



OPEN ACCESS

Original research

Construction of exosome non-coding RNA feature for non-invasive, early detection of gastric cancer patients by machine learning: a multi-cohort study

Ze-Rong Cai,¹ Yong-Qiang Zheng ,¹ Yan Hu,¹ Meng-Yao Ma,² Yi-Jin Wu,¹ Jia Liu,¹ Lu-Ping Yang ,¹ Jia-Bo Zheng,³ Tian Tian,² Pei-Shan Hu,³ Ze-Xian Liu ,¹ Lin Zhang,¹ Rui-Hua Xu ,¹ Huai-Qiang Ju ^{1,4}

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/gutjnl-2024-333522>).

For numbered affiliations see end of article.

Correspondence to

Professor Huai-Qiang Ju; juhq@sysucc.org.cn Professor Rui-Hua Xu; xurh@sysucc.org.cn and Professor Lin Zhang; zhanglin@sysucc.org.cn

Z-RC, Y-QZ, YH and M-YM are joint first authors.

Received 1 August 2024

Accepted 8 December 2024

Published Online First

3 January 2025

ABSTRACT

Background and objective Gastric cancer (GC) remains a prevalent and preventable disease, yet accurate early diagnostic methods are lacking. Exosome non-coding RNAs (ncRNAs), a type of liquid biopsy, have emerged as promising diagnostic biomarkers for various tumours. This study aimed to identify a serum exosome ncRNA feature for enhancing GC diagnosis.

Designs Serum exosomes from patients with GC (n=37) and healthy donors (n=20) were characterised using RNA sequencing, and potential biomarkers for GC were validated through quantitative reverse transcription PCR (qRT-PCR) in both serum exosomes and tissues. A combined diagnostic model was developed using LASSO-logistic regression based on a cohort of 518 GC patients and 460 healthy donors, and its diagnostic performance was evaluated via receiver operating characteristic curves.

Results RNA sequencing identified 182 candidate biomarkers for GC, of which 31 were validated as potential biomarkers by qRT-PCR. The combined diagnostic score (cd-score), derived from the expression levels of four long ncRNAs (RP11.443C10.1, CTD-2339L15.3, LINC00567 and DiGeorge syndrome critical region gene (DGCR9)), was found to surpass commonly used biomarkers, such as carcinoembryonic antigen, carbohydrate antigen 19-9 (CA19-9) and CA72-4, in distinguishing GC patients from healthy donors across training, testing and external validation cohorts, with AUC values of 0.959, 0.942 and 0.949, respectively. Additionally, the cd-score could effectively identify GC patients with negative gastrointestinal tumour biomarkers and those in early-stage. Furthermore, molecular biological assays revealed that knockdown of DGCR9 inhibited GC tumour growth.

Conclusions Our proposed serum exosome ncRNA feature provides a promising liquid biopsy approach for enhancing the early diagnosis of GC.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Early detection is definitely significant to improve prognosis in patients with gastric cancer (GC).
- ⇒ Exosome transcriptome feature based on liquid biopsy is a revolutionary tool for early diagnosis and prognosis of tumours.
- ⇒ The application of exosome transcriptome features in the diagnosis of GC has not been thoroughly studied, and the role of these features in the development of GC is still unknown.

WHAT THIS STUDY ADDS

- ⇒ We conducted a multi-phase study to develop a serum exosome non-coding RNA (ncRNA) feature for diagnosing GC through machine learning.
- ⇒ The exosome ncRNA feature has wide practicality, including diagnosing GC patients with negative traditional gastrointestinal tumour-related biomarkers and those in early stage.
- ⇒ One long non-coding RNA of the exosome feature was demonstrated to play a vital role in the initiation of GC and be a promising intervention target.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ The serum exosome transcriptome feature was expected to be transformed into a novel diagnostic kit for GC.
- ⇒ This study provided a foundation for further mechanism exploration of the gastric carcinogenesis and novel strategies for intervention and treatment of GC.



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

To cite: Cai Z-R, Zheng Y-Q, Hu Y, et al. *Gut* 2025;**74**:884–893.

INSTRUCTION

Gastric cancer (GC) is one of the most prevalent cancers globally. According to the latest statistics from the International Agency for Research on Cancer, there were 1 089 103 new cases of GC and 768 793 deaths due to the disease in 2020, making GC the fifth most common cancer and the

fourth leading cause of cancer-related mortality.¹ While the 5-year survival rate for early-stage GC (EGC) patients who undergo radical surgery may exceed 90%, the median survival for patients with advanced GC receiving chemotherapy and immunotherapy is less than 15 months.^{2,3} Thus, early detection and diagnosis are crucial for improving patient outcomes.⁴ In this regard, gastroscopy with biopsy

remains the standard diagnostic method for GC. However, due to its invasive, costly and dependency on the expertise of endoscopists, which tends to be non-uniform across institutions and even departments, gastroscopy is not ideal for widespread screening, highlighting the urgent need for alternative and less invasive screening methods.

Liquid biopsy is a non-invasive method for detecting circulating tumour cells, circulating tumour DNA and extracellular vesicles in body fluids such as serums.^{5,6} This technique has revolutionised early cancer detection, stage assessment and treatment guidance.⁷ Initially, tumour markers in the blood, such as serum α -fetoprotein for hepatocellular carcinoma, have been used as an adjunct in cancer diagnosis.⁸ For GC, markers such as carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9) and CA72-4 have been used, and later research has shown that they may have limited sensitivity, which can lead to missed diagnoses.⁹ Recently, there has been growing interest in novel biomarkers derived from liquid biopsy for GC.^{10–12} In this regard, Jimmy Bok Yan So *et al* developed a cost-effective serum assay based on 12 microRNAs for GC risk assessment,¹³ and Yangzi Chen *et al* introduced a diagnostic model incorporating 10 metabolites for GC.¹⁴ Overall, current literature suggests that liquid biopsy holds substantial promise for enhancing the early diagnosis of GC.

Transcriptome sequencing studies have demonstrated that non-coding RNAs (ncRNAs) comprise over 90% of the RNA transcribed from the human genome.¹⁵ Long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) play essential roles in gene regulation, thus influencing cell proliferation, metastasis, apoptosis and genome stability.^{16–20} Besides, several lncRNAs and circRNAs have been identified as potential diagnostic markers for GC.^{21–23} Exosomes, which are extracellular vesicles with diameters ranging from 40 to 160 nm, are rich in nucleic acids and proteins. They are readily obtainable from body fluids, which contain high concentrations of exosomes ($\sim 10^9$ particles/mL).^{24,25} Exosomes are secreted by living cells and serve as carriers of abundant biological information from their parental

cells.²⁶ Due to their lipid bilayer, exosomes are highly stable and can circulate in physiological conditions, even within hostile tumour microenvironments. This stability enables exosomes to be isolated and detected reliably, even after long-term storage.²⁷ Therefore, exosomes are highly advantageous for liquid biopsy applications. Specifically, exosome lncRNAs and circRNAs have shown considerable promise as molecular biomarkers for cancer diagnosis.²⁸

Artificial intelligence (AI) has shown considerable promise for widespread application in various medical fields.^{29,30} Machine learning (ML), a key branch of AI, involves the analysis of data to develop effective and efficient detection or diagnostic models.^{31–33} Among the various ML algorithms, logistic regression and LASSO were the most common techniques and have proven valuable in developing diagnostic models for multiple cancers.^{12,34} Thus, integrating ML with liquid biopsy techniques offers a promising approach for identifying and validating biomarkers for GC diagnosis.^{35,36}

Here, we designed this present study to develop and validate a novel ncRNA-based feature derived from serum exosomes for enhancing GC diagnosis using ML techniques. To achieve this, we retrospectively assessed the diagnostic potential of transcriptomic features from serum exosomes. This approach holds promise for offering a new liquid biopsy method for GC diagnosis.

