BMC Medical Genomics

RESEARCH

Open Access

Circulating microRNA panels for multi-cancer detection and gastric cancer screening: leveraging a network biology approach

Leila Kamkar[1,2], Samaneh Saberi[3], Mehdi Totonchi[4,5*] and Kaveh Kavousi[1*]

Abstract

Background Screening tests, particularly liquid biopsy with circulating miRNAs, hold significant potential for non-invasive cancer detection before symptoms manifest.

Methods This study aimed to identify biomarkers with high sensitivity and specificity for multiple and specific cancer screening. 972 Serum miRNA profiles were compared across thirteen cancer types and healthy individuals using weighted miRNA co-expression network analysis. To prioritize miRNAs, module membership measure and miRNA trait significance were employed. Subsequently, for specific cancer screening, gastric cancer was focused on, using a similar strategy and a further step of preservation analysis. Machine learning techniques were then applied to evaluate two distinct miRNA panels: one for multi-cancer screening and another for gastric cancer classification.

Results The first panel (hsa-miR-8073, hsa-miR-614, hsa-miR-548ah-5p, hsa-miR-1258) achieved 96.1% accuracy, 96% specificity, and 98.6% sensitivity in multi-cancer screening. The second panel (hsa-miR-1228-5p, hsa-miR-1343-3p, hsa-miR-6765-5p, hsa-miR-6787-5p) showed promise in detecting gastric cancer with 87% accuracy, 90% specificity, and 89% sensitivity.

Conclusions Both panels exhibit potential for patient classification in diagnostic and prognostic applications, highlighting the significance of liquid biopsy in advancing cancer screening methodologies.

Keywords Multiple cancer detection, Liquid biopsy, miRNA, Biomarker, Network biology, WGCNA, Co-expression analysis

*Correspondence:
Mehdi Totonchi
totonchimehdi@gmail.com
Kaveh Kavousi
kkavousi@ut.ac.ir
[1] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[2] School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran
[3] HPGC Research Group, Department of Medical Biotechnology, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran
[4] Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran
[5] Research Center for Gastroenterology and Liver Diseases, Research Institute For Gastroenterology & Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Introduction

Regular cancer screenings can lead to timely diagnosis, enabling effective intervention and increasing the likelihood of successful treatment. Currently approved cancer detection tests, such as endoscopy, colonoscopy, mammography, and cervical cytology, are often aggressive and costly [1]. On the other hand, while clinical signs of cancer progression are often elusive in the early years, underlying molecular changes in the body are still occurring [2, 3]. Hence, exploring non-invasive molecular biomarkers within a broader cancer diagnostic window emerges as a promising strategy. These biomarker panels can be beneficial in evaluating high-risk individuals and serve as the basis for establishing follow-up strategies. It is crucial that they demonstrate high specificity and accuracy to minimize unnecessary follow-up procedures and reduce anxiety.

Multi-cancer screening is challenging due to the diverse nature of cancers in terms of genetic and molecular characteristics. Liquid biopsy is one of the most promising options in noninvasive cancer detection. The term "liquid biopsy" did not appear in PubMed until 2011 [4]. Since then, liquid biopsy, involving the extraction of non-solid biological material from sources like blood, saliva, urine, cerebrospinal fluid, breast milk, and semen, has consistently grown as a research field [5, 6]. These samples are obtained from the body through non-invasive or minimally invasive methods.

Blood is simply reachable and contains a diverse set of analytes, which can be used for medical purposes such as early diagnosis, treatment monitoring, tumor stage determination, disease recurrence risk assessment, and prognosis [7]. Several studies have dedicated efforts to classify cancers using a combination of blood analytes such as proteins, cell-free DNAs, noncoding RNAs, circulating tumor cells (CTCs), and exosome molecular content [8, 9]. For instance, the CancerSEEK project aimed to diagnose and locate operable cancers (ovarian, liver, stomach, pancreas, intestine, lung, and breast) using a multi-element blood test. They classified different cancer samples with an accuracy of between 69 and 98% and were able to determine healthy samples with a high specificity of 99% [1]. In a recent investigation, a platform(I-Biomarker) has been developed for multi-cancer diagnosis and treatment response monitoring. This platform incorporates bioinformatic pipelines for various omics, including RNA-seq, Small RNA-seq, Variants (SNP and Indels), Copy Number Variation, Single Cell RNA-seq, CHIP-Seq, Methylation, and miRNA. They employed Explainable AI to elucidate predictions based on corresponding biomarkers [10].

Transitioning to a focus on microRNAs (miRNAs) these small noncoding RNAs play crucial posttranslational regulatory roles [11]. Regarding their sequence complementarity miRNA can target the 3′-untranslated region (3′-UTR) of messenger RNA and suppress protein translation [12]. Their expression profiles were indicated to be able to classify human cancers [13]. MiRNAs act as oncogenes or tumor suppressors during tumor development and progression. Experimental evidence demonstrates that correcting specific miRNA alterations using miRNA mimics or antagomirs can normalize the gene regulatory network and signaling pathways and reverse the phenotype in cancerous cells [14]. Despite the absence of standardized protocols for the current clinical use of miRNAs, they constitute a reliable tool for future use [15].

There are different analysis methods for the classification of patient groups. The preference for network biology over previous methods, such as fold change analysis, lies in its ability to capture and analyze complex interactions within biological systems. While fold change analysis focuses on individual gene expression changes, network biology can bridge individual genes to a system-level view, which considers the relationships and dependencies among genes, miRNAs, proteins, and other molecular entities, which are then expected to be enriched in individual pathways, and processes.

Correlation networks are increasingly used in bioinformatics applications to facilitate network-based methods of identifying candidate biomarkers or therapeutic targets [16]. In gene co-expression networks, each gene corresponds to a node, and an edge connects two genes if their expression values are highly correlated. The Weighted Gene Co-Expression Network (WGCNA) assigns a weight to each link to represent the extent of gene co-expression. WGCNA has already been applied to several cancers, and promising results have been obtained [17]. After detecting clusters (modules) of highly correlated genes, this package can summarize such clusters using the module eigengene or an intramodular hub gene, correlate modules to one another or external sample traits (eigengene network methodology), and calculate module membership measures [16]. In network biology, researchers often explore the preservation of modules across multiple networks and assess how gene pathways are affected under specific conditions. Non-preserved modules can also provide valuable information, reflecting species-specific or disease-specific modules [18]. In the second part of the study, a preservation analysis is adopted to detect specific modules in gastric cancer.

Numerous previous studies have explored the expression profiles of miRNAs in various cancers, aiming to identify associated patterns, as summarized in Supplementary Table 1, Additional File 1. However, it remains unclear whether some specific miRNAs can accurately divide the statistical population into two categories, healthy people and individuals at high risk of cancer

in general. Identifying common alterations in miRNAs across multiple cancers could provide valuable insights into the shared regulatory mechanisms of these small RNA molecules.

Considering the important role of multi-cancer screening, this study, conducts an inclusive examination of serum miRNA profiles across thirteen different cancer types, to identify common biomarkers for multi-cancer screening. Employing weighted miRNA co-expression network analysis, the study successfully identifies modules and hub miRNAs that demonstrated shared characteristics across the diverse spectrum of cancers. Subsequently, we assess the capability of this strategy in detecting specific cancers, such as gastric cancer, from other cancers through a preservation analysis. The machine learning outcomes revealed two distinct miRNA panels: one showcasing robust performance in multi-cancer screening (96.1% accuracy, 96% specificity, and 98.6% sensitivity), and the other displaying promise in gastric cancer detection (87% accuracy, 90% specificity, and 89% sensitivity). With potential applications in diagnostic and prognostic contexts, these findings underscore the role of liquid biopsy in advancing cancer screening methodologies with a focus on commonalities among diverse cancer types.

Materials and methods

This study unfolds in two distinct parts. In the initial phase, researchers dedicated their efforts to accurately classify samples into two groups: non-cancer and those suffering from one of thirteen different cancers (regardless of the cancer type). Transitioning into the second phase, the study introduces an approach to pinpointing more precise cancer markers. In this phase, the focus is on identifying miRNAs specifically indicative of gastric cancer as a particular disease. Although gastric cancer ranks seventh among cancers in terms of incidence, it is the fourth leading cause of cancer-related deaths in the world [19]. This might be due to the absence of symptoms in the early stages of the disease, which makes gastric cancer detection difficult. Physicians often do not diagnose the disease until it reaches advanced stages, and at that time, treatment is much more difficult and less responsive. The analysis flow is illustrated in Fig. 1.

Datasets for training and evaluation

This study independently utilized the GSE113486 dataset for training and the GSE112264 dataset for evaluation, both sourced from the Gene Expression Omnibus [20–23]. These datasets comprise serum samples from 13 distinct cancer types.

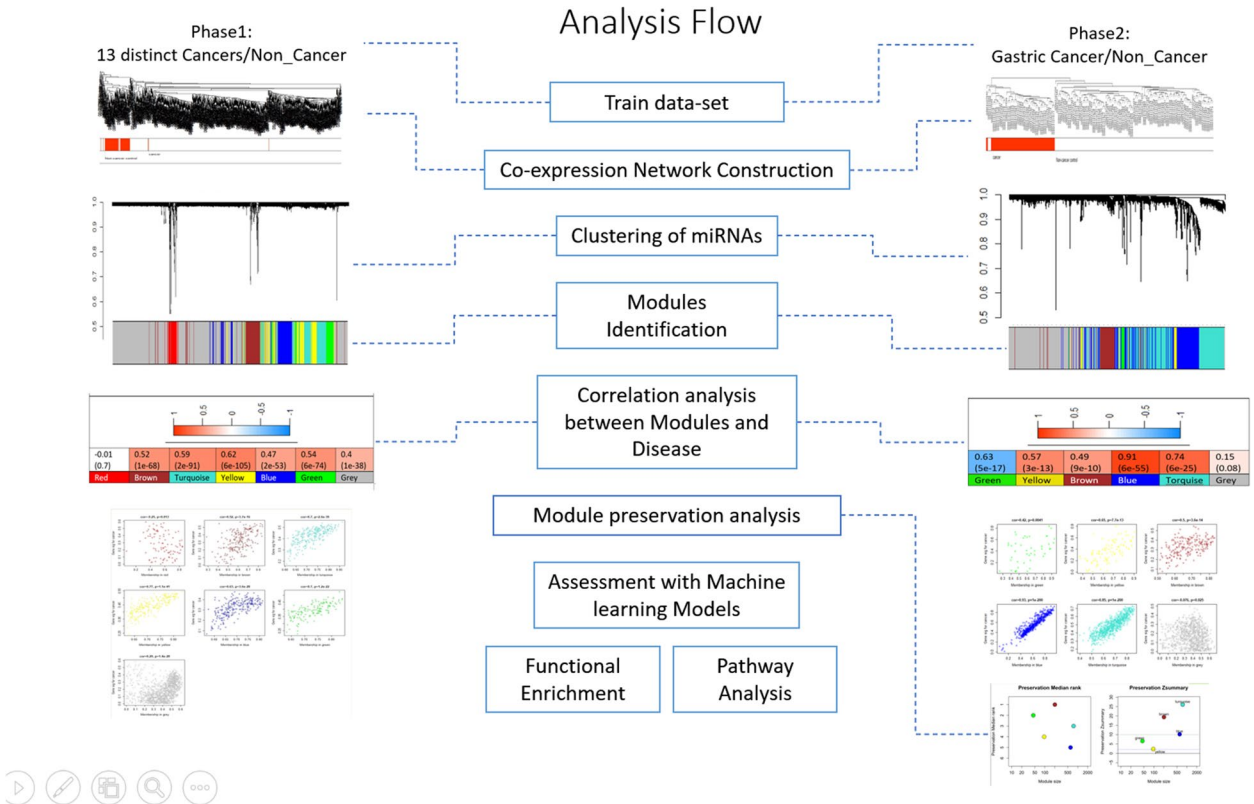

Fig. 1 Study analysis flow describing steps in multi-cancer screening and gastric cancer classification

Before conducting any additional analysis, it is imperative to normalize the data. Over the last decade, Limma has emerged as a widely used tool for microarray data analysis. This tool offers an integrated solution for the comprehensive analysis of data derived from gene expression experiments [24]. Using this package, we downloaded and normalized the data.

