Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Stromal-associated extracellular vesicles snRNAs and piRNAs as non-invasive diagnostic and prognostic biomarkers in colon and pancreatic cancer

Manuel Collado
Instituto de Investigaciones Biomédicas Sols-Morreale
Víctor Medina-Chico
Alfonso X El Sabio University
María E Castillo
University of Alcalá
Mercedes Herrera
Karolinska Institutet
María Jesús Larriba
Instituto de Investigaciones Biomédicas Sols-Morreale
Cristina Galindo-Pumariño
Instituto Ramón y Cajal de Investigación Sanitaria
Carmen Lisset Flores
Instituto de Investigaciones Biomédicas Sols-Morreale
Íñigo Martínez
Hospital Universitario Ramón y Cajal
Blanca Morón
Hospital Universitario Ramón y Cajal
Elisa Conde
Instituto Ramón y Cajal de Investigación Sanitaria
Jorge Villalón
Instituto Ramón y Cajal de Investigación Sanitaria
Laura García-Bermejo
Instituto Ramón y Cajal de Investigación Sanitaria
Félix Hernández-del-Olmo
National University of Distance Education
José Manuel González-Sancho
Instituto de Investigaciones Biomédicas Sols-Morreale
Carolina de la Pinta
Hospital Universitario Ramón y Cajal

Alfredo Carrato
University of Alcalá
Vanesa García-Barberán
Hospital Clínico San Carlos
Cristina Peña
cristina.pena@csic.es

Instituto de Investigaciones Biomédicas Sols-Morreale

Article

Keywords: Cancer-associated fibroblasts, Extracellular Vesicles, Small non-coding RNAs, Liquid biopsy, Colon cancer, Pancreatic cancer, Tumor microenvironment, Risk stratification.

Posted Date: August 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10396814/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Stromal-associated extracellular vesicles snRNAs and piRNAs as non-invasive diagnostic and prognostic biomarkers in colon and pancreatic cancer

Manuel Collado[1,2], Víctor Medina-Chico[3,#], María E. Castillo[2,4,#], Mercedes Herrera[5], María Jesús Larriba[1,6,7], Cristina Galindo-Pumariño[4], Carmen-Lisset Flores[1,8], Íñigo Martinez[4], Blanca Morón[4], Elisa Conde[9], Jorge Villalón[4], Laura García-Bermejo[9], Félix Hernández-del-Olmo[10], José Manuel González-Sancho[1,6,7], Carolina de la Pinta[11], Alfredo Carrato[2,4], Vanesa García-Barberán[12], Cristina Peña[1,6,*]

1. Instituto de Investigaciones Biomédicas Sols-Morreale, Consejo Superior de Investigaciones Científicas - Universidad Autónoma de Madrid, 28029 Madrid, Spain
2. Alcalá University, 28034 Madrid, Spain
3. Universidad Alfonso X el Sabio (UAX), Avenida de la Universidad 1, 28691 Villanueva de la Cañada, Madrid, Spain.
4. Medical Oncology Department, Ramón y Cajal University Hospital, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), 28034 Madrid, Spain
5. Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden.
6. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), 28029 Madrid, Spain.
7. Instituto de Investigación Sanitaria Hospital Universitario La Paz (IdiPAZ), 28046 Madrid, Spain.

8. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), 28029, Madrid, Spain.
9. Biomarkers and Therapeutic Targets Group and Core Facility. RICORS2040, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), EATRIS, Madrid, Spain.
10. Department of Artificial Intelligence, National Distance Education University (UNED), Juan del Rosal 16, 28040 Madrid, Spain
11. Radiation Oncology Department, Ramón y Cajal University Hospital, Biomarkers and Therapeutic Targets Group, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), 28034 Madrid, Spain
12. "Clinical and Translational Research in Oncology" group. Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria Hospital Clínico San Carlos (IdISSC), 28040 Madrid, Spain.

* Corresponding Author: cristina.pena@csic.es (C.P.)

# These authors contributed equally to this work.

ABSTRACT

Background:

Colon cancer (CC) and pancreatic ductal adenocarcinoma (PDAC) remain major causes of cancer mortality due to limitations in early detection and risk stratification. Cancer-associated fibroblasts (CAFs) and stromal remodeling represent underexplored sources of clinically relevant circulating biomarkers in gastrointestinal malignancies.

Methods:

Building on our previous characterization of CAF-derived extracellular vesicle (EV) cargo, we evaluated three stromal-associated small non-coding RNAs (sncRNAs; two snRNAs and one piRNA) as proof-of-concept EV-associated biomarkers in pre-operative peripheral blood. EV-enriched fraction RNA was isolated from serum samples of patients with benign, premalignant, and malignant colon lesions, PDAC patients, and healthy donors. sncRNA expression was quantified by RT–qPCR. Predictive modeling assessed diagnostic and prognostic performance, and the biological relevance of the associated stromal programs was investigated through analysis of predicted sncRNA target genes in a TCGA colon cancer cohort.

Results:

The stromal-associated sncRNA panel discriminated CC patients from healthy individuals (AUC = 0.778) and identified benign and premalignant colonic lesions (AUC = 0.804). In PDAC, the same signature achieved comparable diagnostic performance (AUC = 0.775),

while combined CC/PDAC analysis yielded an AUC of 0.772, supporting the detection of stromal-derived circulating signals across two major gastrointestinal malignancies. In stage I–III CC patients, pre-operative sncRNA levels showed modest prognostic performance for relapse prediction (AUC = 0.561) and were independently associated with relapse risk in multivariable Cox regression analysis. Analysis of 43 predicted target genes in a TCGA cohort of 597 CC patients revealed strong associations with overall survival (AUC = 0.807; log-rank $p < 0.001$).

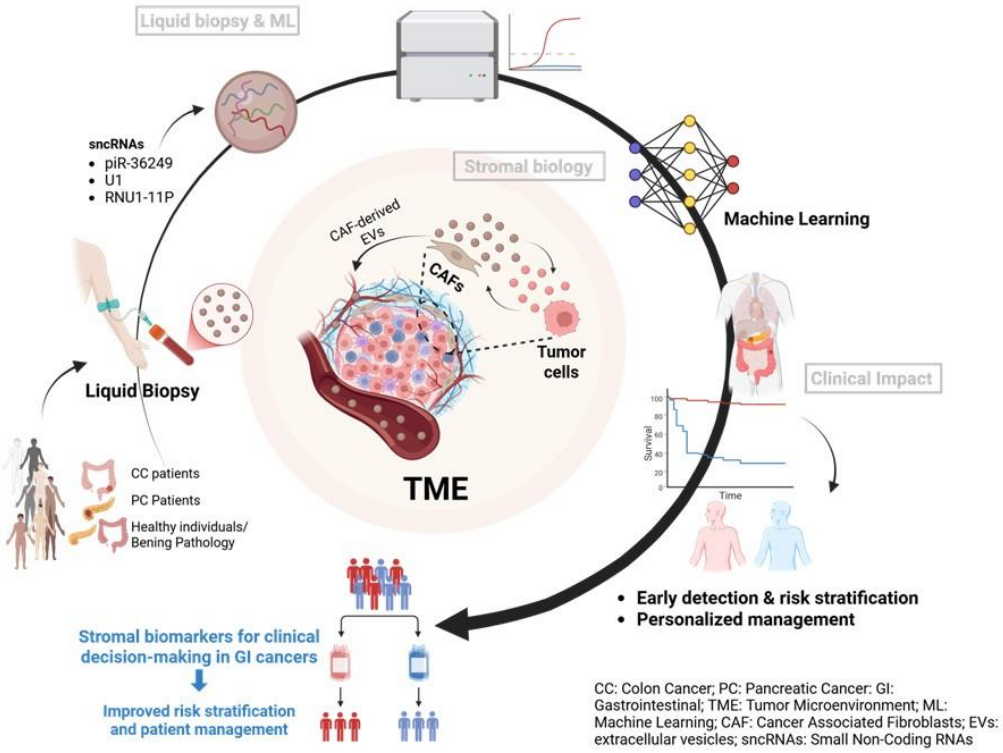
Conclusions:

These findings demonstrate that stromal-associated EV sncRNAs can be detected in peripheral blood and provide clinically informative signals for diagnostic classification and exploratory risk stratification in gastrointestinal malignancies. By capturing tumor stroma–associated biology through a minimally invasive liquid biopsy, these biomarkers complement conventional tumor cell-centered approaches and support the tumor stroma as a promising source of circulating biomarkers for future translational studies.

KEYWORDS

Cancer-associated fibroblasts; Extracellular Vesicles; Small non-coding RNAs; Liquid biopsy; Colon cancer; Pancreatic cancer; Tumor microenvironment; Risk stratification.

GRAPHICAL ABSTRACT



Graphical abstract description

Schematic representation of a non-invasive liquid biopsy strategy based on stromal-associated extracellular vesicles-enriched fraction (EVEF) small non-coding RNAs. Cancer-associated fibroblast–related EV snRNAs and piRNAs are detected in peripheral blood and analyzed using classification approaches to enable early diagnosis, risk stratification, and clinically relevant decision-making in colon and pancreatic cancer. Created with BioRender.com

HIGHLIGHTS

- Tumor stroma represents a measurable source of circulating biomarkers in gastrointestinal cancer.
- Extracellular vesicle–associated stromal sncRNAs are detectable in peripheral blood by liquid biopsy.
- A stromal-derived sncRNA signature supports diagnostic classification and exploratory patient stratification.
- Stromal-associated circulating RNA programs provide complementary biological information beyond tumor cell–derived biomarkers.

BACKGROUND

Gastrointestinal (GI) cancers account for a substantial proportion of cancer incidence and mortality worldwide. In 2022 alone, more than five million new GI cancer cases were diagnosed globally, with close to 3.5 million associated deaths. Among these malignancies, colorectal cancer (CRC) ranks as the third most frequently diagnosed cancer and the second leading cause of cancer-related death, representing the most common GI tumor. Approximately two million new CRC cases and one million deaths are reported each year worldwide. Pancreatic ductal adenocarcinoma (PDAC), while less frequent and ranking twelfth in incidence, remains disproportionately lethal and is currently the sixth leading cause of cancer-related mortality. Together, CRC and PDAC account for nearly 2.5 million diagnoses and 1.5 million deaths annually (1).

Early-stage CRC (stages I–II) is associated with a favorable prognosis, whereas survival declines sharply in advanced disease (2). Despite curative-intent surgery, nearly 30% of patients with stage I–III CRC experience disease recurrence (3,4). Notably, the risk of recurrence is highest during the first postoperative year and decreases thereafter, with 83% of recurrences occurring within three years of surgery (5,6). These observations underscore the clinical importance of diagnosing CRC at early stages, when tumors are more likely to be localized and associated with a lower risk of recurrence and improved long-term outcomes.

In this context, population-based screening programs play a central role in enabling early detection before the onset of advanced disease. Current screening strategies include fecal-based tests followed by colonoscopy when positive (7). Despite their widespread use, their

sensitivity and specificity remain imperfect, false-positive results may lead to unnecessary invasive procedures, and adherence remains suboptimal, highlighting the need for complementary blood-based biomarkers (8–14).