METHOD

Patient enrolment

This study used a retrospective cohort consisting of 1595 participants, recruited from 2003 to 2021. Participants in all phases in [figure 1](#) should meet the following criteria. For GC patients: (1) pathological confirmation of GC, (2) no prior therapy before surgery, (3) availability of complete clinical data and (4) sufficient serum samples collected prior to surgery. For patients with gastric precancerous lesions: (1) pathological confirmation of gastric precancerous lesions as defined by the

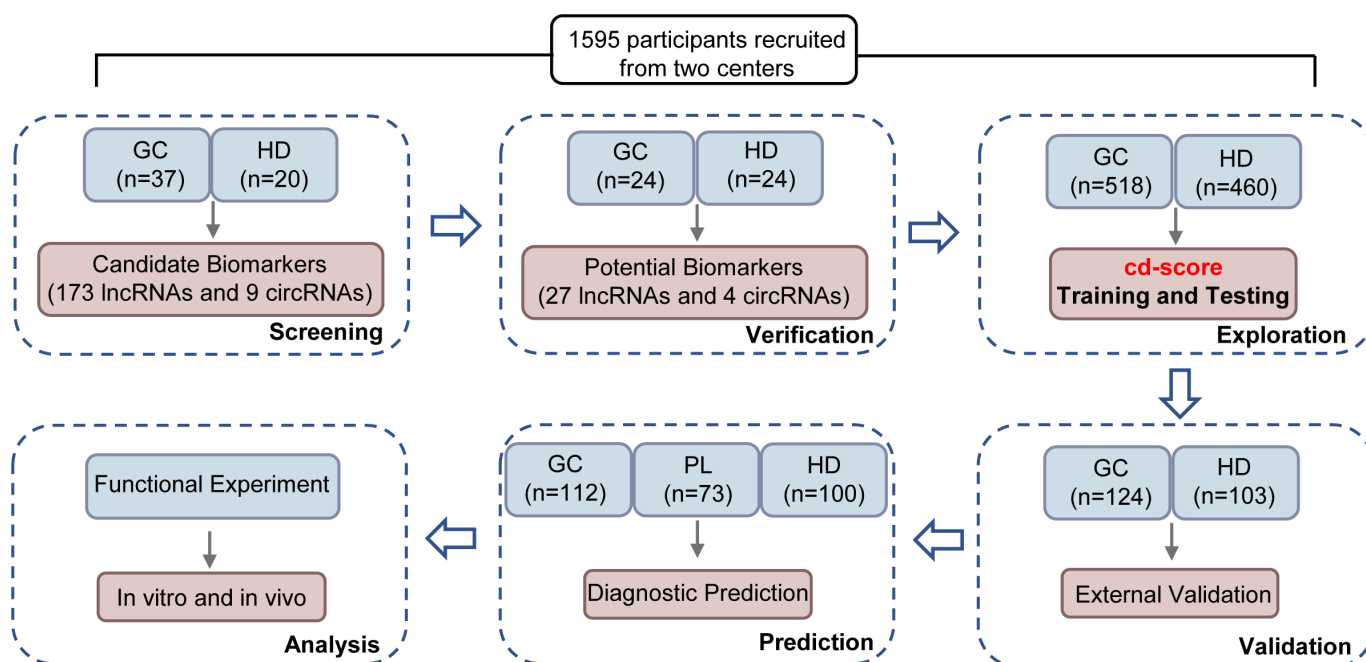


Figure 1 Overview of the study design. The figure illustrates the overall workflow of the study, including the key stages of data collection and analysis. HD, healthy donors; GC, gastric cancer; PL, precancerous lesions.

guideline of the American Society for Gastrointestinal Endoscopy³⁷ and European Society of Gastrointestinal Endoscopy,³⁸ and (2) no history of cancer. Some patients with precancerous lesions presented multiple types of lesions concurrently (online supplemental data 1), complicating subgroup classification. For healthy donors: (1) no precancerous or malignant lesions in the stomach confirmed by endoscopy and (2) no history of GC or precancerous lesions. Baseline clinicopathological data collected included age, gender, pathological tumour-node-metastasis (TNM) stage, maximum tumour diameter, tumour location, tumour differentiation, vascular invasion, nerve infiltration, human epidermal growth factor receptor 2 (HER2) status, smoking history, drinking history, family tumour history, body mass index and levels of CEA, CA19-9, and CA72-4. Pathological staging was based on the Union for International Cancer Control TNM staging system (8th edition). CEA, CA19-9 and CA 72-4 were detected using Elecsys CEA (Roche, 07027079190), Elecsys CA 19-9 (Roche, 07027028190), Elecsys CA 72-4 (Roche, 09005706190), respectively. The cut-off values for CEA, CA19-9 and CA72-4 were set at 5 ng/mL, 35 U/mL and 6.9 U/mL, respectively.

RNA extraction and quantitative reverse transcription PCR

RNA was extracted from exosomes using TRIzol Reagent (15596018, ThermoFisher) according to the manufacturer's instructions. Complementary DNA was synthesised from the extracted RNA using the PrimeScript RT Reagent Kit (RR037A, TAKARA). External Standard Kit (λ polyA) for qPCR (3789, TAKARA) was added as an external control during the reverse transcription on the RNAs extracted from serum exosome samples. Quantitative PCR (qPCR) was conducted using GoTaq qPCR Master Mix (A6002, Promega) and a Fluorescent Quantitative PCR Instrument 480 (ROCHE). Primer sequences for ncRNAs are provided in online supplemental data 3. The expression levels of ncRNAs were normalised using the $2^{-\Delta\Delta C_t}$ method.

Data preprocessing and model construction

The expression levels of 31 ncRNAs were normalised using the mean value of all samples within each cohort. Batch effects among the three cohorts were corrected using the R package 'ComBat'.³⁹

First, the 'superml' R package (<https://github.com/saras-watmks/superml>) and the 'e1071' R package (<https://www.rdocumentation.org/packages/e1071/versions/1.7-16>) were used to evaluate seven common ML algorithms via 100-fold cross validation, including LASSO-logistic regression, logistic regression, extreme gradient boosting (XGBoost), random forest, K nearest neighbours (KMN), support vector machine (SVM) and Naïve Bayes. For each model, the *Trainer* module with default parameters was used to establish the ML models, and the *predict* module was used for model evaluation. 100-fold cross validation was adopted for model comparison. Subsequently, logistic regression analyses of single ncRNA adjusted by age and gender in the training cohort were performed using the 'glm' R package. The performance of each single ncRNA was evaluated by receiver operating characteristic (ROC) curves using the 'pROC' R package,⁴⁰ and 100-fold cross validation was used for feature comparison. Ten top-ranked markers were then included in the final model construction through LASSO-logistic regression algorithm via 'glmnet' packages, and the best model was selected via 10-fold cross validation.

Statistical analysis

Data were analysed using R program V.4.2 and GraphPad Prism V.9.0. Methods for calculating p values are described in the figure legends. A significance threshold of two-tailed $p < 0.05$ was applied, except in the case of verification of potential biomarkers. Univariate and multivariable analyses were conducted using the 'stats' R package. Other detailed methods are included in online supplemental file 4.

RESULTS

Study design and participants

As shown in figure 1, our study involved the analysis of 1595 serum samples from three groups: patients with GC, patients with precancerous lesions and healthy donors. The study was conducted in five distinct phases: the biomarker screening phase, the experimental verification phase, the exploratory phase, the external validation phase and the prediction phase. The detailed clinicopathological characteristics of all participants are summarised in online supplemental table S1.

Screening and verification of potential ncRNAs

During the screening phase, we collected serum samples from 37 patients with GC and 20 healthy donors, which then underwent exosome extraction and subsequent transcriptome sequencing. To seek the ncRNAs in the exosomes derived from tumour cells, we also collected GC tissues and adjacent normal tissues from 20 patients and performed RNA sequencing on these samples. The sequencing results indicated that lncRNAs and circRNAs were relatively enriched in exosomes compared with mRNAs (online supplemental figure S1A). Besides, given that ncRNAs with higher expression levels are more readily detectable, we focused on lncRNAs and circRNAs that were upregulated in GC for further analysis. To identify candidate lncRNAs, we intersected the significantly upregulated lncRNAs in GC serum exosomes with those highly expressed in GC tissues and the RNA-sequencing profiles from stomach adenocarcinoma available in The Cancer Genome Atlas (figure 2A,B) (Dataset).⁴¹ For circRNAs, we selected those with high expression levels in both GC serum exosomes and GC tissues (figure 2C,D), based on which we identified 173 lncRNAs and 9 circRNAs as candidate biomarkers (online supplemental data 2).

