquoted and stored at -80°C until use.

Extracellular vesicles were isolated and characterized using the GlyExo-Capture method, which selectively captures glycosylated exosomes with lectin. They were isolated from the serum samples using the Extracellular vesicles isolation kit (Beijing Hotgen Biotech Co.,Ltd., Beijing, China) following the manufacturer's protocol. This approach utilizes wheat germ agglutinin (WGA)-conjugated magnetic beads to capture extracellular vesicles by taking advantage of the strong affinity between WGA and Neu5Ac α (2,6)-GalNAc α -Thr/Ser (sialyl-Tn, STn) glycan structures on glycoproteins present on the surface of extracellular vesicles, which enhances the enrichment of tumor derived exosomes.¹⁷ To characterize isolated extracellular vesicles, we employed three methods: transmission electron microscopy (TEM), nanoparticle tracking analysis (NTA), and Western blot (WB) analysis.

For TEM analysis, extracellular vesicles were fixed with 2% glutaraldehyde for 1 h at RT and stained with uranyl acetate solution (pH 4.5) to enhance the contrast of the samples. Images were acquired using an FEI Tecnai Spirit microscope (FEI, Eindhoven, The Netherlands) at 120 kV and electron micrographs were captured using a Gatan UltraScan 1000 CCD camera (Pleasanton, CA, USA).

For NTA, extracellular vesicles were diluted in PBS to a certain concentration and analyzed using Nanoparticle Tracking Analyzers ZetaView[®] PMX120 (Particle Metrix, Meerbusch, Germany). Particle size and number were determined by analyzing particle movement using the Software Zeta View (Version 8.05.16 SP2).

For WB analysis, extracellular vesicles were lysed by RIPA Lysis Buffer (Beyotime Biotechnology, Shanghai, China) before protein analysis. Protein concentration was determined using a Qubit Protein Assay (Invitrogen, Pittsburgh, USA) and a Qubit 4.0 fluorometer (Invitrogen, Pittsburgh, USA). They were loaded onto sodium dodecyl sulfate-polyacrylamide gels for electrophoresis. Primary antibodies against commonly used exosome markers, including anti-CD81, anti-CD63, anti-CD9, anti-TSG101, and anti-calnexin (Abcam, Cambridge, MA, USA), were used at 1:1000 dilution. Rabbit anti-mouse IgG-HRP secondary antibody was used to detect primary antibodies.

RNA Isolation and Next-Generation Sequencing

miRNAs were extracted from extracellular vesicles using the miRNeasy Mini Kit (Cat. 217,004, Qiagen, Hilden, Germany), according to the manufacturer's instructions. RNA concentration was measured using the Qubit microRNA assay (Invitrogen, Pittsburgh, USA) and a Qubit 4.0 fluorometer (Invitrogen, Pittsburgh, USA).

Next-generation sequencing was performed according to the manufacturer's instructions. The RNA was compiled into cDNA libraries using the NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible) (Cat No.: E7330, NEB, Ipswich, USA). The size of the sequencing library was determined using the E-Gel SizeSelect II gel of an E-Gel Power Snap electrophoresis system (Thermo Fisher Scientific, Inc., Waltham, USA). The quality and concentration of the cDNA library was assessed using an Agilent Bioanalyzer 2100. Finally, the samples were pooled at the same concentration before sequencing using an Illumina NextSeq 550 Sequencing System (75 nt, single-read).

Marker Discovery and Model Construction

The sequences were subsequently evaluated for quality, and primer-adaptor sequences were removed utilizing Cutadapt software.²⁰ Subsequently, the sequences were aligned to the human reference genome (HG38) using Bowtie 2.²¹ The aligned data were further annotated using HT-seq with the gff3 file sourced from miRBase version 23. The expression of miRNAs was normalized using FPKM. The ratios of expression values between related miRNAs were calculated as new features and screened in the discovery cohort, based on the method introduced in a previous study.²² A total of 32 miRNA pairs were selected and subsequently filtered using the recursive feature elimination (RFE) algorithm based on a random forest model, which yielded eight candidate markers. For the purpose of clustering, we employed the Heatmap package (version 1.0.12) with the ward. D2 method. To visually demonstrate the distinction between GC and non-GC groups, we employed Linear Discriminant Analysis (LDA) on the TPM expression matrix using the MASS package (version 7.3).

The miRNA-targeted mRNAs were predicted using an integrated analysis website starBase (<https://rnasysu.com/encori/>)²³ To visualize the miRNA-mRNA regulatory network, we utilized the “ggraph” package in R. Furthermore, KEGG pathway enrichment analysis was conducted on the miRNA-targeted mRNAs within the network using the “clusterProfiler” package in R to elucidate the functional roles of the DEM.²⁴

Reverse-Transcription Polymerase Chain Reaction (RT-qPCR)

To validate the candidate markers, we performed on an ABI 7500 Real-Time PCR System using a miRcute Plus miRNA qPCR Kit (SYBR Green) (Tiangen Biotechnology Co., Ltd., Beijing, China). Of the eight candidate markers identified, five miRNA pairs showed consistent results between RT-qPCR and NGS and were selected as biomarkers for gastric cancer. The specific primers utilized for qRT-PCR were synthesized by Sangon Biotech (Shanghai) Co., Ltd. The sequence of primers is presented in Table 1.

Prediction Modelling and Validation

Five miRNA pairs were used to construct a predictive model for GC using the random-forest algorithm. The random forest model was trained on the training set, and its predictive capability was evaluated on the test set and an independent validation set using the area under the receiver operating characteristic curve (AUC).

Statistical Analysis

Statistical analysis was performed using Python 3.9.7 software. We evaluated the differences between the two independent sample groups using the Mann-Whitney *U*-test, and all tests were two-tailed.

Results

Study Enrollment and Cohorts

This study aimed to develop a prediction model for miRNA signatures from glycosylated exosomes to distinguish gastric cancer patients from healthy individuals and those with benign gastric disease. As shown in Figure 1, the study design included a discovery cohort comprising 55 GC patients, 16 BN patients, and 30 hCs, and an independent validation cohort comprising 54 GC patients, 35 patients with BN, and 50 hCs. In the discovery cohort, miRNA signatures from glycosylated extracellular vesicles were selected through an integrated process, including univariate analysis and multivariate analysis of next-generation sequencing data and verification with RT-qPCR. The discovery cohort was then randomly divided into a training set and a testing set at a ratio of approximately 7:3. A prediction model was established on the training set, its performance on the testing set was evaluated, and the validation cohort was verified. Finally, the clinical value of the signatures in patient classification was analyzed for all patients with GC.

The clinical characteristics of the participants are presented in Table 2. There were no significant differences in the age distribution between the gastric cancer patients and control groups. Other clinical characteristics such as sex, smoking history, and alcohol consumption were also no significant differences. The TNM staging system was used to estimate the stages of gastric cancer based on tumor, node, and metastasis characteristics. We also categorized GC patients into early-stage gastric cancer (EGC), defined as invasion within the mucosa and submucosa, and advanced-stage gastric cancer (AGC). EGC refers to lesions confined to the mucosal layer regardless of lymph node metastasis. In the TNM staging system, all GC patients with T1 were classified as EGC, whereas the remaining cases were considered AGC. Approximately one-third of patients with GC were observed as EGC in both the discovery and validation cohorts. Most gastric cancer cases show poor- or moderate-grade tumor differentiation. The number of gastric cancer samples with and without lymph node metastasis was found to be equivalent. The most predominant histological type observed was mixed adenocarcinoma, whereas tubular adenocarcinoma and low-grade adhesion adenocarcinoma were also observed in large proportions.

