

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

3 Wenjie Shi<sup>1,\*</sup>, Jianying Xu<sup>2,\*</sup>, Yi Zhu<sup>3,\*</sup>, Chao Zhang<sup>4,\*</sup>, Julia Nagelschmitz<sup>1</sup>, Maximilian Doelling<sup>1</sup>,  
4 Sara Al-Madhi<sup>1</sup>, Ujjwal Mukund Mahajan<sup>2</sup>, Maciej Pech<sup>5</sup>, Georg Rose<sup>6</sup>, Roland Siegfried Croner<sup>1</sup>,  
5 Guo Liang Zheng<sup>7,8,9\*</sup>, Christoph Kahlert<sup>10,\*</sup>, Ulf Dietrich Kahlert<sup>1,6,\*,#</sup>

6  
7 <sup>1</sup> Molecular and Experimental Surgery, Clinic for General-, Visceral -, Vascular- and  
8 Transplantation Surgery, Medical Faculty and University Hospital Magdeburg, Otto-von-Guericke  
9 University, Magdeburg, Germany

10 <sup>2</sup> Department of Medicine II, Hospital of the LMU Munich, 81377, Munich, Germany

11 <sup>3</sup> Department of Gastroenterological Surgery, The Affiliated Hospital of Jiaxing University,  
12 Jiaxing, Zhejiang, China.

13 <sup>4</sup> Department of Radiology, The First Affiliated Hospital of Wannan Medical College, Wuhu City,  
14 Anhui Province, China

15 <sup>5</sup> Clinic for Radiology and Nuclear Medicine, University Hospital Magdeburg, Magdeburg,  
16 Germany

17 <sup>6</sup> Research Campus Stimulate, Otto von Guericke University Magdeburg, Germany

18 <sup>7</sup> Department of Gastric Surgery, Cancer Hospital of China Medical University, Shenyang, China.

19 <sup>8</sup> Cancer Hospital of Dalian University of Technology, Dalian, China

20 <sup>9</sup> Liaoning Cancer Hospital & Institute, Shenyang, China

21 <sup>10</sup> Department of General, Visceral and Transplantation Surgery, Heidelberg University Hospital,  
22 Heidelberg, Germany

23 \* equal contribution

24 # corresponding author: [ulf.kahlert@med.ovgu.de](mailto:ulf.kahlert@med.ovgu.de)

25  
26 **Key words:** PDAC, EV, microRNA, CT, radiogenomics

27 **Running title:** *microRNA signature for the diagnosis of pancreatic cancer*

28 **The authors declare no potential conflicts of interest**

29

30 **Statement of translational relevance**

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

43

44

45

46

47

48

49

50

51

52

53

54

55

56 **Abstract**

57 Purpose

58 Currently there is a lack of effective methods to accurately detect pancreatic cancer. In our study,  
59 we develop a liquid biopsy signature of EV miRNAs based on associated radiomics features of  
60 patients' tumors in order to provide new insights for the early diagnosis of pancreatic cancer.

61 Experimental Design

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

72 Results

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

81 Conclusion

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

85

86

## 87 **Introduction**

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

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

111 Benefiting from interdisciplinary advances in artificial intelligence, the integration of machine  
112 learning and genomics has led to breakthroughs in the early diagnosis and classification of  
113 tumors. For example, we used machine learning algorithms to assist in the development of a  
114 three-serum miRNA signature that effectively provides early warning of premalignant pancreatic  
115 cancer<sup>8</sup>. Another model, based on machine learning algorithms, focuses on the immune  
116 subtypes of triple-negative breast cancer, offering critical insights for identifying patients who  
117 may benefit from immunotherapy<sup>9</sup>. In the current study, we employed radiogenomics technology,  
118 another product of interdisciplinary collaboration between medicine and engineering. This novel  
119 approach integrates the quantification of image features from CT or MRI, which are then  
120 correlated with genomic signatures and allows for a non-invasive prediction of molecular

121 characteristics<sup>10</sup>. Such as, claim to be able to predict the occurrence of p53 mutations using CT  
122 images by radiogenomic analysis and hence make a statement on the prognosis<sup>11</sup>.

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

127  
128 **Materials and Methods**  
129 **Data resources and pre analytics preparation**

130 A total of four hospitals and four public datasets, comprising a total of eight datasets enrolled in  
131 this study, University hospital Magdeburg in Germany (UMMD), and Jiaying Hospital Center,  
132 China (JHC), provided enhanced computed tomography (CT) images of 46 pancreatic benign  
133 lesion (PB) patients and 127 pancreatic cancer (PC) patients as training dataset. For test  
134 datasets, CT data resource from Wanan medical university hospital (WUH) with 27 PB patients  
135 and 72 PC patients. University hospital Dresden, Germany (DUH) center provide miRNA and  
136 mRNA seq data from total plasma extracellular vesicles (EV) of PDAC patients with associated  
137 clinical follow information, including 20 benign pancreatic disease patients and 63 pancreatic  
138 cancer patients. The four public serum miRNA sequence data containing pancreatic cancer (PC)  
139 and healthy control (HC) were GSE106817(2759 PC vs.115 HC), GSE113486(40 PC vs.100HC),  
140 GSE112264(50 PC vs.41 HC), GSE109319(24 PC vs. 21 HC). Before future analysis, we  
141 removed the low quality of images and the low quality of miRNA sequence samples, then we  
142 use Propensity Score Matching (PSM) method to match the begin and tumor patients from DUH  
143 to UMMD & JHC according to age factor. Afterwards, we constructed a new matrix including CT  
144 images, EV miRNAs, mRNAs and patients clinical follow up information. The workflow is  
145 schematically depicted in Figure 1. Ethical approval to conduct the study for UMMD and DUH  
146 after approval by the local Institutional Review Board/ethics committee (UMMD 46/22; 30/01 with  
147 amendment 43/14; DUH: EK76032013). Written informed consent from the patients was  
148 obtained pre-operatively with the disclosure of research purpose.