Multi-cancer screening

In the initial dataset, we categorized all samples into two study groups to build the network based on it. Group 1 comprised 100 non-cancer control samples, while Group 2 consisted of 872 samples representing thirteen cancer types. We constructed the network and prioritized miRNAs based on their significance across various cancers. Subsequently, we rigorously assessed the miRNA candidate list through diverse machine-learning models. Further elaboration on these steps will be provided in subsequent sections.

Co-expression network analysis

Using the WGCNA package, an unsigned weighted miRNA co-expression network was constructed by comparing the miRNA expression profiles of the two mentioned groups. Given that miRNAs are regulatory factors, and any change in their expression may play a critical role, an unsigned network was chosen to capture both positive and negative correlations.

In the network, the nodes represent miRNA expression profiles, and the edges denote pairwise correlations between them. The adjacency matrix, aij, is a symmetric $n \times n$ matrix with continuous values between 0 and 1. Specifically, $aij = |cor(xi,xj)|$ represents the adjacency in the unsigned network, where Pearson's correlation coefficient was used to construct the correlation matrix. Pearson correlation is appropriate for analyzing the relationship between two quantitative variables, assuming they may exhibit a positive, negative, or no linear relationship.

Next, a weighted network adjacency was defined through soft thresholding, which involves raising the absolute value of the correlations to a power greater than one. This approach emphasizes stronger correlations while downplaying weaker ones [25].

Co-expressed miRNAs were then clustered into modules, and the eigengene dendrogram of each module was generated. Following this, the correlation between the module eigengenes and cancer presence was examined using the Pearson correlation.

The filtering of miRNAs was based on two criteria: |mTS| (miRNA Trait Significance) > 0.5 and |mMM| (miRNA Module Membership) > 0.71 within their respective module. This threshold, labeled "Alpha" was applied to the network to yield a list of potential miRNA candidates.

Evaluation by machine learning

Before evaluating the chosen miRNAs through machine learning, balancing sample numbers and adjusting distribution within each group is crucial. This can be achieved by random sampling from each cancer type. Group 1 consisted of 100 non-cancer control samples, and Group 2 labeled as Cancerous, included eight randomly selected samples from each cancer type, totaling 104 cancerous samples.

We assessed the performance of eight algorithms on the candidate miRNA list to identify the most suitable algorithm for our specific purpose and dataset. The algorithms are as follows:

- CART: Classification and Regression Trees
- GLMNET: Regularized Linear Models
- KNN: K Nearest Neighbor
- LDA: Linear Discriminant Analysis
- LR: Linear Regression
- MLP: Multi-Layer Perceptron
- RF: Random Forest
- SVM: Support Vector Machine

To this end, the Classification and Regression Training (Caret) package in the R programming language was used [26, 27]. The evaluation parameters were ROC, sensitivity, and specificity.

Then, by narrowing the threshold alpha, the number of miRNAs was decreased to 25, 17, and 4, and each list was evaluated using the chosen algorithm.

Gastric cancer screening

To pinpoint specific miRNAs associated with a particular cancer type, such as gastric cancer, the second phase of the study focused on constructing a Weighted MiRNA Co-Expression Network (WMiCN) tailored to this specific cancer. The revised sample selection included Group 1 (non-cancer control with 100 samples) and Group 3 (gastric cancer with 40 samples).

Co-expression-network analysis

The construction of the unsigned WMiCN (Weighted miRNA Co-expression Network) was based on the expression profiles of samples from Group 1 and Group 3. As in the initial phase, co-expressed miRNAs were clustered into modules, and the correlation between these modules and the trait of interest—gastric cancer incidence—was explored.

Preservation analysis

In many applications, researchers seek to determine whether network module properties change across different conditions. In our case, we aim to investigate which modules in the gastric cancer network are specifically related to gastric cancer and not preserved in the network representing other cancer types. To identify these gastric cancer-specific modules, we conducted a preservation analysis by comparing circulating miRNA expression profiles in gastric cancer with those in other cancer types.

Various preservation statistics are available to measure this preservation among networks, and since each one measures different aspects, their results may not always align. Therefore, composite preservation metrics offer a more comprehensive and efficient evaluation of module preservation across networks.

The overall significance of the observed preservation statistics can be assessed using Zsummary, which is calculated as the mean of Z.density and Z.connectivity. According to Langfelder et al. [28], the following thresholds apply:

- Zsummary > 10: Strong evidence of module preservation
- 2 < Zsummary < 10: Weak to moderate evidence of preservation
- Zsummary < 2: No evidence of module preservation

It is important to note that Zsummary depends on module size, meaning that the preservation of larger modules tends to be statistically more significant. However, to treat all modules equally, regardless of size, we focused on the medianRank composite statistic. Analogous to Zsummary, medianRank is defined as the mean of medianRank.density and medianRank.connectivity. Unlike Zsummary, medianRank is not affected by module size, making it useful for comparing relative preservation among modules. Modules with lower medianRank values demonstrate stronger evidence of preservation compared to those with higher values.

We used these two statistics to identify non-preserved modules. The hub miRNAs of the non-preserved modules were then selected for further scrutiny based on the criteria $|mTS| > 0.8$ and $|mMM| > 0.71$ within their respective modules. This threshold, referred to as "Beta," was applied to the network, resulting in a refined list of 51 candidate miRNAs.

Evaluation by machine learning

To address the need for balanced sample sizes in classification problems, two groups were carefully formed to ensure an optimal distribution: Group 3, comprising 40 samples representing gastric cancer, and Group 4, consisting of 48 samples randomly selected, with four samples chosen from each of the remaining twelve cancer types.

Similar to the preceding phase, eight distinct machine learning algorithms were scrutinized, and the most effective one was chosen. Employing the selected algorithm, we assessed the suitability of the candidate list for classifying both Group 3 and Group 4. By refining the Beta threshold, we subsequently reduced the number of miRNAs first to 17 and then to 4, evaluating their ensuing performance.

In the next step, we validated both multiple and particular cancer screening models using the independent test dataset. Performance reports of both models determine the panel's ability in cancer classification.

Enrichment analysis

In the next step, functional enrichment and pathway analyses were performed for the candidate miRNA list in both study phases. To investigate the miRNAs' function and disease associations, TAM 2.0 can be used, which performs the enrichment analysis of these miRNAs in the curated miRNA sets using the statistical overrepresentation analysis [29]. The other tools used in this study were miRPathDB, miRNET, Mienturnet, and Enrichr [30–33]. Finally, we reviewed the previously published studies on the two final miRNA panels.

Results

The samples' general characteristics and distribution in both train and test datasets are presented in Tables 1 and 2, respectively. Within the first dataset, the original study focused on bladder cancer, which involved random division of data into two equal groups for training and testing (n = 486 per group). They ultimately proposed a panel of seven miRNAs (miR-6087, miR-6724-5p, miR-3960, miR-1343-5p, miR-1185–1-3p, miR-6831-5p, miR-4695-5p) for classifying bladder cancer samples against normal and other cancer samples, achieving an AUC of 0.97, sensitivity of 95%, and specificity of 87%. Descriptive variables were compared using the $\chi 2$ test, while continuous variables were compared using the independent t-test. Also, Fisher's linear discriminant analysis was used to analyze the expression of multiple miRNAs. The returned values obtained from the discriminant functions were used to define the index. A score of 'zero' or above represented bladder cancer, and scores below zero represented other cancers or clinical conditions [20].

Table 1 The general characteristics of train and test datasets

Usage	Id	Author	Platform	Age	Female	Male
Train	GSE113486	Usuba W, 2019	GPL21263	Avg = 64.38 Min: 17 Max: 93 Median: 66	37%	63%
Test	GSE112264	Urabe F, 2018	GPL21263	Avg: 65.8 Min: 10 Max: 93 Median: 67	0%	100%

Table 2 The distribution of the disease in the samples across the independent training and test datasets

No	Disease	Train	Test
1	Bladder Cancer	392	50
2	Breast Cancer	40	-
3	Biliary tract Cancer	40	50
4	Colorectal cancer	40	50
5	Esophagus cancer	40	50
6	Gastric cancer	40	50
7	Glioma	40	50
8	Hepatocellular carcinoma	40	50
9	Lung Cancer	40	50
10	Non-Cancer Control	100	41
11	Ovarian Cancer	40	-
12	Pancreatic cancer	40	50
13	Prostate Cancer	40	809
14	Sarcoma	40	50
	Total	972	1591

Using the second dataset Urabe et al. investigated the potential of serum miRNAs as a precise diagnostic tool for patients suspected of having prostate cancer. Candidate miRNAs for prostate cancer detection were identified in the discovery set ($n=123$). Using combinations of these candidate miRNAs, a diagnostic model was constructed in the training set ($n=484$), and its performance was assessed in the validation set ($n=484$). The resulting robust diagnostic model featured two miRNAs, namely miR-17-3p and miR-1185–2-3p. It demonstrated high diagnostic performance with a sensitivity of 90% and a specificity of 90% in the validation set, irrespective of the Gleason score and the clinical tumor-node-metastasis stage [20, 21].

In our study, we utilized both datasets differently, with the first used for training and the second for testing. Unlike the original studies, which focused specifically on bladder and prostate cancers, our first phase investigates 13 different cancer types, and the second phase focuses on gastric cancer. While we proposed two distinct panels of four miRNAs for MCD(Multi-Cancer Detection) and GCD(Gastric Cancer Detection), the original studies developed panels for classifying bladder and prostate cancers. Methodologically, we employed co-expression network analysis, whereas the original studies applied statistical techniques such as Fisher's linear discriminant analysis and $\chi 2$ tests to analyze miRNA expression and classify samples.

Multi-cancer screening results

In the first step, patient clustering was performed. To achieve a co-expression network with a balance between scale independence and mean connectivity, both metrics were calculated across power values ranging from 1 to 20 using the soft-connectivity algorithm. The optimal power value was selected when the scale independence reached 0.9. Finally, the co-expression matrix was computed using this determined power value.