Table 1 The Sequence of Primers

Name	Sequence
Forward primer-miR-16	GCCCGTAGCAGCACGTAATATT
Forward primer-miR-186	TGCGTCAAAGAATTCTCCTTTTG
Forward primer-miR-221	AACAAGAGCTACATTGTCTGCTGG
Forward primer-miR-629	CTGGGTTTACGTTGGGAGAACT
Forward primer-miR-363	AGAATTGCACGGTATCCATCTGTA
Forward primer-miR-339	TCCTCGACGACAGAGCCG

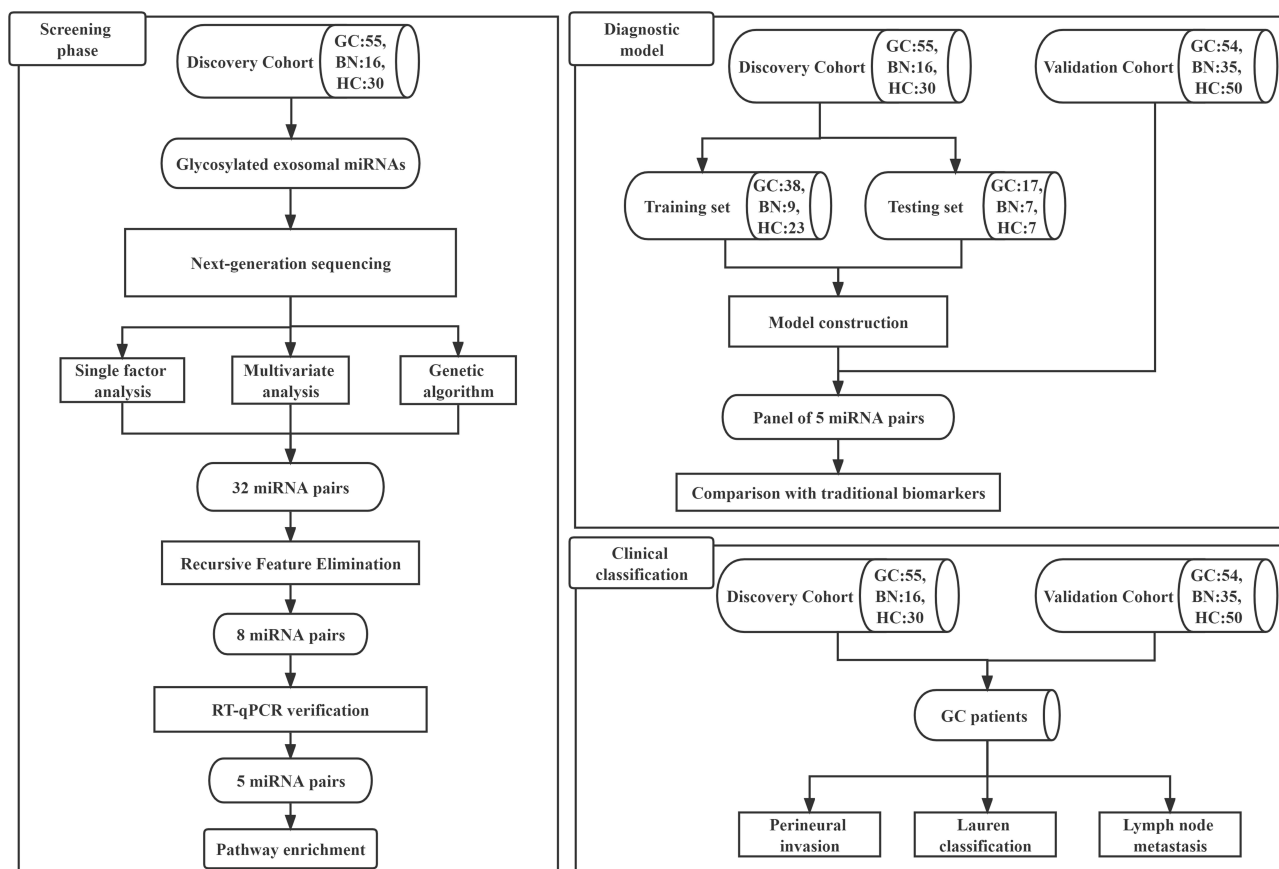


Figure 1 Schematic flowchart illustrating the discovery of EV-miRNA biomarker for GC detection. This study was divided into three stages: screening phase, diagnostic model, and clinical classification.

Abbreviations: GC, Gastric cancer; BN, benign; HC, healthy controls.

Characterization of Serum Glycosylated Extracellular Vesicles

Glycosylated extracellular vesicles were successfully isolated from serum using GlyExo-Capture technology and were characterized using various techniques. TEM images showed that the glycosylated extracellular vesicles had a diameter of approximately 100 nm, with a clear bilayer membrane and a cup-shaped structure (Figure 2A), indicating the presence of intact and functional vesicles. NTA confirmed a size distribution of 177.7 ± 83.2 nm with a concentration of 2.87×10^{10} particles/mL, consistent with the characteristics of typical extracellular vesicles (Figure 2B). WB analysis further validated the isolated exosomes by demonstrating the presence of prominent extracellular vesicles marker proteins, including CD63, CD81, and TSG101 (Figure 2C). Importantly, the absence of the negative control protein marker, calnexin, further underscored the specificity of the isolated vesicles. These comprehensive characterizations provide robust evidence for the efficacy of GlyExo-Capture technology in the successful isolation of extracellular vesicles from serum samples.

miRNA Profiling in the Discovery Cohort

In the discovery cohort, we performed miRNA sequencing analysis of the isolated extracellular vesicles and detected a median read counts of 8.34 million mapped reads per sample, with consistent detection across all groups (Figure 3A). However, significant differences were observed in the number of miRNAs detected between GC patients and healthy controls, indicating a potential bias in the capture method for GC samples (Figure 3B).

Among the detected miRNAs, 32 pairs were screened, and the expression ratio between each pair showed significant differences between GC and non-cancer samples (BN and HC). LDA analysis revealed that the profiles of the 32 miRNA pairs were significantly different among the three groups (Figure 3C). Unsupervised hierarchical clustering analysis showed a distinct separation between GC and non-cancer samples, as illustrated in the heatmap in Figure 3D. These

Table 2 Clinical Characteristics of Subjects in This Study

Characteristic		Discovery Cohort (n=101)									Validation Cohort (n=139)		
		Total (n=101)			Training Cohort (n=70)			Testing Cohort (n=31)					
		GC	Benign	Healthy	GC	Benign	Healthy	GC	Benign	Healthy	GC	Benign	Healthy
Number		55	16	30	38	9	23	17	7	7	54	35	50
Age	Mean ± SD	58.20 ± 9.34	53.94 ± 7.87	51.50 ± 6.80	57.55 ± 9.73	54.56 ± 8.29	51.87 ± 6.40	59.65 ± 8.50	53.14 ± 7.86	50.29 ± 8.44	57.09 ± 7.92	54.55 ± 10.30	50.20 ± 6.18
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Sex	Female	15	5	18	13	2	15	2	3	3	16	17	24
	Male	40	11	12	25	7	8	15	4	4	38	16	26
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Smoking history	Yes	12	3	2	5	1	1	7	2	1	11	2	6
	No	43	13	28	33	8	22	10	5	6	43	29	44
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Drinking history	Yes	14	2	4	10	1	3	4	1	1	15	4	9
	No	41	14	26	28	8	20	13	6	6	39	27	41
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Hypertension history	Yes	5	3	3	3	2	2	2	1	1	6	1	7
	No	50	13	27	35	7	21	15	6	6	48	30	43
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Diabetes history	Yes	4	2	2	3	1	1	1	1	1	6	2	6
	No	51	14	28	35	8	22	16	6	6	48	29	44
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
Family history	Yes	5	2	1	3	1	1	2	1	0	7	2	3
	No	50	14	29	35	8	22	15	6	7	47	29	47
	Unknown	0	0	0	0	0	0	0	0	0	0	2	0
TNM	I	20			15			5			18		
	II	10			6			4			11		
	III	22			15			7			25		
	Unknown	3			2			1			0		
Stage	EGC	17			12			5			14		
	AGC	38			26			12			40		

