eLife

Reviewed Preprint
v2 • July 17, 2025
Revised by authors

Reviewed Preprint
v1 • March 11, 2025

Cancer Biology

Multiethnic radiogenomics reveals low-abundancy microRNA signature in plasma-derived extracellular vesicles for early diagnosis and molecular subtyping of pancreatic cancer

Wenjie Shi, Jianying Xu, Yi Zhu, Chao Zhang, Julia Nagelschmitz, Maximilian Doelling, Sara Al-Madhi, Ujjwal Mukund Mahajan, Maciej Pech, Georg Rose, Roland Siegfried Croner, Guo Liang Zheng, Christoph Kahlert, Ulf Dietrich Kahlert

Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany • Department of Medicine II, Hospital of the LMU Munich, Munich, Germany • Department of Gastroenterological Surgery, The Affiliated Hospital of Jiaxing University, Jiaxing, China • Department of Radiology, The First Affiliated Hospital of Wannan Medical College, Wuhu, China • Clinic for Radiology and Nuclear Medicine, University Hospital Magdeburg, Magdeburg, Germany • Research Campus Stimulate, Otto von Guericke University Magdeburg, Magdeburg, Germany • Department of Gastric Surgery, Cancer Hospital of China Medical University (Liaoning Cancer Hospital and Institute), Shenyang, China • Department of General, Visceral and Transplantation Surgery, Heidelberg University Hospital, Heidelberg, Germany

eLife Assessment

The authors attempt to identify which patients with benign lesions will progress to cancer using a liquid biomarker. Although the study is **valuable**, the evidence provided for the liquid biopsy EV miRNA signature developed based on radiomics features remains **incomplete**. There remain key details missing and validation experiments that would better support the conclusions of the study.

https://doi.org/10.7554/eLife.103737.2.sa3

Abstract

Purpose

Pancreatic cancer is a highly aggressive tumor, and early diagnosis and treatment significantly improves the clinical benefit for patients. However, currently there is a lack of effective methods to accurately detect this disease. In our study, we develop a liquid biopsy signature of EV miRNAs based on associated radiomics features of patients' tumors in order to provide new insights for the early diagnosis of pancreatic cancer.

Experimental Design

A total of eight datasets enrolled in this study, featuring clinical and imaging data from different benign pancreatic lesions and malignant pancreatic cancers as well as small RNAseq data from cargo of plasma extracellular vesicles of tumor patients. Radiomics Feature Extraction and different features analysis performed by limma packages. Feature selection was performed by Boruta algorithms and radiomics related signature model was build and validated by lasso regression algorithms. Radiomic signature related to low abundance EV miRNAs was analyzed by weighted gene co-expression network analysis. The diagnosis ability of above miRNA are validated by ten machine-learning algorithms. The shared target of candidate miRNAs were predicted and clustered followed by subsequently probing for predicting survival benefit of the patient, drug sensitivity of tumor cells and functional differences.

Results

A total of 88 significant radiologic features demonstrate differences between benign lesion and pancreatic cancer. Three radiomics factor related signature related a plasma EV-miRNAs triplet possessing high accuracy in diagnosis cancer from benign lesions. Moreover, clustering miRNA and there predicted molecular signaling partners in tumor tissue identified tow molecular subtypes of pancreatic cancer. Cluster stratification separates low risk tumors in terms of severely prolonged overall survival time of patients, higher sensitivity to immune therapies. We also propose the potential of purposing selected targeted drugs to specifically targeting the molecular activation markers in high-risk tumor cluster.

Conclusion

Our three radiogenomics related blood plasma extracellular vesicle microRNA signature is a useful liquid biopsy tool for early diagnosis and molecular subtyping of pancreatic cancer, which might treatment decision making.

Statement of translational relevance

The identification of a low-abundance microRNA signature in plasma-derived extracellular vesicles offers significant translational potential for the early diagnosis and subtyping of pancreatic cancer, particularly across diverse ethnic populations. This discovery could lead to the development of non-invasive liquid biopsies that improve early detection rates, a critical need for a cancer with notoriously poor prognosis due to late diagnosis. By incorporating this microRNA signature into clinical practice, oncologists may be able to detect pancreatic cancer at earlier, more treatable stages, enhancing patient survival rates. Additionally, the subtyping capability of this signature could guide personalized treatment strategies, allowing for more targeted therapies based on specific cancer subtypes. This could ultimately reduce the need for invasive diagnostic procedures and optimize treatment efficacy, reducing adverse effects and improving outcomes. The integration of radiogenomics and liquid biopsy technologies promises to be a powerful tool in the future of cancer medicine, particularly in underserved populations.

Introduction

Pancreatic cancer is one of the most lethal tumors having a poor prognosis and patients suffering from this disease show one of the lowest 5-year overall survival rate of all cancer patients with approximately 13%. One of the main reasons for this dismissal prognosis is the lack of a proper early detection possibility, resulting in late diagnosing often in advanced, metastatic stage[1].

The detection and diagnosis of pancreatic cancer, of which approximately 90% are classified as pancreatic ductal adenocarcinoma (PDAC), currently relies primarily on modalities of medical imaging, such as computer tomography, magnet resonance imaging, positron emission tomography and transabdominal ultrasonography[2]. The most common biomarker considered for PDAC differential diagnosis is elevated blood abundancy of carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA), though these are only used as prognostic markers and not effective for screening or early diagnosis[3]. Nowadays, new markers based on liquid biopsy such as microRNA are discerned and might pose as promising tools for early detection of PDAC. miRNA are non-coding RNAs that target genes and regulate their expression by inhibiting mRNA translation or enhancing their degradation[4]. Currently, extracellular vesicles (EV) are gaining attention as disease specific marker since they carry the material of their secreting cells and are therefore considered to contain tumor-derived elements, showcasing their molecular fingerprint (Bamankar und Londhe 2023). It has been shown multiple times that miRNA derived from small EVs play a role in differentiation and metastasis of cancer[5,6]. The rapidly developing field of data mining and analytical techniques provide new insights and make the discovery of relevant key players more feasible. Numerous miRNA have been described using co-expression network analysis that might be applicated as diagnostic or prognostic biomarker, for patient stratification or disease recurrence[7].

Benefiting from interdisciplinary advances in artificial intelligence, the integration of machine learning and genomics has led to breakthroughs in the early diagnosis and classification of tumors. For example, we used machine learning algorithms to assist in the development of a three-serum miRNA signature that effectively provides early warning of premalignant pancreatic cancer[8]. Another model, based on machine learning algorithms, focuses on the immune subtypes of triple-negative breast cancer, offering critical insights for identifying patients who may benefit from immunotherapy[9]. In the current study, we employed radiogenomics technology, another product of interdisciplinary collaboration between medicine and engineering. This novel approach integrates the quantification of image features from CT or MRI, which are then correlated with genomic signatures and allows for a non-invasive prediction of molecular characteristics[10]. Such as, claim to be able to predict the occurrence of p53 mutations using CT images by radiogenomic analysis and hence make a statement on the prognosis[11].

In the present study, we aim to develop a liquid biopsy signature of EV-derived miRNA based on the radiogenomic analysis of CT images derived from ethical diverse background and interrogating small RNAseq data of plasma-derived total EVs, in order to advance the diagnostic possibilities and the molecular subtyping of PDAC.

Materials and methods

Data resources and pre analytics preparation

A total of four hospitals and four public datasets, comprising a total of eight datasets enrolled in this study, University hospital Magdeburg in Germany (UMMD), and Jiaxing Hospital Center, China (JHC), provided enhanced computed tomography (CT) images of 46 intraductal papillary mucinous neoplasm (IPMN) patients and 127 pancreatic cancer (PC) patients as training dataset. For test datasets, CT data resource from Wanan medical university hospital (WUH) with 27 PB patients and 72 PC patients. University hospital Dresden, Germany (DUH) center provide miRNA and mRNA seq data from total plasma extracellular vesicles (EV) of PDAC patients with associated clinical follow information, including 20 benign pancreatic disease patients and 63 pancreatic cancer patients. The four public serum miRNA sequence data containing pancreatic cancer (PC) and healthy control (HC) were GSE106817(2759 PC vs.115 HC), GSE113486(40 PC vs.100HC), GSE112264(50 PC vs.41 HC), GSE109319(24 PC vs. 21 HC). All CT images were acquired using a 64-slice multidetector spiral CT scanner, with a standard slice thickness of 1.0–1.5 mm and a reconstruction interval of 1 mm. All pancreatic cancer patients underwent a standard pancreatic protocol triphasic contrast-enhanced CT examination. For radiomics analysis, images from the portal venous phase were selected due to their consistent clarity in delineating pancreatic tumor boundaries and surrounding vasculature. To ensure data consistency, all imaging data underwent preprocessing, including resampling, intensity normalization of grayscale values, and N4 bias field correction to address potential low-frequency signal inhomogeneities. We use Propensity Score Matching (PSM) method to match the begin and tumor patients from DUH to UMMD & JHC according to age factor. Afterwards, we constructed a new matrix including CT images, EV miRNAs, mRNAs and patients clinical follow up information. The workflow is schematically depictured in **Figure 1**. Ethical approval to conduct the study for UMMD and DUH after approval by the local Institutional Review Board/ethics committee (UMMD 46/22; 30/01 with amendment 43/14; DUH: EK76032013). Written informed consent from the patients was obtained pre-operatively with the disclosure of research purpose.

EV isolation and RNA sequencing

Details on the protocols for EV isolation including presentation of optical and molecular characteristics of isolated vesicles proofing high quality isolation performance have been described previously by our groups[12].

EV isolation using ultracentrifugation: 500 µl plasma samples were thawed and mixed with 500 µl PBS. The diluted plasma samples were filtered with 0.2 µm filter and subjected to ultracentrifugation at 100,000×*g*, 2 h, 4 °C in a ultracentrifuge (Sorvall MX150 + micro-ultracentrifuge, Thermo Scientific, Darmstadt, Germany). The supernatant was removed and the pellet was washed once with ice-cold Phosphate Buffered Saline (PBS, Gibco, Carlsbad, California, USA) and ultracentrifuged again at 100,000×*g*, 2 h, 4 °C. The resulting pellet was resuspended 100 µl PBS and transferred to Vivaspin ® 500 filtration (100,000 MWCO, Sartorius, Göttingen, Germany) for centrifugation at 15,000×*g*, 45 min. The concentrated EVs were stored at −20 °C until further use for EV characterization.

EV isolation using precipitation method: 550 µl plasma samples were thawed and first centrifuged at 2000×*g* for 20 min. The supernatant was subjected to another round of centrifugation at 10,000×*g* for 20 min. After the second round of centrifugation, 500 µl supernatant was mixed with 250 µl PBS, vortexed and added with 150 µl Exosome Precipitation Reagent. The mixture was incubated at room temperature for 10 min and centrifuged at 10,000×*g* for 5 min. After removing the supernatant completely, the pellet was resuspended with 500 µl PBS and concentrated with Vivaspin ® 500 filtration (100,000 MWCO, Sartorius, Göttingen, Germany) by centrifugation at 15,000×*g*, 45 min. The resulting EVs were stored at −20 °C until further use for EV characterization.