To verify these biomarkers, we extracted exosomes from serum samples of 24 GC patients and 24 healthy donors and characterised them. Nanoparticle tracking analysis revealed that the isolated exosomes had an average size of approximately 141 nm, consistent with previous reports (online supplemental figure S1B). Additionally, we confirmed the morphological characteristics and membrane proteins of the exosomes (online supplemental figure S1C,D). GC tissues and adjacent normal tissues from these 24 GC patients were also included for analysis. Quantitative reverse transcription PCR (qRT-PCR) assays were performed on RNAs derived from both serum exosome samples and tissue samples. The primers used for candidate biomarkers were listed in online supplemental data 3. Overall, the results showed that 27 lncRNAs and 4 circRNAs were significantly upregulated in both types of samples and defined as potential biomarkers (online supplemental tables S2 and S3).

Construction and validation of combined diagnostic model

To assess the diagnostic performance of the identified ncRNAs, we retrospectively recruited 518 GC patients and 460 healthy donors from Sun Yat-sen University Cancer Center (SYSUCC) to constitute the exploration cohort. The expression levels of

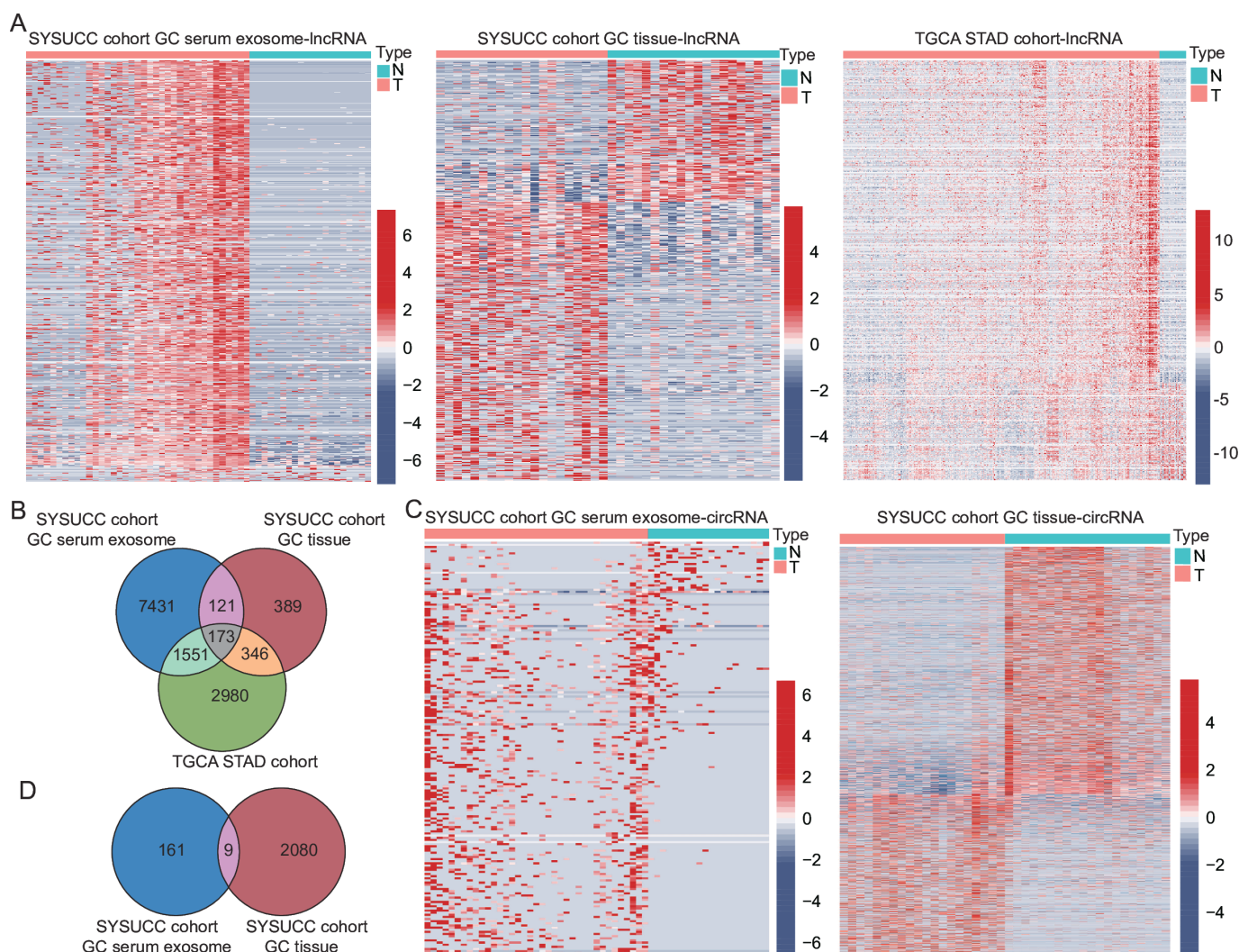


Figure 2 Screening of exosome-derived ncRNAs. (A) Analysis of lncRNAs in gastric cancer patients compared with normal controls using transcriptome data from serum exosomes, tissues and the TCGA STAD cohort. Criteria for selection include $p < 0.05$ and $\log_2 \text{FoldChange} > 1$. (B) Venn diagram depicting the overlap of upregulated lncRNAs across serum exosomes, tissues and the TCGA STAD cohort. (C) Analysis of circRNAs in gastric cancer patients compared with normal controls using transcriptome data from serum exosomes and tissues. Criteria for selection include $p < 0.05$ and $\log_2 \text{FoldChange} > 1$. (D) Venn diagram showing the overlap of upregulated circRNAs in serum exosomes and tissues. Red and blue in A and C represent high and low levels of relative expression of non-coding RNAs. circRNA, circular RNA; GC, gastric cancer; lncRNA, long non-coding RNA; N, non-cancer control; STAD, stomach adenocarcinoma; SYSUCC, Sun Yat-sen University Cancer Center; T, tumour; TCGA, The Cancer Genome Atlas.

the potential biomarkers in serum exosomes from the exploration cohort were quantified by qRT-PCR and 70% of the exploration data were randomly drawn for training. First, multiple common classifier algorithms, including LASSO-logistic regression, logistic regression, XGBoost, random forest, KMN, SVM and Naïve Bayes, were compared when used in constructing the diagnostic model. The result of 100-fold cross validation delineated that LASSO-logistic regression showed the best performance (online supplemental figure S2A). Subsequently, logistic regression analyses of single ncRNA adjusted by age and gender in the training cohort were conducted to screen out the independent predictors, thus avoiding the possible impact caused by the distribution imbalance of clinical characteristics (online supplemental table S4). The performance of the independent predictors in distinguishing GC patients from healthy donors was assessed in the training cohort (online supplemental figure S2B) and 10 ncRNAs with the highest average area under curve (AUC) values were inputted into the LASSO algorithm for dimensionality reduction (online supplemental

figure S2C,D). Finally, with consideration of achieving acceptable performance through a minimum number of ncRNAs, the combined diagnostic model was constructed based on the expression of four lncRNAs (RP11.443C10.1, CTD-2339L15.3, LINC00567 and DGCR9) in the training cohort and logistic regression algorithm, and the combined diagnostic model score (cd-score) was represented using the following formulas: $\text{cd-score} = \text{logit} (0.15527574 \times \text{RP11.443C10.1} + 0.12944201 \times \text{CTD.2339L15.3} + 0.03966617 \times \text{LINC00567} + 0.08059129 \times \text{DGCR9} - 0.19402657)$. Additionally, ROC curve analysis suggested that each of these four lncRNAs performed well in distinguishing GC patients from healthy donors, with AUC values exceeding 0.88 in both the training and testing cohorts (online supplemental figure S2E, online supplemental table S5).