Lymph node metastasis	Positive	25			16			10			28		
	Negative	27			20			6			26		
	Unknown	3			2			1			0		
Perineural invasion	Positive	25			16			9			27		
	Negative	23			17			6			26		
	Unknown	7			5			2			1		
Lauren	Diffuse	22			17			5			21		
	Intestinal	20			13			7			21		
	Mixed	7			4			3			9		
	Unknown	6			4			2			3		
Differentiation	Poor	24			16			8			28		
	Moderate-Poor	9			5			4			13		
	Moderate	11			9			2			8		
	Well	1			1			0			0		
	Unknown	10			7			3			5		

Abbreviations: GC, Gastric cancer; EGC, Early-stage gastric cancer; AGC, Advanced stage gastric cancer.

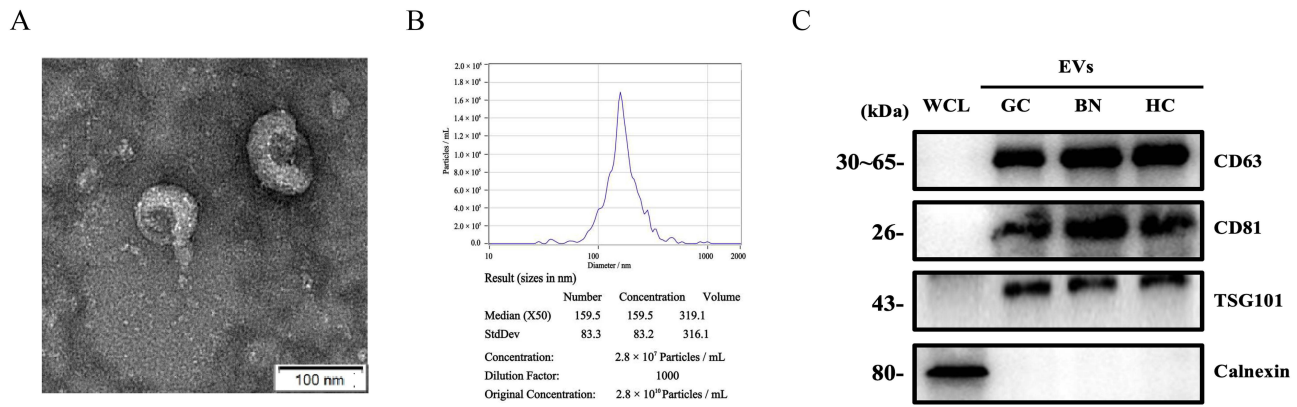


Figure 2 Characterization of glycosylated exosomes. **(A)** Transmission electron microscopy (TEM) image of glycosylated exosomes (bar = 100 nm). **(B)** Nanoparticle tracking analysis (NTA) of glycosylated exosomes. **(C)** Western blot analysis depicting whole cell lysates (WCLs) and EVs from GC, BN, HC. **Abbreviations:** GC, Gastric cancer; BN, benign; HC, healthy controls.

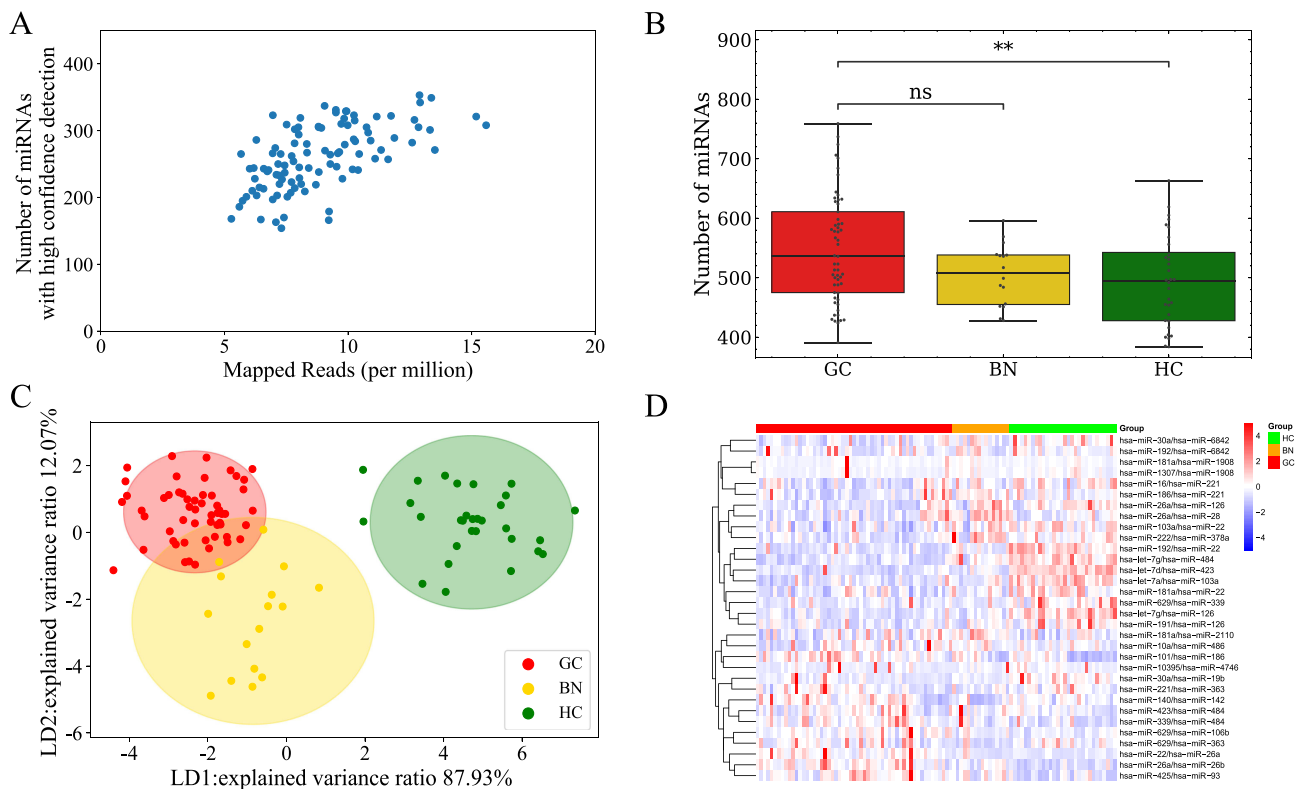


Figure 3 Glycosylated exosomal miRNAs sequencing results. **(A)** Distribution of the total mapped reads to the annotated genes. **(B)** Distribution of glycosylated exosomal miRNAs per sample among gastric cancer (GC) patients, benign (BN) patients, and healthy controls (HC). **(C)** Linear discriminant analysis (LDA) for the 32 miRNA pairs of GC patients compared with BN patients and HC. **(D)** Heatmap of unsupervised hierarchical clustering of the 32 miRNA pairs among GC patients, BN patients, and HC. ***P* < 0.01, ns: nonsignificant.

results suggest that miRNA profiling of glycosylated extracellular vesicles has the potential to serve as a non-invasive diagnostic tool for gastric cancer, and the identified miRNA pairs may serve as potential biomarkers for this disease.