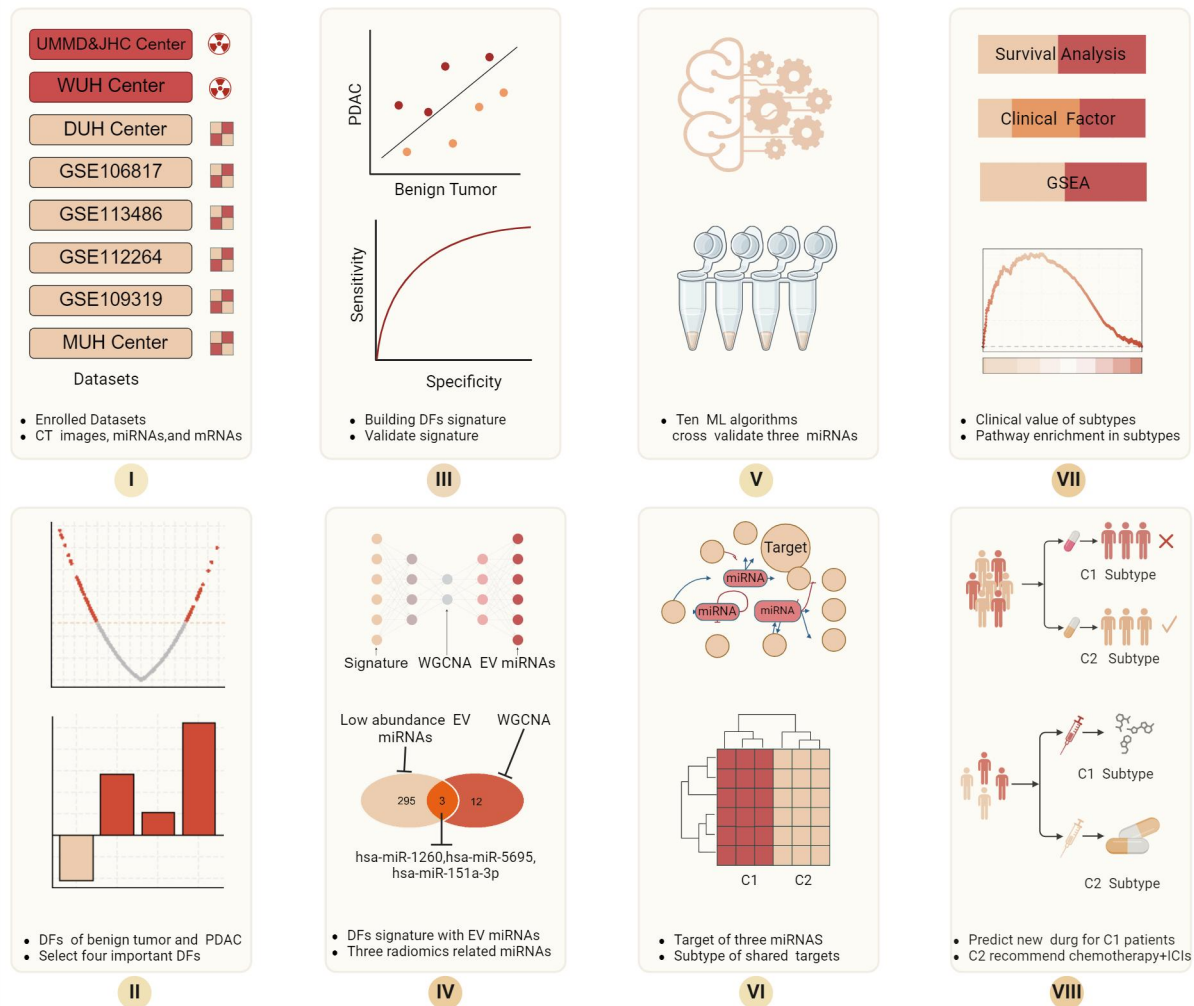
149

150

151

152

Radiogenomics reveals blood plasma extracellular vesicle microRNA signature for early diagnosis and molecular subtyping of pancreatic cancer



153  
 154 **Figure 1:** Schematic presentation of the workflow of this study. The use radiomics to aid finding  
 155 novel EV charged miRNAs to allow PDAC diagnostics.

156 **EV isolation and RNA sequencing**

157 Details on the protocols for EV isolation including presentation of optical and molecular  
 158 characteristics of isolated vesicles proofing high quality isolation performance have been  
 159 described previously by our groups<sup>12</sup>.

160 EV isolation using ultracentrifugation: 500 µl plasma samples were thawed and mixed with  
 161 500 µl PBS. The diluted plasma samples were filtered with 0.2 µm filter and subjected to  
 162 ultracentrifugation at 100,000×g, 2 h, 4 °C in a ultracentrifuge (Sorvall MX150 + micro-  
 163 ultracentrifuge, Thermo Scientific, Darmstadt, Germany). The supernatant was removed and the  
 164 pellet was washed once with ice-cold Phosphate Buffered Saline (PBS, Gibco, Carlsbad,

165 California, USA) and ultracentrifuged again at 100,000×g, 2 h, 4 °C. The resulting pellet was  
166 resuspended 100 µl PBS and transferred to Vivaspin ® 500 filtration (100,000 MWCO, Sartorius,  
167 Göttingen, Germany) for centrifugation at 15,000×g, 45 min. The concentrated EVs were stored  
168 at –20 °C until further use for EV characterization.

169 EV isolation using precipitation method: 550 µl plasma samples were thawed and first  
170 centrifuged at 2000×g for 20 min. The supernatant was subjected to another round of  
171 centrifugation at 10,000×g for 20 min. After the second round of centrifugation, 500 µl  
172 supernatant was mixed with 250 µl PBS, vortexed and added with 150 µl Exosome Precipitation  
173 Reagent. The mixture was incubated at room temperature for 10 min and centrifuged at  
174 10,000×g for 5 min. After removing the supernatant completely, the pellet was resuspended with  
175 500 µl PBS and concentrated with Vivaspin ® 500 filtration (100,000 MWCO, Sartorius,  
176 Göttingen, Germany) by centrifugation at 15,000×g, 45 min. The resulting EVs were stored at  
177 –20 °C until further use for EV characterization.

## 178 **RNA sequencing**

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

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

## 191 **Radiomics Feature Extraction and different features (DFs) analysis**

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

197 **Feature selection and radiomics related signature model build**

198 After obtain the DFs, to avoid multicollinearity of the data, we conduct the dimensionality  
199 reduction analysis by Boruta algorithms and Lasso regression (LR) model. First, the Boruta  
200 algorithms will calculate the importance of each feature and if importance more than shadow  
201 feature, it will be selected as important feature, and then, input above important feature into the  
202 LR model, the features were selected as significant if the model penalty coefficients were  
203 minimized. After selection by machine learning algorithms, we calculate the regression  
204 coefficient of each feature in LR model. After that, by combining the feature expression to build  
205 the signature model in order to predict the images status from large amount of imaging data. The  
206 model formula as listed below, multiplies the regression coefficients of the features with the  
207 corresponding feature expressions separately followed by summation of those. The model  
208 validation was test with applying WUH center CT dataset, whereas the area under curve (AUC)  
209 of Receiver operating characteristic curve (ROC) was used to evaluated the predict ability of  
210 model.