In the first phase of our study, a power of 13, satisfied the criteria for topological scale-freeness, as shown in Supplementary Figure S1A, Additional File 2. Based on this power the adjacency matrix was calculated and the weighted miRNA correlation network was constructed. Finally, a cluster dendrogram of the modules was generated and seven modules were identified with the labels red, brown, turquoise, yellow, blue, green, and gray, containing 99, 213, 258, 206, 242, 145, and 1402 miRNAs, respectively, as shown in Supplementary Figure S1B, Additional File 2.

By examining the correlation between the modules and cancer in general, the highest correlation coefficients belonged to the yellow (0.62), turquoise (0.59), green (0.54), and brown (0.52) modules (Fig. 2A). To confirm the validity of this assessment, we also looked into each module and examined the correlation between the miRNA module membership factor and its significance in cancer. In this evaluation, the highest correlation rates were observed for the yellow (0.77), turquoise (0.7), green (0.7), and blue (0.63) modules (Fig. 2B).

Following the application of the alpha threshold, 70 miRNAs were chosen. Supplementary Table 2, Additional File 3, provides the list of these miRNAs, along with their distribution across different modules. Figure 3

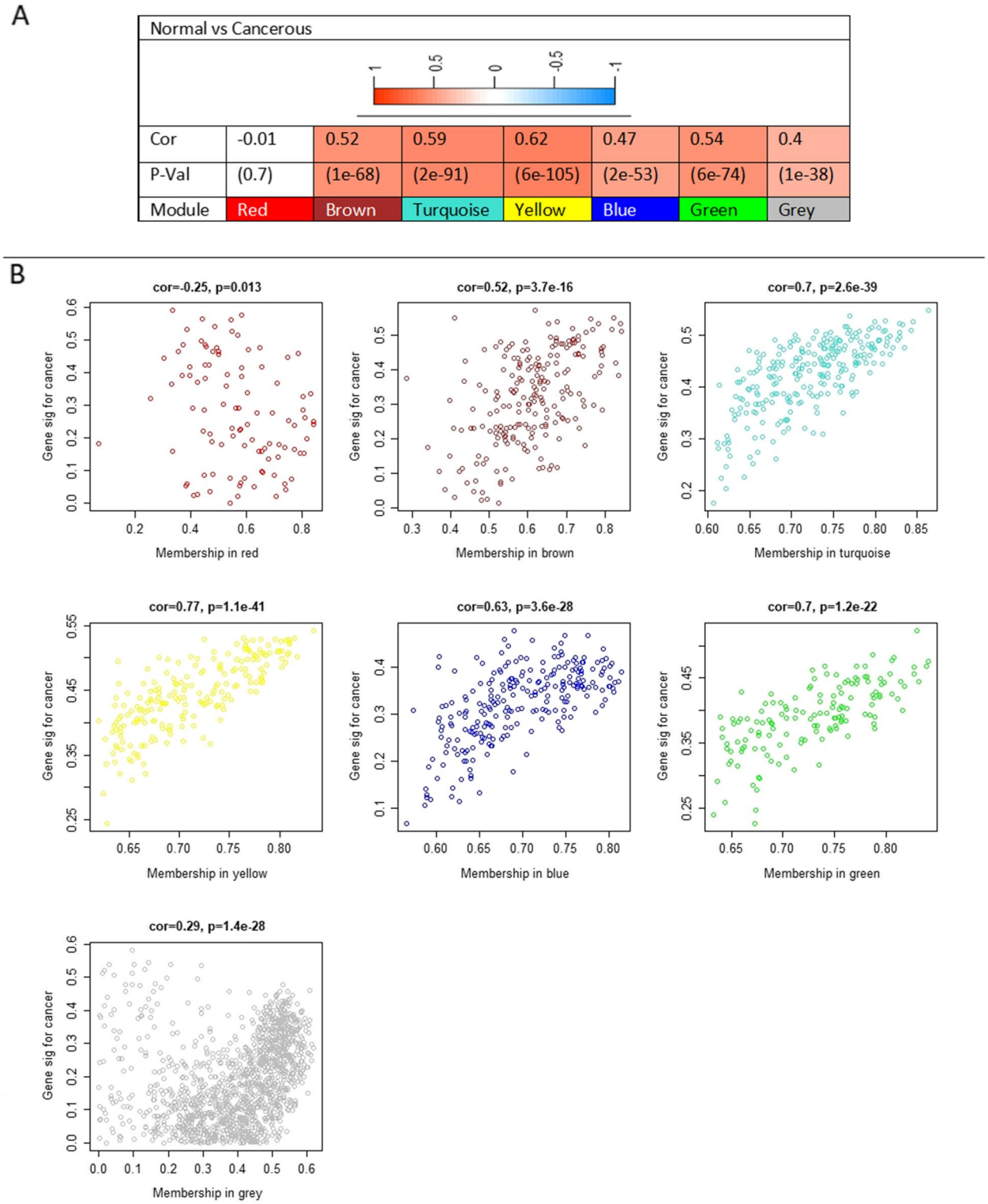

Fig. 2 Correlation analysis of modules and cancer. (**A**) Correlation analysis between each module eigen miRNA and cancer as a trait. (**B**) Correlation analysis between miRNA module membership and miRNA significance in cancer

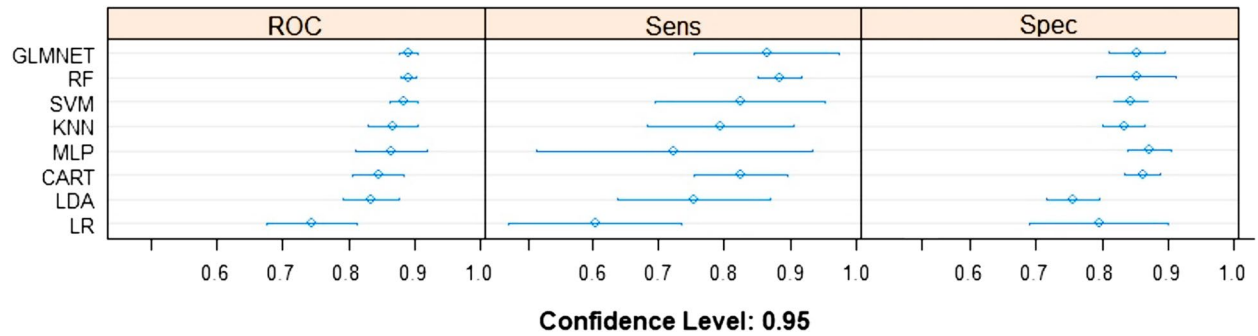

Fig. 3 Comparison of performance between different Machine Learning Algorithms using multi-cancer screening miRNA panel with a confidence level of 0.95

Table 3 The performance results of the multi-cancer screening model on candidate miRNA lists

miRNA count	Data	Accu	Spec	Sens	ROC
70	Train		95.2	98	99
	Test	92.5	92.3	97.5	
25	Train		96	96	98.7
	Test	91	90.8	97.5	
8	Train		97	98	98.9
	Test	92.4	92.3	97.5	
4	Train		96.1	96	98.6
	Test	94.9	94.8	97.5	

illustrates the comparative results of the performance of the selected machine learning algorithms. This evaluation was based on specificity, sensitivity, and the area under the ROC curve. The random forest algorithm was selected according to its appropriate performance regarding the three evaluation criteria for this dataset.

The multi-cancer screening model performance results using the candidate miRNA lists are presented in Table 3. Classifying cancer samples from non-cancer control samples, the best result was achieved by 4-miRNA panel containing "hsa-miR-614", "hsa-miR-1258", "hsa-miR-548ah-5p", and "hsa-miR-8073" with 96.1, 96, and 98.6 specificity, sensitivity, and ROC on original data by fivefold cross-validation and 94.9, 94.8, 97.5 for accuracy, specificity, and sensitivity on the independent dataset.

Considering that the previous evaluation was performed collectively and given that the accuracy might be different for each cancer, further evaluation was performed for different cancers separately. Figure 4 illustrates the model performance results of each cancer type.

Gastric cancer screening results

In the process of classifying gastric cancer, prior to creating the network, we initially detected an outlier among the non-cancer samples, specifically with the ID gsm3107261. Consequently, this outlier was excluded from the analysis. Subsequently, the network was

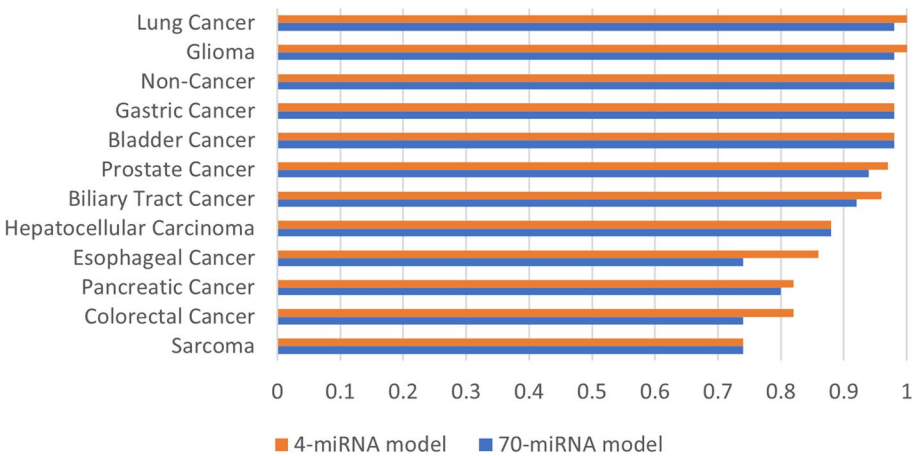

Fig. 4 Comparison of model performance for each cancer type separately using 70-miRNA and 4-miRNA models

established with a power setting of 9 to achieve scale-freeness, as determined by mean connectivity and scale independence metrics (Supplementary Figure S2A, Additional File 4). A dendrogram representing the miRNAs was generated, leading to the identification and labeling of seven distinct modules as green, yellow, brown, blue, turquoise, and gray, each comprising 45, 96, 202, 603, 747, and 872 miRNAs, respectively (Supplementary Figure S2B, Additional File 4).

In the examination of module correlations with gastric cancer, four modules, namely blue, turquoise, green, and yellow, exhibited the strongest correlations (Fig. 5). It's worth noting that, following the guidelines outlined in the WGCNA manuals, the gray module was disregarded in this analysis.

In the preservation analysis, the blue module holds the 5th position and displays the lowest preservation level among other cancer types, regardless of the gray module, as demonstrated in Table 4. Accordingly, it stands out as the most distinctive module specific to gastric cancer. The comparison between the median rank and Zsummary criteria is represented in Fig. 6.

Following applying the Beta threshold, a total of 51 miRNAs were chosen for inclusion in the candidate list. Out of these, 50 miRNAs were sourced from the blue module, recognized as the most distinct module in gastric cancer, while one miRNA originated from the green module (Supplementary Table 3, Additional File 5).