Identification of GC-Specific miRNA Signatures Using the Discovery Cohort

Subsequently, the recursive feature elimination (RFE) algorithm based on the random forest approach was employed to identify candidate markers with significant discriminatory potential from the 32 miRNA pairs. We identified eight candidate miRNAs, five of which (hsa-miR-16-5p/hsa-miR-221-3p, hsa-miR-186-5p/hsa-miR-221-3p, hsa-miR-221-3p/hsa-miR-363-3p, hsa-miR-629-5p/hsa-miR-363-3p, and hsa-miR-629-5p/hsa-miR-339-3p) were potential diagnostic

markers for GC. The expression ratios of these five miRNA pairs were significantly different ($P < 0.05$) between GC patients and non-cancer groups in RT-qPCR, and receiver operating characteristic (ROC) curve analysis demonstrated good classification performance ($AUC > 0.6$), as displayed in Figure 4A and B. Moreover, the expression ratios of the five miRNA pairs also showed differences between cancerous and non-cancerous tissues in the TCGA gastric cancer database, consistent with the miRNA expression trend detected in extracellular vesicles, which further confirmed the reliability of the screening results (Figure 4C).

To further investigate the biological functions and pathways associated with miRNA signatures, we performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the predicted target genes of the miRNA signatures (Figure 5). KEGG results showed that the target genes were enriched in several pathways, including MAPK signaling, FoxO signaling, and proteoglycans in cancer. GO analysis revealed enrichment of target genes in the Wnt signaling process and the cellular components that play crucial roles in regulating the Wnt signaling pathway, such as the nuclear envelope, transport vesicle, focal adhesion, and protein kinase complex. These findings suggest that the identified miRNA signatures may be involved in the regulation of key biological processes and pathways involved in GC development and progression.

Diagnostic Value of the 5 miRNA Pairs Signature

We developed a novel diagnostic model based on the expression of the five miRNA pairs from glycosylated exosomes in the discovery cohort and evaluated its efficacy in the validation cohort (Figure 6A and Table 3). In the testing set within the discovery cohort, the model achieved an AUC of 0.92, a sensitivity of 82.35%, and a specificity of 85.71%. To further scrutinize the diagnostic performance, the model was validated on an independent cohort of 139 participants, comprising 54 GC samples, 35 benign samples, and 50 hC samples. The performance of the model was quite stable, with an AUC of 0.93,

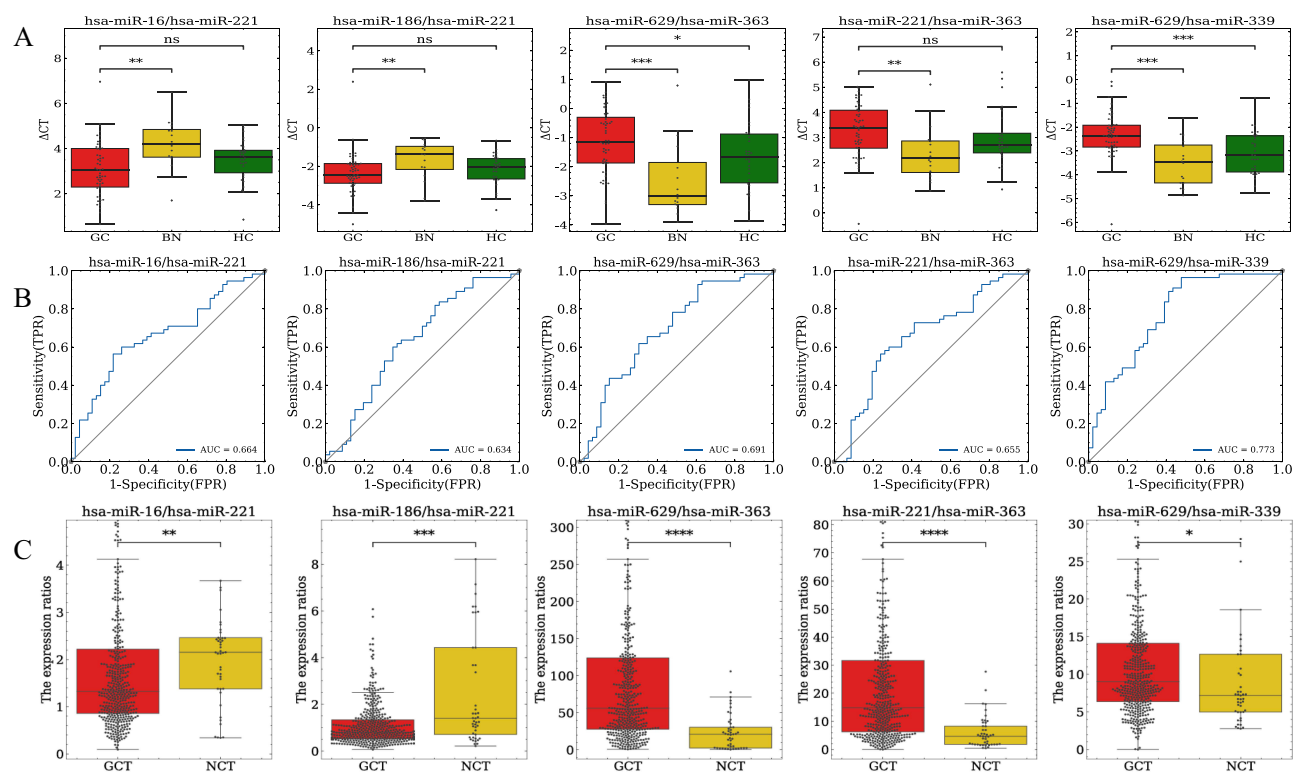


Figure 4 Diagnosis value of the 5 miRNA pairs. **(A)** Quantitative reverse-transcription polymerase chain reaction (RT-qPCR) of the 5 miRNA pairs. **(B)** Receiver-operating characteristic (ROC) curve analysis of the 5 miRNA pairs. **(C)** The expression ratios of the 5 miRNA pairs between gastric cancerous tissues (GCT) and non-cancerous tissues (NCT) in TCGA-STAD database. ΔCT , the relative expression levels of miRNAs were determined by calculating the difference between the CT (cycle threshold) values of the back miRNA and the front miRNA in the CT analysis; The expression ratios, The relative expression levels of miRNAs were determined by calculating the ratios between the front miRNA and the back miRNA; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns: nonsignificant.