211 **Definition and identification of low abundance EV-derived miRNA transcripts.**

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

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

218 According to the radiomics signature, patients will be spilt into low and high-risk group. Low risk  
219 classification means images have high percentage of feature parameters from benign images  
220 while high risk more likely become cancer featuring images. To explore the correlation between  
221 above groups with EV miRNAs, we conduct the WGCNA analysis. The soft-thresholding power  
222 of WGCNA was automatically defined by the model, which was then assisting in calculating the  
223 expression correlation with a) a given miRNA to obtain gene significance (GS), and b) of module  
224 membership with miRNAs to obtain module membership (MM). Based on the cut-off criteria  
225 ( $|MM| > 0.5$  and  $|GS| > 0.1$ ), we obtain the significant miRNAs related to low and high risk  
226 images, respectively.

227

228 **Hub EV miRNAs identify and validation in serum and tissues level**



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

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

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

#### 243 **Common candidate target mRNAs of hub EV miRNAs**

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

#### 251 **Cluster of shared target mRNAs and survival analysis between different clusters**

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

#### 256 **Clustering of tumor subtype with clinical factors**

257 To explore the clinical value of each subtype, we analyze the relationship between the subtypes  
258 and important clinical factors, such as, age, tumor size, number of positive lymph nodes. We

259 also discover the distribution of sex, peri-neural invasion (PNI), and tumor stage the in different  
260 subtypes.

### 261 **Characterization of immune cell infiltration properties and immune check point activation** 262 **in each tumor subtype**

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

### 268 **Potential drug sensitivity for each subtype**

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

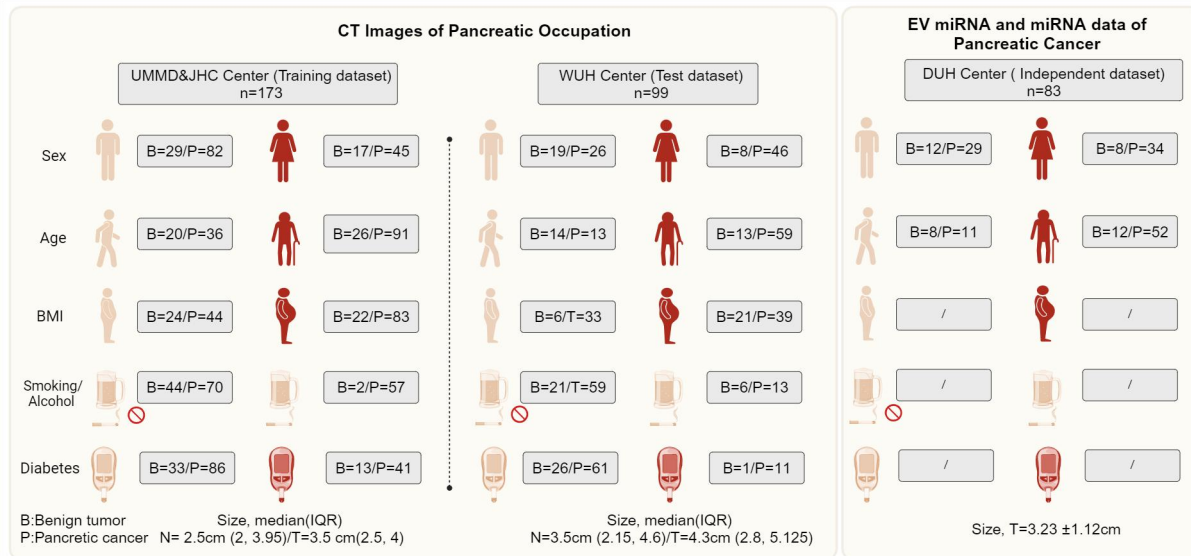
### 272 **Functional enrichment analysis and pathway prediction**

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

## 278 **Results**

### 279 **Enrolled population and baseline information for CT images and EV miRNA**

280 A total of 272 patients enrolled this study providing CT imaging, including 173 in UMMD&JHC  
281 center while 99 in WUH center. In UMMD&JHC center, also including 46 pancreatic benign  
282 lesion and 127 pancreatic cancer CT images. Most of pancreatic patients in this center are older  
283 with obesity, but less smoking or alcohol and less with diabetes. For WUH center, most  
284 pancreatic cancer are female, and also with a high account for older, and obesity. The DUH  
285 provide the patients with EV-miRNAs, EV-mRNAs data and follow up information. About 82.5%  
286 (52/53) patients are older, and 34 patients are female. The clinical characteristics of the data  
287 cohorts are summarized in **Figure 2**.



288

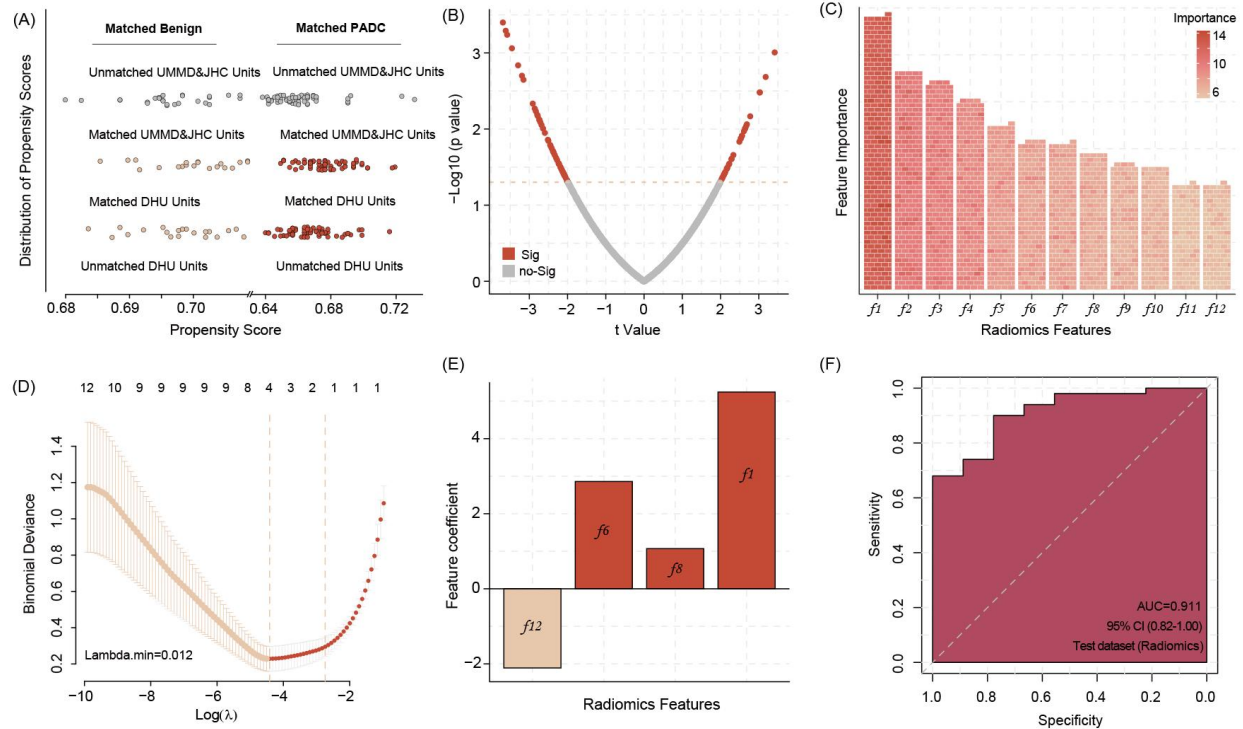
289 **Figure 2:** The base line information of clinical parameters of patients enrolled from four centers  
 290 in this multi-center trial.