PDAC is characterized by an extremely poor prognosis, largely due to late diagnosis, with most tumors detected at advanced stages (15–18). Current screening approaches rely on imaging modalities and are generally restricted to high-risk individuals because of cost, limited performance, and the absence of broadly applicable non-invasive biomarkers (19,20).

Tumor development and progression are strongly influenced by the tumor microenvironment (TME), which comprises both malignant cells and surrounding stromal components (21,22). Among these, cancer-associated fibroblasts (CAFs) represent a major stromal population and play a central role in tumor growth, invasion, and metastatic progression through dynamic interactions with tumor cells (21–24). Increased stromal content and CAF-related molecular signatures have been associated with poor clinical outcomes across multiple tumor types, including colon, pancreatic, gastric, and lung cancers (25–32). Despite their biological and prognostic relevance, clinically applicable stroma-derived diagnostic biomarkers remain largely unavailable.

Although CRC and PDAC are clinically distinct entities, they share important biological features. Both are epithelial adenocarcinomas arising within the gastrointestinal tract and are characterized by a prominent desmoplastic stroma enriched in cancer-associated fibroblasts. In particular, PDAC is well known for its highly fibrotic and CAF-dense microenvironment (33), while increasing evidence highlights the prognostic relevance of stromal content in colon cancer (34). This shared stromal architecture provides a biologically plausible rationale

to explore whether stroma-derived molecular signals may represent common and systemically detectable biomarkers across these malignancies.

In this context, extracellular vesicles (EVs) emerge as key mediators of intercellular communication within the TME and as potential carriers of stromal-derived signals into the systemic circulation. These vesicles, measuring approximately 40–150 nm, transport bioactive cargo, proteins, lipids, RNAs, and DNA fragments that reflect their cellular origin (35,36). Tumor-derived EVs contribute to CAF activation (37–39), while CAF-derived EVs, particularly through their non-coding RNA (ncRNA) content, enhance tumor cell proliferation, stemness, and chemoresistance (40,41). Our group previously demonstrated a distinct ncRNA cargo in CAF-derived versus normal fibroblast (NF)-derived EVs in colon cancer (CC), with enrichment of small nuclear RNAs (snRNAs), PIWI-interacting RNAs (piRNAs), and microRNAs (miRNAs) (42). We further showed that gene expression signatures derived from CAF-related ncRNAs predict prognosis in large cohorts of CC patients (N = 1,235) (43). These findings suggest that CAF-derived ncRNAs detected in EV-enriched fractions (EVEF) may represent clinically informative biomarkers.

Based on our previous identification of CAF-enriched sncRNAs in extracellular vesicles, we investigated whether three selected markers (U1, RNU1-11P, and piR-36249) could be detected in peripheral blood and serve as non-invasive indicators of tumor-associated stromal activity. Specifically, we assessed their ability to distinguish healthy individuals from patients with benign, premalignant, and malignant colon lesions, as well as their applicability in PDAC. We further explored their association with clinical outcomes and evaluated the biological relevance of their predicted target genes in independent cohorts, providing a proof-

of-concept that stromal-derived molecular signals can be captured in circulation and provides a rationale for future liquid biopsy strategies.

MATERIALS AND METHODS

Patient samples and data collection

Samples from patients included in this study were managed by the BioBank Hospital Ramón y Cajal-IRYCIS (National Registry of Biobanks B.0000678), integrated in the Biobanks and Biomodels Platform of the ISCIII (PT23/00098). The study was approved by the Comité de Ética de la Investigación con Medicamentos of Hospital Universitario Ramón y Cajal (Approval ID: PI20/00602; Act No. 414) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to inclusion in the study.

Peripheral blood samples from CC patients were obtained at Ramón y Cajal University Hospital in collaboration with the Nursing Service. The samples were obtained prior to surgical intervention and subsequently processed and stored by the Biobank. Blood samples from sporadic pancreas cancer patients, were collected from the Spanish Familial Pancreatic Cancer Registry (PANGENFAM)(44). In addition, the Hospital Ramón y Cajal Biobank provided blood samples from healthy donors and from individuals diagnosed with colonic hyperplasia or adenoma, recruited through its established sample collections.

Clinicopathological data were retrieved in collaboration with the Medical Oncology, Radiation Oncology, and Pathology Services of Ramón y Cajal University Hospital.

Blood sample processing

All blood samples were collected, in red Vacutainer SST™ II Advance tubes (Becton Dickinson), processed, and stored (at −80 °C), following standardized protocols at the institutional biobank of Hospital Universitario Ramón y Cajal. Peripheral blood was obtained under consistent conditions across all patient groups, and serum preparation was performed using uniform procedures. Samples were processed within comparable time frames after collection and stored under identical conditions until analysis. To minimize pre-analytical variability, all EVEF isolation, RNA extraction, and downstream analyses were performed in a coordinated manner using the same protocols and reagents. No additional freeze–thaw cycles were introduced prior to analysis.

Extracellular vesicles isolation and RNA extraction

EVEF were isolated from 250 µL serum aliquots using polymer-based precipitation with the ExoQuick™ Exosome Precipitation Solution (System Biosciences), following the manufacturer's instructions. This approach yields EVEF that may also contain non-vesicular components such as protein complexes and lipoproteins. Total RNA was extracted from isolated EVEF using the mirVana™ miRNA Isolation Kit (Ambion).

To monitor extraction efficiency, the synthetic RNA spike-in UniSP2 was added to each sample together with MS2 carrier RNA, which improves spike-in stability without interfering with downstream applications. Purified RNA was eluted in nuclease-free water and stored at −80 °C until analysis.

RNA reverse transcription

Reverse transcription was performed using the miRCURY LNA™ RT Kit (Qiagen) according to the manufacturer's protocol. The synthetic *Caenorhabditis elegans* miRNA cel-miR-39-3p (Qiagen) was added to each reaction as an internal control for reverse transcription efficiency. Complementary DNA (cDNA) was stored at −20 °C until quantitative PCR analysis.

Detection of small non-coding RNAs by qPCR

Quantitative PCR (qPCR) assays were performed to assess the efficiency of RNA extraction and reverse transcription using UniSP2 and cel-miR-39-3p controls, respectively, with miRCURY LNA™ miRNA primers (Qiagen). For detection of the small non-coding RNAs (sncRNAs) of interest (piR-36249, U1 and RNU1-11P), custom miRCURY LNA™ miRNA primers specific to each sequence were designed (Qiagen). Importantly, the sncRNAs investigated in this study were not selected from circulating biomarker screens but originated from previous functional analyses demonstrating their enrichment in CAF-derived EVs, thereby providing a biologically informed rationale for their evaluation in peripheral blood (42,43).

cDNA samples were diluted 1:11 with nuclease-free water. qPCR reactions were carried out using miRCURY LNA™ SYBR Green Master Mix (Qiagen) on a LightCycler® 480 system (Roche). UniSP2 was used for data normalization, as it was added at the same concentration to all samples. Relative expression levels of sncRNAs were calculated using the second derivative method implemented in the LightCycler® 480 system. The resulting values, which

are inherently log-transformed, were used for all downstream statistical analyses. Missing expression values due to low transcript abundance were imputed using half of the minimum detected value for the corresponding sncRNA.

Predictive modeling and statistical analysis

Given that sncRNA expression data did not follow a normal distribution, as assessed by Shapiro–Wilk test or distribution inspection, non-parametric statistical tests were applied. Group comparisons were performed using the Kruskal–Wallis test.

Functional enrichment analyses were performed using the Gene Set Enrichment Analysis (GSEA) platform and the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb), focusing on biological process gene sets.

Multiple supervised machine learning algorithms were evaluated to identify the best-performing predictive models for each clinical endpoint and patient cohort.

The following datasets were analyzed:

- Colonic pathology cohort: A total of 183 individuals, including 40 healthy donors, 14 patients with colonic hyperplasia, 19 with adenomas, and 110 with CC were used to analyze the expression of the three CAF-derived EV sncRNAs in peripheral blood.
- Pancreatic pathology cohort: A total of 172 individuals, including 40 healthy donors, 34 patients with familial PDAC, and 98 with sporadic PDAC were used to analyze the expression of the three CAF-derived EV sncRNAs in peripheral blood.
- Public CC cohort: A publicly available dataset of 597 CC patients from TCGA COADREAD cohort (TCGA; COADREAD Pan-Cancer Atlas, 2018) was analyzed

to evaluate overall survival (OS) in relation to the predicted target genes of the studied sncRNAs (42).

Predictive analyses were implemented in Python using scikit-learn and imbalanced-learn libraries. Gene expression variables were used as predictors, whereas the binary outcome variable corresponded to patient relapse or mortality status.

Several supervised machine learning classifiers were evaluated, including Logistic Regression, Gaussian Naive Bayes, k-Nearest Neighbors (k-NN), Support Vector Machine with radial basis function kernel (SVM-RBF), and Random Forest. Prior to model fitting, predictor variables were standardized using z-score normalization (StandardScaler).

To address class imbalance, Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively within the training folds through an integrated machine learning pipeline, thereby avoiding data leakage into validation subsets.

Model optimization and evaluation were performed using nested stratified 5-fold cross-validation, preserving the original class distribution across folds. The outer cross-validation loop was used for unbiased model evaluation, whereas the inner loop performed hyperparameter optimization through GridSearchCV. Hyperparameters were optimized independently for each classifier using predefined parameter grids.

To obtain unbiased patient-level risk estimates, out-of-fold (OOF) predicted probabilities were generated during the outer cross-validation process. For each outer fold, hyperparameter optimization was first performed within the training subset, after which the best-performing model was applied to the corresponding unseen validation subset. This

procedure ensured that each probability estimate was generated from a model that had not previously observed the corresponding patient during training.

Patients were subsequently ranked according to their predicted probabilities of mortality. These ordered probabilities were used for downstream survival analyses, including Kaplan–Meier stratification and log-rank testing. Receiver operating characteristic (ROC) curves were additionally generated from OOF probabilities to evaluate the discriminative capacity of the final model.

Finally, the predicted probabilities were incorporated into the original dataset, which was sorted according to predicted risk and exported for subsequent prognostic and survival analyses.

RESULTS

Clinicopathological characteristics and EVEF sncRNAs profiling of the study cohorts

The cohort of healthy donors included a total of 40 individuals.

The CC cohort comprised 110 patients, with a mean age of 74 years (range: 33–93), including 58 males and 52 females. Tumor location was right-sided in 47 patients (43%) and left-sided in 47 patients (43%). Four tumors were located at the rectosigmoid junction (3%), which were classified according to institutional anatomical and clinical criteria and were included in the colon cancer cohort. Information about tumor location was unavailable for 12 cases (11%). Regarding tumor stage, 22 patients were classified as stage I (20%), 35 as stage II

(32%), 37 as stage III (34%), and 12 as stage IV (11%); staging information was missing for four patients (3%). During follow-up, disease relapse was observed in 22 patients (20%), whereas 88 patients (80%) remained relapse-free. For relapse prediction analyses, stage IV patients were excluded, resulting in a final dataset of 97 individuals, including 16 relapses (17%) and 81 non-relapses (83%). In addition, the study included 14 patients with colonic hyperplasia and 19 with adenomas, considered benign and premalignant lesions, respectively. The pancreatic pathology cohort consisted of 132 patients with PDAC.