In the machine learning assessment, eight different algorithms were compared regarding their classification

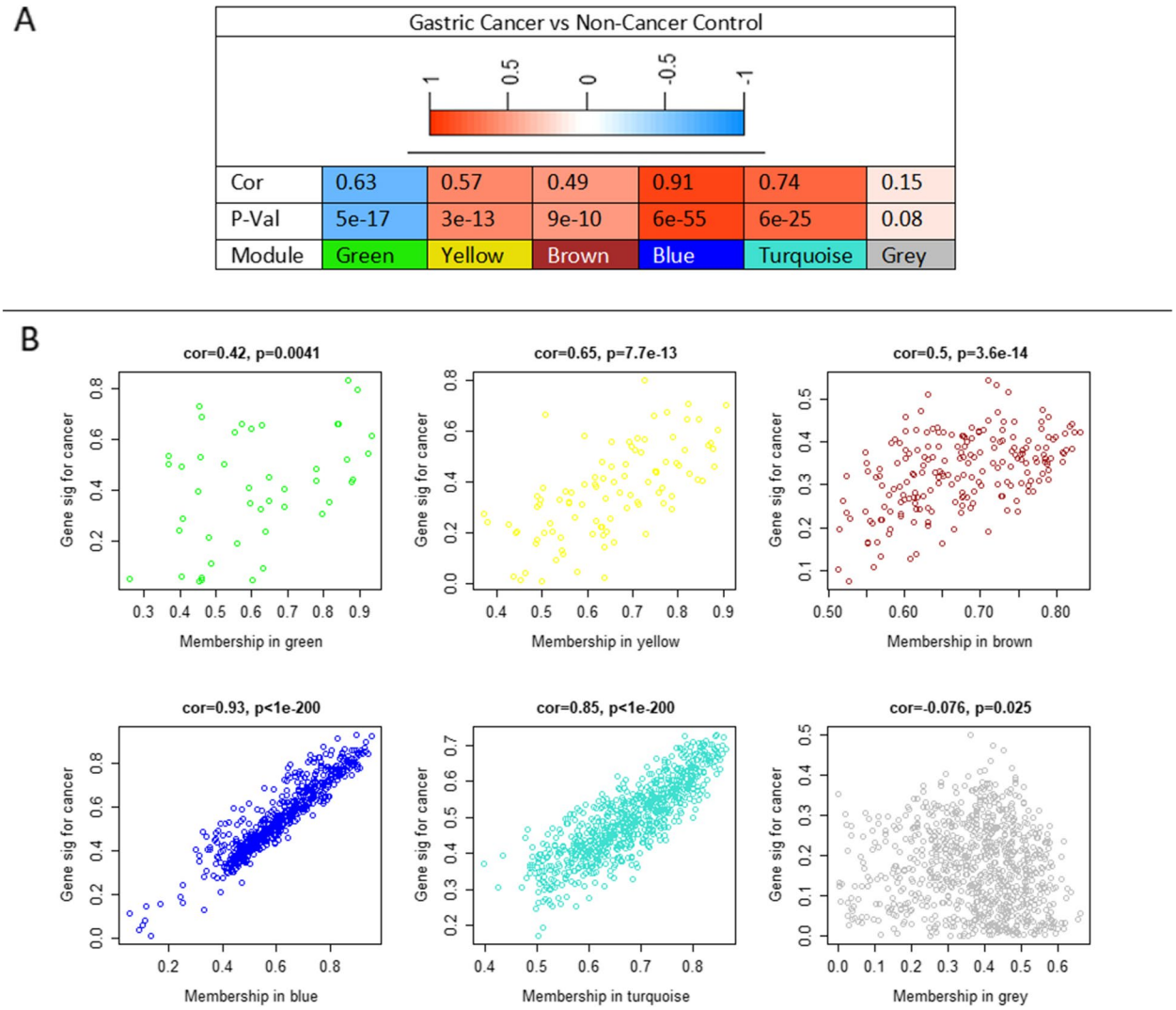

Fig. 5 Detecting most correlated miRNA modules to Gastric cancer. (**A**) Correlation between module eigen miRNAs and Gastric cancer incidence. (**B**) Correlation between miRNAs module membership and their significance in Gastric cancer

Table 4 Preservation analysis results, the blue module shows the lowest level of preservation among other cancers

Module	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
Gray	6	7	12	-15
Blue	5	5	10	21
Yellow	4	3	2.3	9.3
Turquoise	3	1	26	36
Green	2	4	6.6	5.6
Brown	1	2	19	15

performance with the candidate miRNAs as presented in Fig. 7. The GLMNET algorithm showed the best performance overall. Table 5 shows the results for the candidate list consisting of 51, 17, and 4 miRNAs, respectively.

The panel of 4 miRNAs can discriminate between a set of 12 cancers along with normal samples and gastric cancer patients using the GLMNET algorithm with the specificity, sensitivity, and area under ROC of 87, 90, and 89, respectively. The accuracy, specificity, and sensitivity in the independent dataset were 79, 78, and 86, respectively. The model performance results appraisal using the independent test data for each cancer type are presented in Fig. 8.

In the final step, functional enrichment was performed. After examining the enrichment results, sufficient evidence was provided for the effect of selected miRNAs in cancer in both phases of the study.

The miRNA-mRNA interaction network analysis with MiRNet online software revealed that the miRNAs and their target genes were primarily involved in different cancers, including lung, breast, ovarian, and gastrointestinal cancers, especially colorectal and gastric cancers, with the number of microRNAs decreased by four, and these cancers remained prominent on the list. In the second phase of the study, the assessment of microRNAs showed that this specific list was more associated with small intestine, colon, pancreas, and gastric cancers. Detailed results are available in Additional File 6.

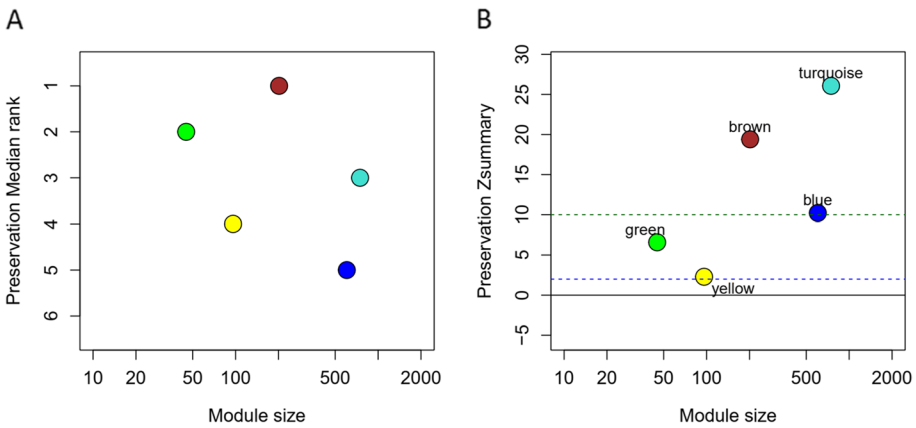

Fig. 6 Preservation analysis indicates which modules are more specific in gastric cancer. (**A**) Comparison based on Median Rank. (**B**) Comparison based on preservation Zsummary

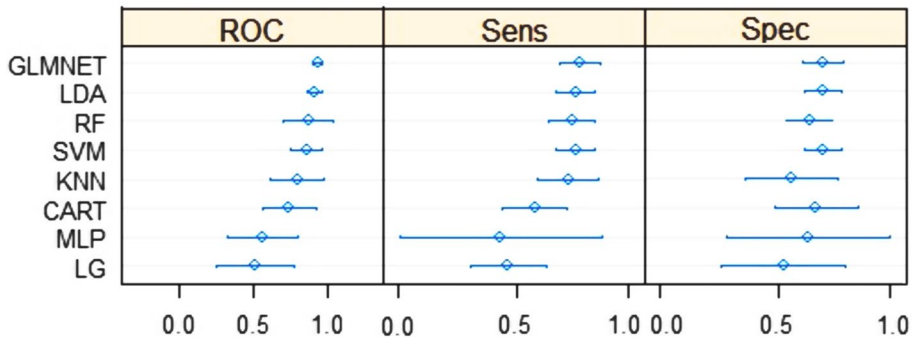

Fig. 7 Performances comparison between the machine learning algorithms using the gastric cancer screening panel with a confidence level of 0.95

Table 5 The performance results of the gastric cancer classification model using different candidate miRNA lists

miRNA list	Data	Accu	Spec	Sens	ROC
51	Train		79.1	82.5	87.8
	Test	82.8	82.6	88	
17	Train		85.3	85	90.2
	Test	81.1	81	82	
4	Train		81	95	92
	Test	78.2	77.9	86	

The miRPathDB analysis showed that the miRNAs and their target genes were mainly implicated in cancer-related pathways, including chronic myeloid leukemia and prostate cancer, and significant cancer signaling pathways such as p53, TGF-beta, and Ras signaling in both study phases.

Discussion

To summarize, our study aimed to develop a Multi-cancer Detection (MCD) method capable of not only distinguishing between patients with specific cancer types and normal individuals but also effectively differentiating gastric cancer from 12 other cancer types. This approach expands beyond single cancer assays, offering a comprehensive multi-cancer detection assay. We compared our methodology and results with three other recent studies. The first study focuses on multi-cancer early detection (MCED), while the two subsequent studies introduce panels specifically for gastric cancer detection.

Study 1: The study by Zhang et al. focused on identifying miRNAs for discriminating lung cancer from normal samples and then utilized the same list as an MCED (multi-cancer early detection) panel to distinguish each of the other cancer types from normal samples [34]. While our study employed network biology and WGCNA, they utilized differential gene expression analysis. Despite methodological differences, our model's performance was comparable to theirs for both multi-cancer and gastric-cancer detection. In our study, we achieved a sensitivity of 98% and a specificity of 98% with the MCD panel for gastric cancer, which made it possible to differentiate between patients with gastric cancer and normal samples. However, for other cancer types, the sensitivity and specificity varied from cancer type to cancer type, as shown in Fig. 9.

Study 2: Abe et al. aimed to identify serum miRNAs distinguishing early gastric cancer (EGC) samples from non-cancer controls [35]. While our study employed network biology, WGCNA, and the Random Forest algorithm, they utilized Fisher's linear discriminant analysis with a greedy algorithm to establish a diagnostic model. Additionally, they employed DeLong's test for selecting the best discrimination model. Despite these methodological differences, our model's performance was comparable to theirs for gastric cancer detection.