291 **Different expression radiomic features between pancreatic benign lesions and aggressive**  
 292 **tumors**

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

298 **Four important radiomic feature was selected to build a related signature**

299 We use Boruta algorithms to select the important features and result show that a total 12  
 300 features were identified, which more than shadow features. In addition, after input above  
 301 features into LR algorithms and we can found that four features are list as key features(Figure  
 302 3C-D). Based on LR model with regression coefficient and feature expression, we build a four  
 303 radiomics feature related signature and validate the prediction ability in WUH center data,  
 304 revealing a signature accuracy of prediction efficiency (AUC=0.911) (Figure 3E-F).



305

306 **Figure 3:** Propensity Score Matching (PSM) allows matching data of benign pancreas lesions  
 307 and PDAC patients from DUH to UMD & JHC patients according to the age factor, all of DUH  
 308 patients successful matched similar patients (A). The different radiomic features between the  
 309 benign lesions and PDAC patients CT images(B). 12 most important radiomic features  
 310 differentiating between the benign pancreatic lesion and PDAC patients' CT images identified by  
 311 the Boruta algorithms (C) Four radiomic features were selected by Lasso Regression to build  
 312 model signature (D&E). Applying the four radiomic features related signature in image analysis  
 313 show high accuracy in predicting the PDAC manifestation in the WUH test dataset(F).

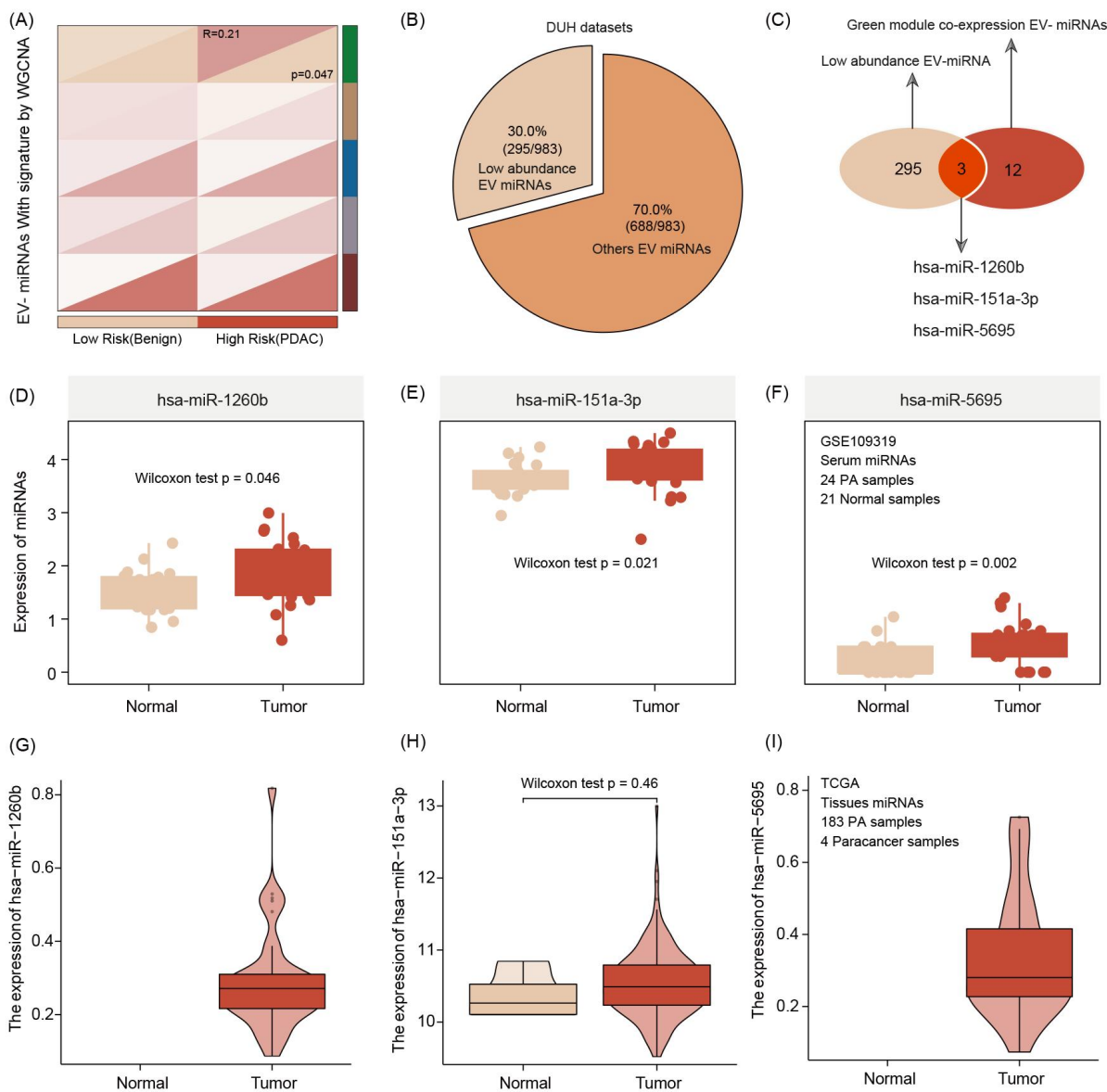
314 **Three EV miRNAs are associated with radiomic features**

315 After radiomics signature was build, each patients presents an individual risk score and we split  
 316 patients into high-risk and low-risk patients, according to median value of risk score. We use  
 317 WGCNA to connect the EV miRNA data and two imaging-featured patient risk groups. The  
 318 WGCNA results shows that the green model is the key module and the hub gene of this module  
 319 co-expresses 12 miRNAs. The number of low abundance of miRNA are calculated was n=295.  
 320 Merging both results, three miRNAs are identified (hsa-miR-1260b, hsa-miR-151a-3p and hsa-  
 321 miR-5695) and selected for subsequent alignment with associated with radiomic features (Figure  
 322 4A-C).