The efficacy of the cd-score was evaluated using ROC curves, and the performance metrics, including AUC, accuracy, sensitivity and specificity, were calculated. The results showed that the cd-score demonstrated high diagnostic performance with an AUC of 0.959 in the training cohort, which was superior to

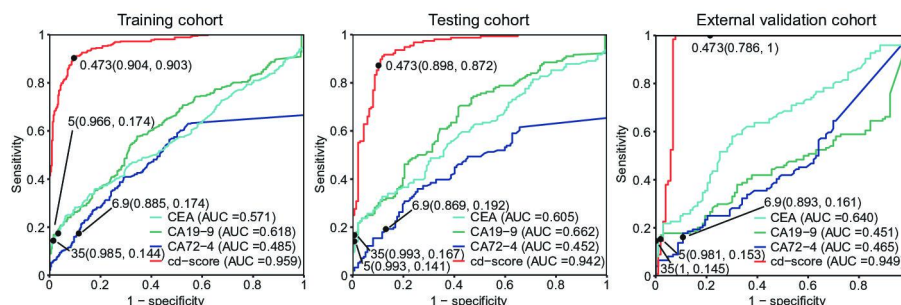


Figure 3 ROC curves for determining diagnostic performance. The ROC curves of the cd-score, CEA, CA19-9 and CA72-4 in the training, testing and external validation cohorts. CA19-9, carbohydrate antigen 19-9; CA72-4, carbohydrate antigen 72-4; cd-score, combined diagnostic model score; CEA, carcinoembryonic antigen; ROC, receiver operating characteristic.

that of single ncRNA. Using the best cut-off of 0.473, which was determined by the maximal Youden index (Youden index=sensitivity+specificity–1), the accuracy was calculated to be 0.904, sensitivity was 0.903, and specificity was 0.904 (figure 3, table 1). In comparison, the performance of traditional biomarkers related to the digestive system—CEA, CA19-9 and CA72-4—were also assessed, and comparative analysis showed that the cd-score significantly outperformed these traditional biomarkers (CEA: AUC=0.571, accuracy=0.547, sensitivity=0.174, specificity=0.966; CA19-9: AUC=0.618, accuracy=0.540, sensitivity=0.144, specificity=0.985; CA72-4: AUC=0.485, accuracy=0.509, sensitivity=0.174, specificity=0.885; figure 3, table 1). Moreover, these findings were successfully validated in the testing cohort (figure 3, table 1).

To evaluate whether the cd-score was specific, univariate and multivariate linear regression analyses were conducted using the GC samples or the health samples in the exploration cohort (online supplemental table S6). The results showed that none of clinical characteristics were identified as prediction factors for the cd-score in the multivariate model. Furthermore, the cd-score was confirmed to be independent of clinical characteristics when used for judging the sample types (OR: 69.28 and $p<0.001$ in the multivariable model, online supplemental table S7). These findings support the use of the cd-score for diagnosing GC and underscore its potential for clinical application.

To further validate the generalisability of the cd-score, we included an external cohort from the Gastrointestinal and Anal Hospital of Guangdong Province, which comprised 124 GC patients and 103 healthy donors. The results demonstrated that the cd-score achieved an AUC of 0.949, with an accuracy

of 0.869, a sensitivity of 1 and a specificity of 0.786 (figure 3; table 1), while traditional biomarkers such as CEA, CA19-9 and CA72-4 displayed significantly lower AUC values, accuracies, sensitivities and specificities (figure 3; table 1). Additionally, the cd-score was also confirmed to be independent of clinical characteristics in the external validation cohort via multivariate linear regression analyses (online supplemental table S8). Collectively, these results reinforce the earlier findings and highlight the potential clinical utility of the cd-score for GC diagnosis.

Practicality of the combined diagnostic model

It has been reported that 20%–40% of GC patients may not have detectable levels of traditional biomarkers such as CEA, CA19-9 and CA72-4 due to the Lewis a[–]b[–] genotype, which can lead to missed diagnoses in these patients, as the traditional biomarkers may be absent or at levels below detectable thresholds.¹¹ To evaluate the diagnostic efficacy of the cd-score in such cases, we assessed GC patients negative for CEA, CA19-9 and CA72-4 from each cohort. Our results showed that the cd-score demonstrated excellent diagnostic performance with an AUC of 0.960, accuracy of 0.902, sensitivity of 0.899 and specificity of 0.904 in the training cohort for distinguishing GC patients with negative traditional biomarkers from healthy donors (figure 4A; online supplemental table S9). This high diagnostic performance was consistent in both the testing cohort and the external validation cohort (figure 4A; online supplemental table S9). These results highlight the potential robustness of the cd-score in identifying GC patients who are negative for conventional biomarkers and indicate the potential clinical efficacy of our proposed cd-score.

Table 1 Performance of cd-score, CEA, CA19-9 and CA72-4 in differentiating patients with gastric cancer from healthy donors

Cohort	Biomarker	Cut-off	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1
Training cohort	cd-score	0.473	0.904 (0.879–0.925)	0.903 (0.873–0.934)	0.904 (0.872–0.936)	0.908
	CEA	5	0.547 (0.509–0.585)	0.174 (0.135–0.213)	0.966 (0.946–0.986)	0.289
	CA19-9	35	0.540 (0.502–0.578)	0.144 (0.108–0.180)	0.985 (0.971–0.998)	0.248
	CA72-4	6.9	0.509 (0.471–0.548)	0.174 (0.135–0.213)	0.885 (0.851–0.920)	0.273
Testing cohort	cd-score	0.473	0.884 (0.842–0.918)	0.872 (0.819–0.924)	0.898 (0.847–0.949)	0.889
	CEA	5	0.539 (0.480–0.597)	0.141 (0.086–0.196)	0.993 (0.978–1.000)	0.246
	CA19-9	35	0.553 (0.494–0.611)	0.167 (0.108–0.225)	0.993 (0.978–1.000)	0.284
	CA72-4	6.9	0.509 (0.450–0.567)	0.192 (0.130–0.254)	0.869 (0.812–0.925)	0.294
External validation cohort	cd-score	0.473	0.903 (0.857–0.938)	1 (1–1)	0.786 (0.707–0.866)	0.919
	CEA	5	0.529 (0.461–0.595)	0.153 (0.090–0.217)	0.981 (0.954–1.000)	0.262
	CA19-9	35	0.533 (0.466–0.599)	0.145 (0.083–0.207)	1 (1–1)	0.254
	CA72-4	6.9	0.493 (0.427–0.560)	0.161 (0.097–0.226)	0.893 (0.834–0.953)	0.258

CA19-9, carbohydrate antigen 19-9; CA72-4, carbohydrate antigen 72-4; cd-score, score of combined diagnostic model; CEA, carcinoembryonic antigen; CI, confidence interval.