The selection of the small non-coding RNAs (sncRNAs) analyzed in this study was based on previous work from our group, which characterized the differential distribution of sncRNAs in cellular and extracellular vesicle (EV) fractions derived from normal fibroblasts (NFs) and CAFs (42). From this dataset, candidate molecules were prioritized according to the magnitude of deregulation between NFs and CAFs and their high read counts, ensuring both biological relevance and technical robustness. Based on these criteria, one piRNA (piR-36249) and two snRNAs (U1 and RNU1-11P) enriched in CAF-derived EVs were selected for evaluation.

Correlation analysis among the three studied sncRNAs showed significant interdependence between variables (Supplementary Table S1). A summary of sncRNA expression levels across all study groups is provided in Supplementary Table S2.

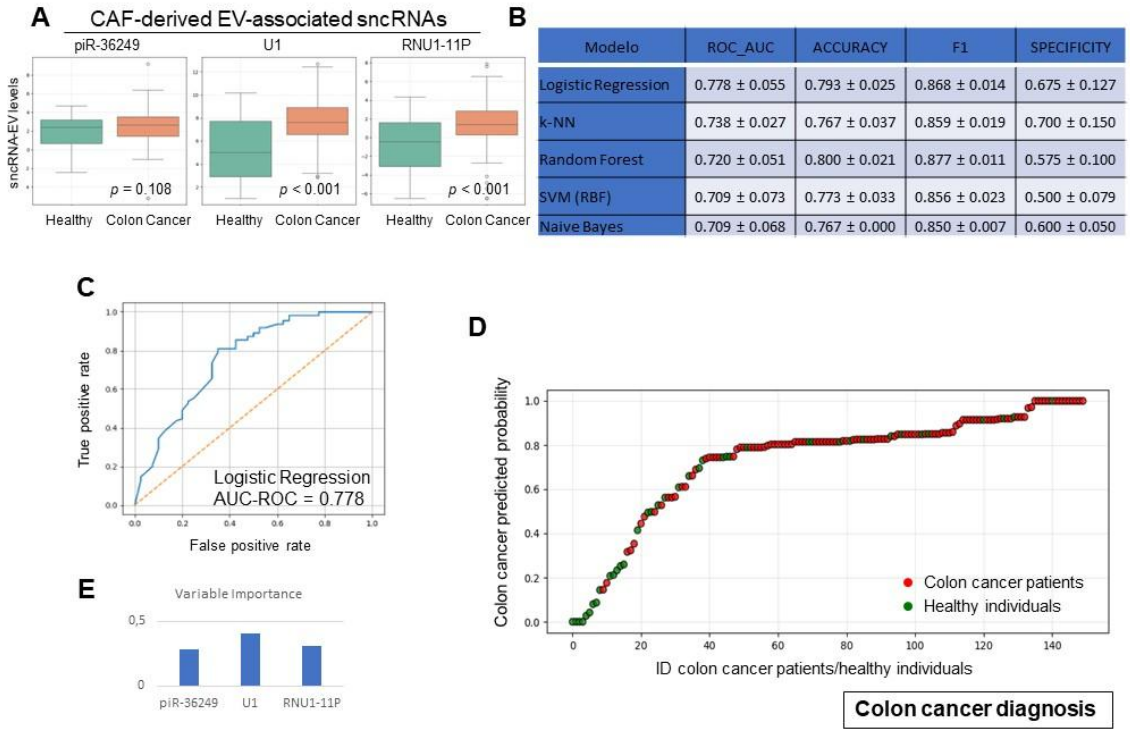
EVEF sncRNAs discriminate colon cancer from healthy individuals

As an initial exploratory approach, Kruskal-Wallis tests were performed to compare healthy donors and CC patients. U1 and RNU1-11P expression levels showed statistically significant differences between groups, whereas piR-36249 did not (Figure 1A).

To further evaluate diagnostic performance, multiple supervised classification algorithms were applied. A five-fold cross-validation scheme was used for both parameter optimization and performance evaluation. All tested models demonstrated consistent performance across cross-validation sets, achieving areas under the receiver operating characteristic curve (AUCs) above 0.70. Detailed performance metrics, including accuracy and specificity, are summarized in Figure 1B.

The Logistic Regression model outperformed other classifiers in terms of accuracy and AUC, leading to its selection for subsequent analyses. The obtained out-of-fold AUC was 0.778, as illustrated by the ROC curve (Figure 1C). Out-of-fold predicted probabilities, derived from Logistic Regression model, for individual samples are shown in Figure 1D, illustrating clear separation between healthy donors and CC patients. Variable importance estimates derived from the Logistic Regression model are presented in Figure 1E.

FIGURE 1



Modelo	ROC_AUC	ACCURACY	F1	SPECIFICITY
Logistic Regression	0.778 ± 0.055	0.793 ± 0.025	0.868 ± 0.014	0.675 ± 0.127
k-NN	0.738 ± 0.027	0.767 ± 0.037	0.859 ± 0.019	0.700 ± 0.150
Random Forest	0.720 ± 0.051	0.800 ± 0.021	0.877 ± 0.011	0.575 ± 0.100
SVM (RBF)	0.709 ± 0.073	0.773 ± 0.033	0.856 ± 0.023	0.500 ± 0.079
Naive Bayes	0.709 ± 0.068	0.767 ± 0.000	0.850 ± 0.007	0.600 ± 0.050

Figure 1. EVEF sncRNAs discriminate colon cancer patients from healthy individuals. (A) Box plots of sncRNAs expression levels detected in EVEF in healthy donors and colon cancer patients, with p-values derived from Kruskal-Wallis tests. (B) Performance metrics of multiple supervised classification models evaluated using internally validated predictive models. Area under the ROC curve (AUC), accuracy, and specificity are shown. (C) Receiver operating characteristic (ROC) curve of the optimized Logistic Regression model distinguishing healthy donors from colon cancer patients. (D) Predicted probabilities assigned by the Logistic Regression for each individual sample. (E) Variable importance scores derived from the Logistic Regression.

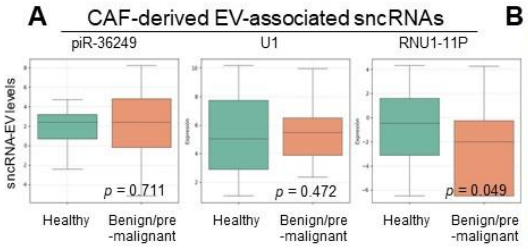
Detection of benign and premalignant colonic lesions using EVEF sncRNAs

Given the clinical relevance of early lesion detection, we next assessed whether the sncRNA panel could distinguish healthy donors from individuals with benign or premalignant colonic lesions (hyperplasia and adenoma).

Kruskal-Wallis test analysis showed modest differences in RNU1-11P expression, whereas no significant differences were observed for piR-36249 or U1 (Figure 2A). Subsequent predictive analyses indicated that the Support Vector Machine (SVM) model achieved the best predictive performance in this setting, with an AUC of 0.804 (Figure 2B).

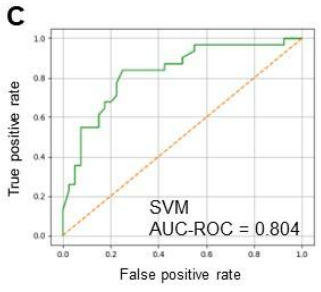
The obtained out-of-fold AUC using the optimized model was 0.804 (Figure 2C). Out-of-fold predicted probabilities for individual samples are shown in Figure 2D. Variable importance rankings are displayed in Figure 2E.

FIGURE 2



B

Modelo	ROC_AUC	ACCURACY	F1	SPECIFICITY
SVM (RBF)	0.804±0.088	0.737±0.087	0.688±0.109	0.800±0.127
k-NN	0.804±0.133	0.724±0.151	0.663±0.196	0.800±0.127
Naive Bayes	0.760±0.159	0.670±0.182	0.642±0.191	0.675±0.187
Random Forest	0.738±0.140	0.681±0.183	0.600±0.249	0.775±0.146
Logistic Regression	0.671±0.133	0.644±0.178	0.547±0.241	0.750±0.209



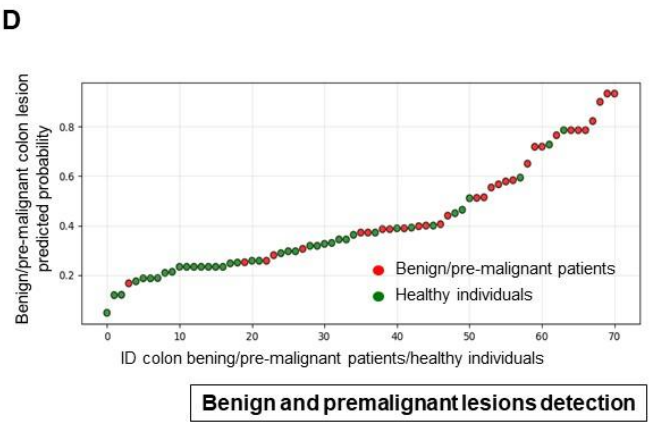


Figure 2. EVEF sncRNAs enable detection of benign and premalignant colonic lesions. (A) Box plots of sncRNA expression levels detected in EVEF in healthy donors and individuals with benign or premalignant colonic lesions (hyperplasia and adenoma), with p-values derived from Kruskal-Wallis tests. (B) Comparative performance of supervised classification models evaluated using cross-validated predictive models, highlighting the discriminative capacity of different algorithms. (C) ROC curve of the optimized Support Vector Machine (SVM) model applied to the same comparison. (D) Predicted probabilities assigned by the SVM to individual samples. (E) Variable importance scores derived from the SVM model.