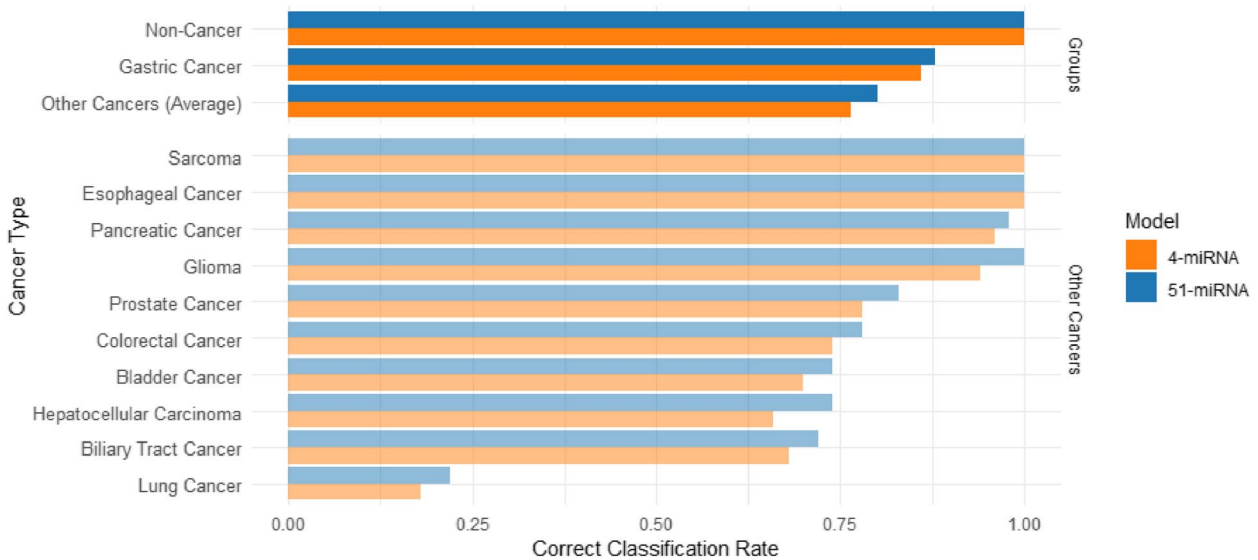

Fig. 8 Comparison of the trained model's performance in distinguishing gastric cancer from other cancer types and non-cancer controls using the 51-miRNA model and the 4-miRNA model

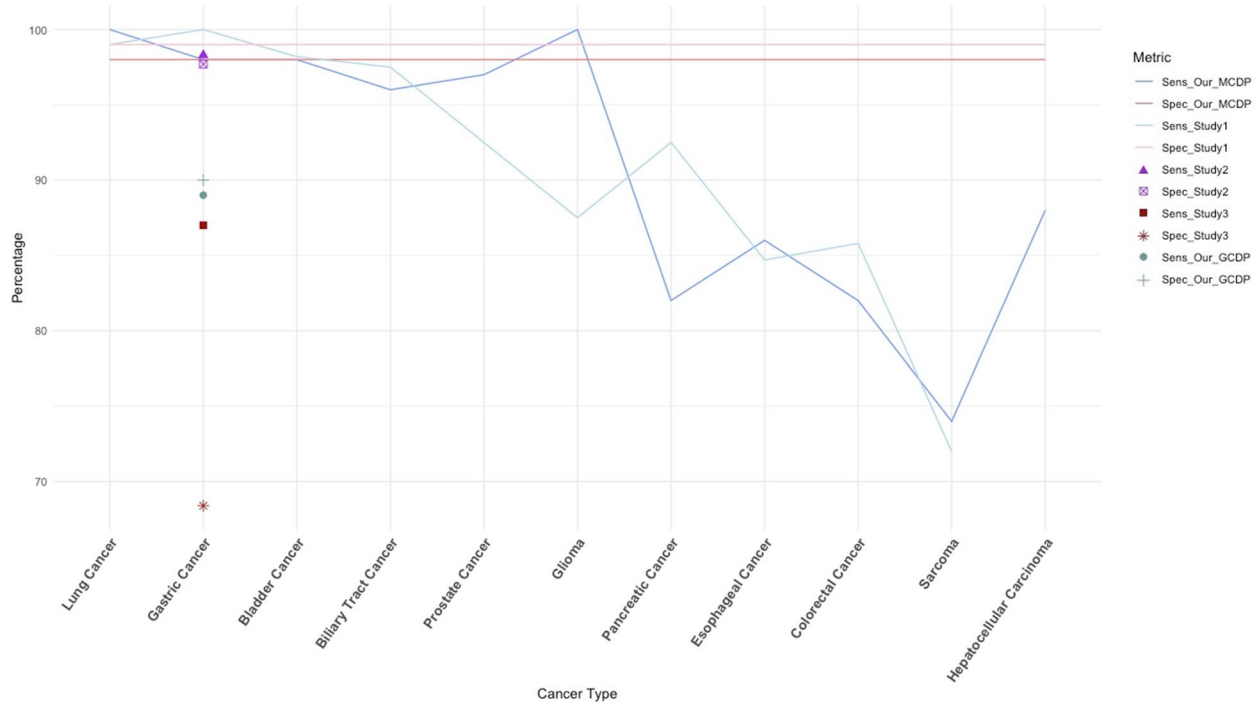

Fig. 9 Performance comparison between our study and three other studies. (MCDP stands for Multi-cancer Detection Panel, and GCP is the abbreviation for Gastric Cancer Detection Panel.)

As their panel is specifically designed to discriminate between gastric cancer samples and normal samples, without consideration for other cancer types, we compared the performance of our MCD panel, designed for general cancer screening, with theirs. Our findings suggest comparable performances between the two panels in discriminating between gastric cancer and normal samples.

Our study achieved a sensitivity of 98% and specificity of 98% using MCD panel for gastric cancer when considering only normal samples, as shown in the following table, In their study's discovery set, they achieved a sensitivity of 98% and a specificity of 97%.

In the second part of our study, we focused on classifying gastric cancer in comparison to various cancer types, while they only evaluated their model's ability to classify gastric cancer in comparison to Esophageal squamous cell carcinoma(ESCC) and Colorectal cancer (CRC). However, they only compared 10 patients with Stage III locally advanced gastric cancer with 50 ESCC and 50 CRC patients, resulting in imbalanced data. They report The AUCs of the EGC index in Stage III gastric cancer, ESCC, and CRC were 1.00, 0.64, and 0.44. While our second proposed panel demonstrates a sensitivity of 86% and specificity of 100% for discriminating between gastric cancer and esophageal cancer, and a specificity of 74% for colorectal cancer as shown in Fig. 8, it is also noteworthy for its ability to differentiate gastric cancer from pancreatic cancer patients, achieving a specificity of 96%. Our model not only performs better in discriminating between gastric cancer and esophageal and colorectal cancer but also demonstrates proficiency in discriminating gastric cancer from 10 other types of cancer.

To further evaluate the diagnostic potential of our circulating miRNA panels in detecting early-stage gastric cancer, we employed the GSE164174 dataset as a third independent validation source. This dataset provides detailed disease stage information, enabling us to examine miRNA expression across different phases of cancer progression. The results demonstrated that the MCD panel achieved a sensitivity of 96.57 and a specificity of 97.18. The Gastric Cancer panel showed a sensitivity of 86.87 and a specificity of 88.33. These findings indicate that both panels maintain robust performance even in the early stages of cancer. Further details are available in Additional File 7.

> Study 3: So et al. developed a serum miRNA panel for identifying patients with all stages of gastric cancer in a high-risk population from normal samples

Table 6 Comparison of miRNAs panels: our study versus three other studies

Study	miRNA panel
Study1 MCED	miR-5100, miR-1343-5p, miR-1290, miR-4787
Study2 GCD	miR-4257, miR-6785-5p, miR-187-5p, miR-5739
Study3 GCD	miR-140, miR-183, miR-30e, miR-103a, miR-126, miR-93, miR-142, miR-21, miR-29c, miR-424 and miR-181a with a reference miRNA (miR-340)
Our MCD	miR-8073, miR-614, miR-548ah-5p, miR-1258
Our GCD	miR-1228-5p, miR-1343-3p, miR-6765-5p, miR-6787-5p

[36]. Their 12-miRNA panel demonstrated a sensitivity of 87.0% at a specificity of 68.4%, with an AUC of 84%, indicating good discriminatory power. They conducted a cross-reactivity test against other common cancers, namely: lung, breast, colorectal, liver, esophageal, prostate, and bladder cancer. They tested for cross-reactivity against these specified cancer types in a total of 90 patients, where 11 of them were mistakenly classified as having a high-risk score for gastric cancer.

As they suggest, their panel can effectively differentiate between gastric cancer samples and normal samples. Therefore, we can compare the performance of both of our 4-miRNA panels with their 12-miRNA panel. Our MCD panel achieves a sensitivity of 98% and a specificity

Table 7 Evaluation of the multi-cancer screening panel in literature

miRNA	Description	Ref[a]
hsa-miR-8073	They demonstrated a significant correlation between microRNA-8073 and both the pathological stage and unfavorable prognosis of Ovarian Cancer	[37]
	It has also been proposed as a novel serum miRNA discriminant model developed for the diagnosis of late-stage esophageal squamous cell carcinoma (ESCC) in analyzing 566 ESCC versus 4965 control patients without cancer	[38]
	It was reported to be present in exosomes and predominantly exported from colorectal cancer cells. Some evidence suggests that miR-8073 can bind to oncogenic target candidates (FOXM1, MBD3, CCND1, KLK10, and CASP2), with antiproliferative effects	[39]
	In their study to discover biomarkers for Breast cancer classification, their models achieved close to 96% accuracy in the independent validation set using miR-8073 as a single biomarker	[40]
hsa-miR-614	The expression of miR-614 in patients with gastric cancer was significantly lower than that in the control group, and the expression of miR-614 in the HGC-27 cell line inhibited the invasion and proliferation of these cells	[41]
	It might serve as an antitumor target for lung cancer in the future	[42]
	The upregulation of miR-614 promotes proliferation and inhibits apoptosis in ovarian cancer by suppressing PPP2R2A expression	[43]
	Furthermore, miR-614 can inhibit the tumorigenesis of tumor cells by targeting RHOT1	[44]
	It is noted that RHOT1, a novel member of the Rho family, induces the proliferation and migration of pancreatic cancer cells	[45]
	It was among the most differentially expressed miRNAs regarding the diagnosis of pancreatic cancer	[46]
miR-548-3p	They showed significantly difference between the means of miR-548-3p expression between Oral squamous cell carcinoma (OSCC) and control groups	[47]
	It was indicated that it plays a critical role in the occurrence and development of NSCLC, thereby affecting the proliferation and apoptosis of NSCLC cells by targeting KLF15	[48]
	It can attenuate the development of CC by targeting TPX2 or inhibiting the PI3K/AKT signaling pathway in lung cancer	[42]
	It serves as an invasion and metastasis regulator targeting heparanase in gastric cancer	[49]
	They reported that it acts as an anti-oncogenic regulator in breast cancer	[50]
	It can proliferate breast cancer cells by regulating the expression of ECHS1	[51]
	It has been revealed that hsa-miR-548ah-5p targets ATG16L1, participates in the autophagy process, and has the potential to be used to manipulate autophagy in 5-FU-based chemotherapy in colon cancer cells	[52]
miR-1258	The upregulation of miR-1258 regulates the cell cycle and inhibits cell proliferation by targeting E2F8 in colorectal cancer directly	[53]
	It was significantly downregulated in hepatocellular carcinoma and associated with poor patient survival. The overexpression of miR-1258 significantly inhibits liver cancer cell growth, proliferation, and tumorigenicity by increasing cell cycle arrest in the G0/G1 phase and promoting cell apoptosis	[54]
	They illustrated how the miR-1258 mechanism was linked to brain metastatic breast cancer through heparanase control. They also offered the development of heparanase-based therapeutics for the treatment of cancer patients with brain metastases, especially BMBC	[55]

[a] The references are organized in descending order based on the publication year

of 98% in the classification of gastric cancer patients versus normal samples. Our gastric cancer detection panel is also promising with a sensitivity of 89% and a specificity of 90%, especially considering that it is a heterogeneous population that includes multiple cancer types and not just gastric cancer and normal samples.