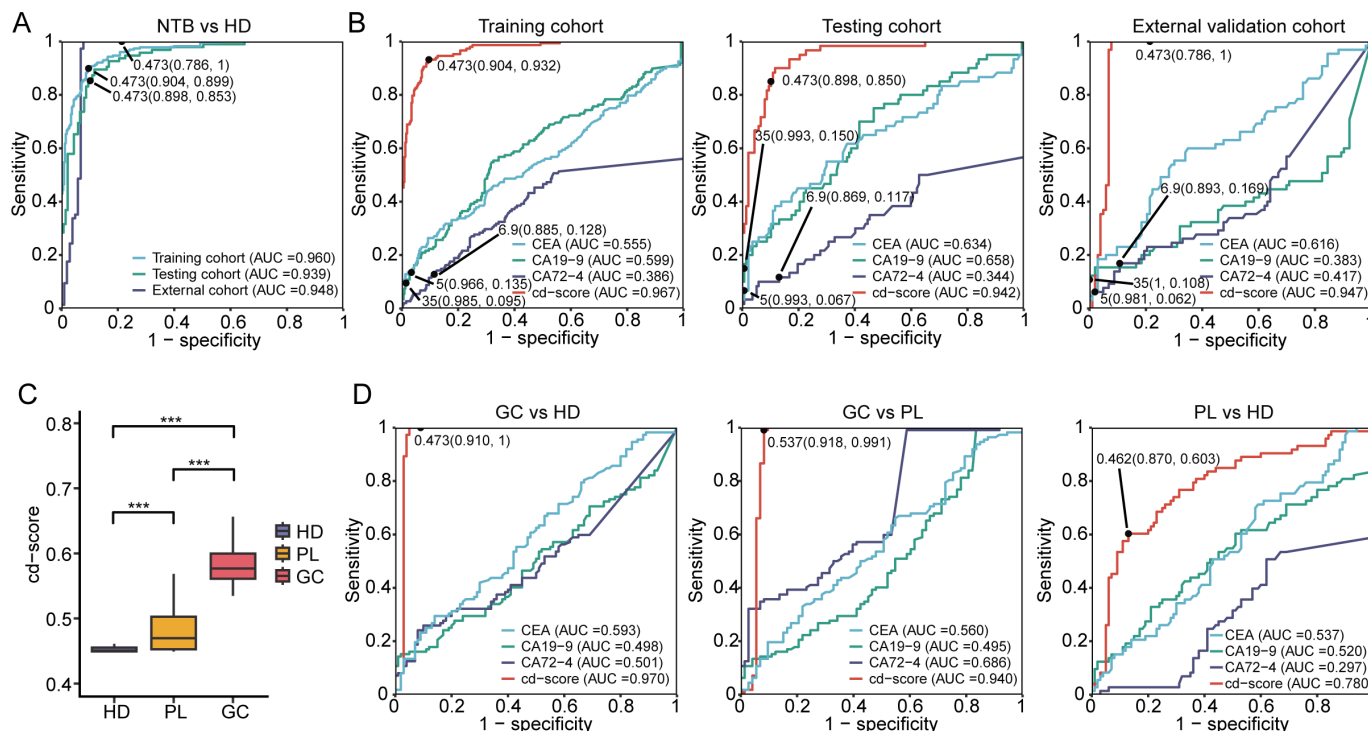


Figure 4 Application of the cd-score. (A) Diagnostic performance of the cd-score for detecting GC patients with negative CEA, CA19-9 and CA72-4 in the training, testing and external validation cohorts. (B) Performance of the cd-score in distinguishing early-stage GC patients from healthy donors compared with CEA, CA19-9 and CA72-4. (C) The cd-score values of the different groups in the prediction cohort. (D) ROC curves for the cd-score, CEA, CA19-9 and CA72-4 in differentiating between any two groups in the prediction cohort. CA19-9, carbohydrate antigen 19-9; CA72-4, carbohydrate antigen 72-4; cd-score, combined diagnostic model score; CEA, carcinoembryonic antigen; GC, gastric cancer; NTB, negative for three biomarkers; ROC, receiver operating characteristic.

Considering that an efficient biomarker should also be capable of identifying patients with EGC (pTNM stage I/II),⁴² we further assessed the diagnostic efficacy of the cd-score in this group. Compared with CEA, CA19-9 and CA72-4, the cd-score exhibited superior AUCs for distinguishing EGC patients from healthy donors, with values of 0.967, 0.942 and 0.947 in the training, testing and external validation cohorts, respectively (figure 4B). Specifically, in the training cohort, the cd-score achieved an accuracy of 0.913, a sensitivity of 0.932 and a specificity of 0.904 (table 2). This performance was similarly validated in the testing cohort and the external validation cohort, with accuracies of 0.883 and 0.869, sensitivities of 0.850 and 1 and specificities of

0.868 and 0.786, respectively (table 2). These findings indicate that the cd-score is highly effective in identifying EGC patients and holds great potential for early screening applications.

Given that GC, particularly the intestinal type, progresses through a series of precancerous lesions, identifying these lesions is essential for effective prevention and management.⁴³ To address this, we included 112 GC patients, 73 patients with gastric precancerous lesions and 100 healthy donors from SYSUCC to form a prediction cohort. The results showed that the cd-score demonstrated a significant increase in a stepwise manner, with higher scores observed in patients with precancerous lesions compared with healthy donors, and the highest scores in patients

Table 2 Performance of cd-score, CEA, CA19-9 and CA72-4 in distinguishing patients with early-stage gastric cancer from healthy donors

Cohort	Biomarker	Cut-off	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1
Training cohort	cd-score	0.473	0.913 (0.884–0.937)	0.932 (0.892–0.973)	0.904 (0.872–0.936)	0.871
	CEA	5	0.705 (0.661–0.746)	0.135 (0.080–0.190)	0.966 (0.946–0.986)	0.223
	CA19-9	35	0.705 (0.661–0.746)	0.095 (0.047–0.142)	0.985 (0.971–0.998)	0.168
	CA72-4	6.9	0.648 (0.603–0.691)	0.128 (0.074–0.182)	0.885 (0.851–0.920)	0.186
Testing cohort	cd-score	0.473	0.883 (0.830–0.925)	0.850 (0.760–0.940)	0.868 (0.847–0.949)	0.816
	CEA	5	0.711 (0.642–0.773)	0.067 (0.004–0.130)	0.993 (0.978–1.000)	0.123
	CA19-9	35	0.736 (0.669–0.796)	0.150 (0.060–0.240)	0.993 (0.978–1.000)	0.257
	CA72-4	6.9	0.640 (0.568–0.707)	0.117 (0.035–0.198)	0.869 (0.812–0.925)	0.165
External validation cohort	cd-score	0.473	0.869 (0.808–0.916)	1 (1–1)	0.786 (0.707–0.866)	0.855
	CEA	5	0.625 (0.547–0.698)	0.062 (0.003–0.120)	0.981 (0.954–1.000)	0.113
	CA19-9	35	0.655 (0.578–0.726)	0.108 (0.032–0.183)	1 (1–1)	0.194
	CA72-4	6.9	0.613 (0.535–0.687)	0.169 (0.078–0.260)	0.893 (0.834–0.953)	0.253

CA19-9, carbohydrate antigen 19-9; CA72-4, carbohydrate antigen 72-4; cd-score, score of combined diagnostic model; CEA, carcinoembryonic antigen; CI, confidence interval.

with GC (figure 4C). Furthermore, performance analysis of the cd-score in the prediction cohort, using ROC curves, showed high diagnostic performance with an AUC of 0.970 in distinguishing GC patients from healthy donors, which is consistent with our previous result (figure 4D). Additionally, in the ROC curve of the cd-score for differentiating between GC patients and patients with precancerous lesions, the cut-off could be optimised as 0.537 according to the maximal Youden index. In this instance, the cd-score achieved acceptable performance with an AUC of 0.940, accuracy of 0.962, sensitivity of 0.991 and specificity of 0.918 (figure 4D; online supplemental table S10). Taken together, these results suggest that the cd-score could be valuable for diagnosing gastric precancerous lesions and holds promise for broader clinical application.

Knockdown of DGCR9 inhibits GC cell proliferation in vitro and in vivo

Moreover, we evaluated the expression levels of the four lncRNAs included in the combined diagnostic model in the prediction cohort. The results revealed that DGCR9 was the only lncRNA that showed a significant increase in expression from healthy donors to patients with precancerous lesions and then to GC patients (figure 5A; online supplemental figure S3A). Based on these findings, we focused on investigating the role of DGCR9 in GC.

DGCR9 has been shown to be dysregulated in various cancers, and its detection in liquid biopsies holds promise for non-invasive early diagnosis of malignant tumours.⁴⁴ In our preliminary investigations, we also observed that DGCR9 was upregulated in GC tissues (online supplemental table S3, online supplemental figure S3B). To investigate the functional role of DGCR9 in GC, we used short hairpin RNA to knock down DGCR9 in GC cells, including HGC27 and MKN74 cells. This intervention significantly inhibited cell proliferation and colony formation in vitro (figure 5B–D; online supplemental figure S3C–E), consistent with previous reports.⁴⁵ Metabolic reprogramming of tumour cells was essential to meet the energy requirement during the abnormal proliferation⁴⁶ and it has been reported that knockdown of DGCR9 in GC cells suppressed the glucose uptake.³⁸ Thus, we first analysed the correlation of DGCR9 and glycolysis-related genes in our transcriptome data of tissues and calculated the Pearson correlation coefficients. The result exhibited that the expression of glycolysis-related genes was positively related to the expression of DGCR9 (online supplemental figure S3F). Notably, *GPI*, *GAPDH*, *PGK1*, *PKM*, *LDHA* and *PFKFB3* showed a significant correlation with DGCR9. Then, we determined the glycolytic flux of GC cells with or without knockdown of DGCR9 via detecting the glycolytic proton efflux rate (glycoPER). The results delineated that the glycoPER of GC cells was remarkably inhibited after knocking down DGCR9 (figure 5E, online supplemental figure S3G). In general, our experiments demonstrated that DGCR9 played a vital role in the growth of GC in vitro.