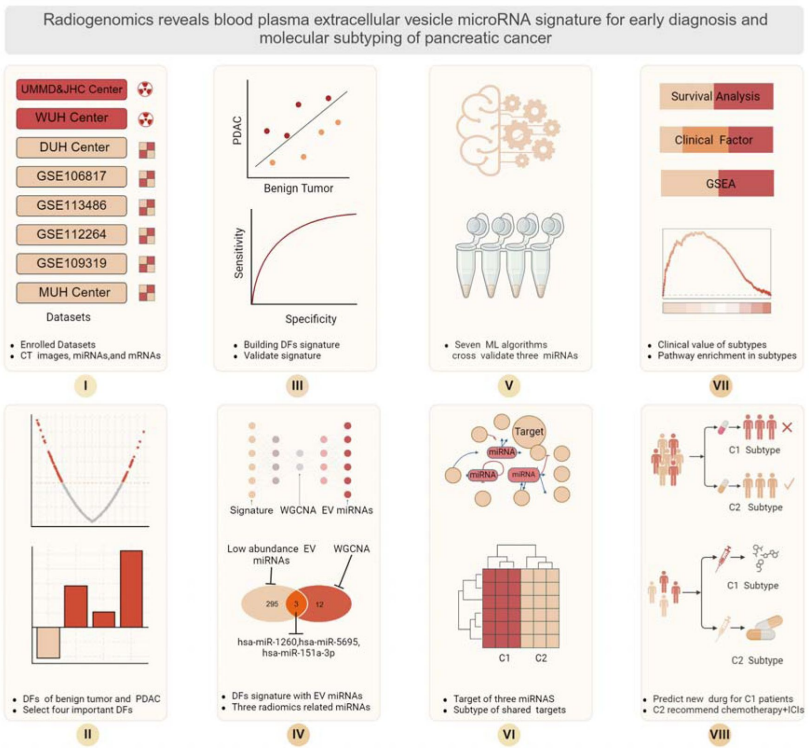


Figure 1

Schematic presentation of the workflow of this study.

The use radiomics to aid finding novel EV charged miRNAs to allow PDAC diagnostics.

RNA sequencing

The eluted EV RNAs were first analyzed for their integrity and concentration using Agilent Fragment Analyzer 5200™ with DNF-472 High Sensitivity RNA Analysis Kit, 15 nt (Agilent Technologies, Santa Clara, California, United States). A range of 1 ng–2 µg RNA was used for complementary DNA (cDNA) synthesis as a preparation for EV RNA sequencing (RNA-seq) libraries with SMARTer smRNA-Seq Kit for Illumina (Takara Bio Inc, Mountain View, California, USA) and were sequenced on an Illumina sequencing platform (NextSeq® 500/550 Mid Output Kit v2, San Diego, California, USA) with run configurations of single read, read 1:51 cycles, index 1:8 cycles, index 2:8 cycles and an average of 3.7 million reads per sample.

Raw reads were first converted from bcl to fastq format using bcl2fastq (v2.20.0.4.422) and subsequently filtered using FastQ Screen to remove potential contaminations by microorganisms or artefacts due to technical issues. The reads were mapped to a phase II reference genome of the 1000 Genomes Project.

Radiomics Feature Extraction and different features (DFs) analysis

We use 3DSlice soft (*www.slicer.org*) to mark pancreatic benign and pancreatic cancer in CT images as mask and use original figure as reference. Subsequently, we use python (Version 3.8) soft to extract radiomics feature, respectively. Afterwards, we use limma packages analysis to conduct different feature analysis to identify the significant features between benign and cancer[13]. The p-value ≤ 0.05 was defined as statistical significant.

Feature selection and radiomics related signature model build

After obtain the DFs, to avoid multicollinearity of the data, we conduct the dimensionality reduction analysis by Boruta algorithms and Lasso regression (LR) model. First, the Boruta algorithms will calculate the importance of each feature and if importance more than shadow feature, it will be selected as important feature, and then, input above important feature into the LR model, the features were selected as significant if the model penalty coefficients were minimized. After selection by machine learning algorithms, we calculate the regression coefficient of each feature in LR model. After that, by combining the feature expression to build the signature model in order to predict the images status from large amount of imaging data. The model formula as listed below, Each regression coefficient of the features is multiplied by its corresponding feature expression level, and then these products are summed. The resulting sum is the Risk Score. The model validation was test with applying WUH center CT dataset, whereas the area under curve (AUC) of Receiver operating characteristic curve (ROC) was used to evaluated the predict ability of model.

Definition and identification of low abundance EV-derived miRNA transcripts

Low abundance miRNAs are defined as the EV miRNAs with the lowest 30% Counts Per Million (CPM) value across all samples. First, we calculate the CPM value for each miRNA across all samples using the edgeR function[14]. Then, the non-zero miRNAs with CPM values ranked in the lowest 30% of all samples were selected and defined as low abundance miRNAs

Weighted gene co-expression network analysis (WGCNA) analysis to identify imaging feature related EV miRNAs

According to the radiomics signature, patients will be spilt into low and high-risk group. Low risk classification means images have high percentage of feature parameters from benign images while high risk more likely become cancer featuring images. To explore the correlation between

above groups with EV miRNAs, we conduct the WGCNA analysis. The soft-thresholding power of WGCNA was automatically defined by the model, which was then assisting in calculating the expression correlation with a) a given miRNA to obtain gene significance (GS), and b) of module membership with miRNAs to obtain module membership (MM). Based on the cut-off criteria (|MM| > 0.5 and |GS| > 0.1), we obtain the significant miRNAs related to low and high risk images, respectively.

Hub EV miRNAs identify and validation in serum and tissues level

We merge WGCNA results and low abundance EV miRNAs to identify the candidate hub miRNAs, then we use GSE109319 to validate this EV miRNAs different expression between healthy participants and PADC patients in serum level. In addition, we also validate this EV miRNAs different expression between pancreatic tissue of healthy individuals and PADC tissues in TCGA-sample cohort.

Screening the best models to aid hub EV miRNAs in blood based diagnosis

We collect serum dataset from GSE106817, GSE113486, GSE112264, our hospital center (UMMD) and GSE109319 dataset to perform this procedure. First, the GSE106817 as training dataset, GSE113486 and GSE112264 are used as independent test dataset. Then we use ten machine learning algorithms (Gradient Boosting Machine GBM ,KNN ,Lasso ,XGBoost ,ENR ,SVM ,LR ,RR ,StepWise ,and QDA) composition to select the best model for prediction. After selection based on highest performance values, we apply our hospital (UMMD) dataset and GSE109319 to validate the model accuracy for ability of clinical prediction.

Common candidate target mRNAs of hub EV miRNAs

To discover the candidate regulation mechanisms of hub EV miRNAs, we use miRPathDB v2.0 database (*https://mpd.bioinf.uni-sb.de/overview.html*) to predict the candidate target mRNAs of hub EV miRNAs. Subsequently we select the pass experiment validation mRNAs from the evidence column, as candidate targets of each miRNAs. then, we explore candidate targets of above which were regulated by this three miRNAs at the same time. Finally, we merge above targets and DHU patient's mRNA sequence to identify the final shared mRNAs which were regulated by three miRNAs at the same time and for the future analysis.

Cluster of shared target mRNAs and survival analysis between different clusters

We use R package ConsensusClusterPlus to perform the cluster analysis of common target mRNAs, and rank the best cluster results according to the Consensus value output received. Afterwards, we also analyze the survival difference between identified subtypes by R package survival. OS and DFS were used to as event endpoint.

Clustering of tumor subtype with clinical factors

To explore the clinical value of each subtype, we analyze the relationship between the subtypes and important clinical factors, such as, age, tumor size, number of positive lymph nodes. We also discover the distribution of sex, peri-neural invasion (PNI), and tumor stage the in different subtypes.

Characterization of immune cell infiltration properties and immune check point activation in each tumor subtype

We use the Microenvironment Cell Populations-counter (MCP-counter) algorithms to calculate immune cell infiltration of each samples and calculated the difference in the two tumor subtypes for each immune cell type separately. We also conduct the relationship between the subtypes and immune check point activation, to predict the candidate subtype could benefit from the immune checkpoint inhibitors.

Potential drug sensitivity for each subtype

We download drug sensitivity data of molecular characterized cell lines to FDA approved, clinical drugs from GDSC database *https://www.cancerrxgene.org/*, and then use pRRophetic package to estimate the drug sensitivity of the two discovered subtypes.

Functional enrichment analysis and pathway prediction

To explore the biological function difference in Biological Process (BP), Cellular Component (CC) and Molecular Function (MF), we conduct the Gene Set Enrichment Analysis (GSEA) analysis. We also use this method to analysis the pathway enrichment difference between the two subtypes. Cluster profile package perform this operation and set p-value ≤ 0.05 as significant enrichment results.

Results

Enrolled population and baseline information for CT images and EV miRNA

A total of 272 patients enrolled this study providing CT imaging, including 173 in UMMD&JHC center while 99 in WUH center. In UMMD&JHC center, also including 46 pancreatic benign lesion and 127 pancreatic cancer CT images. Most of pancreatic patients in this center are older with obesity, but less smoking or alcohol and less with diabetes. For WUH center, most pancreatic cancer are female, and also with a high account for older, and obesity. The DUH provide the patients with EV-miRNAs, EV-mRNAs data and follow up information. About 82.5% (52/63) patients are older, and 34 patients are female. The clinical characteristics of the data cohorts are summarized in **Figure 2**.

Different expression radiomic features between pancreatic benign lesions and aggressive tumors

Before the analysis, we conduct the PSM procedure to match the benign and aggressive tumor from DUH to JHC, respectively, according to age. After match, we found both center baseline difference are removed (**Figure 3A**). Then the difference expression radiomics features was conduct, the results indicate that a total of n=88 significant features demonstrate differences between groups (**Figure 3B**).

Figure 2

The base line information of clinical parameters of patients enrolled from four centers in this multi-center trial.The * means p less than 0.05,** means p less than0.01,and ***means p less than0.001.