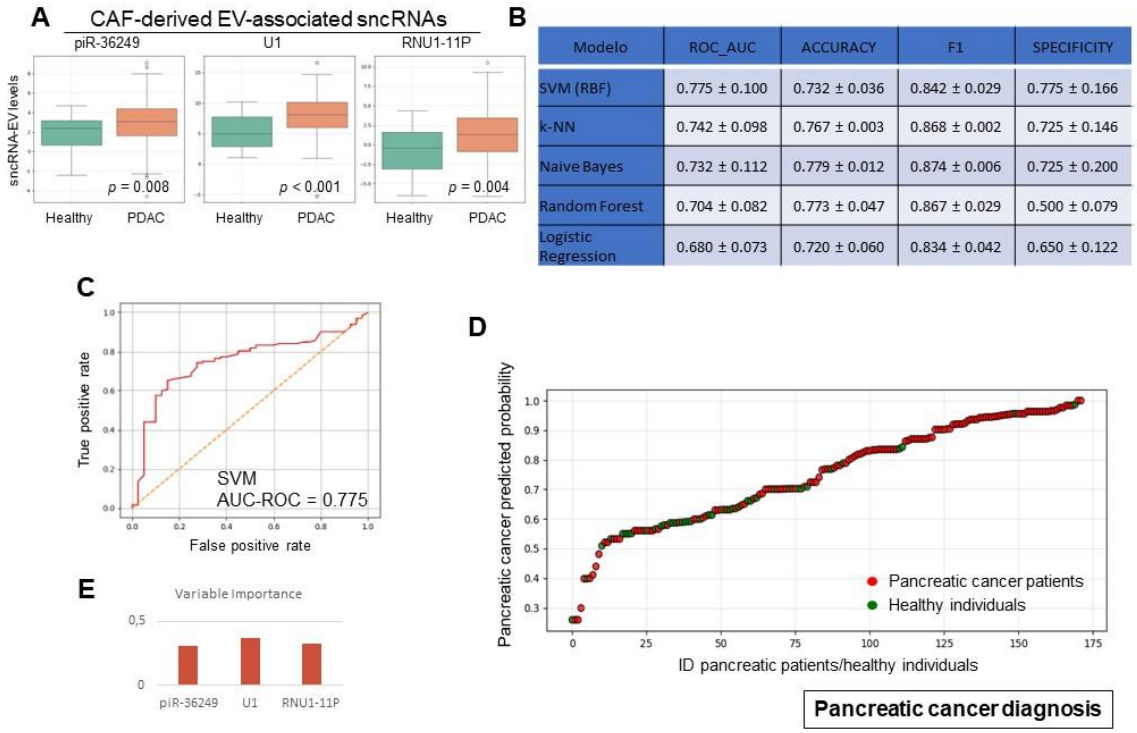
EVEF sncRNAs as diagnostic biomarkers in pancreatic cancer

To evaluate whether the diagnostic potential of the selected sncRNAs extended beyond colon malignancy, their expression was analyzed in EVEF isolated from peripheral blood samples of PDAC patients.

Kruskal-Wallis tests demonstrated statistically significant differences for all three sncRNAs when comparing PDAC patients with healthy donors (Figure 3A). Following the same analytical workflow established for CC, classification models were evaluated and compared using a five-fold cross-validation approach (Figure 3B).

Support Vector Machine again showed the best overall performance. The obtained out-of-fold AUC was 0.775 (Figure 3C). Out-of-fold predicted probabilities, derived by the best-performing classifier, are shown in Figure 3D, with clear separation between healthy individuals and PDAC patients. Variable importance scores are presented in Figures 3E.

FIGURE 3



Modelo	ROC_AUC	ACCURACY	F1	SPECIFICITY
SVM (RBF)	0.775 ± 0.100	0.732 ± 0.036	0.842 ± 0.029	0.775 ± 0.166
k-NN	0.742 ± 0.098	0.767 ± 0.003	0.868 ± 0.002	0.725 ± 0.146
Naive Bayes	0.732 ± 0.112	0.779 ± 0.012	0.874 ± 0.006	0.725 ± 0.200
Random Forest	0.704 ± 0.082	0.773 ± 0.047	0.867 ± 0.029	0.500 ± 0.079
Logistic Regression	0.680 ± 0.073	0.720 ± 0.060	0.834 ± 0.042	0.650 ± 0.122

Figure 3. EVEF sncRNAs signature shows diagnostic value in pancreatic ductal adenocarcinoma (PDAC). (A) Box plots of sncRNA expression levels detected in EVEF in healthy donors and PDAC patients, with p-values derived from Kruskal-Wallis tests. (B) Performance comparison of supervised classification models evaluated using internally validated predictive models. (C) ROC curve of the optimized Support Vector Machine (SVM) model discriminating pancreatic cancer patients from healthy individuals. (D) Predicted probabilities assigned to each sample by the SVM model. (E) Variable importance scores from the SVM model.

EVEF sncRNAs signature identifies gastrointestinal cancer

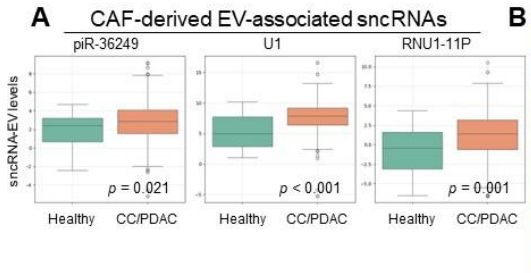
To explore the ability of the sncRNA panel to detect GI malignancy more broadly, CC and PDAC patients were combined into a single cancer group and compared with healthy donors.

All three sncRNAs showed significant differential expression between healthy individuals and cancer patients in Kruskal-Wallis test analyses (Figure 4A). Classification models

evaluated via five-fold cross-validation consistently yielded AUC values exceeding 0.700 (Figure 4B).

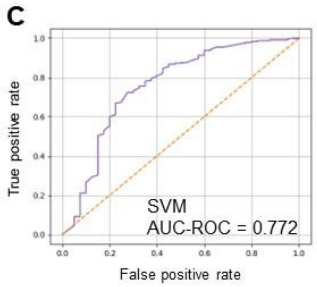
The obtained out-of-fold AUC derived by Support Vector Machine model was 0.772 (Figure 4C). Out-of-fold predicted probabilities for individual samples are shown in Figure 4D. Variable importance rankings are shown in Figures 4E.

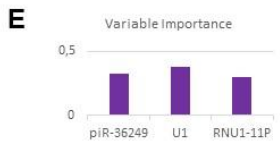
FIGURE 4



B

Modelo	ROC_AUC	ACCURACY	F1	SPECIFICITY
SVM (RBF)	0.772 ± 0.090	0.869 ± 0.027	0.928 ± 0.015	0.525 ± 0.122
k-NN	0.740 ± 0.084	0.855 ± 0.008	0.921 ± 0.004	0.600 ± 0.122
Logistic Regression	0.729 ± 0.038	0.833 ± 0.014	0.906 ± 0.007	0.700 ± 0.100
Random Forest	0.717 ± 0.069	0.851 ± 0.017	0.919 ± 0.010	0.350 ± 0.146
Naive Bayes	0.710 ± 0.083	0.847 ± 0.014	0.917 ± 0.008	0.625 ± 0.209





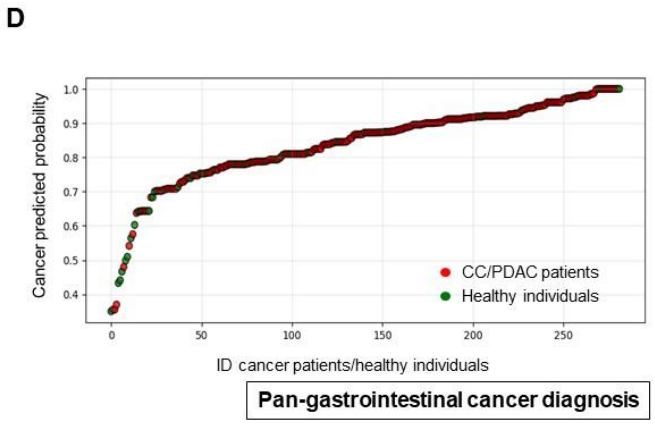


Figure 4. An EVEF sncRNAs signature identifies gastrointestinal malignancy. (A) Box plots of sncRNA expression levels detected in EVEF in healthy donors and patients with gastrointestinal cancer (colon cancer [CC] and pancreatic ductal adenocarcinoma [PDAC]), with p-values derived from Kruskal-Wallis tests. (B) Performance metrics of classification models distinguishing healthy individuals from gastrointestinal cancer patients using cross-validated predictive models. (C) ROC curve of the optimized Support Vector Machine

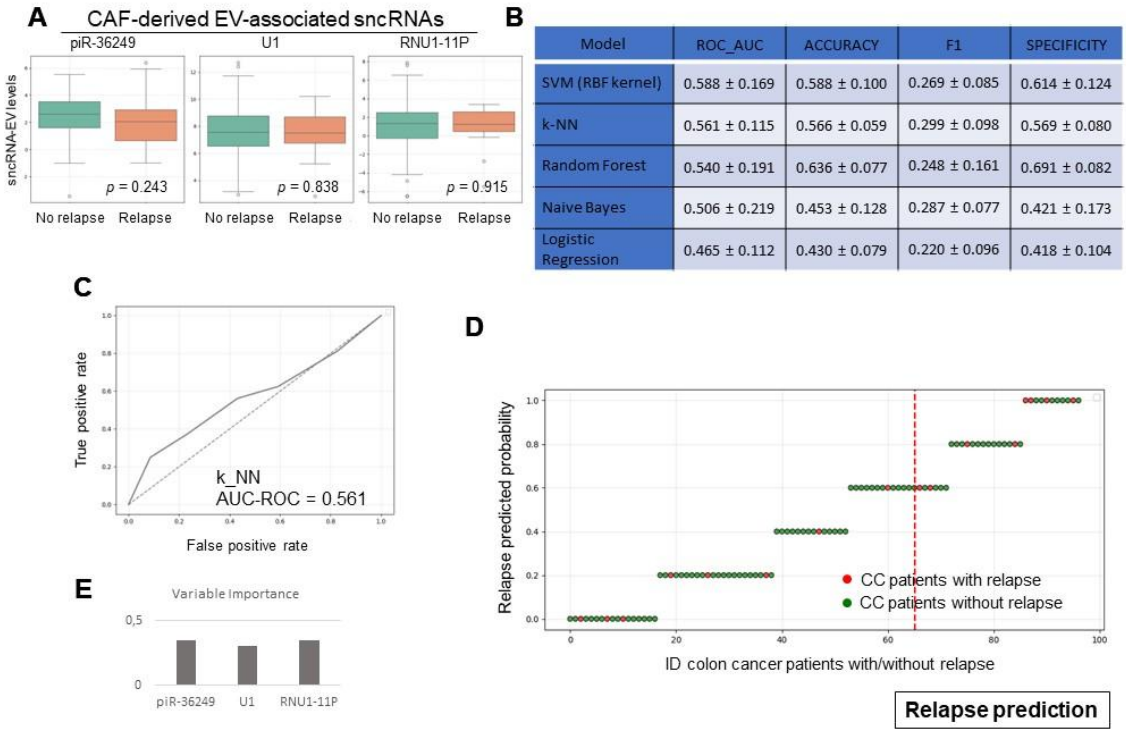
(SVM) model for gastrointestinal cancer detection. (D) Predicted probabilities for individual samples classified by the SVM model. (E) Variable importance scores derived from the SVM model.

Prognostic value of EVEF sncRNAs in colon cancer relapse

We next assessed whether sncRNAs expression levels detected in EVEF were associated with relapse in CC patients. Kruskal-Wallis test analysis did not reveal significant associations between relapse status and expression of all three sncRNAs (Figure 5A). To develop an exploratory predictive model for three-year relapse, analyses were restricted to the 97 stage I–III CC patients included in the relapse cohort. The dataset comprised 81 non-relapsing patients and 16 patients who experienced relapse. Classification algorithms were evaluated using five-fold cross-validation, and performance metrics are shown in Figure 5B.