The exosome-packaged DGCR9 has been verified to be an effective biomarker for diagnosing GC and its functionality has also aroused our interest. First, we determined that DGCR9 could be encapsulated by exosomes and secreted out of GC cells since the expression of exosome-packaged DGCR9 was decreased after knocking down the regulatory enzyme of exosome secretion, Rab27a (online supplemental figure S3H).⁴⁷ Subsequently, it could be found that knockdown of DGCR9 in GC cells could also reduce the expression of DGCR9 in exosomes (online supplemental figure S3I). Moreover, we collected the exosomes

in culture medium from GC cells without or with depletion of DGCR9 and labelled them as DGCR-rich or DGCR9-deficient exosomes.⁴⁸ The results showed that in comparison with DGCR9-deficient exosomes, the incubation with DGCR9-rich exosomes showed a greater advantage in promoting cell growth through replenishing DGCR9 (online supplemental figure S3J,K). In conclusion, exosome-packaged DGCR9 could expand the DGCR9 pool in cells, thus accelerating cell proliferation.

To further investigate the effect of DGCR in the GC growth in vivo, we constructed the xenograft models by subcutaneously injecting the DGCR9-knockdown cells or control cells into the dorsal flanks of severe combined immunodeficiency mice. The results further supported our previous experimental findings in vitro, showing that DGCR9 knockdown significantly reduced tumour growth in vivo (figure 5F, online supplemental figure S3L). Moreover, antisense oligonucleotides (ASOs) have been reported to show great potential as therapeutics since they could bind to RNA and thereby modulate the expression of target RNA.^{49 50} Therefore, we synthesised ASO which targeted the lncRNA DGCR9 and preliminarily demonstrated that intervention with this ASO could inhibit the growth of HGC27 cells through in-vitro experiments (online supplemental figure S3M–P). Subsequently, we established patient-derived tumour xenograft (PDX) models from tumour tissues of two GC patients at SYSUCC and explored the therapeutic effect of ASO in vivo. The outcome indicated that DGCR9 depletion using an in vivo-optimised DGCR9 inhibitor led to a dramatic reduction in tumour growth (figure 5G, online supplemental figure S3Q), implicating DGCR9 as a potential therapeutic target in GC. Additionally, histological analysis of PDX sections via H&E staining and immunohistochemistry (IHC) demonstrated that DGCR9 knockdown led to reduced levels of the cell proliferation marker Ki67, while apoptosis was increased as indicated by the TUNEL assay (figure 5H,I). Finally, we assessed DGCR9 expression in serum exosomes from a cohort of 46 GC patients at SYSUCC and classified them into a high or low DGCR9 expression group based on their median value. Tumour tissue sections from these patients were manufactured into GC chips, followed by H&E and IHC staining. Consistent with our earlier findings, the low-DGCR9 expression group showed reduced cell proliferation and increased apoptosis (figure 5J,K). Collectively, our data underscored the pivotal role of DGCR9 in regulating tumour growth in GC and suggested its potential as a target for therapeutic intervention.

DISCUSSION

Recent studies have highlighted the potential of liquid biopsy in cancer detection and prognosis evaluation, emphasising the role of ncRNAs, such as lncRNAs and circRNAs, in cancer diagnosis and prognosis prediction.^{23 34 51–53} Notably, lncRNA PCA3, a urinary biomarker, has been approved by the US Food and Drug Administration for prostate cancer diagnosis.⁵⁴ Similarly, the diagnostic potential of lncRNAs and circRNAs for GC has been supported by various studies.^{21 55} However, the application of these markers has been limited due to the high levels of ribonuclease activity in blood, which can degrade RNA.² To overcome this challenge, our study focused on ncRNAs derived from serum exosomes, which offer increased stability in blood circulation. Our sequencing data revealed that lncRNAs and circRNAs are significantly enriched in serum exosomes from GC patients (online supplemental figure S1A). Therefore, both types of ncRNAs were included in our investigation. Although the final combined diagnostic model did not incorporate circRNAs, our