Panels comparison

The panels introduced in the above studies are summarized in Table 6. The only direct commonality is the presence of hsa-miR-1343 in different forms (5p and 3p) between the Study1 MCED panel and Our GC panel. Other miRNAs do not seem to overlap directly between the studies. Each study has identified distinct miRNAs for their respective panels, reflecting possibly different methodologies, target populations, and cancer types being considered. Differences in sample collection, processing techniques, and statistical analysis methods can also contribute to these inconsistencies.

Upon assessing the multi-cancer screening panel, we observed that several studies in the literature have also documented their significance across a range of cancer types, including breast, ovarian, colorectal, and esophageal squamous cell carcinoma (ESCC). A comprehensive review of these studies can be found in Table 7.

In parallel with these established findings, our study concurs, as we identified a significant increase in the expression of miR-8073, miR-614, miR-548-3p, and miR-1258 among individuals with cancer, compared with the non-cancer control group. Figure 10 illustrates this consistent overexpression across the spectrum of cancers examined in this study. In our investigation, we

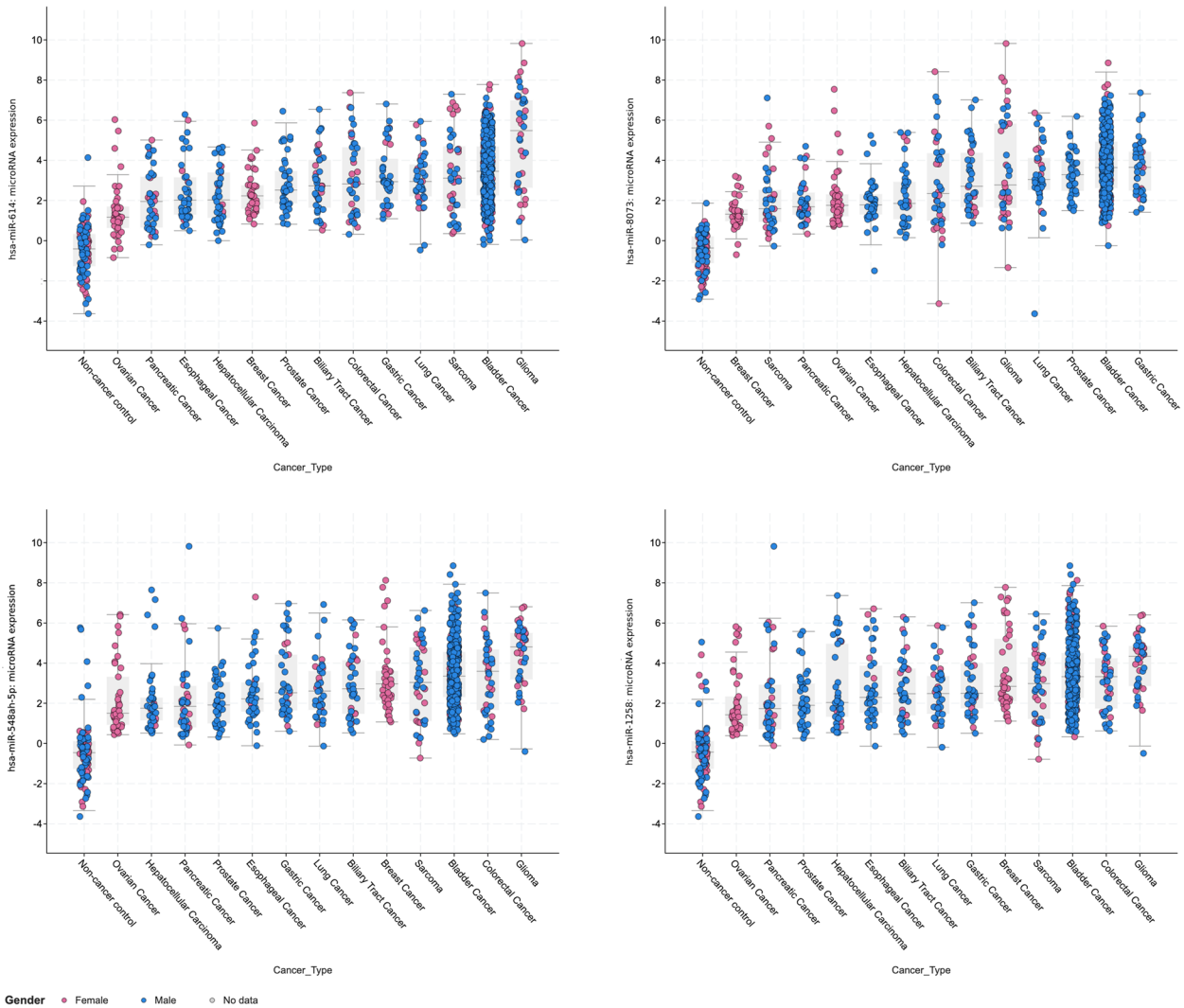

Fig. 10 Multi-cancer screening panel Expression values in different cancer types and non-cancer control group. The cancers-types legend is sorted ascending by expression median [55]

demonstrated that the multi-cancer screening model effectively classifies cancer patients when compared to the non-cancer control group.

Furthermore, the model exhibits varying effectiveness in distinguishing between different cancer types and the non-cancer control group. Glioma, lung cancer, bladder cancer, and gastric cancer exhibit the highest classification performance, with correct classification rates of 1, 1, 0.98, and 0.98, respectively, as illustrated in Fig. 4. Conversely, the weakest classification performance is observed for sarcoma, with a correct classification rate of 0.74.

Regarding the proposed panel for gastric cancer, it became apparent that some studies in existing literature have highlighted their relevance in the context of gastric cancer. Table 8 provides an in-depth overview of these studies for reference.

In line with the three previous studies reported in Table 8, we also showed that the downregulation of miR-1228 was associated with gastric cancer progression compared to other cancers and non-cancer controls. Since miR-1228 plays an important role in the stages of angiogenesis and EMT and can inhibit these stages, its reduction in the individual can be one of the first signs of threat in tumor formation. The downregulation of miR-1228 is a promising candidate in gastric cancer patients. Nonetheless, by exploring the correlation between miR-1228 and other miRNAs, a more precise understanding of the onset of gastric cancer can be obtained. As a result, we propose a panel of four miRNAs, including miR-1228, to enhance the accuracy of the predictive model.

Multiple studies presented in Table 8 provide substantial evidence regarding the involvement of miR-1343-3p in Gastric cancer. We also showed that the upregulation of miR-1343-3p was associated with gastric cancer compared to the non-cancer control group. Furthermore, this molecule was more expressed in gastric cancer compared to other cancers; hence, it can differentiate gastric cancer from other cancers.

In our investigation, the downregulation of hsa-miR-6765-5p in gastric cancer compared to other cancers and non-cancer controls was significantly associated with gastric cancer controls, and this molecule was down-expressed in gastric cancer compared to other cancers. Limited evidence was discovered concerning the involvement of hsa-miR-6765-5p in gastric cancer, with only one study addressing its association with venous thrombosis. However, additional research investigating the correlation between hsa-miR-6765-5p and various cancers, particularly gastric cancer, is recommended.

The second panel effectively classifies gastric cancer among other cancers and non-cancer controls. The group of four microRNAs could accomplish this classification with high sensitivity and specificity, while individual

Table 8 Evaluation of the gastric cancer screening panel in literature

miRNA	Description	Ref[a]
miR-1228	It has been documented that miR-1228 is highly expressed in the serum exosomes of patients with gastric cancer and its upregulation can downregulate the expression of MMP-14 and remarkably hinder the development and progression of gastric cancer	[56]
	They indicated that the expression of miR-1228 was reduced in human gastric cancer tissues compared to normal tissues. Moreover, miR-1228 acts as a negative regulator of NF-κB activity in the xenograft tumor model of gastric cancer, targets CK2A2, decreases the expression of mesenchymal markers, and increases the epithelial marker E-cadherin, suggesting it as a potential target for antiangiogenic therapy against gastric cancer	[57]
	It has been proved that miR-1228 targets macrophage migration inhibitory factor (MIF), which negatively regulates gastric cancer growth and angiogenesis by downregulating MIF The upregulation of miR-1228 decreased the expression of mesenchymal markers and increased the epithelial marker E-cadherin, suggesting its potential role in suppressing epithelial-mesenchymal transition; hence, miR-1228 plays a key role in regulating gastric cancer growth, and the selective restoration of miR-1228 may be beneficial for gastric cancer therapy	[58]
miR-1343-3p	It has earned the foremost position in the recommended panel for gastric cancer classification, showcasing outstanding precision (AUC: 100%, sensitivity: 100%, specificity: 100%, ROC: 100%)	[59]
	It is also shown that LINC01559 (long intergenic non-protein coding RNA 1559) is upregulated in GC tissues, where it can be transmitted from MSCs to GC cells via exosomes and target miR-1343-3p to promote GC progression by activating the PI3K/AKT pathway	[60]
	It was reported that TEAD1/4 oncogenic factors were overexpressed in GC cell lines and primary GC tissues, where TEAD4 was negatively regulated by miR-1343-3p, and its aberrant activation was mediated by silencing miR-1343-3p	[61]
hsa-miR-6787-5p	It was ranked 17th in the recent machine learning variable selection approach for gastric cancer classification	[59]
	It was significantly upregulated in extracellular vesicle-miRNA profiles in metastatic cell lines under RAPA treatment compared to CsA and untreated conditions. It can be a potential epigenetic mechanism induced by RAPA therapy in regulating the premetastatic niche of posttransplant colorectal cancer	[62]
hsa-miR-6765-5p	The significant upregulation of which was the main regulator of messenger RNAs involved in the pathology of venous thrombosis	[63]

[a] The references are arranged in descending order based on the publication year