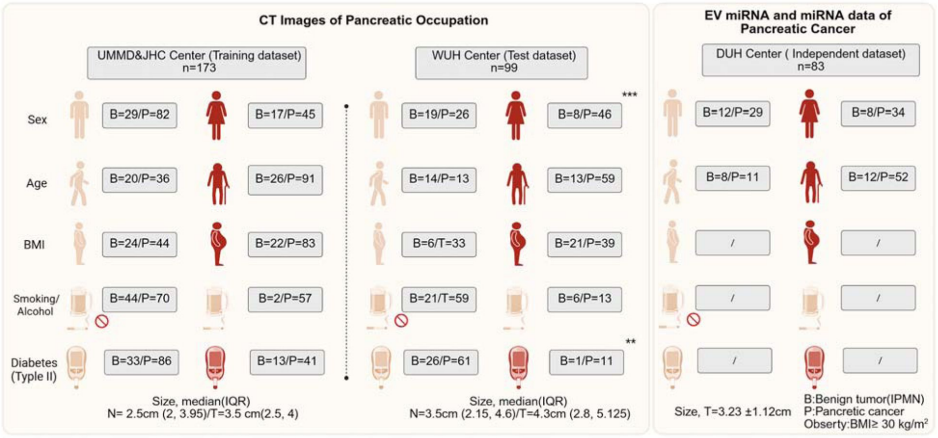
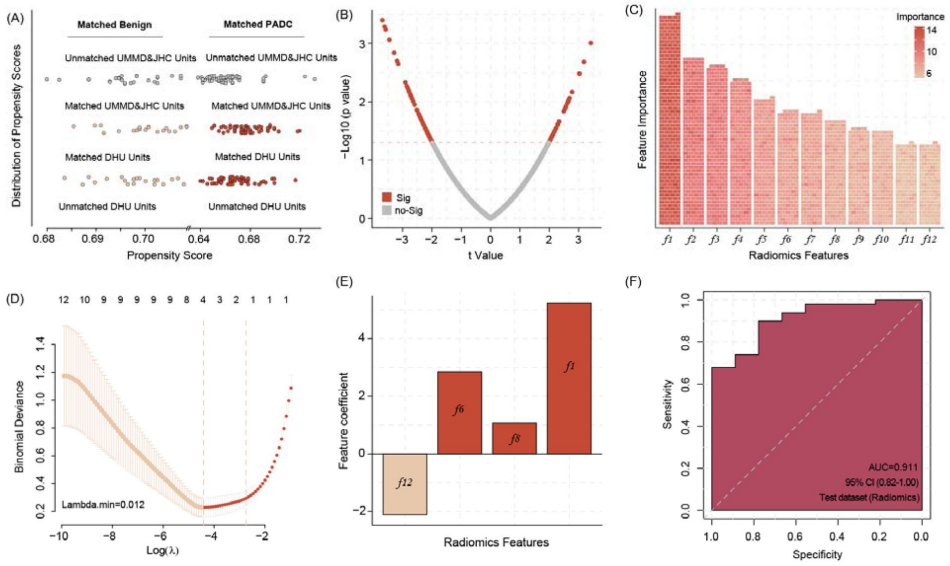


Figure 3

Propensity Score Matching (PSM) allows matching data of benign pancreas lesions and PDAC patients from DUH to UMMD & JHC patients according to the age factor, all of DUH patients successful matched similar patients (A). The different radiomic features between the benign lesions and PDAC patients CT images(B). 12 most important radiomic features differentiating between the benign pancreatic lesion and PDAC patientś CT images identified by the Boruta algorithms (C) Four radiomic features were selected by Lasso Regression to build model signature (D&E). Applying the four radiomic features related signature in image analysis show high accuracy in predicting the PDAC manifestation in the WUH test dataset(F).



Four important radiomic feature was selected to build a related signature

We use Boruta algorithms to select the important features and result show that a total 12 features were identified(**Supplement Table 1**), which more than shadow features. In addition, after input above features into LR algorithms and we can found that four features are list as key features(**Figure 3C-D**). Based on LR model with regression coefficient and feature expression, we build a four radiomics feature related signature and validate the prediction ability in WUH center data, revealing a signature accuracy of prediction efficiency (AUC=0.911) (**Figure 3E-F**).

Three EV miRNAs are associated with radiomic features

After radiomics signature was build, each patients presents an individual risk score and we split patients into high-risk and low-risk patients, according to median value of risk score. We use WGCNA to connect the EV miRNA data and two imaging-featured patient risk groups. According to the WGCNA analysis, the green module was identified as the key module and includes 12 hub co-expressed miRNAs. The number of low abundance of miRNA are calculated was n=295. Merging both results, three miRNAs are identified (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) and selected for subsequent alignment with associated with radiomic features (**Figure 4A-C**).

Expression validation of three EV miRNAs

We use serum and tissues sample to validate the expression of three hub EV miRNAs and the results show that compare with healthy serum sample, this three miRNAs are enriched in tumor patient serum sample. Interestingly, this correlation of upregulation in tumor conditions was also true when comparing tumor tissue with non-tumor tissue (**Figure 4D-I**).

Three EV miRNAs predictd pancreatic cancer with high accuracy

We use seven machine-learning combos to train and test the ability of three EV miRNAs levels to predict tumor manifestation and clinical course of patient. The results show that in the GBM-default (cutoff=0.75) model, three EV miRNAs show a high accuracy to predict the pancreatic cancer with the training accuracy was 0.978 with AUC=0.978, and two test dataset accuracy are 0.923 with AUC=0.919, and 0.871 with AUC=0.857,respectively. Then, we choose GBM model for the extend validation by our hospital data(MUH) and GSE109319.Before this procedure, we use combat package to remove the batch effect allowing the merge of the two datasets. The results of the extend dataset validation via GBM further highlights the high diagnostic accuracy of the three EV miRNAs (accuracy =0.894, and AUC=0.897) (**Figure 5**), while the CA19-9 predict auc is 0.843(95% CI, 0.762−0.924).

Identification of two molecular subtypes of PDAC with significant clinical predictive value

After predicting molecular targets of three miRNAs, we obtain a number of 117 mRNA are shared candidate targets (**Supplement Table 2**). Stratifying the tumors according to the level of expression of those allows the differentiation of tumors in cluster into two subtypes termed Cluster 1 (C1) and C2 (**Figure 6A** and Supplement Figure 1). The survival analysis show that compared with C1, the C2 patients feature a prolonged survival outcome, no matter in OS or DFS(**Figure 6 B&C**). In addition, analysis the relationship with clinical factors of patients, we also found the C1 subtype patients are older in age and carry bigger tumor sizes and higher number of tumor cell positive lymph nodes(**Figure 6D-F**), while compared with C2 subtype

Id	Detail inforamtion of features
f_1	wavelet.HHH_ngtdm_Contrast
f_2	wavelet.HHH_glcm_SumAverage
f_3	wavelet.HHH_glcm_JointAverage
f_4	wavelet.LLL_firstorder_Kurtosis
f_5	wavelet.HLH_firstorder_Uniformity
f_6	wavelet.HHH_firstorder_Minimum
f_7	wavelet.HLH_firstorder_Entropy
f_8	wavelet.LLH_firstorder_Mean
f_9	wavelet.LLH_firstorder_Median
f_{10}	wavelet.HLL_firstorder_Median
f_{11}	log.sigma.1.mm.3D_glszm_GrayLevelNonUniformity
f_{12}	wavelet.HLH_firstorder_10Percentile

Supplement Table 1

Important radiomic features

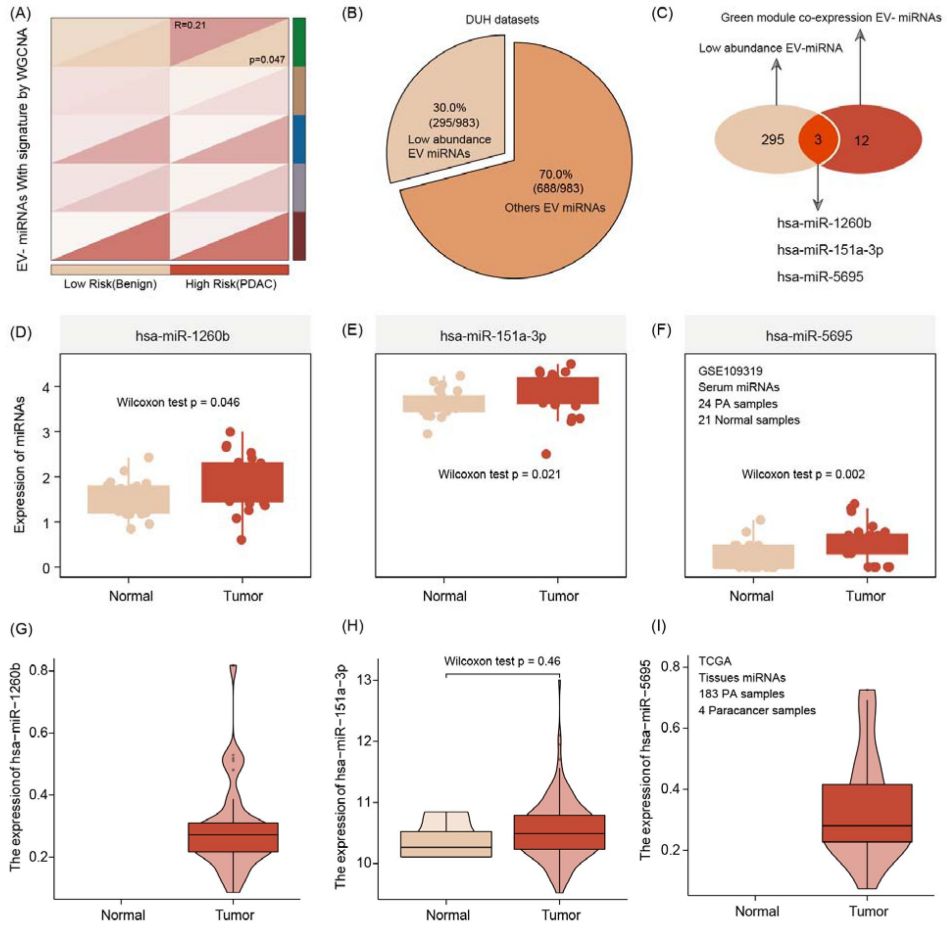


Figure 4

EV miRNA presenting the risk group stratification based on radiomics signature by WGCNA analysis featuring green mode discovering our key module for further analysis (r=0.21, p=0.047) (A). The number of low abundance miRNA in the entire EVseq dataset cohort is n= 295 (B). Out of those low abundance miRNAs, n=12 present matching candidates differentially expressed in high risk group patients. Alignment to our radiomics feature parameters identified three core miRNAs (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) (C). The three key miRNAs show significantly different expression levels in tumor condition, both for serum (D-F) and tissue (G-I).

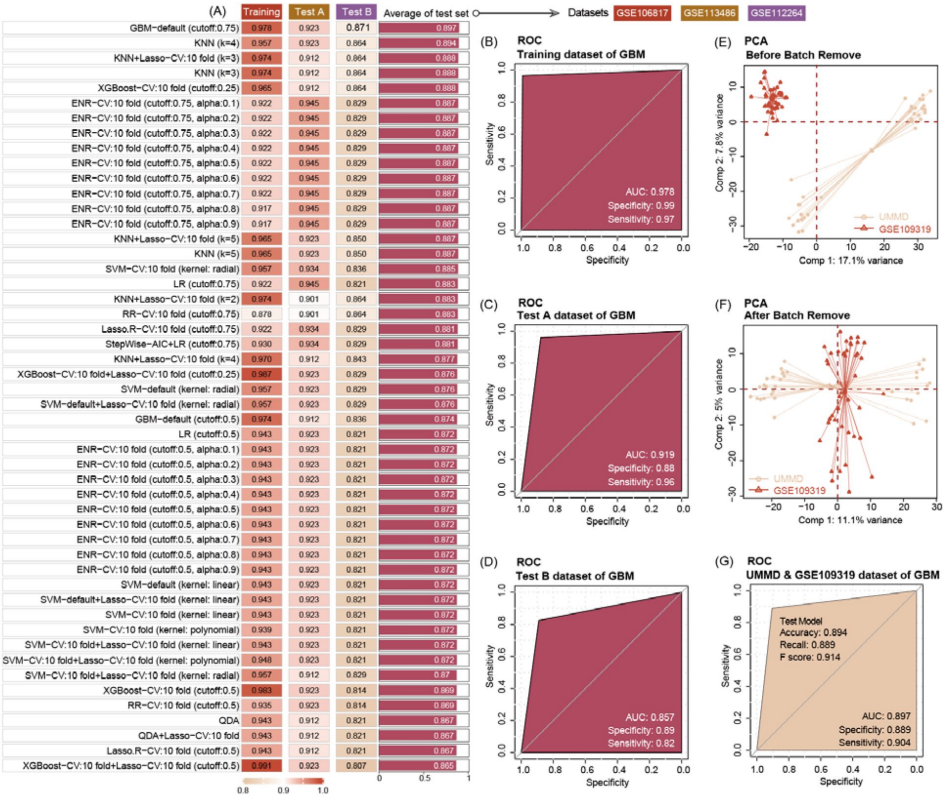


Figure 5

Ten machine learning algorithms demonstrate that three key miRNAs show a high accuracy to diagnosis PDAC in early stage, no matter in training or test datasets. The best machine learning algorithms is GBM (cutoff:0.75) (A). Three miRNAs prediction ability of training dataset (GSE10106817) in GBM model is 0.978(B). Three miRNAs prediction ability of test dataset (GSE113486 and GSE112264) in GBM model is 0.919, and 0.857, respectively(C&D). Data distribution before removal of batch effect of our center data and GSE109319 dataset (E). Data distribution after remove batch effect of our center and GSE109319 dataset (F). Three EV miRNAs prediction ability to identify cancer of our center data and GSE109319 in GBM model is 0.897(G).

patients. We also demonstrate C1 patients particular aggressive in female patients, feature high number of tumors containing perineural invasion (PNI) and advanced pathological tumor staging (**Figure 6 G-I**).