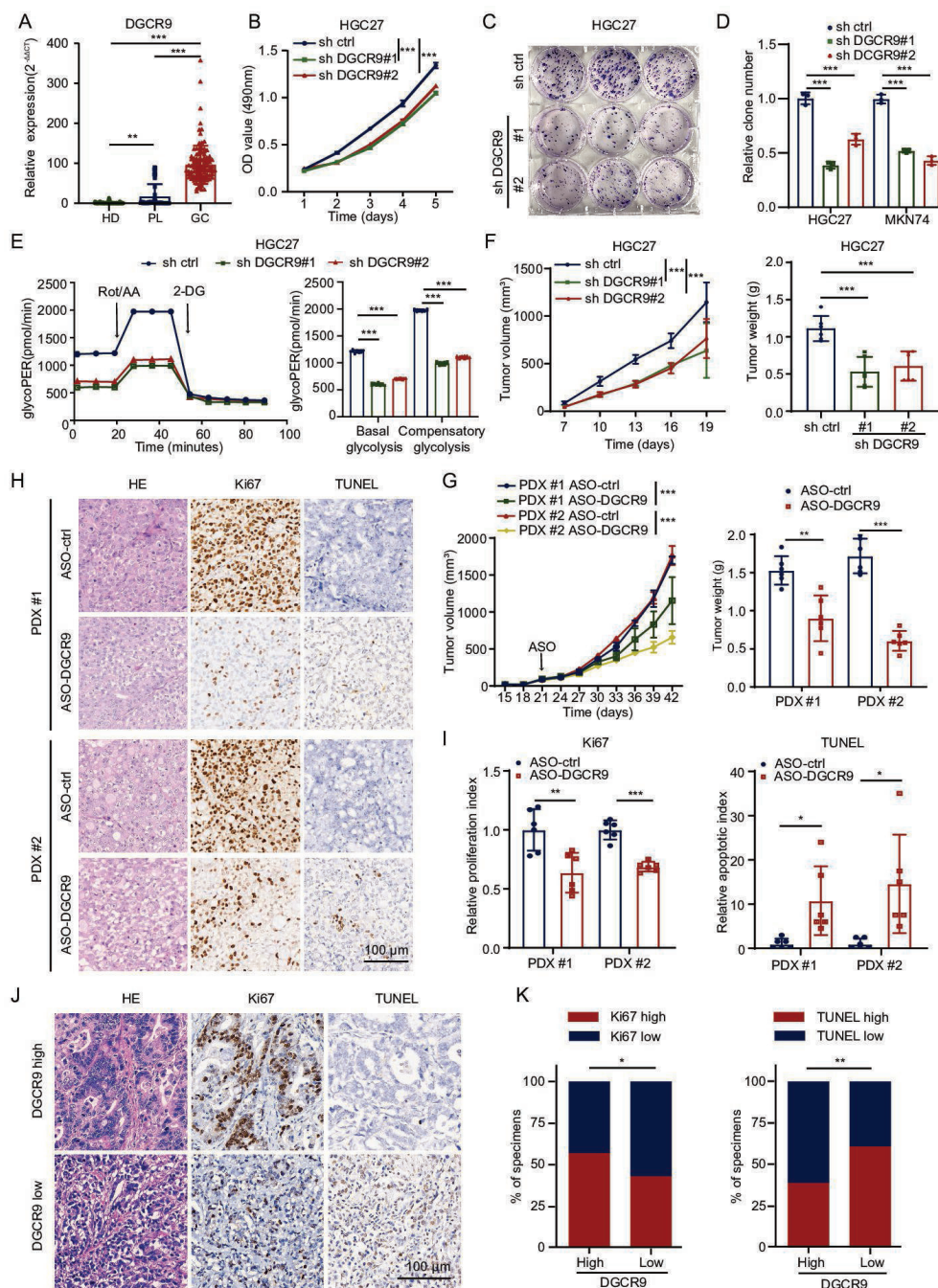


Figure 5 Impact of DGCR9 knockdown on GC cell proliferation. (A) Expression levels of DGCR9 in exosomes from GC patients (n=112), those with precancerous lesions (n=73) and healthy donors (n=100). (B) MTS assay assessing the effect of DGCR9 knockdown on HGC27 cell growth. (C) Clonogenic assay showing colony formation in DGCR9-knockdown HGC27 cells compared with control cells. (D) Statistical analysis of the clonogenic assays on HGC27 and MKN74 cells. (E) Glycolytic proton efflux rate (glycoPER) changes in DGCR9-knockdown HGC27 cells compared with control cells and statistical analysis. (F) Growth curves and tumour weights of DGCR9-knockdown HGC27 cells versus control cells in subcutaneous tissues of SCID mice (NCG mice). (G) Growth curve and tumour weights of PDX models in SCID mice (NCG mice) that were subcutaneously implanted with tumour tissues from two GC patients and intratumorally injected with scrambled or DGCR9 inhibitors (5 nmol per injection) every 3 days for five times. (H) Representative H&E staining and IHC images (Ki67 and TUNEL) of PDX tissue sections from each group. Scale bar: 100 μ m. (I) Quantification of Ki67 and TUNEL expression in PDX tissues treated with scrambled or DGCR9 inhibitors. (J) Representative H&E and IHC images of Ki67 and TUNEL in GC tissues from patients with high (n=23) or low (n=23) serum exosome DGCR9 levels. Scale bar: 100 μ m. (K) Percentage of specimens with high or low Ki67 and TUNEL index in high versus low DGCR9 groups. The numbers of biological replicates were three (B–D), four (E) and six (F–I), respectively. Data in A–B, D–G and I were presented as mean \pm SD. P values were determined by one-way ANOVA (A, D, E, tumour weight in F), two-way ANOVA (B, tumour volume in F–G), two-tailed unpaired Student's t-test (tumour weight in G, I) and χ^2 test (K). ASO, antisense oligonucleotide; DGCR9, DiGeorge Critical Region 9; GC, gastric cancer; GlycoPER, glycolytic proton efflux rate; HD, healthy donor; IHC, immunohistochemistry; Ki67, a proliferation marker; MTS, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; PDX, patient-derived xenograft; PL, precancerous lesion; Rot/AA, rotenone and antimycin A; SCID, severe combined immunodeficiency; TUNEL, terminal deoxynucleotidyl transferase dUTP nick end labelling; 2-DG, 2-deoxy-D-glucose; *p<0.05; **p<0.01; ***p<0.001.

analysis demonstrated that each circRNA individually showed promising performance in GC detection, with AUC values around 0.8 in the training and testing cohort. The combined diagnostic model, incorporating lncRNAs, achieved exceptional performance with AUCs exceeding 0.94 in the training and testing cohorts. Our model, based on comprehensive transcriptome features in serum exosomes of GC and with overall consideration of lncRNAs and circRNAs, demonstrates the potential of exosome-derived ncRNAs as effective diagnostic tools, with the cd-score proving particularly effective in distinguishing GC patients from healthy individuals.

Accumulating evidence highlights the critical role of precancerous lesions in the stepwise progression to GC.⁵⁶ Patients with precancerous lesions often present with symptoms similar to those with early GC, such as stomachache, loss of appetite, nausea and vomiting. Early detection and differentiation of these lesions from GC are essential for timely and effective treatment, which is vital for the prevention and management of GC. Recent studies have identified plasma metabolomic signatures in precancerous gastric lesions that progress to GC,⁵⁷ and a plasma lipid profile consisting of 11 lipids has been shown to differentiate accurately between GC and gastric precancerous lesions.⁵⁸ Our proposed cd-score demonstrated promising clinical use, thereby broadening its potential applications. Additionally, DGCR9 expression showed a significant increase from healthy donors to patients with precancerous lesions, with a further notable rise in GC patients, indicating that this lncRNA may play an essential role in the initiation and progression of malignancy.

Consistent with the previous report that DGCR9 detection in body fluids is expected to provide a non-invasive diagnosis method for malignant tumours, our study has demonstrated that DGCR9 could act as a biomarker for GC diagnosis. Additionally, Chao Ni *et al* have revealed that DGCR9 is overexpressed in GC pathological tissues and exogenous silencing of DGCR9 by siRNA could inhibit the proliferation as well as suppressing the extracellular glucose uptake of GC cells.⁴⁵ In this regard, our study has supported the opinion of Chao Ni *et al* via a series of powerful experimental evidence, including the cell-based xenograft model, PDX model and IHC staining in GC chips. Furthermore, we have refreshed the knowledge that DGCR9 hold promise for an intervention target since antagonism of DGCR9 via ASO in vivo could suppress the growth of GC. Although our findings could be preliminary, they provide a foundation for further exploration of DGCR9 in GC, and continued research in this area could lead to novel strategies for intervention and treatment of GC.

Despite the interesting findings reported, this study had several limitations that should be clarified. First, while the cd-score showed substantial promise as a non-invasive diagnostic tool for GC and could offer significant clinical value before conducting invasive endoscopic examination, our study relied solely on retrospective cohorts. Therefore, prospective validation in larger populations remains essential. Second, the extraction method of exosomes in this study could not rule out the interference of small size ectosomes. It should be acknowledged that the term 'exosome' in our study was actually small extracellular vesicles with diameters of 40~160 nm and confirmed by the markers of CD9, CD63 and TSG101.

In conclusion, we identified and developed a novel exosome ncRNA feature for the non-invasive detection of GC using a multi-step procedure, in which the feature was validated across several cohorts and demonstrated high efficacy in distinguishing GC patients from healthy individuals. The feature also showed strong performance in detecting patients with

negative gastrointestinal tumour-related biomarkers, those with EGC and those with precancerous lesions. The combination of serum exosome-derived ncRNAs holds promise for enhancing early GC screening and may contribute to reducing mortality rates.

Author affiliations

¹State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China

²Department of Medical Biochemistry and Molecular Biology, School of Medicine, Jinan University, Guangzhou, People's Republic of China

³Guangdong Institute of Gastroenterology, The Sixth Affiliated Hospital of Sun Yat-Sen University, Sun Yat-Sen University, Guangzhou, People's Republic of China

⁴Department of Clinical Oncology, Shenzhen Key Laboratory for Cancer Metastasis and Personalized Therapy, The University of Hong Kong-Shenzhen Hospital, Shenzhen, People's Republic of China

Contributors H-QJ designed the study. Z-RC, Y-QZ, YH, M-YM, JL, J-BZ, L-PY, P-SH and R-HX collected the data. Z-RC, Y-QZ, YH, Y-JW and Z-XL analysed and interpreted the data. Z-RC, Y-QZ and H-QJ performed the statistical analysis. Z-RC, Y-QZ and H-QJ wrote the manuscript. Z-RC, TT, LZ and H-QJ contributed to discussion and data interpretation. Z-RC, Y-QZ, LZ, R-HX and H-QJ revised the manuscript. All authors reviewed the manuscript and approved the final version, and H-QJ is the guarantor. We have used the LASSO-logistic regression, one of the machine learning methods, for the construction of diagnostic model.

Funding This project was supported by the National Key R&D Program of China (2023YFC3402100), National Natural Science Foundation of China (82372920, 82203732), Huizhou and Hong Kong/Macau Science and Technology Cooperation Project (2024E0050021) and Young Talents Program of Sun Yat-sen University Cancer Center (YTP-SYSUCC-0019).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. All samples were collected in accordance with the approval granted by the Human Genetic Resources Management Office of China (permit number (2023) CJ0426). This study was conducted following the clinical protocol approved by the Institutional Research Ethics Committee of Sun Yat-sen University Cancer Center (SYSUCC, permit number B2022-769-01), which included exemptions from informed consent. All animal studies were performed with approval of the Institutional Ethics Committee for Clinical Research and Animal Trials of the SYSUCC (permit number L025501202212006).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The raw sequence data generated in the screening phase have been deposited in the Genome Sequence Archive (GSA-Human: HRA008261) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Yong-Qiang Zheng <http://orcid.org/0000-0003-3246-2800>