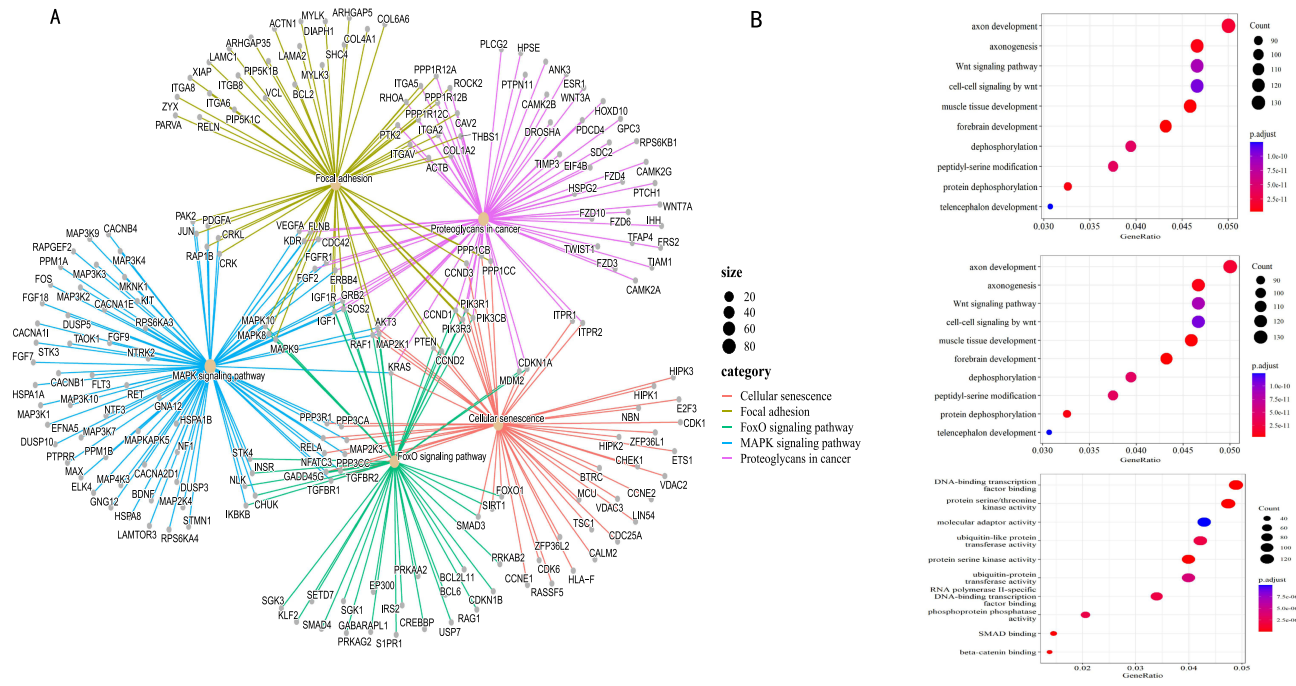


Figure 5 KEGG and GO enrichment analyses of the 5 miRNA pairs. **(A)** KEGG analysis of the 5 miRNA pairs. **(B)** GO analysis of the 5 miRNA pairs.

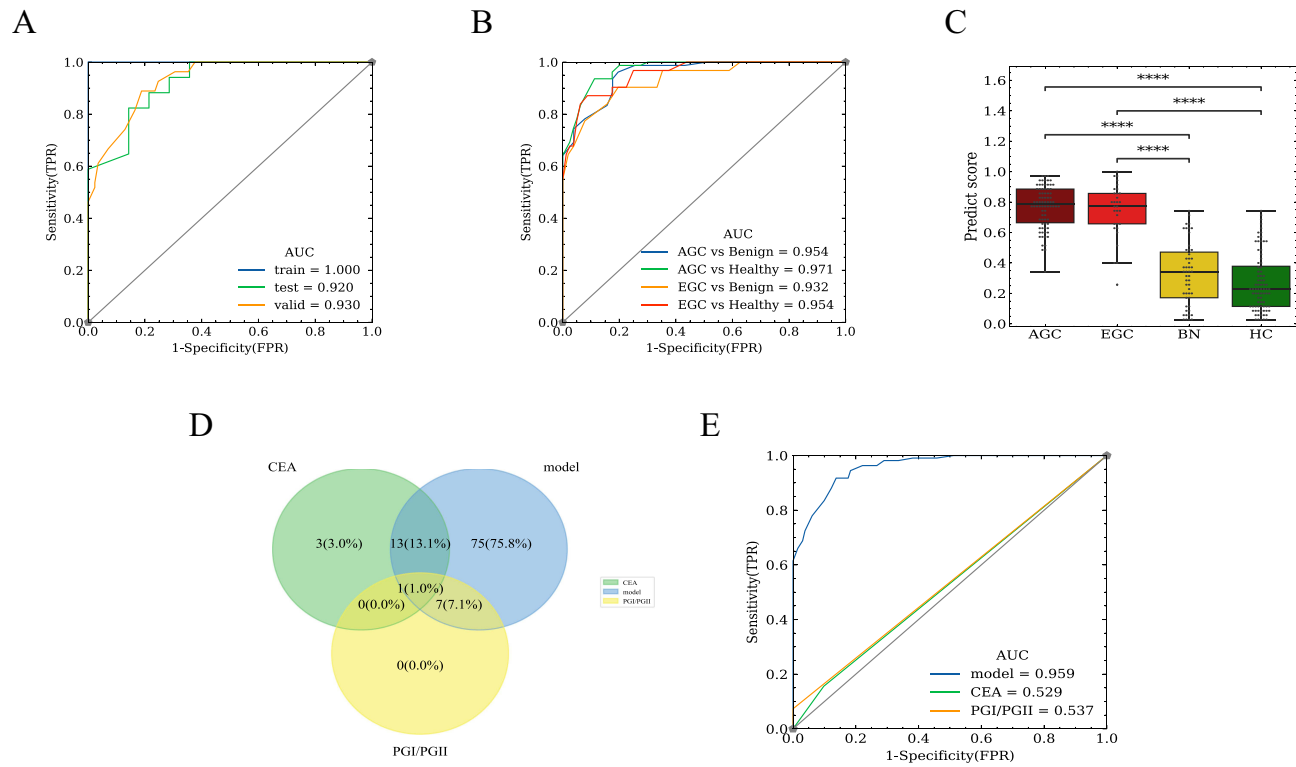


Figure 6 The performance of the prediction model in sub-group analysis and compared with protein markers. **(A)** ROC curve analysis for the performance of the prediction model in the training data set, the testing data set and the validation data set. **(B)** ROC curve for the performance of the prediction model for advanced stage gastric cancer (AGC) patients (n = 78), early-stage gastric cancer (EGC) patients (n = 31), benign patients (n = 51), and healthy controls (n = 80). **(C)** The prediction score of the prediction model in AGC, EGC, BN and HC. **(D)** Venn diagram of the prediction model and traditional biomarkers (CEA, PGI/PGII) for GC patients (n = 109). **(E)** ROC curve for the performance of the prediction model and traditional biomarkers (CEA, PGI/PGII). ****P < 0.0001.

Table 3 Confusion Table and Performance of the Prediction Model on the Discovery Cohort and the Validation Cohort

Cohort	Discovery Cohort (n=101)				Validation Cohort (n=139)	
	Training data set		Testing data set		Validation data set	
Performance	Real GC	Real NC	Real GC	Real NC	Real GC	Real NC
Predict GC	38	0	14	2	44	14
Predict NC	0	32	3	12	10	71
Totals	38	32	17	14	54	85
Correct	38	32	14	12	44	71
Sensitivity	100.00%		82.35%		81.48%	
Specificity		100.00%		85.71%		83.53%

Abbreviations: GC, Gastric cancer; NC, Negative control.

a sensitivity of 81.48%, and a specificity of 83.53%. These findings demonstrate the robustness and reliability of our model in distinguishing patients with GC from non-cancer controls, indicating its potential as a reliable diagnostic tool for GC.

Furthermore, the model achieved impressive diagnostic performance in distinguishing both early stage (EGC) and advanced-stage gastric cancers (AGC) from BN and HC. According to the prediction scores of all the samples in the discovery and validation cohorts, this model yielded an AUC of 0.932 for EGC vs BN, 0.954 for AGC vs BN, 0.954 for EGC vs HC, and 0.971 for AGC vs HC, as shown in [Figure 6B](#). The boxplot of model scores also demonstrated significant differences between GC patients and benign patients, as well as healthy controls ([Figure 6C](#)). These findings highlight the excellent performance of the model in the early diagnosis of gastric cancer, distinguishing GC from both healthy individuals and those with benign gastric disease.

Moreover, we compared the diagnostic performance of our model with that of traditional markers, such as CEA and PGI/PGII, in 109 patients with GC. Compared to CEA (17.1%) and PGI/PGII (8.1%), our model (97.0%) demonstrated a higher ability to identify patients with GC ([Figure 6D](#)). The model also had a higher AUC of 0.959 than CEA (AUC = 0.529) and PGI/PGII (AUC = 0.537), complementing their diagnostic results ([Figure 6E](#)).