323

324 **Expression validation of three EV miRNAs**

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

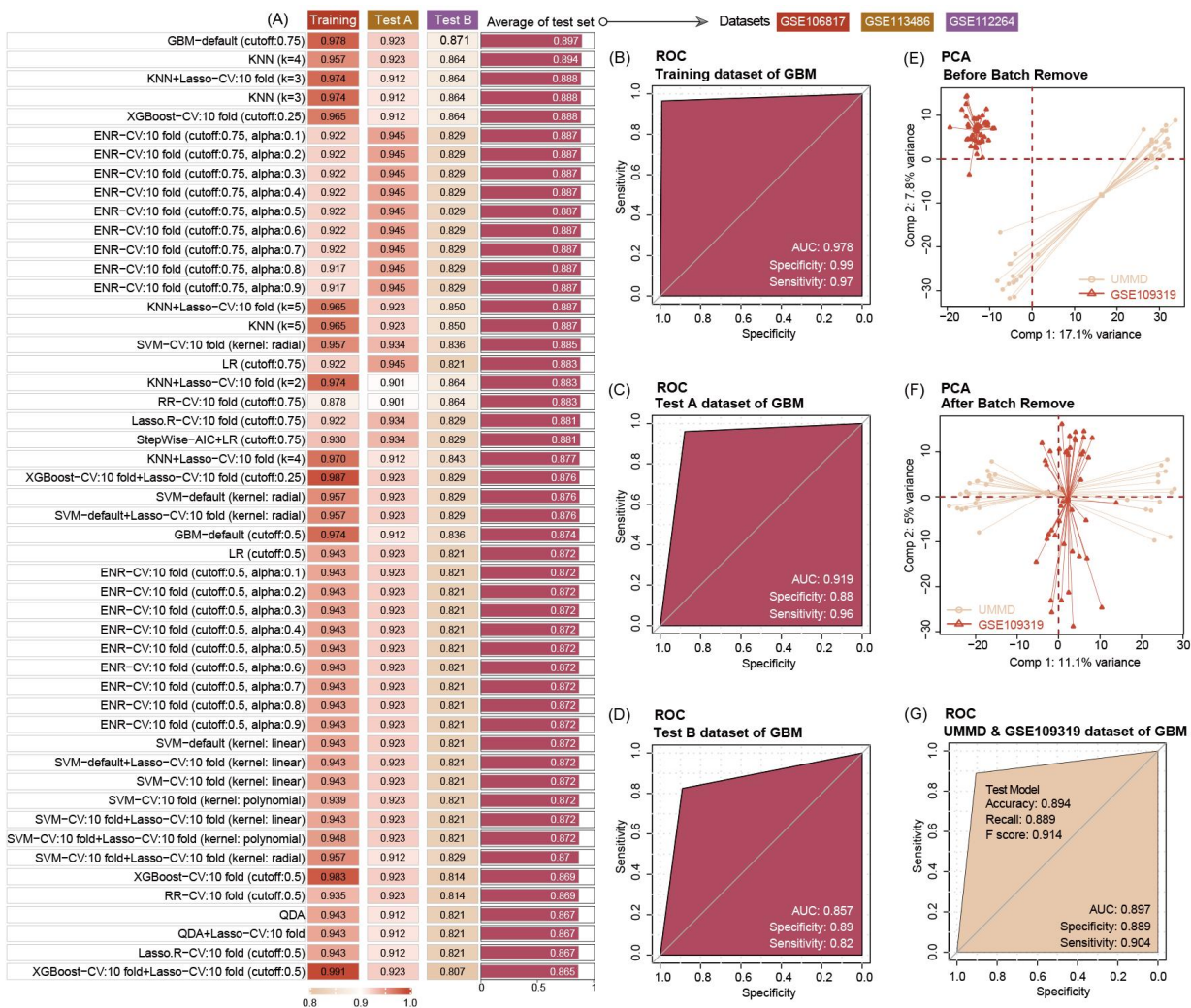


329  
 330 **Figure 4:** EV miRNA presenting the risk group stratification based on radiomics signature by  
 331 WGCNA analysis featuring green mode discovering our key module for further analysis ( $r=0.21$ ,  
 332  $p=0.047$ ) (A). The number of low abundance miRNA in the entire EVseq dataset cohort is  $n=$   
 333 295 (B). Out of those low abundance miRNAs,  $n=12$  present matching candidates differentially  
 334 expressed in high risk group patients. Alignment to our radiomics feature parameters identified

335 three core miRNAs (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) (C). The three key  
 336 miRNAs show significantly different expression levels in tumor condition, both for serum (D-F)  
 337 and tissue (G-I).