C2 patients with high immune infiltration and positive with immune checkpoints

Testing for gene signatures that indicate immune cell existence, we show that C2 tumors are significantly more enriched for immune cell signals, including those for CD8 T cells, Cytotoxic lymphocytes and NK cells (**Figure 7 A&B**). In addition, samples of patients of that subtype show also a more physiological expression signature in terms of immune checkpoints as compared to C1 tumors (**Figure 7C**).

Prediction of drug sensitivity of C1 patients

We use Genomics of Drug Sensitivity in Cancer (GDSC) data to merge cell line response data and expression data with expression signals in C1 and C2 tumors. This analysis indicated that C1 patients may benefit from therapy with FDA approved AKT inhibitor VIII, Bleomycin, Dasatinib, GNF-2, PF −562271, Refametinib, BMS-509744 (**Figure 7E**).

Functional annotation and pathway enrichment for C1 subtype

GO functional enrichment analysis indicated that the C1 subtype is enriched for intermediate filament organization, GOCC ribosome, as well as GOMF symporter activity (**Figure 8 A&B**). Pathway enrichment analysis showed that the C1 subtype was activated with the Reactome fatty acids, Reactome diseases of metabolism, as well as Reactome biological oxidations pathways, and may be inhibited with Reactome apoptosis, Reactome DNA repair, and Reactome signaling by Hippo pathways (**Figure 8C**).

Discussion

Using multicenter radiomics of Asian and Western world patients, we identified 3 plasma total extracellular vesicle fraction microRNAs that demonstrated strong predictive performance and prognostic value, offering new insights for noninvasive early diagnosis and prognosis of pancreatic cancer. This study is the first to distinguish between benign lesions and pancreatic cancer by identifying novel and significant plasma extracellular vesicle microRNAs based on differential radiomics features. We focused on low-abundance microRNAs, which are often overlooked in routine RNA sequencing analyses. These 3 microRNA candidates were validated across multiple cohorts, achieving excellent predictive performance in both testing and external validation sets. Additionally, these microRNAs demonstrated prognostic value, potentially aiding in treatment selection.

Pancreatic cancer remains one of the most lethal diseases, with insufficient early detection methods and limited treatment options[15,16]. This study aimed to develop a liquid biopsy nucelaic acid diagnsotic test based in EV-derived microRNA using radiogenomic analysis to enhance diagnostic accuracy and molecular subtyping of pancreatic ductal adenocarcinoma. By focusing on imaging features to distinguish benign lesions from pancreatic cancer, we identified underlying genetic factors contributing to these differences. Unlike previous studies on serum microRNA[17,18], our research identified three microRNAs through radiomics that reflect tumor-specific changes rather than systemic variations. Additionally, our findings reveal relationships between imaging features and tumor biology, improving diagnostic accuracy and offering a more comprehensive tool for early detection and personalized disease management.

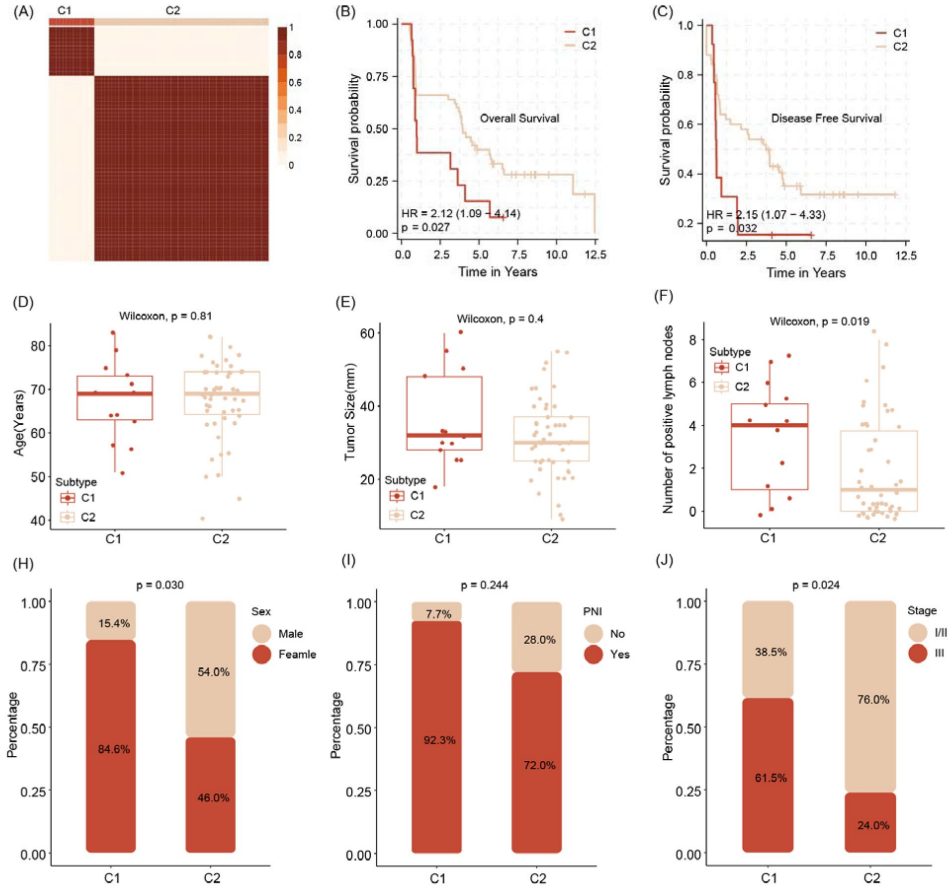


Figure 6

Stratification of abundancy levels of shared mRNAs by Non-negative matrix factorization method allows the clustering of patients into two subtypes (C1 and C2) (A). Patients of the C1 with a poor out outcome in OS and DFS(B&C). C1 subtype patients are characterized with older age and bigger average tumor size, higher number of tumor cell positive lymph nodes as compared to C2 patients (D-F). C1 patients are predominantly female, have tumors with pathological classification marks of perineural invasion and advanced tumor stage (G-I).

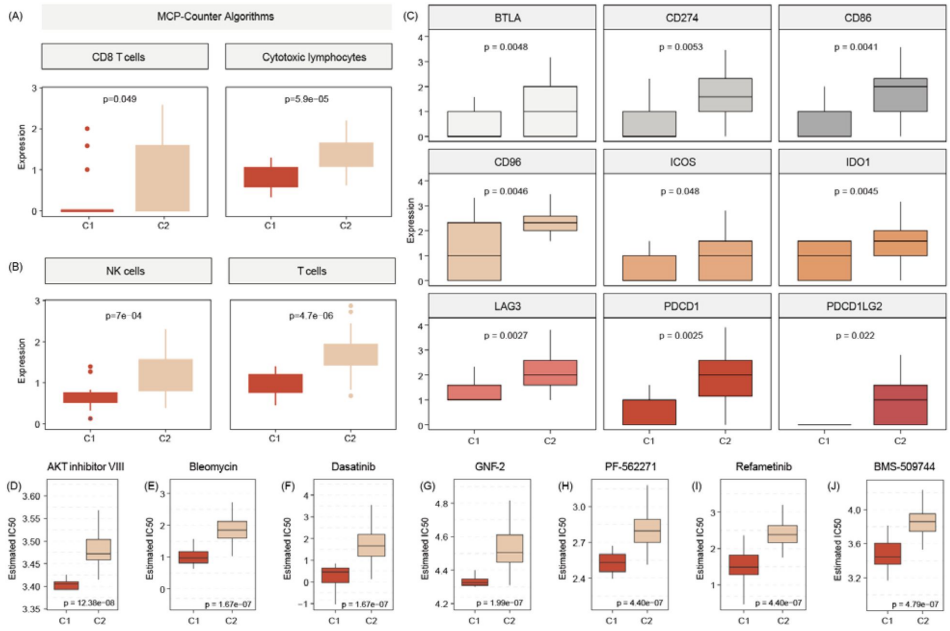
Supplement Table 2

Shared targets of three EV miRNAs

hsa-miR-1260b	hsa-miR-151a-3p	hsa-miR-5695
Candidate Targets		
AATF,AZIN2,CABIN1,CECR2,DENND4B,EIF3B,EMB,FKBP15,GLMN,HADH,IKBKB,MLH3,MRPS21,		
NDE1,PIM3,PLEKHA2,PSMF1,PUM1,RNF125,SAE1,SAP30BP,SLC6A6,TCOF1,USP36,YWHAQ,A SCL5,CCDC140,CHEK2,DAPK2,FAM71F2,GIGYF2,HEATR5A,HIST1H2BJ,MPHOSPH9,NFASC,PC BD2,PRR13,SGK3,SH3PXD2A,SLC22A3,STON2,TOMM40,ZNF398,ZNF791,ABI2,ACAT1,ALDOA,A PLP2,ATAD1,AUTS8,C16orf58,C2orf48,CCT7,CDCA4,CDKN1A,CKB,CLCN2,COL4A3BP,COPA,CT PS1,DDX3Y,DHX30,DNAH10OS,DNMT1,EEF1A1,EIF4G1,ENO1,EPHB4,EZH2,GAN,HIPK1,HNRNP A2B1,HNRNPA3,IKZF2,LGALS3BP,MCM7,MDN1,MKNK2,MRPL24,MRPL37,MYO5C,NBPF15,NCA PG2,NFE2L1,NOL8,OGG1,PALM2,PARP4,PLOD2,POLR2A,POLR3D,PPIAL4A,PRC1,PRKRIP1,PT PRF,RAB19,RAB5IF,RBMXL1,RPE,RPL3,RPL36A,RPL4,RPS15,RPS6,RPS7,SEC61G,SETDB1,SF 3B1,SLC25A3,SLC39A8,SLC3A2,SLC7A5,SOGA3,SP3,TMEM59,USP12,WDR62		

Figure 7

C2 subtype is positively associated with elevated levels of transcripts regulating gene pathways encoding for CD8 T cells, cytotoxic lymphocytes, NK cells(A&B). Commonlyknown immune checkpoints also higher expressed in C2 subtypes(D). Aligning in vitro drug sensitvtiy and expression data from the GDSC database to subtype signature, predicts that tumors of C1 patients might be more sensitive to AKT INHIBITOR VIII, Bleomycin, Dasatinib, GNF-2, PF - 562271, Refametinib, BMS-509744 as C2 subtypes.