Given its overall performance across evaluation metrics, also supported by survival analysis, the K-NN algorithm was selected for further characterization. When assessed by five-fold cross-validation, the model achieved a modest mean AUC of 0.561 (Figure 5C), likely reflecting the limited number of patients experiencing relapse in this cohort. Notably, despite the reduced discriminative capacity, the model retained a relatively higher specificity, indicating a consistent ability to correctly identify patients unlikely to relapse. Out-of-fold predicted probabilities, based on final predictive model, were ranked and visualized to facilitate interpretation (Figure 5D). Variable importance estimates are shown in Figure 5E.

FIGURE 5



409

410 **Figure 5. EVEF sncRNAs stratify colon cancer patients according to relapse risk.** (A) Box plots of
411 sncRNA expression levels detected in EVEF in relapsing and non-relapsing stage I–III colon cancer patients,
412 with p-values derived from Kruskal-Wallis tests. (B) Performance metrics of supervised classification models
413 predicting three-year relapse using cross-validated predictive models. (C) ROC curve of the optimized K-NN
414 model for relapse prediction. (D) Ranked predicted relapse probabilities for individual patients classification
415 by K-NN. (E) Variable importance scores derived from the K-NN.

416 To further evaluate the association between EVEF sncRNA levels and relapse risk, a
417 multivariable Cox proportional hazards regression model including clinicopathological
418 variables (sex, stage, mucinous histology, perineural invasion, and lymphatic permeation)
419 was applied. The sncRNA-based variable showed an association with relapse risk after
420 adjustment for clinicopathological variables (HR = 3.206, 95% CI: 1.046–9.819; p = 0.041)
421 (Figure 6A). Based on this model, a data-driven threshold corresponding to patient position

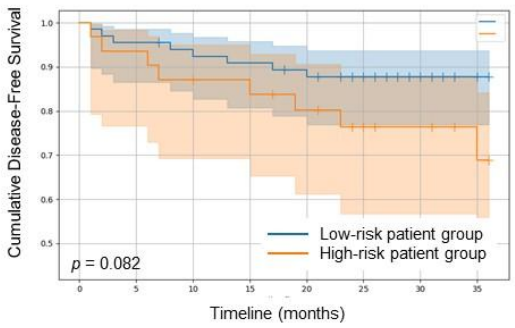
65 was selected for stratification into low- and high-risk groups. Kaplan–Meier survival analysis using this threshold showed a trend toward differences in relapse-free survival between groups, although statistical significance was not reached (log-rank p = 0.082; Figure 6B).

FIGURE 6

A

Covariate	COEF	HAZARD_RATIO	SE	P_VALUE	CI_LOWER	CI_UPPER
Snc-RNA-variable	1.165	3.206	0.571	0.041	1.046	9.819
Sex	0.424	1.529	0.572	0.458	0.497	4.696
Stage	1.061	2.890	0.550	0.053	0.982	8.506
Mucinous histology	0.808	2.243	0.592	0.172	0.702	7.169
Perineural invasión	2.506	12.26	0.627	6.433	3.587	41.93
Lymphatic permeation	-1.77	0.169	0.743	0.017	0.039	0.728
CONCORDANCA	0.82					

B



Relapse risk stratification

Figure 6. Prognostic evaluation of EVEF sncRNA levels in stage I–III colon cancer patients. (A) Multivariable Cox proportional hazards regression analysis evaluating the association between EVEF sncRNA levels and relapse-free survival in stage I–III colon cancer patients. The model included clinicopathological variables and the sncRNA-based variable. (B) Kaplan–Meier relapse-free survival analysis of stage I–III colon cancer patients stratified into low- and high-risk groups according to a data-driven threshold derived from the Cox proportional hazards model.

The extracellular vesicles sncRNAs target gene signature supports biological relevance and is associated with overall survival in public TCGA colon cancer cohorts

To support the biological and clinical relevance of the studied sncRNAs, we analyzed the expression of their predicted target genes, which were previously identified and functionally characterized by our group in CAF-derived EVs (42). In that study, a total of 64 target genes regulated by these sncRNAs were defined. Functional enrichment analysis of these genes revealed significant involvement in pathways related to cancer progression, including tyrosine kinase signaling, apoptosis, cell proliferation, and DNA repair (Supplementary Table S3). Based on this enrichment analysis, a prognostic gene signature comprising 43 genes was constructed, corresponding to those included in the ten most significantly represented biological process pathways (Supplementary Table S4).

Gene expression and overall survival data were obtained from a TCGA CC cohort of 597 patients. Among these, 514 patients had overall survival exceeding 36 months, while 83 patients died within 36 months of diagnosis. After data preprocessing, 43 uncorrelated variables were retained. Their individual associations with survival were evaluated using Kruskal-Wallis tests (Supplementary Table S5).

A cross-validated predictive scheme was used for both parameter optimization and performance evaluation. Random Forest model achieved out-of-fold AUC value above 0.80, with specificity approaching 1, indicating a strong ability to identify patients with particularly poor overall survival (Figure 7A). Although a small subset of patients who died within 36

months were classified as low risk, the near-perfect specificity indicates that patients classified by the model as high risk almost invariably experienced death within this time frame. This finding supports the potential utility of the model for identifying patients with particularly unfavorable prognosis.

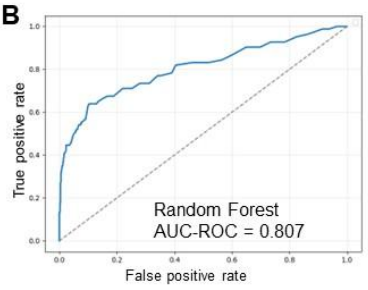
Using the Random Forest model, the out-of-fold AUC was 0.807 (Figure 7B). Ranked out-of-fold predicted probabilities are shown in Figure 7C, while variable importance scores are reported in Supplementary Tables S6.

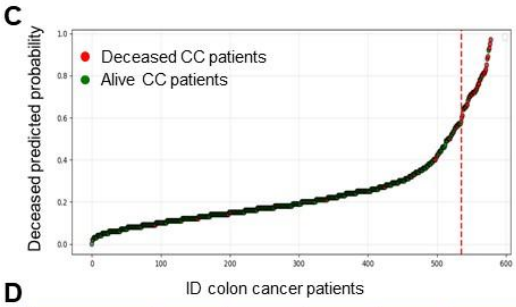
To further evaluate the association between the target gene signature and overall survival, Cox proportional hazards regression analysis was performed. The gene signature showed a strong association with overall survival (HR = 61.42, 95% CI: 25.03–151.41; $p < 0.001$). Based on this model, a data-driven threshold corresponding to patient position 520 of ranked predicted probability was selected for stratification into low- and high-risk groups (Figure 7D). Kaplan–Meier analysis using this threshold showed significant differences in overall survival between risk groups (log-rank $p < 0.001$; Figure 7E).

FIGURE 7

A

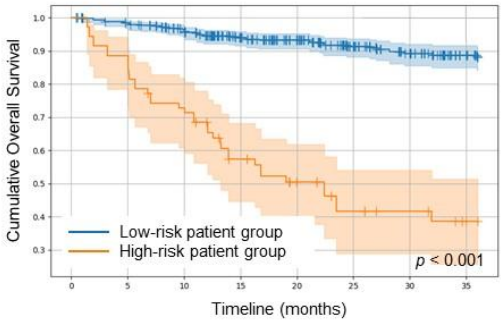
Modelo	ROC_AUC	ACCURACY	F1	SPECIFICITY
Random Forest	0.807±0.063	0.884±0.028	0.532±0.149	0.946±0.017
SVM (RBF kernel)	0.764±0.088	0.889±0.023	0.560±0.120	0.947±0.019
k-NN	0.722±0.079	0.633±0.053	0.329±0.055	0.630±0.072
Logistic Regression	0.634±0.056	0.724±0.020	0.351±0.056	0.753±0.019
Naive Bayes	0.577±0.073	0.663±0.033	0.250±0.077	0.704±0.034





D

Covariate	COEF	HAZARD_ RATIO	SE	P_VALUE	CI_LOWER	CI_UPPER
Snc-RNA-variable	4.12	61.42	0.46	<0.001	25.03	151.41
Stage	0.77	2.16	0.15	<0.001	1.60	2.97
CONCORDANCE	0.85					



Overall survival prediction

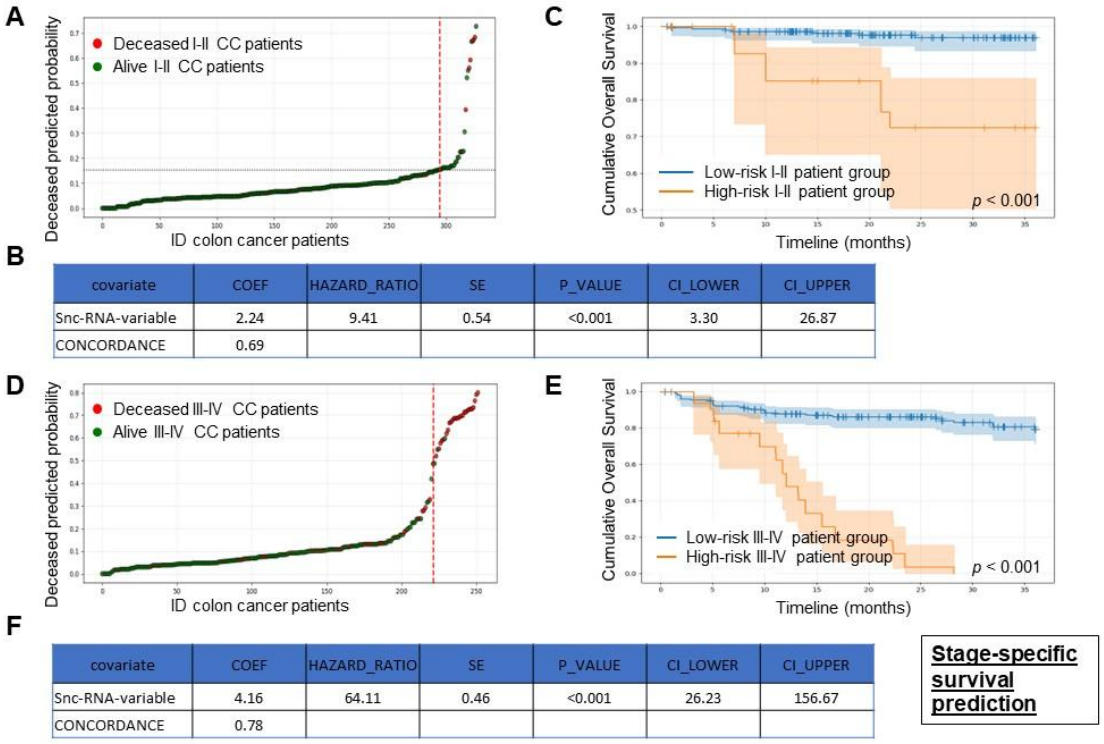
Figure 7. EV sncRNA–based risk modeling identifies colon cancer patients with adverse clinical outcome. (A) Performance metrics of classification models predicting overall survival in a TCGA colon cancer cohort based on the sncRNA-regulated target gene expression. (B) ROC curve of the optimized Ramdom Forest model for overall survival prediction. (C) Ranked predicted probabilities for mortality within 36 months by SVM. (D) Cox proportional hazards regression analysis evaluating the association between the sncRNA-regulated target gene signature and overall survival in the TCGA colon cancer cohort. The data-driven threshold used for patient stratification was derived from this model. (E) Kaplan–Meier overall survival curves stratifying patients into low- and high-risk groups based on the data-driven threshold derived from the Cox proportional hazards model.