### 338 Three EV miRNAs predict pancreatic cancer with high accuracy

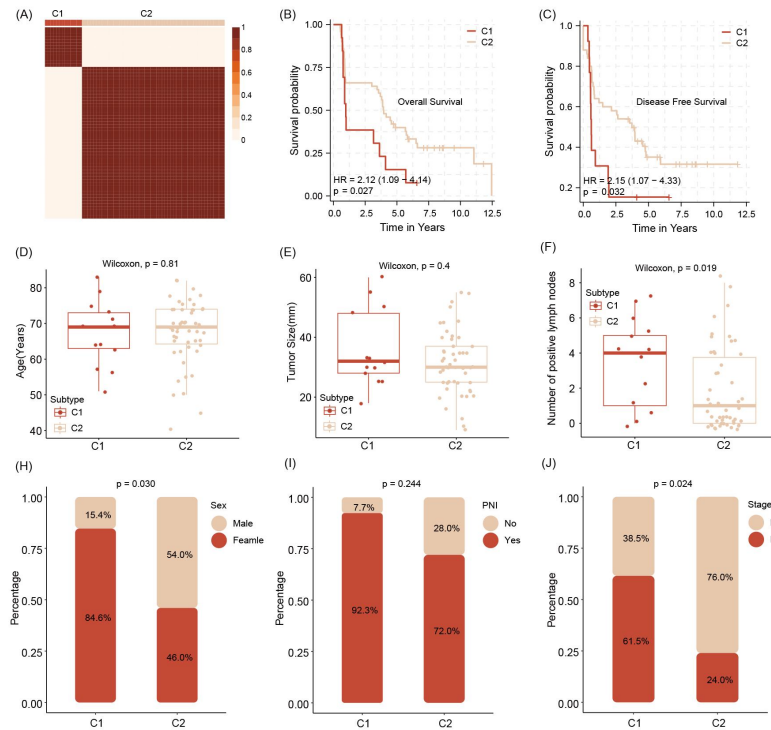
339 We use seven machine-learning combos to train and test the ability of three EV miRNAs levels  
 340 to predict tumor manifestation and clinical course of patient. The results show that in the GBM-  
 341 default (cutoff=0.75) model, three EV miRNAs show a high accuracy to predict the pancreatic  
 342 cancer with the training accuracy was 0.978 with AUC=0.978, and two test dataset accuracy are  
 343 0.923 with AUC=0.919, and 0.871 with AUC=0.857, respectively. Then, we choose GBM model  
 344 for the extend validation by our hospital data(MUH) and GSE109319. Before this procedure, we  
 345 use combat package to remove the batch effect allowing the merge of the two datasets. The  
 346 results of the extend dataset validation via GBM further highlights the high diagnostic accuracy  
 347 of the three EV miRNAs (accuracy =0.894, and AUC=0.897) (Figure 5).



349 **Figure 5:** Ten machine learning algorithms demonstrate that three key miRNAs show a high  
350 accuracy to diagnosis PDAC in early stage, no matter in training or test datasets. The best  
351 machine learning algorithms is GBM (cutoff:0.75) (A). Three miRNAs prediction ability of training  
352 dataset (GSE10106817) in GBM model is 0.978(B). Three miRNAs prediction ability of test  
353 dataset (GSE113486 and GSE112264) in GBM model is 0.919, and 0.857, respectively(C&D).  
354 Data distribution before removal of batch effect of our center data and GSE109319 dataset (E).  
355 Data distribution after remove batch effect of our center and GSE109319 dataset (F). Three EV  
356 miRNAs prediction ability to identify cancer of our center data and GSE109319 in GBM model is  
357 0.897(G).

### 358 **Identification of two molecular subtypes of PDAC with significant clinical predictive value**

359 After predicting molecular targets of three miRNAs, we obtain a number of 117 mRNA are  
360 shared candidate targets (Supplement Table 1). Stratifying the tumors according to the level of  
361 expression of those allows the differentiation of tumors in cluster into two subtypes termed  
362 Cluster 1 (C1) and C2 (Figure 6A and Supplement Figure 1). The survival analysis show that  
363 compared with C1, the C2 patients feature a prolonged survival outcome, no matter in OS or  
364 DFS(Figure 6 B&C). In addition, analysis the relationship with clinical factors of patients, we also  
365 found the C1 subtype patients are older in age and carry bigger tumor sizes and higher number  
366 of tumor cell positive lymph nodes(Figure 6D-F), while compared with C2 subtype patients. We  
367 also demonstrate C1 patients particular aggressive in female patients, feature high number of  
368 tumors containing perineural invasion (PNI) and advanced pathological tumor staging (Figure 6  
369 G-I).



370

371 **Figure 6:** Stratification of abundance levels of shared mRNAs by Non-negative matrix  
 372 factorization method allows the clustering of patients into two subtypes (C1 and C2) (A). Patients  
 373 of the C1 with a poor outcome in OS and DFS (B&C). C1 subtype patients are characterized  
 374 with older age and bigger average tumor size, higher number of tumor cell positive lymph nodes  
 375 as compared to C2 patients (D-F). C1 patients are predominantly female, have tumors with  
 376 pathological classification marks of perineural invasion and advanced tumor stage (G-I).

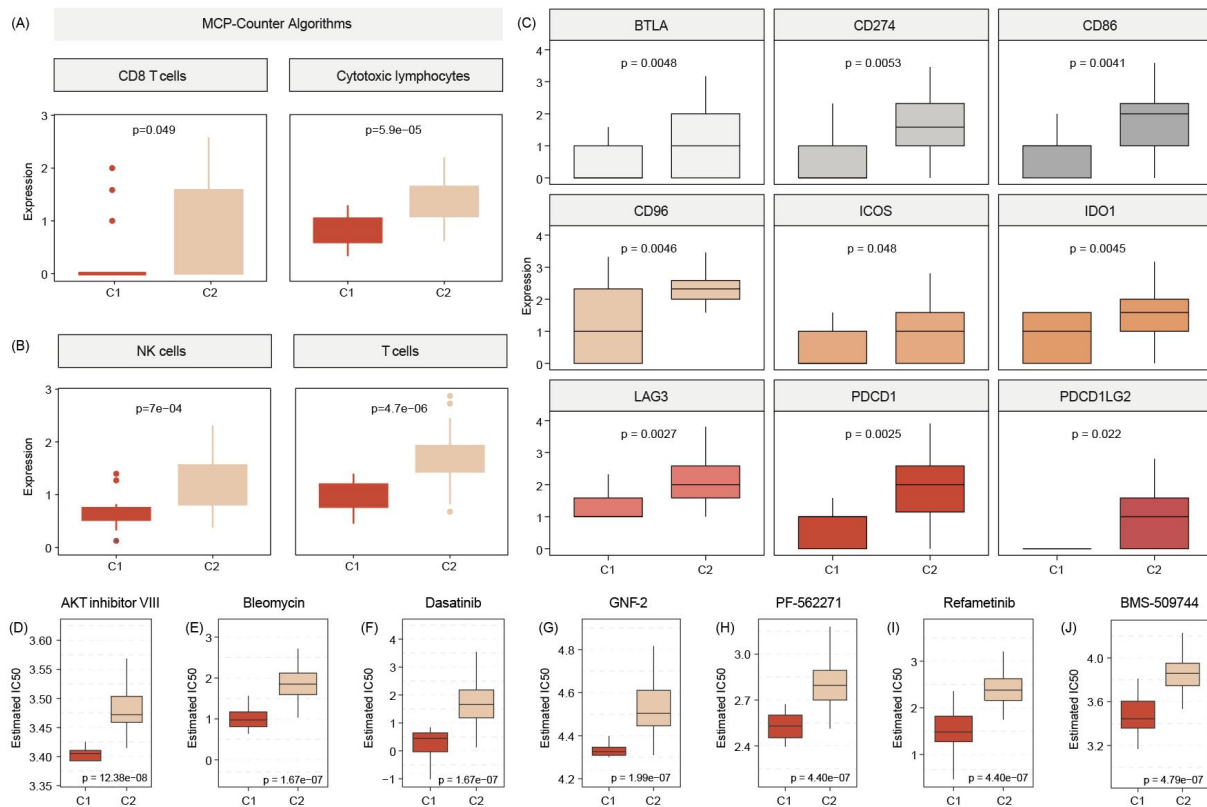
377 **C2 patients with high immune infiltration and positive with immune checkpoints**