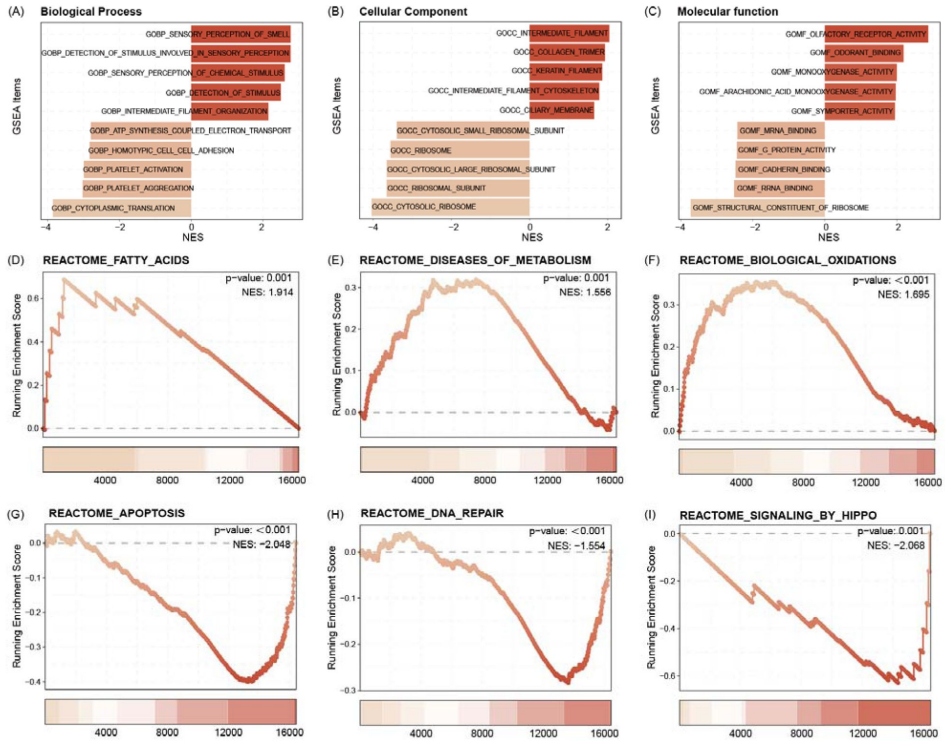


Figure 8

GO functional enrichment analysis indicated that the C1 subtype is enriched for intermediate filament organization, CC ribosome, as well as MF symporter activity(A-C). Pathway enrichment analysis showed that the C1 subtype was activated with the Reactome fatty acids, Reactome diseases of metabolism, as well as Reactome biological oxidations pathways, and may be inhibited with Reactome apoptosis, Reactome DNA repair, and Reactome signaling by Hippo pathways (D-I).

Although numerous studies have reported on microRNA sequencing results[19], highly abundant microRNAs often mask the signals of many others that are present at very low abundance during routine analysis. Furthermore, if a microRNA is associated with tumor development, it is likely to be in low abundance during early detection and may even be undetectable in benign lesions. In our study, after identifying 12 microRNAs related to radiomic differences, we further defined three (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) of them as being of low abundance. These microRNAs showed significantly higher expression in pancreatic cancer blood than in non malignant cancer patients' plasma, with hsa-miR-1260b and hsa-miR-5695 undetectable in pancreatic cancer samples. The model based on these 3 microRNAs demonstrated strong performance in the training group (AUC, 0.978; sensitivity, 0.97; specificity, 0.99), testing group (AUC, 0.919; sensitivity, 0.96; specificity, 0.88), and external testing group (AUC, 0.897; sensitivity, 0.90; specificity, 0.89).

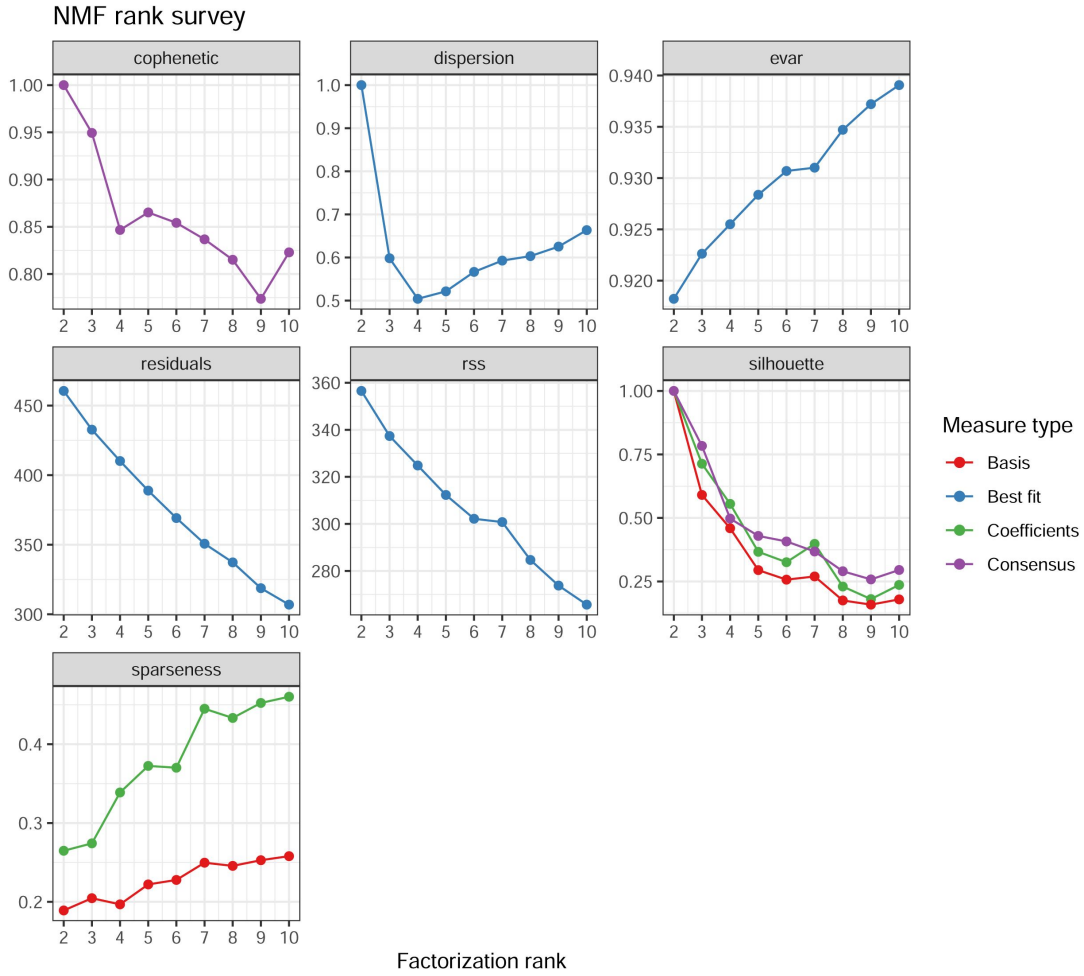
Previous studies on microRNAs as diagnostic biomarkers for pancreatic cancer vs healthy controls report AUCs of 0.78-0.96 and sensitivities of 0.62-0.88[19], but many lack external validation[20,21]. One case-control study identified 2 diagnostic panels based on microRNA expression (index I: 4 microRNAs; AUC, 0.86; sensitivity, 0.85; specificity, 0.64; index II: 10 microRNAs; AUC, 0.93; sensitivity, 0.85; specificity, 0.85) [22]. Our study achieved similar performance using only 3 microRNAs, making it more practical for clinical application.

To the best of our knowledge, hitherto none of the three identified microRNAs have been reported in the context of diagnostic biomarkers in pancreatic cancer. Moreover, to our knowledge, hsa-miR-5695 only indirectly been linked to breast carcinogenesis[23] and subtypes of metastatic prostate cancer[24], while hsa-miR-1260b has been implicated in promoting tumorigenesis in lung cancer[25], breast cancer[26], sarcoma[27] and prostate cancer[28]. Interestingly, in a screen of microRNAs that are differentially expressing in total plasma fraction of patients suffering from intraductal papillary mucinous neoplasm (IPMN) as compared to non-diseased controls, hsa-miR-1260b was identified as one of the top upregulated signals in plasma of IPMN cases[29]. Our work further enforces the potential of hsa-miR-1260b abundancy in plasma to detect onset of abnormal pancreas genesis. Exosomal hsa-miR-151a-3p has emerged as a potential novel biomarker for predicting bone metastasis in lung cancer[30]. A very recent report of results of multicenter trial in Asian participants also identified circulating 2'-O-methylated form of hsa-miR-151a-3p is upregulated in total fraction of plasma of PDAC patients compared to those derived from healthy donors or patients with chronic pancreatitis[31]. Focusing on EV fraction, our results verifies the potential of this newly discovered biomarker in pancreatic cancer blood diagnosis.

Pancreatic cancer currently has only two standard systemic treatments with limited efficacy, and the patient population responsive to immunotherapy and targeted therapy remains unclear[32]. Therefore, we explored the prognostic value of these microRNAs, which allowed for patient stratification. Patients classified as C2 had better prognoses, with higher immune cell infiltration and more immune checkpoint expression, suggesting they may be suitable for immunotherapy after standard treatment. In contrast, patients classified as C1 had poorer prognoses, related to inhibited pathways such as DNA repair and apoptosis.

The findings suggest potential for developing kits for early detection and treatment stratification, and highlight these three candidates for further exploration. However, despite thorough matching of imaging, genomic, and clinical data, biases may still exist. Additionally, using the MCP-counter algorithm for blood samples is a limitation of our study, as it was originally designed for solid tissues, which may affect the accuracy of cell population quantification in the blood. Furthermore, when validating the three microRNAs in tissue, the sample size of paracancerous tissue was limited. Lastly, since our EV purification did not use measures to enrich for cancer cell derived EVs such as CD147 membrane presenting amplitude[33], we cannot exclude contamination of our miRNA data due to signals from non-cancer cell background, which might impact the cancer specificity of our proposed blood test. Moreover, our data is based on OMICS data acquisition. Real

world application of our proposed test would benefit from proofing the predictability and sensitivity of the miRNAs by targeted methods, such as RT-qPCR or CRISPR/Cas diagnostics and in prospective, multicenter trial. In addition, We acknowledge the absence of third-phase validation results, which will limit application in the clinical practice.



Data availability

All data generated from this study, if not included in this article, are available from the corresponding authors on reasonable request.

Additional information

Funding

The MD fellowship program of the medical faculty of University Magdeburg supported the project.