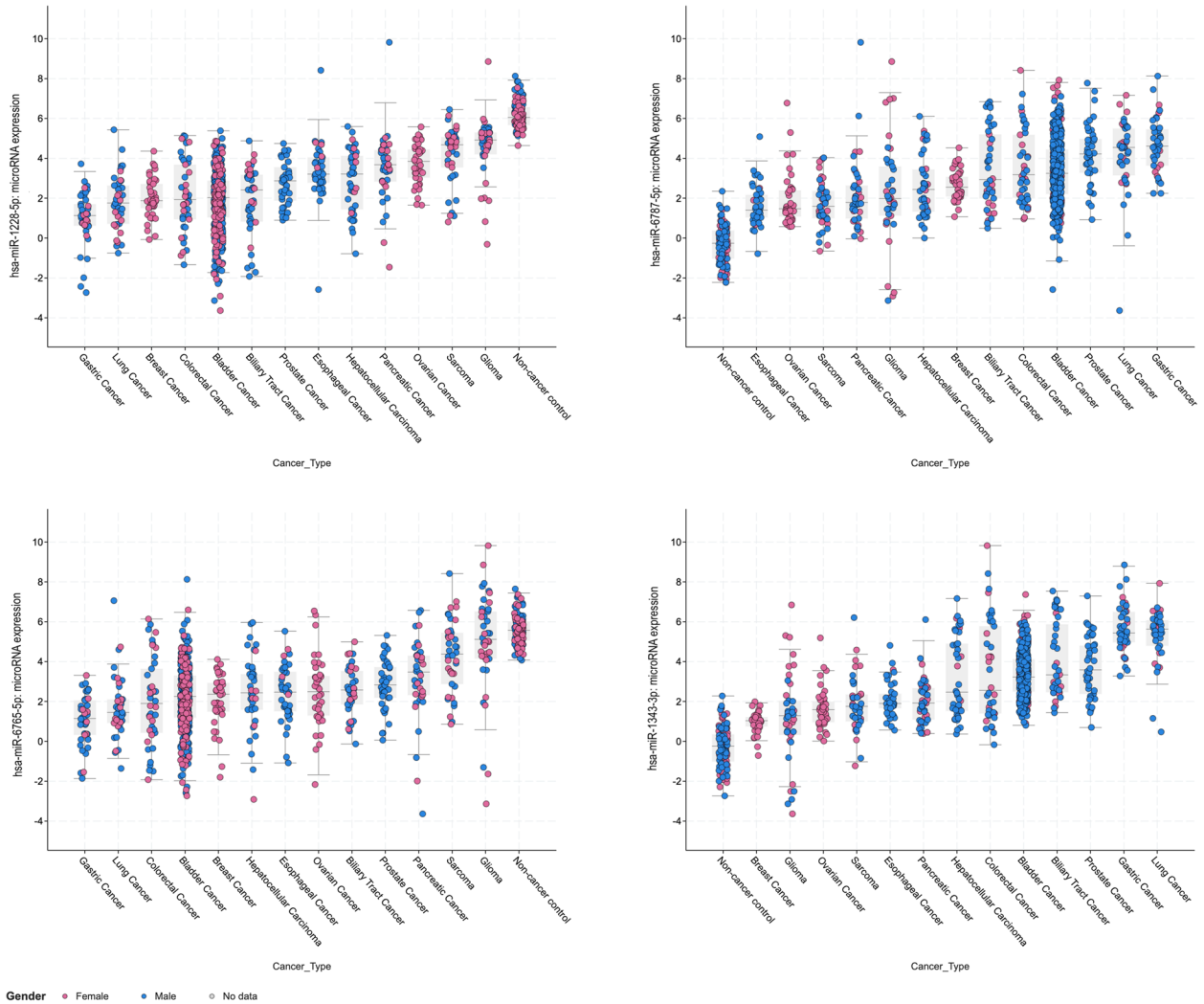

Fig. 11 Gastric cancer screening panel expression level in different cancer types and non-cancer control group. The cancers-types legend is sorted ascending by expression level median [55]

microRNAs alone did not exhibit this characteristic. As depicted in Fig. 11, miR-1228-5p and miR-6765-5p exhibited comparatively lower expression levels, whereas miR-6787 and miR-1343-3p demonstrated relatively higher expression levels compared to other cancer types and the control group. The proximity in serum miRNA profile patterns between Lung cancer and gastric cancer is also evident in this illustration.

Moreover, the performance of the gastric cancer model varies across different cancer types. The best classification performance is observed for esophageal cancer, non-cancer controls, sarcoma, and pancreatic cancer, with correct classification rates of 1, 1, 1, and 0.98, respectively (Fig. 8). Conversely, the weakest classification performance is observed for lung cancer, with a correct classification rate of 0.18. This discrepancy could be attributed to the close similarity in serum miRNA expression profiles between lung and gastric cancers. We attempted to construct a distinct weighted miRNA co-expression network just between gastric and lung cancer cases. However, the network criteria were not met, and no relevant modules could be detected. The serum miRNA profiles of these two cancers appear to be closely related. The other reason might be the heightened likelihood of metastasis from the lung to the gastric region in cancer patients.

Conclusions

The multi-cancer screening panel comprises four miRNAs with common expression altered in the thirteen studied cancers. The gastric cancer screening panel can differentiate the blood samples of gastric cancer patients from other cancers and non-cancer control groups. In

this context, candidate miRNA panels identified through this in silico study hold potential for patient screening and other therapeutic purposes. The approach utilized for gastric cancer discrimination can also be adapted and generalized to other cancers and diseases.

According to the pathogenesis of gastric cancer, imaging methods like endoscopy have achieved high screening efficiency for advanced cases. Over the past decade, this has shifted the focus toward identifying novel biomarkers for early detection. Given the limited clinical data available in the two primary datasets used in this study, we adopted an alternative approach by validating our miRNA panel using a third independent dataset(Additional File 7). Our findings demonstrate that both miRNA panels exhibit robust diagnostic performance, even in the early stages of at least three cancer types: colorectal cancer, esophageal cancer, and gastric cancer. These results highlight the potential of the panels in early cancer detection.

However, a comprehensive assessment of the panels' performance across all studied cancer types could not be conducted due to limitations in the datasets. This limitation presents an opportunity for future studies to investigate diagnostic performance based on other clinical information. To facilitate this, the publication of standardized datasets that integrate clinical and genomic data across various platforms is essential and could significantly enhance research in the field of liquid biopsy.It's important to note that our research primarily relies on bioinformatics analysis of publicly available datasets, utilizing the standard platform of 3D-gene microarray. While we have successfully developed and evaluated two models — one for multi-cancer detection and another for gastric cancer detection —the absence of experimental studies under wet laboratory conditions limits our ability to gain a deeper understanding of the predictive role of these miRNAs in carcinogenesis. Furthermore, additional validation of these two models using an independent cohort, ideally employing alternative technical platforms such as qRT-PCR, followed by ROC and survival curve analysis is necessary before considering their potential clinical application. Addressing these aspects will be the focus of our future work, as they are beyond the scope of the current study.

Abbreviations

GCD	Gastric Cancer Detection
GCDP	Gastric Cancer Detecion Panel
MCD	Multi-Cancer Detection
MCDP	Multi-Cancer Detection Panel
MCED	Multi-Cancer Early Detection
mMM	MiRNA Module Membership
mMM_pVal	MiRNA Module Membership P-Value
mTS	MiRNA Trait Significance
mTS_pVal	MiRNA Trait Significance P-Value
WGCNA	Weighted Gene Co-Expression Network Analysis
WMiCN	Weighted MiRNA Co-Expression Network

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-025-02091-x.

Additional file 1. Blood circulating miRNAs as biomarkers: An overview of some studies in the cancer domain.

Additional file 2: Supplementary Figure S1. Title: Weighted gene co-expression network analysis for Multiple cancer detection. (A) The left panel illustrates the relationship between soft-threshold power and the scale-free topology fit index, while the right panel shows how different soft-threshold powers affect mean connectivity. The optimal power value, selected at a scale independence of 0.9, is 13, which is used to calculate the co-expression matrix. This value is ideal for producing a scale-free network based on both mean connectivity and scale independence measures. (B) Clustering dendrogram of miRNAs, categorized into seven module colors: Red, Brown, Turquoise, Yellow, Blue, Green, and Grey. The number of miRNAs clustered within each module is indicated on the right.

Additional file 3. miRNA list capable of discriminating between normal and cancerous patients.

Additional file 4: Supplementary Figure S2. Title: Weighted gene co-expression network analysis for Gastric Cancer Detection. (A) Similar to Supplementary Figure 1, the left panel shows the relationship between soft-threshold power and the scale-free topology fit index, and the right panel illustrates the effect of different soft-threshold powers on mean connectivity. A power value of 9, determined at a scale independence of 0.9, is used to compute the co-expression matrix, producing a scale-free network. (B) Clustering dendrogram of miRNAs, grouped into six module colors: Green, Yellow, Brown, Blue, Turquoise, and Grey. The number of miRNAs clustered in each module is displayed on the right.

Additional file 5. miRNA list capable of distinguishing Gastric Cancer from other cancerous patients and non-cancer controls.

Additional file 6. miRNA Interaction Networks and Disease Associations for MCDP and GCDP: Results from MiRNet Analysis.

Additional file 7. Assessment of the diagnostic performance of the circulating miRNA panels in early gastric cancer detection.

Acknowledgements

Not applicable.

Authors' contributions

Conceptualization, LK, SS, MT, KK; Methodology, LK; Validation, LK, SS; Formal analysis, LK; Resources, LK, KK, MT; Data curation, LK; Writing original draft, LK, SS; Review and editing, LK, SS, MT, KK; Visualization, LK; Supervision, KK, MT; Project administration, KK, MT; All authors read and approved the final manuscript.

Funding

This research received no external funding.

Data availability

The datasets used during the current study are available at Gene Expression Omnibus – NCBI with GSE113486 and GSE112264 accession codes [21, 22]. The programming codes used in this research are available through GitHub under the project name: Cancer Screening. The project home page address is https://github.com/Kamkarl/Cancer_Screening. The code scripts are written in R Programming language.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 January 2024 Accepted: 22 January 2025
Published online: 06 February 2025