Clinical Classification of GC Patients

Lymph node metastasis critically influences the progression and staging of GC. Accurate detection of lymph node metastasis is crucial for determining the appropriate treatment strategy and predicting patient outcomes. In the results, these five miRNA pairs exhibited excellent diagnostic accuracy for lymph node metastasis in GC. The model achieved an impressive AUC of 0.998 in the training set, while it maintained a high AUC of 0.953 in the independent validation set ([Figure 7A and B](#)). These results confirmed the model's effectiveness in identifying lymph node involvement in GC.

Perineural invasion is recognized as a prognostic indicator for poor outcomes in GC patients. Among the 109 patients enrolled in our study, 53 presented with perineural invasion, 53 were without, and three cases remained undetermined. Notably, while the five miRNA pairs were not entirely effective in distinguishing perineural invasion status, two specific pairs showed significant differentiation between positive and negative groups ([Figure 7C and D](#)). This finding suggests that the discovered biomarkers may have potential implications for GC patients with perineural invasion.

The Lauren classification categorizes gastric cancers into intestinal, diffuse, and mixed types and is a key clinical tool. However, our analysis revealed that the five miRNA pairs did not show discriminative value for different Lauren classifications in GC patients.

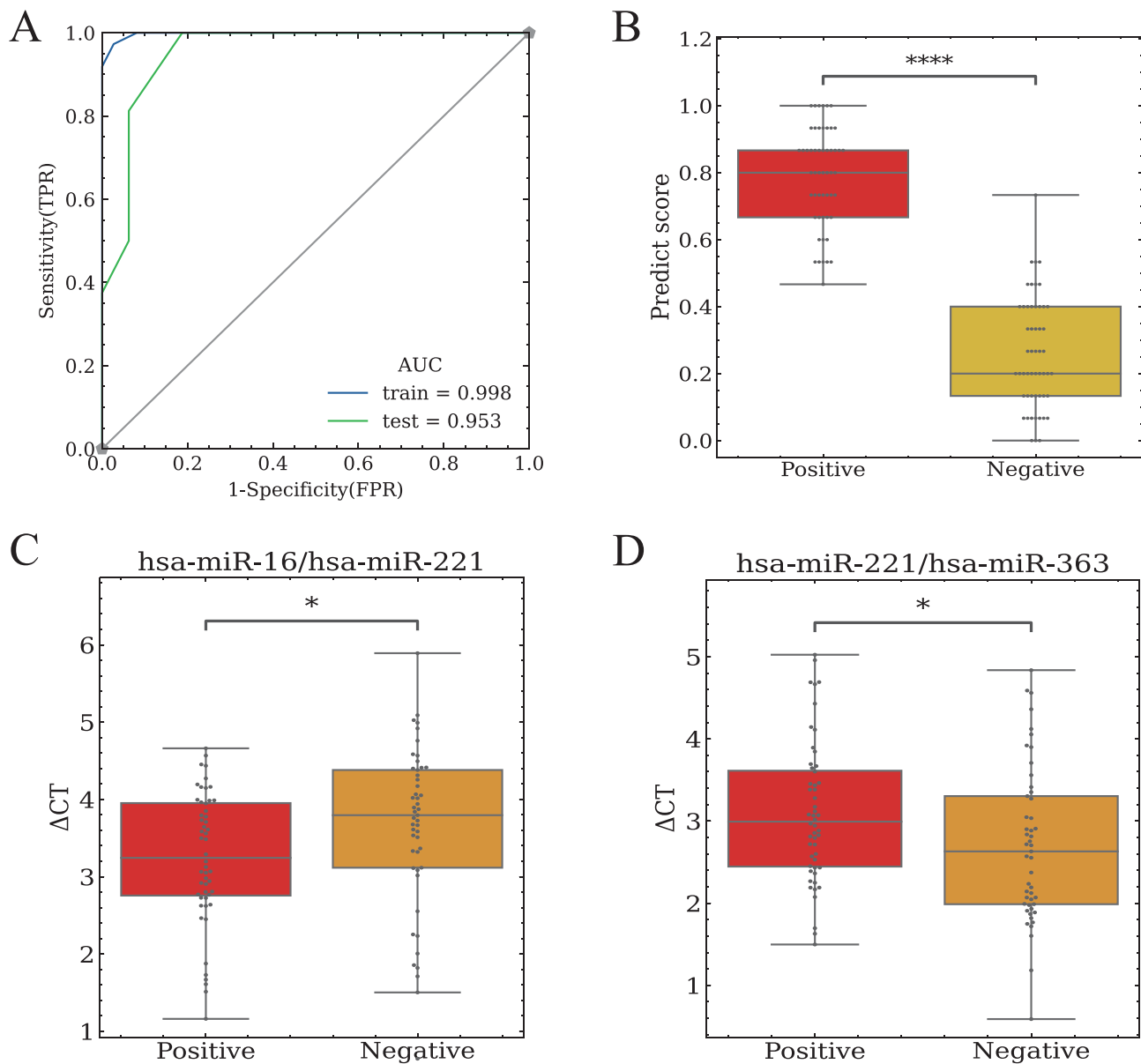


Figure 7 Diagnostic value of the 5 miRNA pairs for GC patient classification. **(A and B)** ROC curve analysis and the prediction score of 5 miRNAs pairs in GC with lymph node metastasis. **(C and D)** Quantitative reverse-transcription polymerase chain reaction (RT-qPCR) of two specific pairs in perineural invasion group. * $P < 0.05$, **** $P < 0.0001$.

Discussion

Early diagnosis is critical for improving the prognosis and selecting appropriate treatment options for patients with gastric cancer. In this study, we developed a novel diagnostic model based on five miRNA pairs derived from glycosylated extracellular vesicles and evaluated its effectiveness in gastric cancer diagnosis. Our diagnostic model demonstrated high accuracy, sensitivity, specificity, and stable performance in distinguishing GC patients from non-cancer controls, particularly in differentiating early stage GC from benign patients, indicating its potential as a diagnostic tool. Additionally, the markers showed excellent performance in indicating lymph node metastasis status of patients, which are crucial for treatment planning.

MiRNAs have proven to be powerful markers compared to traditional protein markers such as CEA or CA19-9 owing to their higher sensitivity and specificity in distinguishing GC patients.²⁵ Circulating miRNA biomarkers primarily originate from tumor-derived exosomes in the blood. However, there is still a lack of efficient isolation methods for these exosomes at

the technical level,^{17,18} despite miRNAs from exosomes being reported as biomarkers of gastric cancer in a large number of published articles.^{26,27} Considering that tumor exosomes are usually covered with abnormal glycans,^{17,28} we utilized a novel GlyExo-capture technology that employs lectin to isolate glycosylated extracellular vesicles from serum samples of GC patients and non-GC individuals. Our TEM, NTA, and WB results demonstrated that extracellular vesicles could be successfully captured by this method. More miRNAs were sequenced in vesicles from the GC group than in those from non-GC individuals, suggesting the advantage of GlyExo-capture technology in preferentially capturing extracellular vesicles from tumors and its potential to improve the sensitivity and specificity of miRNA-based biomarkers for GC diagnosis.