Authors contribution

Data acquisition and experiments, data analysis, layout: WS, IZ, GZ,

Resources: UDK, CK; RSC, GR, MP, GLZ

Data interpretation, writing and review manuscript: all authors

References

1 Bamankar S., Londhe V. Y (2023) **The rise of extracellular vesicles as new age biomarkers in cancer diagnosis: promises and pitfalls** *Technology in Cancer Research & Treatment* **22**:15330338221149266 Google Scholar

2 Wang S., et al. (2021) **The molecular biology of pancreatic adenocarcinoma: translational challenges and clinical perspectives** *Signal transduction and targeted therapy* **6**:249 Google Scholar

3 Bestari M. B., Joewono I. R., Syam A. F (2024) **A Quest for Survival: A Review of the Early Biomarkers of Pancreatic Cancer and the Most Effective Approaches at Present** *Biomolecules* **14**:364 Google Scholar

4 Rashid S., et al. (2024) **Elucidating the Role of miRNA-326 Modulating Hedgehog Signaling in Pancreatic Carcinoma** *Pancreas* **53**:e42–e48 Google Scholar

5 Mok E. T., Chitty J. L., Cox T. R (2024) **miRNAs in pancreatic cancer progression and metastasis** *Clinical & Experimental Metastasis* :1–24 Google Scholar

6 Min L., et al. (2019) **Evaluation of circulating small extracellular vesicles derived miRNAs as biomarkers of early colon cancer: a comparison with plasma total miRNAs** *Journal of extracellular vesicles* **8**:1643670 Google Scholar

7 Tiwari P. K., et al. (2024) **Extracellular Vesicular miRNA in Pancreatic Cancer: From Lab to Therapy** *Cancers* **16**:2179 Google Scholar

8 Shi W., et al. (2024) **Integrating a microRNA signature as a liquid biopsy-based tool for the early diagnosis and prediction of potential therapeutic targets in pancreatic cancer** *British Journal of Cancer* **130**:125–134 Google Scholar

9 Chen Z., et al. (2021) **A machine learning model to predict the triple negative breast cancer immune subtype** *Frontiers in immunology* **12**:749459 Google Scholar

10 Casà C., et al. (2022) **The impact of radiomics in diagnosis and staging of pancreatic cancer** *Therapeutic advances in gastrointestinal endoscopy* **15**:26317745221081596 Google Scholar

11 Iwatate Y., et al. (2020) **Radiogenomics for predicting p53 status, PD-L1 expression, and prognosis with machine learning in pancreatic cancer** *British Journal of Cancer* **123**:1253–1261 Google Scholar

12 Han Y., et al. (2023) **Plasma extracellular vesicle messenger RNA profiling identifies prognostic EV signature for non-invasive risk stratification for survival prediction of patients with pancreatic ductal adenocarcinoma** *Journal of Hematology & Oncology* **16**:7 Google Scholar

13 Ritchie M. E., et al. (2015) **limma powers differential expression analyses for RNA-sequencing and microarray studies** *Nucleic Acids Res* **43**:e47 Google Scholar

14 Robinson M. D., McCarthy D. J., Smyth G. K (2010) **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data** *Bioinformatics* **26**:139–140 Google Scholar

15 Halbrook C. J., Lyssiotis C. A., Pasca di Magliano M., Maitra A. (2023) **Pancreatic cancer: Advances and challenges** *Cell* **186**:1729–1754 Google Scholar

16 Fahrmann J. F., et al. (2021) **Lead-Time Trajectory of CA19-9 as an Anchor Marker for Pancreatic Cancer Early Detection** *Gastroenterology* **160**:1373–1383 Google Scholar

17 Wei J., Yang L., Wu Y. N., Xu J. (2020) **Serum miR-1290 and miR-1246 as Potential Diagnostic Biomarkers of Human Pancreatic Cancer** *J Cancer* **11**:1325–1333 Google Scholar

18 Shams R., et al. (2020) **Identification of potential microRNA panels for pancreatic cancer diagnosis using microarray datasets and bioinformatics methods** *Sci Rep* **10**:7559 Google Scholar

19 Yang J., et al. (2021) **Early screening and diagnosis strategies of pancreatic cancer: a comprehensive review** *Cancer Commun (Lond)* **41**:1257–1274 Google Scholar

20 Liu J., et al. (2012) **Combination of plasma microRNAs with serum CA19-9 for early detection of pancreatic cancer** *Int J Cancer* **131**:683–691 Google Scholar

21 Wang Z (2020) **Diagnostic performance for declined microRNA-133a in pancreatic cancer** *J Cell Biochem* **121**:3882–3886 Google Scholar

22 Schultz N. A., et al. (2014) **MicroRNA biomarkers in whole blood for detection of pancreatic cancer** *Jama* **311**:392–404 Google Scholar

23 Søkilde R., et al. (2019) **Refinement of breast cancer molecular classification by miRNA expression profiles** *BMC genomics* **20**:1–12 Google Scholar

24 Watahiki A., et al. (2011) **MicroRNAs associated with metastatic prostate cancer** *PloS one* **6**:e24950 Google Scholar

25 Kim D. H., et al. (2021) **Exosomal miR-1260b derived from non-small cell lung cancer promotes tumor metastasis through the inhibition of HIPK2** *Cell Death Dis* **12**:747 Google Scholar

26 Park S., et al. (2022) **Promotion of tumorigenesis by miR-1260b-targeting CASP8: Potential diagnostic and prognostic marker for breast cancer** *Cancer Sci* **113**:2097–2108 Google Scholar

27 Morita T., et al. (2020) **Clinical relevance and functional significance of cell-free microRNA-1260b expression profiles in infiltrative myxofibrosarcoma** *Scientific reports* **10**:9414 Google Scholar

28 Hirata H., et al. (2014) **Genistein downregulates onco-miR-1260b and upregulates sFRP1 and Smad4 via demethylation and histone modification in prostate cancer cells** *Br J Cancer* **110**:1645–1654 Google Scholar

29 Permuth-Wey J., et al. (2015) **Plasma MicroRNAs as Novel Biomarkers for Patients with Intraductal Papillary Mucinous Neoplasms of the Pancreas** *Cancer Prev Res (Phila)* **8**:826–34 https://doi.org/10.1158/1940-6207 | PubMed | Google Scholar

30 Zhao K., et al. (2023) **Exosomal hsa-miR-151a-3p and hsa-miR-877-5p are potential novel biomarkers for predicting bone metastasis in lung cancer** *Aging (Albany NY)* **15**:14864–14888 Google Scholar

31 Yang Z., et al. (2024) **Contribution of a Circulating 2'-O-methylated MicroRNA Panel to the Diagnosis of Pancreatic Ductal Adenocarcinoma** *Journal of Cancer* **15**:1583 Google Scholar

32 Neoptolemos J. P., et al. (2018) **Therapeutic developments in pancreatic cancer: current and future perspectives** *Nat Rev Gastroenterol Hepatol* **15**:333–348 Google Scholar

33 Ko S. Y., et al. (2023) **The glycoprotein CD147 defines miRNA-enriched extracellular vesicles that derive from cancer cells** *Journal of Extracellular Vesicles* **12**:12318 Google Scholar

Author information

Wenjie Shi*
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany
ORCID iD: 0009-0001-7345-579X

*equal contribution

Jianying Xu*
Department of Medicine II, Hospital of the LMU Munich, Munich, Germany

*equal contribution

Yi Zhu*
Department of Gastroenterological Surgery, The Affiliated Hospital of Jiaxing University, Jiaxing, China

*equal contribution

Chao Zhang*
Department of Radiology, The First Affiliated Hospital of Wannan Medical College, Wuhu, China

*equal contribution

Julia Nagelschmitz
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany

Maximilian Doelling
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany

Sara Al-Madhi
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany

Ujjwal Mukund Mahajan
Department of Medicine II, Hospital of the LMU Munich, Munich, Germany

Maciej Pech
Clinic for Radiology and Nuclear Medicine, University Hospital Magdeburg, Magdeburg, Germany

Georg Rose
Research Campus Stimulate, Otto von Guericke University Magdeburg, Magdeburg, Germany

Roland Siegfried Croner
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany

Guo Liang Zheng*
Department of Gastric Surgery, Cancer Hospital of China Medical University (Liaoning Cancer Hospital and Institute), Shenyang, China

*equal contribution

Christoph Kahlert*
Department of General, Visceral and Transplantation Surgery, Heidelberg University Hospital, Heidelberg, Germany

*equal contribution

Ulf Dietrich Kahlert*
Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke University, Magdeburg, Germany, Research Campus Stimulate, Otto von Guericke University Magdeburg, Magdeburg, Germany
ORCID iD: 0000-0002-6021-1841

For correspondence: ulf.kahlert@med.ovgu.de
*equal contribution

Editors

Reviewing Editor
Anirban Maitra
University of Texas MD Anderson Cancer Center, Houston, United States of America

Senior Editor
Yamini Dalal
National Cancer Institute, Bethesda, United States of America

Reviewer #1 (Public review):

Summary:

The study aimed to develop a liquid biopsy EV miRNA signature associated with radiomics features for early diagnosis of pancreatic cancer. Flawed study design and inadequate description of clinical characteristics of the enrolled samples makes the findings unconvincing.

Strengths:

The concept of developing EV miRNA signature associated with disease relevant radiomics features is a strength.

Weaknesses:

There are many weaknesses in this manuscript, which include drawing association of data derived from unmatched sample sets, selection of low abundance miRNAs for developing the signature with inadequate rationale, incomplete description of experimental methods and confusing statements in the text.

https://doi.org/10.7554/eLife.103737.2.sa2

Reviewer #2 (Public review):

Summary:

This study investigates a low abundance microRNA signature in extracellular vesicles to subtype pancreatic cancer and for early diagnosis. In this revision, there remain several major and minor issues.

Strengths:

The authors did a comprehensive job with numerous analyses of moderately sized cohorts to describe the clinical and translational significance of their miRNA signature.

Weaknesses:

The weaknesses of the study largely revolve around a lack of clarity about the methodology used and the validation of their findings.

(1) The WGCNA analysis was critical to identify the EV miRNAs associated with imaging features, but the "cut-off criteria" for MM and GS have no clear justification. How were these cut-offs determined? How sensitive were the results to these cut-offs?

(2) The authors now clarify that patients for the sub-study on differentiating early stage from benign pancreatic lesions were matched by age and that the benign pancreatic lesions were predominantly IPMNs. This scientific design is flawed. The CT features extracted likely differentiate solid from cystic pancreatic lesions, and the miRNA signature is doing the same. The authors need to incorporate the following benign controls into their imaging analysis and their EV miRNA analysis: pancreatitis and normal pancreata.

(3) For the radiomics features, the authors should include an additional external validation set to better support the ability to use these features reproducibly, especially given that the segmentation was manual and reliant on specific people.

(4) The DF selection process still lacks cited references as originally requested in the first review.

(5) In Figure 2, more quantitative details are needed in the manuscript. The reviewers failed to incorporate this and only responded in their rebuttal. Add details to the manuscript as originally requested.

(6) It is still not clear what Figure 4A is illustrating as regards to model performance. The authors need to state in the manuscript very clearly what they are showing in the figure and what the modules represent.