References

1. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science (1979). 2018;359(6378):926–30. Available from: http://science.sciencemag.org/content/early/2018/01/17/science.aar3247.abstract.
2. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science (1979). 2013;339(6127):1546–58.
3. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, et al. Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci. 2008;105(11):4283–8.
4. Pachmann K, Camara O, Kroll T, Gajda M, Gellner AK, Wotschadlo J, et al. Efficacy control of therapy using circulating epithelial tumor cells (CETC) as "liquid biopsy": trastuzumab in HER2/neu-positive breast carcinoma. J Cancer Res Clin Oncol. 2011;137(9):1317–27. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21739182. 2011/07/08.
5. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. Nat Rev Clin Oncol. 2017;14(9):531–48.
6. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2018;20(2):71–88. https://doi.org/10.1038/s41576-018-0071-5.
7. Heitzer E, Perakis S, Geigl JB, Speicher MR. The potential of liquid biopsies for the early detection of cancer. NPJ Precis Oncol. 2017;1(1):36. https://doi.org/10.1038/s41698-017-0039-5.
8. Lennon AM, Buchanan AH, Kinde I, Warren A, Honushefsky A, Cohain AT, et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science (1979). 2020;369(6499):9601. Available from: https://www.science.org/doi/abs/10.1126/science.abb9601. Cited 2022 Apr 18.
9. Cohen JD, Javed AA, Thoburn C, Wong F, Tie J, Gibbs P, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci. 2017;114(38):10202–7.
10. Floares A, Zety A, Floares C, Kreutz E, Calin G, Manolache F. i-biomarker - multi cancer early detection as a data science with ai problem. In: Molecular Diagnostics of Cancer [Working Title]. IntechOpen; 2023.
11. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. J Clin Oncol. 2009;27(34):5848–56.
12. Ying SY, Chang DC, Lin SL. The MicroRNA (miRNA): Overview of the RNA genes that modulate gene function. Mol Biotechnol. 2008;38(3):257–68. Available from: https://link.springer.com/article/10.1007/s12033-007-9013-8. Cited 2022 Apr 18.
13. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834–8.
14. Wang V, Wu W. MicroRNA-based therapeutics for cancer. BioDrugs. 2009;23(1):15–23.
15. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. Cells. 2020;9(2):276.
16 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19114008%0A, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2631488.
17. Giulietti M, Occhipinti G, Principato G, Piva F. Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. Cell Oncol. 2017;40(2):181–92.
18. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7(1):e1001057. https://doi.org/10.1371/journal.pcbi.1001057.
19. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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):209–49.
20. Usuba W, Urabe F, Yamamoto Y, Matsuzaki J, Sasaki H, Ichikawa M, et al. Circulating miRNA panels for specific and early detection in bladder cancer. Cancer Sci. 2019;110(1):408–19.
21. Urabe F, Matsuzaki J, Yamamoto Y, Kimura T, Hara T, Ichikawa M, et al. Large-scale circulating microRNA profiling for the liquid biopsy of prostate cancer. Clin Cancer Res. 2019;25(10):3016–25. Available from: https://pubmed.ncbi.nlm.nih.gov/30808771/. Cited 2022 Apr 18.
22. Usuba W, Urabe F, Yamamoto Y, Matsuzaki J, Sasaki H, Ichikawa M, et al. A serum miRNA combination could be a powerful classifier for the detection of patients with bladder cancer. GEO. GSE113486. 2018.
23. Urabe F, Matsuzaki J, Yamamoto Y, Hara T, Kimura T, Takizawa S, et al. Large-scale and high-confidence serum circulating miRNA biomarker discovery in prostate cancer. GEO. GSE112264. 2019.
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
25. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics. 2007;8(1):22.
26. Kuhn M. Building predictive models in r using the caret package. J Stat Softw. 2008;28(5):1–26. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v028i05. Cited 2022 Jun 6.
27. R: The R Project for Statistical Computing. Available from: https://www.r-project.org/. Cited 2022 Jun 6.
28. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7(1):e1001057. https://doi.org/10.1371/journal.pcbi.1001057.
29. Li J, Han X, Wan Y, Zhang S, Zhao Y, Fan R, et al. TAM 2.0: tool for MicroRNA set analysis. Nucleic Acids Res. 2018;46(W1):W180-5. Available from: https://academic.oup.com/nar/article/46/W1/W180/5033528. Cited 2019 Aug 10.
30. Kehl T, Kern F, Backes C, Fehlmann T, Stöckel D, Meese E, et al. miRPathDB 2.0: a novel release of the miRNA pathway dictionary database. Nucleic Acids Res. 2020;48(D1):D142-7. https://doi.org/10.1093/nar/gkz1022.
31. Chang L, Zhou G, Soufan O, Xia J. miRNet 2.0: Network-based visual analytics for miRNA functional analysis and systems biology. Nucleic Acids Res. 2020;48(W1):W244-51.
32 Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27141961%0A, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4987924.
33. Licursi V, Conte F, Fiscon G, Paci P. MIENTURNET: An interactive web tool for microRNA-target enrichment and network-based analysis. BMC Bioinform. 2019;20(1):1–10. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3105-x. Cited 2022 Apr 18.
34. Zhang A, Hu H. A novel blood-based microRNA diagnostic model with high accuracy for multi-cancer early detection. Cancers (Basel). 2022;14(6):1450.
35. Abe S, Matsuzaki J, Sudo K, Oda I, Katai H, Kato K, et al. A novel combination of serum microRNAs for the detection of early gastric cancer. Gastric Cancer. 2021;24(4):835–43.
36. So JBY, Kapoor R, Zhu F, Koh C, Zhou L, Zou R, et al. Development and validation of a serum microRNA biomarker panel for detecting gastric cancer in a high-risk population. Gut. 2021;70(5):829–37.
37. Zhang L, Wang YH, Wang L. MiRNA-8073 targets ZnT1 to inhibit malignant progression of ovarian cancer. Eur Rev Med Pharmacol Sci. 2019;23(14):6062–9. Available from: https://pubmed.ncbi.nlm.nih.gov/31364107/. Cited 2022 Apr 16.
38. Sudo K, Kato K, Matsuzaki J, Boku N, Abe S, Saito Y, et al. Development and Validation of an Esophageal Squamous Cell Carcinoma Detection Model by Large-Scale MicroRNA Profiling. JAMA Netw Open. 2019;2(5). Available from: https://pubmed.ncbi.nlm.nih.gov/31125107/. Cited 2022 Apr 18.

39. Mizoguchi A, Takayama A, Arai T, Kawauchi J, Sudo H. MicroRNA-8073: Tumor suppressor and potential therapeutic treatment. PLoS One. 2018;13(12). Available from: https://pubmed.ncbi.nlm.nih.gov/30589909/. Cited 2022 Apr 16.
40. Cui X, Li Z, Zhao Y, Song A, Shi Y, Hai X, et al. Breast cancer identification via modeling of peripherally circulating miRNAs. PeerJ. 2018;6(3). Available from: https://pubmed.ncbi.nlm.nih.gov/29607263/. Cited 2022 Apr 16.
41. Jing F, Li H, Gao Q. Expression of miR-614 in gastric cancer and its effect on invasion and proliferation of gastric cancer cells HGC-27. J BUON. 2021;26(2):483–9.
42. Wang Z, Wu X, Hou X, Zhao W, Yang C, Wan W, et al. miR-548b-3p functions as a tumor suppressor in lung cancer. Lasers Med Sci. 2020;35(4):833–9.
43. Zhang J, Gao D, Zhang H. Upregulation of miR-614 promotes proliferation and inhibits apoptosis in ovarian cancer by suppressing PPP2R2A expression. Mol Med Rep. 2018;17(5):6285–92. Available from: http://www.spandidos-publications.com/10.3892/mmr.2018.8714/abstract. Cited 2022 Apr 18.
44 Saidijam M, Afshar SS, Afshar SS, Warden E, Manochehri H, Saidijam M. Application of artificial neural network in miRNA biomarker selection and precise diagnosis of colorectal cancer. Iran Biomed J. 2018;23(x):1–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30056689.
45. Li Q, Yao L, Wei Y, Geng S, He C, Jiang H. Role of RHOT1 on migration and proliferation of pancreatic cancer. Am J Cancer Res. 2015;5(4):1460. Available from: https://pubmed.ncbi.nlm.nih.gov/PMC4473323/. Cited 2022 Apr 16.
46. Schultz NA, Werner J, Willenbrock H, Roslind A, Giese N, Horn T, et al. MicroRNA expression profiles associated with pancreatic adenocarcinoma and ampullary adenocarcinoma. Mod Pathol. 2012;25(12):1609–22. Available from: https://pubmed.ncbi.nlm.nih.gov/22878649/. Cited 2022 Apr 16.
47. Garajei A, Parvin M, Mohammadi H, Allameh A, Hamidavi A, Sadeghi M, et al. Evaluation of the Expression of miR-486–3p, miR-548–3p, miR-561–5p and miR-509–5p in Tumor Biopsies of Patients with Oral Squamous Cell Carcinoma. Pathogens. 2022;11(2).
48. Liu XW, Zhang CC, Zhang T. MiR-376b-3p functions as a tumor suppressor by targeting KLF15 in non-small cell lung cancer. Eur Rev Med Pharmacol Sci. 2020;24(18):9480–6.
49. Shi J, Chen P, Sun J, Song Y, Ma B, Gao P, et al. MicroRNA-1258: an invasion and metastasis regulator that targets heparanase in gastric cancer. Oncol Lett. 2017;13(5):3739–45. Available from: http://www.spandidos-publications.com/10.3892/ol.2017.5886/abstract. Cited 2022 Apr 18.
50. Shi Y, Qiu M, Wu Y, Hai L. MiR-548-3p functions as an anti-oncogenic regulator in breast cancer. Biomed Pharmacother. 2015;75:111–6.
51. Hou N, Han J, Li J, Liu Y, Qin Y, Ni L, et al. MicroRNA Profiling in human colon cancer cells during 5-Fluorouracil-Induced Autophagy. PLoS ONE. 2014;9(12):e114779.
52. Zhang Z, Li J, Huang Y, Peng W, Qian W, Gu J, et al. Upregulated miR-1258 regulates cell cycle and inhibits cell proliferation by directly targeting E2F8 in CRC. Cell Prolif. 2018;51(6):e12505.
53. Hu M, Wang M, Lu H, Wang X, Fang X, Wang J, et al. Loss of miR-1258 contributes to carcinogenesis and progression of liver cancer through targeting CDC28 protein kinase regulatory subunit 1B. Oncotarget. 2016;7(28):43419. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5190034/. Cited 2022 Apr 16.
54. Zhang L, Sullivan PS, Goodman JC, Gunaratne PH, Marchetti D. MicroRNA-1258 suppresses breast cancer brain metastasis by targeting heparanase. Cancer Res. 2011;71(3):645–54. Available from: https://pubmed.ncbi.nlm.nih.gov/21266359/. Cited 2022 Apr 18.
55. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4.
56. Chang L, Gao H, Wang L, Wang N, Zhang S, Zhou X, et al. Exosomes derived from miR-1228 overexpressing bone marrow-mesenchymal stem cells promote growth of gastric cancer cells. Aging. 2021;13(8):11808–21.
57. Jia L, Chen J, Xie C, Shao L, Xu Z, Zhang L. microRNA-1228* impairs the pro-angiogenic activity of gastric cancer cells by targeting macrophage migration inhibitory factor. Life Sci. 2017;1(180):9–16.
58. Jia L, Wu J, Zhang L, Chen J, Zhong D, Xu S, et al. Restoration of miR-1228* expression suppresses epithelial-mesenchymal transition in gastric cancer. PLoS ONE. 2013;8(3):e58637.
59. Gilani N, ArabiBelaghi R, Aftabi Y, Faramarzi E, Edgünlü T, Somi MH. Identifying potential miRNA biomarkers for gastric cancer diagnosis using machine learning variable selection approach. Front Genet. 2022;10(12):2538.
60. Wang L, Bo X, Yi X, Xiao X, Zheng Q, Ma L, et al. Exosome-transferred LINC01559 promotes the progression of gastric cancer via PI3K/AKT signaling pathway. Cell Death Dis. 2020;11(9):723.
61. Zhou Y, Huang T, Zhang J, Wong CC, Zhang B, Dong Y, et al. TEAD1/4 exerts oncogenic role and is negatively regulated by miR-4269 in gastric tumorigenesis. Oncogene. 2017;36(47):6518–30.
62. Tubita V, Segui-Barber J, Lozano JJ, Banon-Maneus E, Rovira J, Cucchiari D, et al. Effect of immunosuppression in miRNAs from extracellular vesicles of colorectal cancer and their influence on the pre-metastatic niche. Sci Rep. 2019;9(1):11177.
63. Gabler J, Basílio J, Steinbrecher O, Kollars M, Kyrle PA, Eichinger S. MicroRNA signatures in plasma of patients with venous thrombosis: a preliminary report. Am J Med Sci. 2021;361(4):509–16.

Publisher's Note