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

383 **Prediction of drug sensitivity of C1 patients**

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



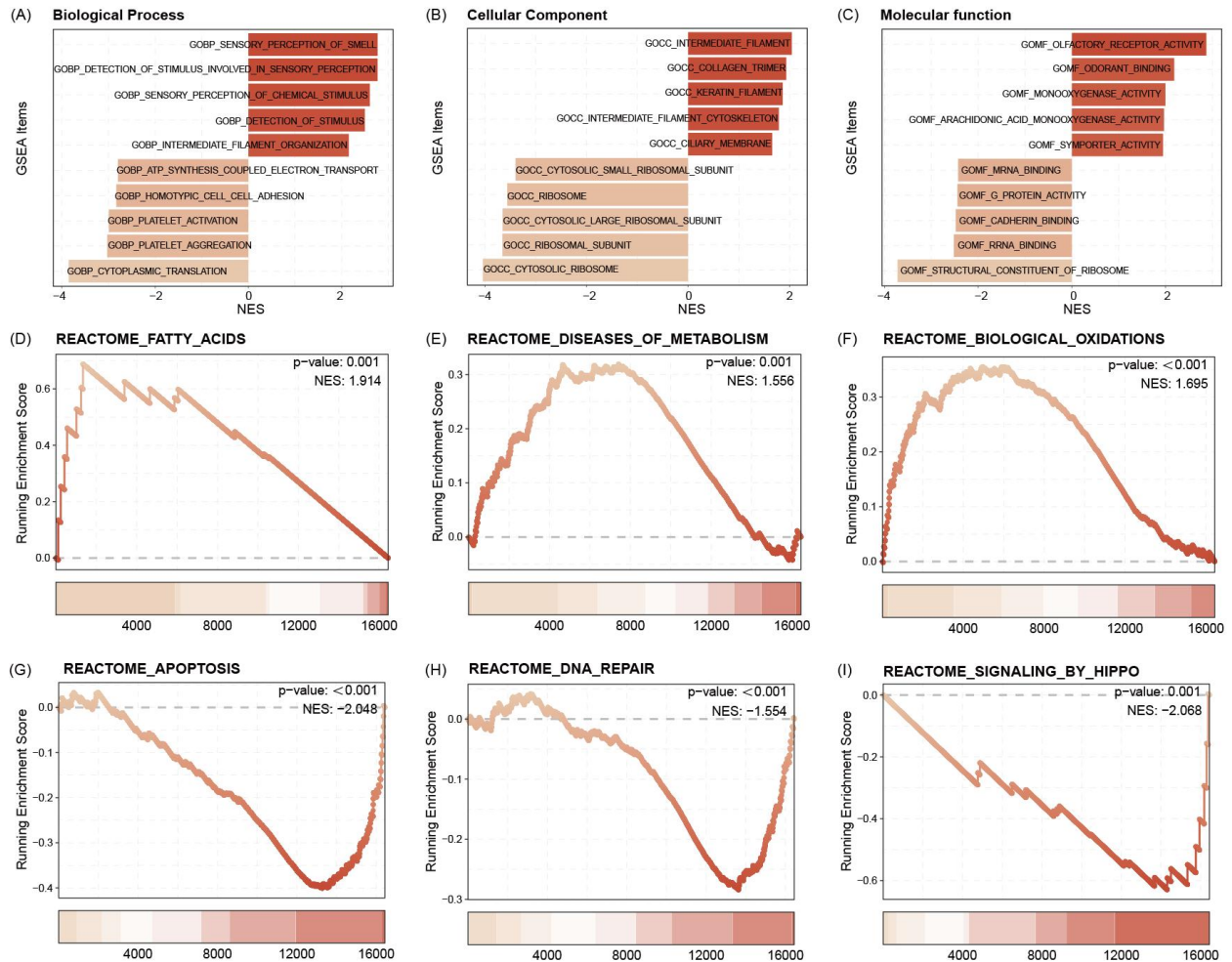


388

389 **Figure 7:** C2 subtype is positively associated with elevated levels of transcripts regulating gene  
 390 pathways encoding for CD8 T cells, cytotoxic lymphocytes, NK cells(A&B). Commonlyknown  
 391 immune checkpoints also higher expressed in C2 subtypes(D). Aligning in vitro drug sensitvity  
 392 and expression data from the GDSC database to subtype signature, predicts that tumors of C1  
 393 patients might be more sensitive to AKT INHIBITOR VIII, Bleomycin, Dasatinib, GNF-2, PF -  
 394 562271, Refametinib, BMS-509744 as C2 subtypes.

395 **Functional annotation and pathway enrichment for C1 subtype**

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



402

403 **Figure 8:** GO functional enrichment analysis indicated that the C1 subtype is enriched for  
 404 intermediate filament organization, CC ribosome, as well as MF symporter activity(A-C).  
 405 Pathway enrichment analysis showed that the C1 subtype was activated with the Reactome fatty  
 406 acids, Reactome diseases of metabolism, as well as Reactome biological oxidations pathways,  
 407 and may be inhibited with Reactome apoptosis, Reactome DNA repair, and Reactome signaling  
 408 by Hippo pathways (D-I).

409

## 410 Discussion

411

412 Using multicenter radiomics of Asian and Western world patients, we identified 3 plasma total  
 413 extracellular vesicle fraction microRNAs that demonstrated strong predictive performance and  
 414 prognostic value, offering new insights for noninvasive early diagnosis and prognosis of  
 415 pancreatic cancer. This study is the first to distinguish between benign lesions and pancreatic  
 416 cancer by identifying novel and significant plasma extracellular vesicle microRNAs based on  
 417 differential radiomics features. We focused on low-abundance microRNAs, which are often

418 overlooked in routine RNA sequencing analyses. These 3 microRNA candidates were validated  
419 across multiple cohorts, achieving excellent predictive performance in both testing and external  
420 validation sets. Additionally, these microRNAs demonstrated prognostic value, potentially aiding  
421 in treatment selection.