(7) Figure 5 and the descriptions for the public serum miRNA datasets need more details. Were these pancreatic cancers all adenocarcinoma, what stage, age range, sex distribution, comorbid conditions were the cases? Were the controls all IPMNs or were there other conditions in the controls?

(8) The subtype results in figures 6 and 7 are not convincing. An association on univariate analysis is not sufficient. The explanation that clinical data is not available to do a multivariable analysis indicates that the authors do not have the ability to claim that they have identified unique subtypes that have clinical relevance. A thorough evaluation of the prognostic significance and the associated molecular features of these tumors is needed.

Summary:

There remain key details and validation experiments to better support the conclusions of the study.

https://doi.org/10.7554/eLife.103737.2.sa1

Author Response:

The following is the authors' response to the original reviews.

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The manuscript by Shi et al, has utilized multiple imaging datasets and one set of samples for analyzing serum EV-miRNAs & EV-RNAs to develop an EV miRNA signature associated with disease-relevant radiomics features for early diagnosis of pancreatic cancer. CT imaging features (in two datasets (UMMD & JHC and WUH) were derived from pancreatic benign disease patients vs pancreatic cancer cases), while circulating EV miRNAs were profiled from samples obtained from a different center (DUH). The EV RNA signature from external public datasets (GSE106817, GSE109319, GSE113486, GSE112264) were analyzed for differences in healthy controls vs pancreatic cancer cases.*

> *The miRNAs were also analyzed in the TCGA tissue miRNA data from normal adjacent tissue vs pancreatic cancer.*
>
> *Strengths:*
>
> *The concept of developing EV miRNA signatures associated with disease relevant radiomics features is a strength.*
>
> *Weaknesses:*
>
> *While the overall concept of developing EV miRNA signature associated with radiomics features is interesting, the findings reported are not convincing for the reasons outlined below:*
>
> *(1) Discrepant datasets for analyzing radiomic features with EV-miRNAs: It is not justified how CT images (UMMD & JHC and WUH) and EV-miRNAs (DUH) on different subjects and centers/cohorts shown in Figures 1 &2 were analyzed for association. It is stated that the samples were matched according to age but there is no information provided for the stages of pancreatic cancer and the kind of benign lesions analyzed in each instance.*

Thank you to the reviewer for the valuable comments. We acknowledge that the radiomics data and EV-miRNA data were derived from different patient cohorts. The primary aim of this study was to explore the integration of data from different omics sources in an exploratory manner to identify potential shared biological features.

We have revised the Methods section accordingly. Regarding the imaging data, we mainly performed batch effect correction on CT images from different centers to eliminate variability. As you correctly pointed out, the EV-miRNA data and CT images from DUH were matched by age. Since all the patients we included had early-stage pancreatic cancer, and the benign pancreatic lesions were predominantly IPMN, we did not specifically highlight this aspect. However, we have now clarified this approach in the data collection section. Thank you for your attention.

> *(2) The study is focused on low-abundance miRNAs with no adequate explanation of the selection criteria for the miRNAs analyzed.*

We used MAD (Median Absolute Deviation) to filter low-abundance miRNAs in the manuscript, as this concept was introduced by us for the first time in this context, and we acknowledge that there is still considerable room for refinement and improvement.

> *(3) While EV-miRNAs were profiled or sequenced (not well described in the Methods section) with two different EV isolation methods, the authors used four public datasets of serum circulating miRNAs to validate the findings. It would be better to show the expression of the three miRNAs in the additional dataset(s) of EV-miRNAs and compare the expressions of the three EV-miRNAs in pancreatic cancer with healthy and benign disease controls.*

Thank you for your suggestion. We have attempted to identify available EV-miRNA datasets; however, due to current limitations in data access, we opted to use serum samples for validation. In our follow-up studies, we are already in the process of collecting relevant EV samples for further validation.

> *(4) It is not clear how the 12 EV-miRNAs in Figure 4C were identified.*

These 12 EV-miRNAs were identified through WGCNA analysis and are associated with the high-risk group.

> *(5) Box plots in Figures 4D-F and G-I of three miRNAs in serum and tissue should show all quantitative data points.*

We have completed the revisions. Kindly review them at your convenience.

> *(6) What is the GBM model in Figure 5?*

Thank you to the reviewer for raising this question. The "GBM model" referred to in Figure 5 is a classification model built using the Gradient Boosting Machine (GBM) algorithm, designed to predict the diagnostic status of pancreatic cancer by integrating EV-miRNA expression and radiomics features. We implemented the model using the GradientBoostingClassifier from the scikit-learn library (version 1.2.2), and optimized the model's hyperparameters—including learning rate, maximum depth, and number of trees—within a five-fold cross-validation framework. The training process and performance evaluation of the model, including the ROC curve and AUC values, are presented in Figure 5.

> *(7) What are the AUCs of individual EV-miRNAs integrated as a panel of three EV-miRNAs?*

Thanks for your comments, Our GBM model integrates the panel of these three EV-miRNAs.

> *(8) The authors could have compared the performance of CA19-9 with that of the three EV-miRNAs.*

Since our main focus is on the panel of three EV-miRNAs, we did not present the AUC for each individual miRNA separately. However, we have included the performance of CA19-9 in our dataset as a reference. The predictive AUC for CA19-9 is 0.843 (95% CI, 0.762–0.924).

> *(9) How was the diagnostic performance of the three EV-miRNAs in the two molecular subtypes identified in Figure 6&7? Do the C1 & C2 clusters correlate with the classical/basal subtypes, staging, and imaging features?*

Thank you to the reviewer for raising this important question. In fact, our EV panel is primarily designed to distinguish between normal and tumor samples, whereas both C1 and C2 represent tumor subtypes, and thus the panel is not applicable for diagnostic purposes in this context. Additionally, our subtypes are novel and do not align with the conventional classical and basal-like gene expression profiles. Furthermore, the C1 subtype is more frequently observed in stage III tumors (Figure 6J) and is associated with distinct imaging features such as higher texture heterogeneity and lower CT density.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *This study investigates a low abundance microRNA signature in extracellular vesicles to subtype pancreatic cancer and for early diagnosis. There are several major questions that need to be addressed. Numerous minor issues are also present.*
>
> *Strengths:*
>
> *The authors did a comprehensive job with numerous analyses of moderately sized cohorts to describe the clinical and translational significance of their miRNA signature.*
>
> *Weaknesses:*

> *There are multiple weaknesses of this study that should be addressed:*
>
> *(1) The description of the datasets in the Materials and Methods lacks details. What were the benign lesions from the various hospital datasets? What were the healthy controls from the public datasets? No pancreatic lesions? No pancreatic cancer? Any cancer history or other comorbid conditions? Please define these better.*

We sincerely thank the reviewer for the detailed and important suggestions regarding sample definition. Indeed, the source of the datasets and the definition of control groups are critical for ensuring the rigor and interpretability of the study. In response to this comment, we have added clarifications in the revised "Materials and Methods" section.

First, for the benign lesion group derived from various clinical centers (DUH, UMMD, WUH, etc.), we have carefully reviewed the pathological and clinical records and defined these samples as histologically confirmed non-malignant pancreatic lesions, primarily IPMN. All patients in the benign lesion group had no diagnosis of pancreatic cancer at the time of sample collection, and for cohorts with available follow-up data, no evidence of malignant progression was observed within at least six months.

Second, the healthy control group from public databases was derived from healthy individuals.

Finally, to eliminate potential confounding factors, we excluded any samples with a history of other malignancies (e.g., breast cancer, colorectal cancer, etc.) from all datasets with available clinical information, to ensure the specificity of the EV-miRNA expression analysis.

> *(2) It is unclear how many of the controls and cases had both imaging for radiomics and blood for biomarkers.*

Due to limitations in resource availability, our study does not include samples with both CT imaging and serological data from the same individuals. Instead, we integrated blood samples and CT imaging data collected from different clinical centers.

> *(3) The authors should define the imaging methods and protocols used in more detail. For the CT scans, what slice thickness? Was a pancreatic protocol used? What phase of contrast is used (arterial, portal venous, non-contrast)? Any normalization or pre-processing?*

Thank you to the reviewer for the professional suggestions regarding the imaging section. We have added detailed technical information on CT imaging in the revised Materials and Methods section. All CT images were acquired using a 64-slice multidetector spiral CT scanner, with a standard slice thickness of 1.0–1.5 mm and a reconstruction interval of 1 mm. All pancreatic cancer patients underwent a standard pancreatic protocol triphasic contrast-enhanced CT examination, which included non-contrast, arterial phase (approximately 25–30 seconds), and portal venous phase (approximately 65–70 seconds) imaging.

For the radiomics analysis, images from the portal venous phase were selected, as this phase provides consistent clarity in delineating tumor boundaries and surrounding vasculature. To ensure data consistency, all imaging data underwent preprocessing, including resampling, intensity normalization of grayscale values (standardized using z-score normalization to a mean of 0 and a standard deviation of 1), and N4 bias field correction to address potential low-frequency signal inhomogeneities.

> *(4) Who performed the segmentation of the lesions? An experienced pancreatic radiologist? A student? How did the investigators ensure that the definition of the lesions*

> *was performed correctly? Raidomics features are often sensitive to the segmentation definitions.*

All lesion segmentations were performed on portal venous phase contrast-enhanced CT images. Manual delineation was conducted using 3D Slicer (version 4.11) by two radiologists with extensive experience in pancreatic tumor diagnosis. A consensus was reached between the two radiologists on the ROI definition criteria prior to analysis.

To further assess the robustness of radiomic features to segmentation boundary variations, we selected a subset of representative cases and created “expanded/shrunk ROIs” by adding or subtracting a 2-pixel margin at the lesion boundary. Feature extraction was then repeated, and the coefficient of variation (CV) for the main features included in the model was found to be below 10%, indicating that the model is stable with respect to minor boundary fluctuations.

> *(5) Figure 1 is full of vague images that do not convey the study design well. Numbers from each of the datasets, a summary of what data was used for training and for validation, definitions of all of the abbreviations, references to the Roman numerals embedded within the figure, and better labeling of the various embedded graphs are needed. It is not clear whether the graphs are real results or just artwork to convey a concept. I suspect that they are just artwork, but this remains unclear.*

We thank the reviewer for the detailed feedback on Figure 1. We would like to clarify that Figure 1 is a conceptual schematic intended to visually illustrate the overall design of the study, the relationships among different data modules, and the logical sequence of the analytical strategy. It is not meant to present actual results or quantitative details.

Regarding the reviewer’s concerns about sample sizes, the division between training and validation cohorts, explanations of specific abbreviations, and the precise meaning of each panel, we have provided comprehensive and detailed clarifications in Figure 2.

> *(6) The DF selection process lacks important details. Please reference your methods with the Boruta and Lasso models. Please explain what machine learning algorithms were used. There is a reference in the "Feature selection.." section of "the model formula listed below" but I do not see a model formula below this paragraph.*

We thank the reviewer for the thoughtful and detailed comments on the feature selection strategy. We first applied the Boruta algorithm (based on random forests, implemented using the Boruta R package) to the original feature set—which included both radiomics and EV-miRNA features—to identify variables that consistently demonstrated importance across multiple rounds of random resampling.