Lu-Ping Yang <http://orcid.org/0000-0003-4697-6853>

Ze-Xian Liu <http://orcid.org/0000-0001-9698-0610>

Rui-Hua Xu <http://orcid.org/0000-0001-9657-4380>

Huai-Qiang Ju <http://orcid.org/0000-0003-1713-5465>

REFERENCES

- 1 Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clinicians* 2018;68:394-424.
- 2 Quackenbush JF, Cassidy PB, Pfeffer LM, et al. Isolation of circulating microRNAs from microvesicles found in human plasma. *Methods Mol Biol* 2014;1102:641-53.
- 3 Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet* 2021;398:27-40.
- 4 Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. *Science* 2022;375:eaay9040.
- 5 Heidrich I, Pantel K. Liquid biopsy: blood-based analyses of circulating cell-free DNA in xenografts. *EMBO Mol Med* 2022;14:e16326.
- 6 van der Pol Y, Tanyto NA, Evander N, et al. Real-time analysis of the cancer genome and fragmentome from plasma and urine cell-free DNA using nanopore sequencing. *EMBO Mol Med* 2023;15:e17282.
- 7 Alix-Panabières C, Pantel K. Liquid Biopsy: From Discovery to Clinical Application. *Cancer Discov* 2021;11:858-73.
- 8 Johnson P, Zhou Q, Dao DY, et al. Circulating biomarkers in the diagnosis and management of hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2022;19:670-81.
- 9 Shimada H, Noie T, Ohashi M, et al. Clinical significance of serum tumor markers for gastric cancer: a systematic review of literature by the Task Force of the Japanese Gastric Cancer Association. *Gastric Cancer* 2014;17:26-33.
- 10 Xu Z, Huang Y, Hu C, et al. Efficient plasma metabolic fingerprinting as a novel tool for diagnosis and prognosis of gastric cancer: a large-scale, multicentre study. *Gut* 2023;72:2051-67.
- 11 Guo X, Peng Y, Song Q, et al. A Liquid Biopsy Signature for the Early Detection of Gastric Cancer in Patients. *Gastroenterology* 2023;165:402-13.
- 12 Izumi D, Zhu Z, Chen Y, et al. Assessment of the Diagnostic Efficiency of a Liquid Biopsy Assay for Early Detection of Gastric Cancer. *JAMA Netw Open* 2021;4:e2121129.
- 13 So JBY, Kapoor R, Zhu F, et al. Development and validation of a serum microRNA biomarker panel for detecting gastric cancer in a high-risk population. *Gut* 2021;70:829-37.
- 14 Chen Y, Wang B, Zhao Y, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun* 2024;15:1657.
- 15 Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 2018;172:393-407.
- 16 Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;10:155-9.
- 17 Huarte M. The emerging role of lncRNAs in cancer. *Nat Med* 2015;21:1253-61.
- 18 Salzman J. Circular RNA Expression: Its Potential Regulation and Function. *Trends Genet* 2016;32:309-16.
- 19 Li J, Sun D, Pu W, et al. Circular RNAs in Cancer: Biogenesis, Function, and Clinical Significance. *Trends Cancer* 2020;6:319-36.
- 20 Kristensen LS, Andersen MS, Stagsted LVW, et al. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019;20:675-91.
- 21 Zhang K, Shi H, Xi H, et al. Genome-Wide lncRNA Microarray Profiling Identifies Novel Circulating lncRNAs for Detection of Gastric Cancer. *Theranostics* 2017;7:213-27.
- 22 Wu Y, Wang Y-Q, Weng W-W, et al. A serum-circulating long noncoding RNA signature can discriminate between patients with clear cell renal cell carcinoma and healthy controls. *Oncogenesis* 2016;5:e192.
- 23 Xie M, Yu T, Jing X, et al. Exosomal circSHKBP1 promotes gastric cancer progression via regulating the miR-582-3p/HUR/VEGF axis and suppressing HSP90 degradation. *Mol Cancer* 2020;19:112.
- 24 Cai X, Janku F, Zhan Q, et al. Accessing Genetic Information with Liquid Biopsies. *Trends Genet* 2015;31:564-75.
- 25 Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science* 2020;367:eaau6977.
- 26 Hoshino A, Costa-Silva B, Shen T-L, et al. Tumour exosome integrins determine organotropic metastasis. *Nature New Biol* 2015;527:329-35.
- 27 Yu W, Hurley J, Roberts D, et al. Exosome-based liquid biopsies in cancer: opportunities and challenges. *Ann Oncol* 2021;32:466-77.
- 28 Yu D, Li Y, Wang M, et al. Exosomes as a new frontier of cancer liquid biopsy. *Mol Cancer* 2022;21:56.
- 29 Kermansy DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* 2018;172:1122-31.
- 30 Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature New Biol* 2018;555:469-74.
- 31 Greener JG, Kandathil SM, Moffat L, et al. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;23:40-55.
- 32 Deo RC. Machine Learning in Medicine. *Circulation* 2015;132:1920-30.
- 33 Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature New Biol* 2019;570:385-9.
- 34 Nakamura K, Zhu Z, Roy S, et al. An Exosome-based Transcriptomic Signature for Noninvasive, Early Detection of Patients With Pancreatic Ductal Adenocarcinoma: A Multicenter Cohort Study. *Gastroenterology* 2022;163:1252-66.
- 35 Song Z, Zou S, Zhou W, et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nat Commun* 2020;11:4294.
- 36 Dong D, Fang M-J, Tang L, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol* 2020;31:912-20.
- 37 Evans JA, Chandrasekhara V, ASGE Standards of Practice Committee. The role of endoscopy in the management of premalignant and malignant conditions of the stomach. *Gastrointest Endosc* 2015;82:1-8.
- 38 Pimentel-Nunes P, Libânio D, Marcos-Pinto R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. *Endoscopy* 2019;51:365-88.
- 39 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118-27.
- 40 Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
- 41 Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nat New Biol* 2017.
- 42 Ma S, Zhou M, Xu Y, et al. Clinical application and detection techniques of liquid biopsy in gastric cancer. *Mol Cancer* 2023;22:7.
- 43 Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 1992;52:6735-40.
- 44 Xia H, Huang Z, Liu S, et al. lncRNA DiGeorge syndrome critical region gene 5: A crucial regulator in malignant tumors. *Biomed Pharmacother* 2021;141:111889.
- 45 Ni C, Yang P, Guo J, et al. Role of DiGeorge syndrome critical region gene 9, a long noncoding RNA, in gastric cancer. *Onco Targets Ther* 2018;11:2259-67.
- 46 Martinez-Reyes I, Chandel NS. Cancer metabolism: looking forward. *Nat Rev Cancer* 2021;21:669-80.
- 47 Ostrowski M, Carmo NB, Krumeich S, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat Cell Biol* 2010;12:19-30.
- 48 Fu Z, Zhang X, Zhou X, et al. In vivo self-assembled small RNAs as a new generation of RNAi therapeutics. *Cell Res* 2021;31:631-48.
- 49 Evers MM, Toonen LJA, van Roon-Mom WMC. Antisense oligonucleotides in therapy for neurodegenerative disorders. *Adv Drug Deliv Rev* 2015;87:90-103.
- 50 Diazzi S, Baeri A, Fassio J, et al. Blockade of the pro-fibrotic reaction mediated by the miR-143/-145 cluster enhances the responses to targeted therapy in melanoma. *EMBO Mol Med* 2022;14:e15295.
- 51 Ju H-Q, Zhao Q, Wang F, et al. A circRNA signature predicts postoperative recurrence in stage II/III colon cancer. *EMBO Mol Med* 2019;11:e10168.
- 52 Luo X-J, Zhao Q, Liu J, et al. Novel Genetic and Epigenetic Biomarkers of Prognostic and Predictive Significance in Stage II/III Colorectal Cancer. *Mol Ther* 2021;29:587-96.
- 53 Wu X, Heng J, Li M, et al. The prognostic value of microRNA-200 family expression in digestive system tumors: A meta-analysis and validation. *MedComm - Oncol* 2024;3:e86.
- 54 Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 2016;29:452-63.
- 55 Dong L, Qi P, Xu M-D, et al. Circulating CUDR, LSINCT-5 and PTENP1 long noncoding RNAs in sera distinguish patients with gastric cancer from healthy controls. *Int J Cancer* 2015;137:1128-35.
- 56 Sugano K, Moss SF, Kuipers EJ. Gastric Intestinal Metaplasia: Real Culprit or Innocent Bystander as a Precancerous Condition for Gastric Cancer? *Gastroenterology* 2023;165:1352-66.
- 57 Huang S, Guo Y, Li Z-W, et al. Identification and Validation of Plasma Metabolomic Signatures in Precancerous Gastric Lesions That Progress to Cancer. *JAMA Netw Open* 2021;4:e2114186.
- 58 Liu Z-C, Wu W-H, Huang S, et al. Plasma lipids signify the progression of precancerous gastric lesions to gastric cancer: a prospective targeted lipidomics study. *Theranostics* 2022;12:4671-83.