The discovery of the five miRNA pairs involved an integrated process, combining high-throughput sequencing on miRNAs, screening using a sophisticated bioinformatics algorithm, RT-qPCR validation, and the construction of a diagnostic model. In an independent validation cohort, the panel of five miRNA pairs achieved high diagnostic efficiency and significantly outperformed traditional markers, such as CEA and PGI/II. As the validation cohort was from the same center as the discovery cohort, this panel should be further validated or transferred to other centers. To ensure robustness across diverse datasets, we utilized the expression ratio of each miRNA pair as a model feature, enhancing the generalizability of our findings as discussed in reference.²²

Although the combination of these five miRNA pairs as a diagnostic model has never been reported in the literature, each miRNA identified has been previously associated with the development and progression of gastric cancer by regulating various biological processes and signaling pathways. Specifically, miR-221 has been reported to be significantly upregulated in GC²⁹ and has been implicated in promoting cell proliferation and inhibiting apoptosis in various cancers.³⁰ miR-221-3p is an oncogenic factor in gastric cancer that modulates PTEN expression, robustly stimulates cell proliferation, and facilitates *in vitro* invasion.³¹ Overexpression of miR-363-3p in GC cells exerts inhibitory effects on cell growth and migration; conversely, its downregulation exhibits opposite effects and is significantly associated with tumor differentiation in GC tissues.^{32,33} Previous studies have also observed upregulation of miR-629³⁴ and a significant reduction in miR-339³⁵ in primary GC tissues. Recent evidence highlights the downregulation of miR-186, which functions as a tumor suppressor in GC development by inhibiting Twist1 and exerting influence on the proliferation, invasion, and migration of human gastric cancer cells.^{36,37} These previous findings are consistent with the behavior of miRNA markers in our results, suggesting the significant biological significance and reliability of our markers. As for other cancer types, these miRNAs have also been verified to exert an important function during cancer progression. For example, miR-16-5p and miR-221-3p can be used as the differential diagnosis marker in nervous system lymphoma and mesenchymal stromal cells.^{38,39} miR-363-3p promotes prostate cancer tumor progression.⁴⁰ And miR-186-5p regulates docetaxel resistance in triple negative breast cancer, and promotes metastasis in lung cancer.^{41,42} What's more, miR-339-3p could promote osteosarcoma progression, and inhibits cell growth and EMT in nasopharyngeal carcinoma.^{43,44}

Additionally, we found that the five miRNA pairs showed good performance in discriminating early gastric cancer from benign controls, which may be a breakthrough in the early diagnosis of GC. Analysis of TCGA samples demonstrated the significant differential expression of these miRNAs in GC patients as well. Extracellular vesicles miRNAs can also act as prognostic biomarkers in GC.⁴⁵ In this study, the subgroup analysis also revealed the efficacy of the discovered biomarkers in assessing lymph node metastasis and perineural invasion status in GC patients, which provides valuable information for clinical decision-making and patient management. Nonetheless, further studies with larger sample sizes are required to confirm our findings.

Despite the high mortality associated with late-stage GC, early diagnosis continues to pose significant challenge. The potential of tumor extracellular vesicles in early GC detection lies in their widespread expression and stable structure, which effectively preserves miRNAs. Our adoption of the GlyExo-capture technology allows for the high-throughput and easy acquisition of tumor extracellular vesicles in clinical settings. Furthermore, the non-invasive liquid biopsy technology based on exosome miRNAs is expected to become a promising tool for the early diagnosis of GC in the future. While our novel panel of five miRNA pairs necessitates further validation in larger population cohorts, its diagnostic potential for GC patients is clear and holds significant promise for future advancements in cancer diagnostics.

Conclusion

In conclusion, we have made significant strides in the early diagnosis of GC by developing a novel diagnostic model using miRNA pairs derived from tumor extracellular vesicles, employing the GlyExo-capture method for efficient isolation. This approach has demonstrated promising diagnostic accuracy, particularly for early-stage GC, paving the way for advances in non-invasive cancer diagnostics through liquid biopsy technology. While our findings offer substantial promise, further validation in larger cohorts is essential to ensure the model's clinical applicability.

Abbreviations

GC, Gastric cancer; CEA, Carcinoembryonic antigen; EV, Extracellular vesicles; TEM, Transmission electron microscopy; NTA, Nanoparticle tracking analysis; WB, Western blot; RFE, Recursive feature elimination; ROC, receiver operating characteristic; AUC, Area under curve; LDA, Linear discriminant analysis; BN, Benign gastric disease; HC, Healthy controls; EGC, Early stage gastric cancer; AGC, Advanced stage gastric cancer; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Data Sharing Statement

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding authors.

Ethics Approval

This study was approved by the Institutional Ethics Committee on Scientific Research of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College, and informed consent was obtained from all patients.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by Beijing Chaoyang District Science and technology plan 2020 (CYSF2033), National Natural Science Foundation of China (82174531), National High Level Hospital Clinical Research Funding (XK2023-13) and National Natural Science Foundation of China (GrantNo.U23A20463).

Disclosure

The authors declare no conflicts of interest regarding the publication of this paper.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249. doi:10.3322/caac.21660
2. Luo M, Li L. Clinical utility of miniprobe endoscopic ultrasonography for prediction of invasion depth of early gastric cancer: a meta-analysis of diagnostic test from PRISMA guideline. *Medicine.* 2019;98(6):e14430. doi:10.1097/MD.00000000000014430
3. Tsai MM, Wang CS, Tsai CY, et al. Potential diagnostic, prognostic and therapeutic targets of MicroRNAs in human gastric cancer. *Int J mol Sci.* 2016;17(6):945. doi:10.3390/ijms17060945
4. Raab-Traub N, Dittmer DP. Viral effects on the content and function of extracellular vesicles. *Nat Rev Microbiol.* 2017;15(9):559–572. doi:10.1038/nrmicro.2017.60
5. Wahlgren J, Karlson TD, Brisslert M, et al. Plasma exosomes can deliver exogenous short interfering RNA to monocytes and lymphocytes. *Nucleic Acids Res.* 2012;40(17):e130. doi:10.1093/nar/gks463
6. Huang T, Song C, Zheng L, et al. The roles of extracellular vesicles in gastric cancer development, microenvironment, anti-cancer drug resistance, and therapy. *mol Cancer.* 2019;18(1):62. doi:10.1186/s12943-019-0967-5