To further explore the prognostic performance of the signature across different stages of disease, subgroup analyses were performed according to tumor stage. In patients with stage I–II disease (n = 327, including 14 deaths), Cox proportional hazards analysis showed a significant association between the gene signature and overall survival (HR = 9.41, 95% CI: 3.30–26.87; $p < 0.001$) (Figure 8A and B). Based on this model, a data-driven threshold corresponding to patient position 293 was selected for stratification into low- and high-risk groups. Kaplan–Meier analysis using this threshold also showed significant differences in overall survival between groups (log-rank $p < 0.001$; Figure 8C).

Similarly, in patients with stage III–IV disease (n = 251, including 69 deaths), Cox proportional hazards analysis also demonstrated a significant association with overall survival (HR = 64.11, 95% CI: 26.23–156.67; $p < 0.001$) (Figure 8D and 8E). Based on this model, a threshold corresponding to patient position 220 was selected for risk stratification. Kaplan–Meier analysis using this threshold showed significant differences between low- and high-risk groups (log-rank $p < 0.001$; Figure 8F).

Although this analysis is based on tumor tissue gene expression rather than circulating sncRNAs, it provides independent evidence supporting the biological relevance of the pathways associated with the studied sncRNAs.

FIGURE 8



covariate	COEF	HAZARD_RATIO	SE	P_VALUE	CI_LOWER	CI_UPPER
Snc-RNA-variable	2.24	9.41	0.54	<0.001	3.30	26.87
CONCORDANCE	0.69					

covariate	COEF	HAZARD_RATIO	SE	P_VALUE	CI_LOWER	CI_UPPER
Snc-RNA-variable	4.16	64.11	0.46	<0.001	26.23	156.67
CONCORDANCE	0.78					

Figure 8. Stage-specific prognostic evaluation of the EV sncRNA-regulated target gene signature in the TCGA colon cancer cohort. (A) Ranked predicted probabilities for mortality within 36 months in patients with stage I–II colon cancer. (B) Cox proportional hazards regression analysis evaluating the association between the sncRNA-regulated target gene signature and overall survival in patients with stage I–II colon cancer. The data-driven threshold used for patient stratification was derived from this model. (C) Kaplan–Meier overall survival curves stratifying patients with stage I–II colon cancer into low- and high-risk groups based on the data-driven threshold derived from the Cox proportional hazards model. (D) Ranked predicted probabilities for mortality within 36 months in patients with stage III–IV colon cancer. (E) Cox proportional hazards regression analysis evaluating the association between the sncRNA-regulated target gene signature and overall survival in patients with stage III–IV colon cancer. The data-driven threshold used for patient stratification was derived from this model. (F) Kaplan–Meier overall survival curves stratifying patients with stage III–IV colon cancer into low- and high-risk groups based on the data-driven threshold derived from the Cox proportional hazards model.

DISCUSSION

In this study, we demonstrate that tumor stroma–associated molecular signals carried by extracellular vesicles can be detected in peripheral blood and provide clinically informative signals for diagnostic classification and exploratory patient risk stratification in two highly lethal gastrointestinal malignancies, CC and PDAC. Using three CAF-associated sncRNAs (U1, RNU1-11P and piR-36249), together with the analysis of their predicted target genes in an independent public CC cohort, we show that stromal-derived circulating signals capture biologically relevant features associated with disease presence and patient outcome. Building on our previous work demonstrating that these sncRNAs are deregulated in CAFs and enriched in CAF-derived EVs, this study extends those findings into a clinically relevant liquid biopsy setting and supports the tumor stroma as a measurable source of circulating biomarkers. These findings further demonstrate that tumor stroma–associated molecular signals can be captured through peripheral blood samples, supporting diagnostic classification and exploratory patient risk stratification through minimally invasive liquid biopsy.

Most circulating biomarkers currently used or under evaluation in GI oncology are derived directly from tumor cells, including circulating tumor DNA (ctDNA), tumor-derived EVs, serum protein biomarkers such as carcinoembryonic antigen and CA19-9, and circulating miRNAs (35,45–47). Although these approaches have contributed substantially to clinical management, their performance in early-stage disease remains limited and their specificity may be compromised in non-malignant conditions (48–50). These limitations highlight the value of complementary biomarkers that capture biological processes not reflected by tumor

cell–intrinsic signals. Rather than replacing existing biomarkers, the stromal-derived sncRNA panel described here provides complementary information on tumor-associated stromal activity, broadening the biological landscape that can be interrogated through liquid biopsy.

Within the tumor microenvironment, CAFs actively shape tumor behavior through the transfer of ncRNAs via EVs. CAF-derived EV ncRNAs have been implicated in stemness, invasion, metastasis, and treatment resistance in both CC and PDAC (40,41,51–55). While most studies have focused on miRNAs and long non-coding RNAs, other classes of sncRNAs, including snRNAs and piRNAs, remain comparatively understudied despite increasing evidence linking them to cancer progression and prognosis (56–58). Our previous work identified specific snRNAs and piRNAs enriched in CAF-derived EVs and linked them to distinct gene expression programs in CC (42,43). The present study demonstrates that these stromal-associated sncRNAs are detectable in peripheral blood, providing direct evidence that tumor-associated stromal activity can be non-invasively monitored through liquid biopsy.

Importantly, these markers discriminated not only patients with overt malignancy but also individuals with benign and premalignant colonic lesions. An intriguing observation was the non-linear behavior of RNU1-11P, which showed reduced levels in benign and premalignant lesions compared with both healthy individuals and cancer patients. Although seemingly counterintuitive, this pattern may reflect dynamic changes in stromal activation during tumor development. Early tissue remodeling in the absence of fully activated CAF populations may result in reduced secretion of specific EV-associated signals, whereas progressive stromal activation during malignant transformation could promote their increased release into

559 circulation (22,59). In addition, EV cargo composition is selectively regulated and may vary
560 according to cellular context and disease stage (60,61). Together, these observations suggest
561 that stromal remodeling may occur before overt malignancy develops and may be detectable
562 through circulating EV-associated signals, a finding of particular interest in gastrointestinal
563 oncology where early detection remains a major unmet need.

564 Beyond their diagnostic performance, the three-sncRNA panel also showed exploratory
565 prognostic relevance in CC. Given the limited number of relapse events, the objective was
566 not to develop a clinically deployable predictor but to determine whether circulating stromal-
567 derived sncRNA patterns contain biologically relevant information associated with patient
568 outcome. The association between sncRNA-derived risk estimates and relapse risk, together
569 with the exploratory patient stratification achieved through probability-based modeling,
570 indicates that these markers capture biological information beyond disease detection alone.
571 Additional support came from the analysis of sncRNA-regulated target genes in a large
572 independent TCGA cohort, where the derived transcriptional signature showed robust
573 associations with overall survival. Although this analysis does not constitute an external
574 validation of the circulating biomarker model, it independently supports the biological
575 relevance of the underlying stromal programs.

576 Importantly, all prognostic analyses were performed using peripheral blood samples
577 collected before surgery, indicating that the biological information captured by these
578 biomarkers is available before pathological staging. Although exploratory, these findings
579 support the rationale for evaluating stromal-derived circulating biomarkers as part of future
580 risk-adapted management strategies, including patient surveillance and treatment planning,
581 once validated in larger prospective cohorts. In this context, EVEF-associated stromal

582 sncRNAs represent promising candidates for incorporation into future liquid biopsy–based
583 decision-support approaches.

584 The diagnostic potential of the sncRNA signature extended to PDAC, a malignancy for which
585 early detection remains a major clinical challenge. The ability of the same stromal-associated
586 signature to discriminate both CC and PDAC supports the concept that stromal remodeling
587 constitutes a shared biological feature across distinct gastrointestinal malignancies that can
588 be captured through circulating EV-associated sncRNAs. Although the present study is
589 limited to these two tumor types, these findings provide a rationale for exploring stromal-
590 derived circulating biomarkers across a broader spectrum of gastrointestinal diseases.

591 Several limitations should be acknowledged. The retrospective single-center design and the
592 absence of external validation limit the immediate generalizability of the findings despite the
593 use of nested cross-validation and strict separation of training and validation subsets. In
594 addition, the limited number of benign and premalignant lesions, the relatively small number
595 of relapse events, and the lack of longitudinal follow-up in the PDAC cohort restrict the
596 strength of prognostic conclusions. EV isolation was performed using a polymer-based
597 precipitation method that yields EVEF and may co-isolate non-vesicular components.
598 Furthermore, although previous experimental evidence supports the enrichment of the
599 studied sncRNAs in CAF-derived EVs, the precise cellular origin of circulating signals
600 cannot be definitively established in this clinical setting. Finally, additional functional studies
601 will be required to further clarify the mechanistic roles of these molecules and support future
602 clinical translation.

Despite these limitations, our findings demonstrate that tumor stroma–associated EV snRNAs and piRNAs can be detected in peripheral blood and provide clinically informative circulating signals in gastrointestinal cancer. The consistent associations observed across CC and PDAC support the concept that stromal activation is systemically reflected and can be interrogated through liquid biopsy. More broadly, this work provides proof-of-concept that the tumor stroma constitutes a measurable source of circulating biomarkers, complementing conventional tumor cell–centered approaches and expanding the biological information accessible through minimally invasive sampling. These findings establish stromal-derived EV-associated sncRNAs as a promising class of stromal-derived translational biomarkers and provide a rationale for incorporating stromal biology into future liquid biopsy strategies aimed at improving early detection, pre-operative risk assessment, and personalized patient management in gastrointestinal cancers.

CONCLUSIONS

This study demonstrates that tumor stroma–associated molecular signals carried by extracellular vesicles can be detected in peripheral blood, using three CAF-associated EV sncRNAs (two snRNAs and one piRNA) as proof-of-concept biomarkers. By capturing stromal-derived signals through a minimally invasive liquid biopsy approach, these biomarkers provide complementary information on tumor-associated stromal remodeling that is not reflected by conventional tumor cell–centered strategies.

In CC, the stromal-associated sncRNA panel discriminated healthy individuals from patients with benign and premalignant lesions as well as overt malignancy, supporting the relevance of stromal-derived signals during early stages of colon tumorigenesis. Pre-operative sncRNA levels also showed exploratory associations with relapse risk, while analysis of sncRNA-regulated target genes in a large independent TCGA cohort independently supported the biological and prognostic relevance of the underlying stromal programs. The ability of the same stromal-associated signature to discriminate PDAC patients, a malignancy for which non-invasive biomarkers remain scarce, further supports stromal remodeling as a shared biological feature across distinct gastrointestinal cancers.

Together, these findings establish the tumor stroma as a measurable source of clinically informative circulating biomarkers and identify stromal-derived EV-associated sncRNAs as a promising class of translational biomarkers for gastrointestinal cancer. Although prospective multicenter validation will be required before clinical implementation, this work provides a strong rationale for incorporating stromal biology into future liquid biopsy strategies aimed at improving early detection, pre-operative risk assessment, and personalized patient management.