Subsequently, we used LASSO regression with five-fold cross-validation to further reduce the dimensionality of the Boruta-selected features and to construct the final feature set used for modeling. The formula for the model is as follows: each regression coefficient is multiplied by the corresponding feature expression level, and the resulting products are summed to generate the Risk Score.

> *(7) In Figure 2, more quantitative details are needed. How are patients dichotomized into non-obese and obese? What does alcohol/smoking mean? Is it simply no to both versus one or the other as yes? These two risk factors should be separated and pack years of smoking should be reported. The details of alcohol use should also be provided. Is it an alcohol abuse history? Any alcohol use, including social drinking? Similarly, "diabetes" needs to be better explained. Type I, type II, type 3c? P values should be shown to*

> *demonstrate any statistically significant differences in the proportions of the patients from one dataset to another.*

Our definition of obesity was based on the standard BMI threshold (30 kg/m²). A history of smoking or alcohol consumption was defined as continuous use for more than one year. Specific details regarding smoking and alcohol use were recorded at baseline under the category of “smoking/alcohol history”; unfortunately, we did not collect follow-up data on these variables. As for diabetes, only type II diabetes was documented. Statistically significant p-values have been added. Thank you.

> *(8) In the section "Different expression radiomic features between pancreatic benign lesions and aggressive tumors", there is a reference to "MUJH" for the first time. What is this? There is also the first reference to "aggressive tumors" in the section. Do the authors just mean the cases? Otherwise there is no clear definition of "aggressive" (vs. indolent) pancreatic cancer. This terminology of tumor "aggressiveness" either needs to be removed or better defined.*

We have corrected the abbreviation (MUJH); it should in fact be JHC. Additionally, regarding the term "aggressive," we have reviewed the literature and used it to convey the highly malignant nature of pancreatic cancer.

> *(9) Figure 3 needs to have the specific radiomic features defined and how these features were calculated. Labeling them as just f1, f2, etc is not sufficient for another group to replicate the results independently.*

We have presented these features in Supplementary Table 1. Kindly refer to it for details.

> *(10) It is not clear what Figure 4A illustrates as regards model performance. What do the different colors represent, and what are the models used here? This is very confusing.*

This represents the correlation between WGCNA modules and miRNAs. Different module colors indicate distinct miRNA clusters—for example, the green module contains 12 miRNAs grouped together. The colors themselves do not carry any intrinsic meaning.

> *(11) Figure 5 shows results for many more model runs than the described 10, please explain what you are trying to convey with each row. What are "Test A" and "Test B"? There is no description in the manuscript of what these represent. In the figure caption, there is a reference to "our center data" which is not clear. Be more specific about what that data is.*

We have indicated this using arrows in Figure 5 from Test A/B/C. Please check.

> *(12) Figure 6 describes the subtypes identified in this study, but the authors do not show a multi-variable cox proportional hazards model to show that this subtype classification independently predicts DFS and OS when incorporating confounding variables. This is essential to show the subtypes are clinically relevant. In particular, the authors need to account for the stage of the patients, and receipt of chemotherapy, surgery, and radiation. If surgery was done, we need to know whether they had R1 or R0 resection. The details about the years in which patients were included is also important.*

We sincerely thank the reviewer for this critical comment. We fully agree that incorporating a multivariate Cox proportional hazards model to control for potential confounding factors would provide a more robust validation of the independent prognostic value of our proposed subtypes for DFS and OS.

However, as the clinical data used in this study were retrospectively collected and access to certain variables is currently restricted, we were only able to obtain limited clinical information. At this stage, we are unable to systematically include key variables such as tumor staging, adjuvant chemoradiotherapy regimens, and resection margin status (R0 vs. R1), which prevents us from performing a rigorous multivariate Cox analysis.

Similarly, regarding the postoperative resection status, after reviewing the original surgical reports and pathology records, we regret to confirm that margin status (R0 vs. R1) is missing in a substantial portion of cases, making it unsuitable for reliable statistical analysis.

We fully acknowledge this as a limitation of the current study and have explicitly addressed it in the Discussion section. To address this gap, we are currently designing a more comprehensive prospective cohort study, which will allow us to validate the clinical independence and utility of the proposed subtypes in future research.

> *(13) How do these subtypes compare to other published subtypes?*

We sincerely thank the reviewer for raising this important point. Clusters 1 and 2 represent a novel molecular classification proposed for the first time in this study, driven by EV-miRNA profiles. This classification approach is conceptually independent from traditional transcriptome-based subtyping systems, such as the classical/basal-like subtypes, as well as other existing classification schemes. Comparisons with previously reported subtypes and validation of clinical relevance will require further investigation in future studies.

> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *The authors appear to be attempting to identify which patients with benign lesions will progress to cancer using a liquid biomarker. They used radiomics and EV miRNAs in order to assess this.*
>
> *Strengths:*
>
> *It is a strength that there are multiple test datasets. Data is batch-corrected. A relatively large number of patients is included. Only 3 miRNAs are needed to obtain their sensitivity and specificity scores.*
>
> *Weaknesses:*
>
> *This manuscript is not clearly written, making interpretation of the quality and rigor of the data very difficult. There is no indication from the methods that the patients in their cohorts who are pancreatic cancer patients (from the CT images) had prior benign lesions, limiting the power of their analysis. The data regarding the cluster subtypes is very confusing. There is no discussion or comparison if these two clusters are just representing classical and basal subtypes (which have been well described).*

Sorry,we don't have the data of record from patients, in addition, Regarding the relationship between Cluster 1/Cluster 2 and classical subtypes:We are very grateful for the reviewer's insightful question. We would like to clarify that Clusters 1 and 2, as shown in Figures 6 and 7, are derived from a novel EV-miRNA–driven molecular classification proposed for the first time in this study. This classification system is constructed independently of the traditional transcriptome-based classical/basal-like subtypes.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *There are errors in reference citations and several typos, misspellings, and grammatical errors throughout the manuscript.*

We have made the necessary revisions.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) Were the radiomic features associated with the subtypes and prognostic in the subset of patients who had CT scans?*

Unfortunately, there are no corresponding CT imaging results available for these cases, as the genes were identified based on predicted miRNA targets and were not derived from patients who had undergone CT scans.

> *(2) There is a whole body of literature on prognostic imaging-based subtypes of pancreatic cancer that needs to be cited.*

Thank you for your suggestion. We have cited the relevant references accordingly in the manuscript.

> *(3) Similarly, the authors should be more comprehensive about prognostic and early detection markers for miRNAs for pancreatic cancer. Early detection markers really should be described separately from prognostic markers. The authors did not do a PROBE phase 3 study, so early detection is not really relevant. Please see https://edrn.nci.nih.gov/about-edrn/five-phase-approach-and-prospective-specimen-collection-retrospective-blinded-evaluation-study-design/*

The primary objective of our study is early detection. We acknowledge the absence of third-phase validation results, which we will address in the limitations section. Additionally, the subtype classification represents our secondary objective.

> *(4) If they want to couch this as a PROBE phase 2 study, then they should review the PROBE guidelines and ensure they are meeting standards. Many of the comments above regarding methodologies, definitions, and patient cohort descriptions would address this concern.*

We have revised the Methods section accordingly. Please kindly review the updated version.

> *(5) The entire manuscript needs to have a review for the use of the English language. There are numerous typos and grammatical errors that make this manuscript difficult to follow and hard to interpret.*

We have revised the Methods section accordingly. Please kindly review the updated version.

> *(6) In the section on "Definition and identification of low abundance EV-derived miRNA transcripts", provide a reference for the "edger" function.*

We have revised the Methods section accordingly. Please kindly review the updated version.

> *(7) In the Abstract: The purpose section only mentions early diagnosis as the goal of this study. It seems subtyping is also a major goal, but it is not mentioned.*

The primary objective of our study is early detection.Additionally, the subtype classification represents our secondary objective.so,we didn't add it in the purpose.

> *(8) The experimental design fails to describe any of the 8 datasets that were used. How many patients? What were the ethnic and racial backgrounds, which is one of the key aspects of this study and mentioned in the title? What range of stages? When were the images and the blood collected in relation to diagnosis? Over what time frame were the patients included? What patients were excluded, if any? These details are important to understand the materials used, along with the methods to design the signatures and models.*

We have revised the Methods section accordingly. Please kindly review the updated version.

> *(9) Again, the purpose section of the abstract does not align with the rest of the study, including the description of the experimental design. The last sentence of the experimental design section mentions predicting drug sensitivity and survival, which is unrelated to the aim of early diagnosis.*

We have revised the Methods section accordingly. Please kindly review the updated version.

> *(10) The results section lacks key details to indicate the impact of the work. Vague descriptions of the findings are not sufficient. The performance of the biomarkers to differentiate benign from malignant lesions, hazard ratios, survival times, and p values should be reported for key results.*

Our aim was to develop an integrated panel for diagnostic purposes; therefore, we provided the AUC to evaluate its performance. However, since this is a diagnostic model, we did not include hazard ratios or survival time data.

> *(11) What are "tow" molecular subtypes of pancreatic cancer? Did you mean "two"? What system was used to subtype the pancreatic cancers? Is some new subtyping or a previously published method to subtype the disease?*

Yes, it means two, previously published method.In method part, we have describe it.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *The writing of this manuscript needs extensive re-wording and clarification to increase the readability and interpretability of the data presented. The authors could include a dataset of pancreatic cancer patient imaging data where the status of prior benign lesions was detected (as opposed to patients with benign lesions that do not develop pancreatic cancer). The authors could also address if their clusters 1 and 2 are representing (or are correlated with) the classical and basal subtypes that have been well described for pancreatic cancer.*

Thank you to the reviewer for the constructive comments. We sincerely appreciate your careful review, particularly regarding language clarity, data interpretability, and subtype correlation. To enhance the readability and scientific precision of the manuscript, we have conducted a thorough revision and language polishing throughout the text, improving logical structure, terminology consistency, and clarity in result descriptions. We have especially reinforced the Methods and Discussion sections to better explain key analytical steps and data interpretation.

We fully understand the reviewer's suggestion to include information on "the presence of benign lesions prior to pancreatic cancer diagnosis." However, due to the retrospective nature of our study, the current imaging and EV-miRNA datasets do not contain systematically collected follow-up annotations of this type. Therefore, it is not feasible to incorporate such data into the present manuscript.

That said, we fully recognize the importance of this direction. In future studies, we plan to evaluate longitudinal samples to investigate the dynamic changes in EV-miRNAs and imaging features during the progression from premalignant to malignant states, aiming to clarify their potential value for early cancer warning.

Regarding the relationship between Cluster 1/Cluster 2 and classical subtypes:We are very grateful for the reviewer's insightful question. We would like to clarify that Clusters 1 and 2, as shown in Figures 6 and 7, are derived from a novel EV-miRNA–driven molecular classification proposed for the first time in this study. This classification system is constructed independently of the traditional transcriptome-based classical/basal-like subtypes.

Although we attempted a cross-comparison with existing TCGA subtypes, differences in data origin, analysis modality (EV-miRNA vs. tissue transcriptome), and limitations in sample matching prevent us from establishing a direct correspondence. In the revised Discussion, we have emphasized that these two classification approaches are complementary rather than equivalent, reflecting different dimensions of tumor heterogeneity. Further integrative multi-omics studies will be needed to validate their biological significance and clinical utility.

https://doi.org/10.7554/eLife.103737.2.sa0