7. Sakaue T, Koga H, Iwamoto H, et al. Glycosylation of ascites-derived exosomal CD133: a potential prognostic biomarker in patients with advanced pancreatic cancer. *Med Mol Morphol.* 2019;52(4):198–208. doi:10.1007/s00795-019-00218-5
8. Bellingham SA, Guo BB, Coleman BM, et al. Exosomes: vehicles for the transfer of toxic proteins associated with neurodegenerative diseases? *Front Physiol.* 2012;3:124.
9. Shi Y, Wang Z, Zhu X, et al. Exosomal miR-1246 in serum as a potential biomarker for early diagnosis of gastric cancer. *Int J Clin Oncol.* 2020;25(1):89–99. doi:10.1007/s10147-019-01532-9
10. Tang S, Cheng J, Yao Y, et al. Combination of four serum exosomal mirnas as novel diagnostic biomarkers for early-stage gastric cancer. *Front Genet.* 2020;11:237. doi:10.3389/fgene.2020.00237
11. Wang N, Wang L, Yang Y, et al. A serum exosomal microRNA panel as a potential biomarker test for gastric cancer. *Biochem Biophys Res Commun.* 2017;493(3):1322–1328. doi:10.1016/j.bbrc.2017.10.003
12. Zheng GD, Xu ZY, Hu C, et al. Exosomal miR-590-5p in serum as a biomarker for the diagnosis and prognosis of gastric cancer. *Front Mol Biosci.* 2021;8:636566. doi:10.3389/fmolb.2021.636566
13. Necula L, Matei L, Dragu D, et al. Recent advances in gastric cancer early diagnosis. *World J Gastroenterol.* 2019;25(17):2029–2044. doi:10.3748/wjg.v25.i17.2029
14. Jiang X, Wang W, Yang Y, et al. Identification of circulating microRNA signatures as potential noninvasive biomarkers for prediction and prognosis of lymph node metastasis in gastric cancer. *Oncotarget.* 2017;8(39):65132–65142. [J]. doi:10.18632/oncotarget.17789
15. Zhu C, Ren C, Han J, et al. A five-microRNA panel in plasma was identified as potential biomarker for early detection of gastric cancer. *Br J Cancer.* 2014;110(9):2291–2299. doi:10.1038/bjc.2014.119
16. Tokuhisa M, Ichikawa Y, Kosaka N, et al. Exosomal miRNAs from peritoneum lavage fluid as potential prognostic biomarkers of peritoneal metastasis in gastric cancer. *PLoS One.* 2015;10(7):e0130472. doi:10.1371/journal.pone.0130472
17. Martins AM, Ramos CC, Freitas D, et al. Glycosylation of cancer extracellular vesicles: capture strategies, functional roles and potential clinical applications. *Cells.* 2021;10(1):109.
18. Williams C, Royo F, Aizpurua-Olaizola O, et al. Glycosylation of extracellular vesicles: current knowledge, tools and clinical perspectives. *J Extracell Vesicles.* 2018;7(1):1442985. doi:10.1080/20013078.2018.1442985
19. Freitas D, Balmaña M, Poças J, et al. Different isolation approaches lead to diverse glycosylated extracellular vesicle populations. *J Extracell Vesicles.* 2019;8(1):1621131. doi:10.1080/20013078.2019.1621131
20. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–12. doi:10.14806/ej.17.1.200
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth.* 2012;9(4):357–359. doi:10.1038/nmeth.1923
22. Zhang Y, Ma C, Ding R, Chen H, Xu L, Yu C. The ratio of interacting miRNAs' expressions is a robust biomarker for disease classification in multi-center data. *medRxiv.* 2023; 2023:2023–06.
23. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42(D1):D92–D97. doi:10.1093/nar/gkt1248
24. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation.* 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
25. Guo X, Lv X, Ru Y, et al. Circulating exosomal gastric cancer-associated long noncoding rna1 as a biomarker for early detection and monitoring progression of gastric cancer: a multiphase study. *JAMA Surg.* 2020;155(7):572–579. doi:10.1001/jamasurg.2020.1133
26. Ohzawa H, Kumagai Y, Yamaguchi H, et al. Exosomal microRNA in peritoneal fluid as a biomarker of peritoneal metastases from gastric cancer. *Ann Gastroenterol Surg.* 2020;4(1):84–93. doi:10.1002/ags3.12296
27. Jamali L, Tofigh R, Tutunchi S, et al. Circulating microRNAs as diagnostic and therapeutic biomarkers in gastric and esophageal cancers. *J Cell Physiol.* 2018;233(11):8538–8550. doi:10.1002/jcp.26850
28. Block TM, Comunale MA, Lowman M, et al. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc Natl Acad Sci U S A.* 2005;102(3):779–784. doi:10.1073/pnas.0408928102
29. Liu K, Li G, Fan C, et al. Increased Expression of MicroRNA-221 in gastric cancer and its clinical significance. *J Int Med Res.* 2012;40(2):467–474. doi:10.1177/147323001204000208
30. Garofalo M, Quintavalle C, Romano G, et al. miR221/222 in cancer: their role in tumor progression and response to therapy. *Curr Mol Med.* 2012;12(1):27–33. doi:10.2174/156652412798376170
31. Shi J, Zhang Y, Jin N, et al. MicroRNA-221-3p plays an oncogenic role in gastric carcinoma by inhibiting PTEN expression. *Oncol Res.* 2017;25(4):523–536. doi:10.3727/096504016X14756282819385
32. Chen Z, Liu X, Hu Z, et al. Identification and characterization of tumor suppressor and oncogenic miRNAs in gastric cancer. *Oncol Lett.* 2015;10(1):329–336. [J]. doi:10.3892/ol.2015.3179
33. Song B, Yan J, Liu C, et al. Tumor suppressor role of miR-363-3p in gastric cancer. *Med Sci Monit.* 2015;21:4074–4080. doi:10.12659/MSM.896556
34. Li M, Wang Y, Liu X, et al. miR-629 targets FOXO3 to promote cell apoptosis in gastric cancer. *Exp Ther Med.* 2020;19(1):294–300. doi:10.3892/etm.2019.8168
35. Shen B, Zhang Y, Yu S, et al. MicroRNA-339, an epigenetic modulating target is involved in human gastric carcinogenesis through targeting NOVA1. *FEBS Lett.* 2015;589(20 Pt B):3205–3211. doi:10.1016/j.febslet.2015.09.009
36. Cao C, Sun D, Zhang L, et al. miR-186 affects the proliferation, invasion and migration of human gastric cancer by inhibition of Twist1. *Oncotarget.* 2016;7(48):79956–79963. [J]. doi:10.18632/oncotarget.13182
37. Liu L, Wang Y, Bai R, et al. MiR-186 inhibited aerobic glycolysis in gastric cancer via HIF-1alpha regulation. *Oncogenesis.* 2016;5(5):e224. doi:10.1038/onsis.2016.35
38. Sromek M, Rymkiewicz G, Paziewska A, et al. A set of 17 microRNAs common for brain and cerebrospinal fluid differentiates primary central nervous system lymphoma from non-malignant brain tumors. *Biomolecules.* 2021;11(9):1395. doi:10.3390/biom11091395
39. Ragni E, Perucca Orfei C, Silini AR, et al. miRNA reference genes in extracellular vesicles released from amniotic membrane-derived mesenchymal stromal cells. *Pharmaceutics.* 2020;12(4):347. doi:10.3390/pharmaceutics12040347
40. Xu L-Z, Ning J-Z, Ruan Y, et al. MiR-363-3p promotes prostate cancer tumor progression by targeting Dickkopf 3. *J Clin Lab Anal.* 2022;36(4):e24360. doi:10.1002/jcla.24360

41. Jing L, Lan L, Mingxin Z, et al. METTL3/LINC00662/miR-186-5p feedback loop regulates docetaxel resistance in triple negative breast cancer. *Scient Rep.* 2022;12(1):16715. doi:10.1038/s41598-022-20477-0
42. Hsu X-R, Wu J-E, Wu -Y-Y, et al. Exosomal long noncoding RNA MLETA1 promotes tumor progression and metastasis by regulating the miR-186-5p/EGFR and miR-497-5p/IGF1R axes in non-small cell lung cancer. *J Exp Clin Cancer Res.* 2023;42(1):283. doi:10.1186/s13046-023-02859-y
43. Liu L, Huang W. hsa_circ_0020378 regulating miR-339-3p/COL1A1 promotes osteosarcoma progression. *Cancer Biol Ther.* 2023;24(1):2274120. doi:10.1080/15384047.2023.2274120
44. Gao P, Zhao K, Lu W, et al. miR-339-3p inhibits cell growth and epithelial-mesenchymal transition in nasopharyngeal carcinoma by modulating the KAT6A/TRIM24 axis. *Int J Clin Oncol.* 2022;27(11):1684–1697. doi:10.1007/s10147-022-02231-8
45. Kahroba H, Hejazi MS, Samadi N. Exosomes: from carcinogenesis and metastasis to diagnosis and treatment of gastric cancer. *Cell mol Life Sci.* 2019;76(9):1747–1758. doi:10.1007/s00018-019-03035-2

Cancer Management and Research

Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>

Dovepress
Taylor & Francis Group