DECLARATIONS

Ethics approval and consent to participate

The study protocol was approved by the Comité de Ética de la Investigación con Medicamentos of Hospital Universitario Ramón y Cajal (Approval ID: PI20/00602; Act No. 414) and was conducted in accordance with the Declaration of Helsinki.

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

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable scientific request. Publicly available colon cancer gene expression and clinical data were obtained from The Cancer Genome Atlas (TCGA) COADREAD cohort.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by PI20/00602 from the Instituto de Salud Carlos III and co-financed by the European Development Regional Fund (FEDER) “A way to achieve Europe” (ERDF); “CIBER de Cáncer” (CB16/12/00273); “CIBER Enfermedades Raras” (CB06/07/1029), and Biobank and Biomodels Platform PT20/0045 and PT23/00098 from the Instituto de Salud Carlos III - FEDER “A way to achieve Europe” (ERDF); CNS2023-144882 from Agencia Estatal de Investigación “Plan de Recuperación, Transformación y Resiliencia; Next Generation EU”; PID2022-136729OB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER, UE”; and P2022/BMD7212, Comunidad de Madrid. MC was supported by FI21/00132 from Instituto de Salud Carlos III - FEDER “A way to achieve Europe” (ERDF). MJL and JMG-S belong to the Spanish National Research Council (CSIC)'s Cancer Hub.

Author Contributions

MC, VMC, MEC, MH, CGP and CP made substantial contribution to conception and design of the work; the acquisition, analysis and interpretation of the data; and draft and revision of the work. MJL, CLF, LGB, FHdO, JMGS, AC and VGB have contributed to the conception and design of the work. IM, BM, EC, JV and CdlP contributed to the acquisition of the data. All authors reviewed the manuscript.

Acknowledgments

The authors want to particularly acknowledge the patients and the BioBank Hospital Ramón y Cajal-IRYCIS (B.0000678), integrated in the Biobanks and Biomodels Platform of the

ISCIII for its collaboration. We also thank the biobank staff for their professionalism and technical support. We are grateful to lab members for help and advice throughout this research. ChatGPT (OpenAI) was used to assist with English language editing.

SUPPLEMENTARY TABLE

Supplementary Table S1. Pairwise correlation analysis of EVEF sncRNAs expression levels in all studied individuals. *Spearman correlation coefficients and significance values for U1, RNU1-11P, and piR-36249.*

Supplementary Table S2. Distribution of EVEF sncRNA expression levels across all study groups. *Summary statistics for sncRNA expression detected in EV-enriched fractions in healthy donors, patients with benign and premalignant colonic lesions, and colon and pancreatic cancer.*

Supplementary Table S3. Functional enrichment analysis of CAF-associated EV sncRNA target genes. *20 most significantly enriched Gene Ontology biological processes terms.*

Supplementary Table S4. List of predicted target genes regulated by CAF-associated EV sncRNAs used to construct the prognostic gene signature.

Supplementary Table S5. Univariate association of sncRNA target gene expression with overall survival in the TCGA colon cancer cohort. *Kruskal-Wallis test results for genes retained after preprocessing.*

Supplementary Table S6. Variable importance scores for sncRNA target genes derived from the Support Vector Machine predicting overall survival in the TCGA colon cancer cohort.

BIBLIOGRAPHY

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024 May;74(3):229–63. doi:10.3322/CAAC.21834 PubMed PMID: 38572751.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020 Jan;70(1):7–30. doi:10.3322/CAAC.21590 PubMed PMID: 31912902.

3. Van Der Stok EP, Spaander MCW, Grünhagen DJ, Verhoef C, Kuipers EJ. Surveillance after curative treatment for colorectal cancer. Nat Rev Clin Oncol. 2017 May 1;14(5):297–315. doi:10.1038/NRCLINONC.2016.199 PubMed PMID: 27995949.

4. Boute TC, Swartjes H, Greuter MJE, Elferink MAG, van Eekelen R, Vink GR, et al. Cumulative Incidence, Risk Factors, and Overall Survival of Disease Recurrence after Curative Resection of Stage II-III Colorectal Cancer: A Population-based Study. Cancer research communications. 2024 Feb 29;4(2):607–16. doi:10.1158/2767-9764.CRC-23-0512 PubMed PMID: 38363145.

5. Shi Q, Andre T, Grothey A, Yothers G, Hamilton SR, Bot BM, et al. Comparison of outcomes after fluorouracil-based adjuvant therapy for stages II and III colon cancer between 1978 to 1995 and 1996 to 2007: evidence of stage migration from the ACCENT database. J Clin Oncol. 2013 Oct 10;31(29):3656–63. doi:10.1200/JCO.2013.49.4344 PubMed PMID: 23980089.

6. Elferink MAG, de Jong KP, Klaase JM, Siemerink EJ, de Wilt JHW. Metachronous metastases from colorectal cancer: a population-based study in North-East Netherlands. Int J Colorectal Dis. 2015 Feb 1;30(2):205–12. doi:10.1007/S00384-014-2085-6 PubMed PMID: 25503801.

7. Méndez G, Rivera-Matos L, Shuja A. Faecal occult blood testing: a review of its use and common misutilisation. BMJ Open Gastroenterol. 2025 Jul 17;12(1). doi:10.1136/BMJGAST-2025-001876 PubMed PMID: 40675631.

8. Fecal Occult Blood Test - PubMed [Internet]. [cited 2026 Jan 23]. Available from: https://pubmed.ncbi.nlm.nih.gov/30725823/

9. Soin S, Akanbi O, Ahmed A, Kim Y, Pandit S, Wroblewski I, et al. Use and abuse of fecal occult blood tests: a community hospital experience. BMC Gastroenterol. 2019 Sep 3;19(1). doi:10.1186/S12876-019-1079-9 PubMed PMID: 31481027.

10. Fisher DA, Maple JT, Ben-Menachem T, Cash BD, Decker GA, Early DS, et al. Complications of colonoscopy. Gastrointest Endosc. 2011;74(4):745–52. doi:10.1016/j.gie.2011.07.025 PubMed PMID: 21951473.

11. Coelho-Prabhu N, Forbes N, Thosani NC, Storm AC, Pawa S, Kohli DR, et al. Adverse events associated with EGD and EGD-related techniques. Gastrointest Endosc. 2022 Sep 1;96(3):389-401.e1. doi:10.1016/J.GIE.2022.04.024 PubMed PMID: 35843754.

12. Colorectal Cancer — Cancer Stat Facts [Internet]. [cited 2026 Jan 23]. Available from: https://seer.cancer.gov/statfacts/html/colorect.html

13. Udaikumar J, Nimmagadda R, Ingawale S, Lella VV, Vijayakumar K, Faye AS, et al. Current Perspectives on Colorectal Cancer Screening and Surveillance in the Geriatric Population. Curr Gastroenterol Rep. 2025 Dec 1;27(1). doi:10.1007/S11894-025-00985-5 PubMed PMID: 40455318.

14. Shaukat A, Ladabaum U, Kanth P, Lieberman D. AGA Clinical Practice Update on Current Role of Blood Tests for Colorectal Cancer Screening: Commentary. Clinical Gastroenterology and Hepatology. 2025 Aug 1;23(9):1486–91. doi:10.1016/j.cgh.2025.04.003 PubMed PMID: 40267995.

15. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014 Jun 1;74(11):2913–21. doi:10.1158/0008-5472.CAN-14-0155 PubMed PMID: 24840647.

16. Wood LD, Hruban RH. Pathology and molecular genetics of pancreatic neoplasms. Cancer J. 2012 Nov;18(6):492–501. doi:10.1097/PPO.0B013E31827459B6 PubMed PMID: 23187835.

17. Wood LD, Canto MI, Jaffee EM, Simeone DM. Pancreatic Cancer: Pathogenesis, Screening, Diagnosis, and Treatment. Gastroenterology. 2022 Aug 1;163(2):386-402.e1. doi:10.1053/j.gastro.2022.03.056 PubMed PMID: 35398344.

18. He J, Ahuja N, Makary MA, Cameron JL, Eckhauser FE, Choti MA, et al. 2564 resected periampullary adenocarcinomas at a single institution: Trends over three decades. HPB. 2014;16(1):83–90. doi:10.1111/hpb.12078 PubMed PMID: 23472829.

19. Ballehaninna UK, Chamberlain RS. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. J Gastrointest Oncol. 2012 Jun;3(2):105–19. doi:10.3978/j.issn.2078-6891.2011.021 PubMed PMID: 22811878.

20. Lewis D, Jiménez L, Chan KK, Horton S, Wong WWL. A Systematic Review of Cost-Effectiveness Studies on Pancreatic Cancer Screening. Curr Oncol. 2025 Apr 1;32(4). doi:10.3390/CURRONCOL32040225 PubMed PMID: 40277782.

21. Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. Nature Reviews Gastroenterology and Hepatology [Internet]. Nature Publishing Group; 2019 [cited 2020 Oct 26]. p. 282–95. Available from: https://pubmed.ncbi.nlm.nih.gov/30778141/ doi:10.1038/s41575-019-0115-0 PubMed PMID: 30778141.

22. Kalluri R. The biology and function of fibroblasts in cancer. Nature Reviews Cancer [Internet]. Nature Publishing Group; 2016 [cited 2021 Feb 24]. p. 582–98. Available from: https://pubmed.ncbi.nlm.nih.gov/27550820/ doi:10.1038/nrc.2016.73 PubMed PMID: 27550820.

23. Singh AJ, Gray JW. Chemokine signaling in cancer-stroma communications. J Cell Commun Signal. 2021 Sep 1;15(3):361–81. doi:10.1007/S12079-021-00621-7 PubMed PMID: 34086259.

24. Zhang W, Wang J, Liu C, Li Y, Sun C, Wu J, et al. Crosstalk and plasticity driving between cancer-associated fibroblasts and tumor microenvironment: significance of breast cancer metastasis. J Transl Med. 2023 Dec 1;21(1). doi:10.1186/S12967-023-04714-2 PubMed PMID: 37978384.

25. Mesker WE, Liefers GJ, Junggeburt JMC, Van Pelt GW, Alberici P, Kuppen PJK, et al. Presence of a high amount of stroma and downregulation of SMAD4 predict for worse survival for stage I-II colon cancer patients. Cell Oncol. 2009;31(3):169–78. doi:10.3233/CLO-2009-0478 PubMed PMID: 19478385.

26. Downey CL, Simpkins SA, White J, Holliday DL, Jones JL, Jordan LB, et al. The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. Br J Cancer. 2014 Apr 1;110(7):1744–7. doi:10.1038/BJC.2014.69 PubMed PMID: 24548861.

27. He R, Li D, Liu B, Rao J, Meng H, Lin W, et al. The prognostic value of tumor-stromal ratio combined with TNM staging system in esophagus squamous cell carcinoma. J Cancer. 2021 Jan 1;12(4):1105–14. doi:10.7150/JCA.50439 PubMed PMID: 33442408.

28. Zheng H, Liu H, Ge Y, Wang X. Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer. Cancer Cell Int. 2021 Dec 1;21(1). doi:10.1186/S12935-021-02252-9 PubMed PMID: 34670584.

29. Wang F, Li Z, Xu T, Zhang Q, Ma T, Li S, et al. A comprehensive multi-omics analysis identifies a robust scoring system for cancer-associated fibroblasts and intervention targets in colorectal cancer. J Cancer Res Clin Oncol. 2024 Mar 1;150(3). doi:10.1007/S00432-023-05548-7 PubMed PMID: 38478111.

30. Zhang J, Chen M, Fang C, Luo P. A cancer-associated fibroblast gene signature predicts prognosis and therapy response in patients with pancreatic cancer. Front Oncol. 2022 Nov 18;12. doi:10.3389/FONC.2022.1052132/PDF

31. Ucaryilmaz Metin C, Ozcan G. Comprehensive bioinformatic analysis reveals a cancer-associated fibroblast gene signature as a poor prognostic factor and potential therapeutic target in gastric cancer. BMC Cancer. 2022 Dec 1;22(1). doi:10.1186/S12885-022-09736-5 PubMed PMID: 35739492.

32. Si Y, Zhao Z, Meng X, Zhao K. RNA-seq and bulk RNA-seq data analysis of cancer-related fibroblasts (CAF) in LUAD to construct a CAF-based risk signature. Sci Rep. 2024 Dec 1;14(1). doi:10.1038/S41598-024-74336-1 PubMed PMID: 39369095.

33. Liot S, Balas J, Aubert A, Prigent L, Mercier-Gouy P, Verrier B, et al. Stroma Involvement in Pancreatic Ductal Adenocarcinoma: An Overview Focusing on Extracellular Matrix Proteins. Front Immunol. 2021 Apr 6;12. doi:10.3389/fimmu.2021.612271 PubMed PMID: 33889150.

34. Qian C, Ren L, Zhu S, Chen F, Shen H, Hu X, et al. Current perspectives and trends in colorectal cancer and cancer-associated fibroblasts: a review and bibliometric analysis. Front Immunol. 2025;16. doi:10.3389/fimmu.2025.1618742 PubMed PMID: 40677714.

35. Kalluri R, LeBleu VS. The biology , function , and biomedical applications of exosomes. Science. 2020 Feb 7;367(6478). doi:10.1126/SCIENCE.AAU6977 PubMed PMID: 32029601.

36. Mathieu M, Martin-Jaular L, Lavieu G, Théry C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. Nat Cell Biol. 2019 Jan 1;21(1):9–17. doi:10.1038/S41556-018-0250-9 PubMed PMID: 30602770.

838 37. Yoshii S, Hayashi Y, Iijima H, Inoue T, Kimura K, Sakatani A, et al. Exosomal
839 microRNAs derived from colon cancer cells promote tumor progression by
840 suppressing fibroblast TP53 expression. Cancer Sci. 2019. doi:10.1111/cas.14084

841 38. Mito I, Takahashi H, Kawabata-Iwakawa R, Horikawa M, Ida S, Tada H, et al.
842 Tumor-derived exosomes elicit cancer-associated fibroblasts shaping inflammatory
843 tumor microenvironment in head and neck squamous cell carcinoma. Oral Oncol.
844 2023 Jan 1;136. doi:10.1016/j.oraloncology.2022.106270 PubMed PMID:
845 36462328.

846 39. Ringuette Goulet C, Bernard G, Tremblay S, Chabaud S, Bolduc S, Pouliot F.
847 Exosomes Induce Fibroblast Differentiation into Cancer-Associated Fibroblasts
848 through TGFβ Signaling. Mol Cancer Res. 2018;16(7):1196–204. doi:10.1158/1541-
849 7786.MCR-17-0784 PubMed PMID: 29636362.

850 40. Ren J, Ding L, Zhang D, Shi G, Xu Q, Shen S, et al. Carcinoma-associated
851 fibroblasts promote the stemness and chemoresistance of colorectal cancer by
852 transferring exosomal lncRNA H19. Theranostics. 2018;8(14):3932–48.
853 doi:10.7150/thno.25541 PubMed PMID: 30083271.

854 41. Hu JL, Wang W, Lan XL, Zeng ZC, Liang YS, Yan YR, et al. CAFs secreted
855 exosomes promote metastasis and chemotherapy resistance by enhancing cell
856 stemness and epithelial-mesenchymal transition in colorectal cancer. Mol Cancer.
857 2019 May 7;18(1). doi:10.1186/s12943-019-1019-x PubMed PMID: 31064356.

858 42. Herrera M, Llorens C, Rodríguez M, Herrera A, Ramos R, Gil B, et al. Differential
859 distribution and enrichment of non-coding RNAs in exosomes from normal and
860 Cancer-associated fibroblasts in colorectal cancer. Mol Cancer. 2018 Aug
861 3;17(1):114. doi:10.1186/s12943-018-0863-4 PubMed PMID: 30075793.

862 43. Herrera M, Berral-González A, López-Cade I, Galindo-Pumariño C, Bueno-Fortes S,
863 Martín-Merino M, et al. Cancer-associated fibroblast-derived gene signatures
864 determine prognosis in colon cancer patients. Molecular Cancer [Internet]. BioMed
865 Central Ltd; 2021 [cited 2021 Jun 15]. Available from:
866 https://pubmed.ncbi.nlm.nih.gov/33926453/ doi:10.1186/s12943-021-01367-x
867 PubMed PMID: 33926453.

868 44. Mocci E, Guillen-Ponce C, Earl J, Marquez M, Solera J, Salazar-López MT, et al.
869 PanGen-Fam: Spanish registry of hereditary pancreatic cancer. Eur J Cancer. 2015
870 Sep;51(14):1911–7. doi:10.1016/j.ejca.2015.07.004 PubMed PMID: 26212471.

871 45. Hall C, Clarke L, Pal A, Buchwald P, Eglinton T, Wakeman C, et al. A Review of
872 the Role of Carcinoembryonic Antigen in Clinical Practice. Ann Coloproctol.
873 2019;35(6):294–305. doi:10.3393/AC.2019.11.13 PubMed PMID: 31937069.

874 46. Ballehaninna UK, Chamberlain RS. Serum CA 19-9 as a Biomarker for Pancreatic
875 Cancer-A Comprehensive Review. Indian J Surg Oncol. 2011 Jun 1;2(2):88–100.
876 doi:10.1007/S13193-011-0042-1 PubMed PMID: 22693400.

47. Cui M, Wang H, Yao X, Zhang D, Xie Y, Cui R, et al. Circulating MicroRNAs in Cancer: Potential and Challenge. Front Genet. 2019 Jul 18;10. doi:10.3389/FGENE.2019.00626 PubMed PMID: 31379918.

48. Oh HJ, Imam-Aliagan AB, Kim Y Bin, Kim HJ, Izaguirre IA, Sung CK, et al. Clinical applications of circulating biomarkers in non-small cell lung cancer. Front Cell Dev Biol. 2024;12. doi:10.3389/FCELL.2024.1449232 PubMed PMID: 39239557.

49. Duffy MJ, Crown J. Circulating Tumor DNA as a Biomarker for Monitoring Patients with Solid Cancers: Comparison with Standard Protein Biomarkers. Clin Chem. 2022 Nov 1;68(11):1381–90. doi:10.1093/CLINCHEM/HVAC121 PubMed PMID: 35962648.

50. Fiala C, Diamandis EP. Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. BMC Med. 2018 Oct 2;16(1). doi:10.1186/S12916-018-1157-9 PubMed PMID: 30285732.

51. Zhang HW, Shi Y, Liu J Bin, Wang HM, Wang PY, Wu ZJ, et al. Cancer-associated fibroblast-derived exosomal microRNA-24-3p enhances colon cancer cell resistance to MTX by down-regulating CDX2/HEPH axis. J Cell Mol Med. 2021 Apr 1;25(8):3699–713. doi:10.1111/JCMM.15765 PubMed PMID: 33621425.

52. Zhao J, Zhang J, Liu J, Zan L. Exosomal transfer of miR-223-3p from carcinoma-associated fibroblasts promotes the malignant properties and chemoresistance of colon cancer cells by targeting NF2/Hippo signaling. Med Oncol. 2025 Nov 1;42(11). doi:10.1007/S12032-025-03063-Y PubMed PMID: 41026317.

53. Bhome R, Goh RW, Bullock MD, Pillar N, Thirdborough SM, Mellone M, et al. Exosomal microRNAs derived from colorectal cancer-associated fibroblasts: role in driving cancer progression. Aging. 2017;9(12):2666–94. doi:10.18632/aging.101355 PubMed PMID: 29283887.

54. Qi R, Bai Y, Li K, Liu N, Xu Y, Dal E, et al. Cancer-associated fibroblasts suppress ferroptosis and induce gemcitabine resistance in pancreatic cancer cells by secreting exosome-derived ACSL4-targeting miRNAs. Drug Resist Updat. 2023 May 1;68. doi:10.1016/J.DRUP.2023.100960 PubMed PMID: 37003125.

55. Richards KE, Zeleniak AE, Fishel ML, Wu J, Littlepage LE, Hill R. Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. Oncogene. 2017 Mar 30;36(13):1770–8. doi:10.1038/onc.2016.353

56. Escrich V, Romero-Aranda C, López R, de Toro M, Metola Á, Ezcurra B, et al. Unprocessed snRNAs Are a Prognostic Biomarker and Correlate with a Poorer Prognosis in Colorectal Cancer. Cancers (Basel). 2024 Jul 1;16(13). doi:10.3390/CANCERS16132340 PubMed PMID: 39001402.

914 57. Garcia-Borja E, Siegl F, Mateu R, Slaby O, Sedo A, Busek P, et al. Critical appraisal
915 of the piRNA-PIWI axis in cancer and cancer stem cells. Biomark Res. 2024 Dec
916 1;12(1). doi:10.1186/S40364-024-00563-3 PubMed PMID: 38303021.

917 58. Sohn EJ, Han ME, Park YM, Kim YH, Oh SO. The potential of piR-823 as a
918 diagnostic biomarker in oncology: A systematic review. PLoS One. 2023 Dec
919 1;18(12). doi:10.1371/JOURNAL.PONE.0294685 PubMed PMID: 38060527.

920 59. Biffi G, Tuveson DA. Diversity and Biology of Cancer-Associated Fibroblasts.
921 Physiol Rev. 2021 Jan 1;101(1):147–76. doi:10.1152/PHYSREV.00048.2019
922 PubMed PMID: 32466724.

923 60. Van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of
924 extracellular vesicles. Nature Reviews Molecular Cell Biology. Nature Publishing
925 Group; 2018. p. 213–28. doi:10.1038/nrm.2017.125

926 61. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of
927 exosomes. Science [Internet]. American Association for the Advancement of
928 Science; 2020 [cited 2020 Dec 28]. Available from:
929 /pmc/articles/PMC7717626/?report=abstract doi:10.1126/science.aau6977 PubMed
930 PMID: 32029601.

931

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ColladoMetalSupplementaryTables.ods