422 Pancreatic cancer remains one of the most lethal diseases, with insufficient early detection  
423 methods and limited treatment options<sup>13,14</sup>. This study aimed to develop a liquid biopsy nucleic  
424 acid diagnostic test based in EV-derived microRNA using radiogenomic analysis to enhance  
425 diagnostic accuracy and molecular subtyping of pancreatic ductal adenocarcinoma. By focusing  
426 on imaging features to distinguish benign lesions from pancreatic cancer, we identified  
427 underlying genetic factors contributing to these differences. Unlike previous studies on serum  
428 microRNA<sup>15,16</sup>, our research identified three microRNAs through radiomics that reflect tumor-  
429 specific changes rather than systemic variations. Additionally, our findings reveal relationships  
430 between imaging features and tumor biology, improving diagnostic accuracy and offering a more  
431 comprehensive tool for early detection and personalized disease management.

432 Although numerous studies have reported on microRNA sequencing results<sup>17</sup>, highly abundant  
433 microRNAs often mask the signals of many others that are present at very low abundance  
434 during routine analysis. Furthermore, if a microRNA is associated with tumor development, it is  
435 likely to be in low abundance during early detection and may even be undetectable in benign  
436 lesions. In our study, after identifying 12 microRNAs related to radiomic differences, we further  
437 defined three (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) of them as being of low  
438 abundance. These microRNAs showed significantly higher expression in pancreatic cancer  
439 blood than in non malignant cancer patients' plasma, with hsa-miR-1260b and hsa-miR-5695  
440 undetectable in pancreatic cancer samples. The model based on these 3 microRNAs  
441 demonstrated strong performance in the training group (AUC, 0.978; sensitivity, 0.97; specificity,  
442 0.99), testing group (AUC, 0.919; sensitivity, 0.96; specificity, 0.88), and external testing group  
443 (AUC, 0.897; sensitivity, 0.90; specificity, 0.89).

444 Previous studies on microRNAs as diagnostic biomarkers for pancreatic cancer vs healthy  
445 controls report AUCs of 0.78-0.96 and sensitivities of 0.62-0.88<sup>17</sup>, but many lack external  
446 validation<sup>18,19</sup>. One case-control study identified 2 diagnostic panels based on microRNA  
447 expression (index I: 4 microRNAs; AUC, 0.86; sensitivity, 0.85; specificity, 0.64; index II: 10  
448 microRNAs; AUC, 0.93; sensitivity, 0.85; specificity, 0.85)<sup>20</sup>. Our study achieved similar  
449 performance using only 3 microRNAs, making it more practical for clinical application.

450 To the best of our knowledge, hitherto none of the three identified microRNAs have been  
451 reported in the context of diagnostic biomarkers in pancreatic cancer. Moreover, to our  
452 knowledge, hsa-miR-5695 only indirectly been linked to breast carcinogenesis<sup>21</sup> and subtypes of  
453 metastatic prostate cancer<sup>22</sup>, while hsa-miR-1260b has been implicated in promoting  
454 tumorigenesis in lung cancer<sup>23</sup>, breast cancer<sup>24</sup>, sarcoma<sup>25</sup> and prostate cancer<sup>26</sup>. Interestingly,  
455 in a screen of microRNAs that are differentially expressing in total plasma fraction of patients  
456 suffering from intraductal papillary mucinous neoplasm (IPMN) as compared to non-diseased  
457 controls, hsa-miR-1260b was identified as one of the top upregulated signals in plasma of IPMN  
458 cases<sup>27</sup>. Our work further enforces the potential of hsa-miR-1260b abundancy in plasma to  
459 detect onset of abnormal pancreas genesis. Exosomal hsa-miR-151a-3p has emerged as a  
460 potential novel biomarker for predicting bone metastasis in lung cancer<sup>28</sup>. A very recent report of  
461 results of multicenter trial in Asian participants also identified circulating 2'-O-methylated form of  
462 hsa-miR-151a-3p is upregulated in total fraction of plasma of PDAC patients compared to those  
463 derived from healthy donors or patients with chronic pancreatitis<sup>29</sup>. Focusing on EV fraction, our  
464 results verifies the potential of this newly discovered biomarker in pancreatic cancer blood  
465 diagnosis.

466 Pancreatic cancer currently has only two standard systemic treatments with limited efficacy, and  
467 the patient population responsive to immunotherapy and targeted therapy remains unclear<sup>30</sup>.  
468 Therefore, we explored the prognostic value of these microRNAs, which allowed for patient  
469 stratification. Patients classified as C2 had better prognoses, with higher immune cell infiltration  
470 and more immune checkpoint expression, suggesting they may be suitable for immunotherapy  
471 after standard treatment. In contrast, patients classified as C1 had poorer prognoses, related to  
472 inhibited pathways such as DNA repair and apoptosis.

473 The findings suggest potential for developing kits for early detection and treatment stratification,  
474 and highlight these three candidates for further exploration. However, despite thorough matching  
475 of imaging, genomic, and clinical data, biases may still exist. Additionally, using the MCP-  
476 counter algorithm for blood samples is a limitation of our study, as it was originally designed for  
477 solid tissues, which may affect the accuracy of cell population quantification in the blood.  
478 Furthermore, when validating the three microRNAs in tissue, the sample size of paracancerous  
479 tissue was limited. Lastly, since our EV purification did not use measures to enrich for cancer  
480 cell derived EVs such as CD147 membrane presenting amplitude<sup>31</sup>, we cannot exclude  
481 contamination of our miRNA data due to signals from non-cancer cell background, which might  
482 impact the cancer specificity of our proposed blood test. Moreover, our data is based on OMICS

483 data acquisition. Real world application of our proposed test would benefit from proofing the  
484 predictability and sensitivity of the miRNAs by targeted methods, such as RT-qPCR or  
485 CRISPR/Cas diagnostics and in prospective, multicenter trial.

486

487 **References**

- 488 1 Bamankar, S. & Londhe, V. Y. The rise of extracellular vesicles as new age biomarkers  
489 in cancer diagnosis: promises and pitfalls. *Technology in Cancer Research & Treatment*  
490 **22**, 15330338221149266 (2023).
- 491 2 Wang, S. *et al.* The molecular biology of pancreatic adenocarcinoma: translational  
492 challenges and clinical perspectives. *Signal transduction and targeted therapy* **6**, 249  
493 (2021).
- 494 3 Bestari, M. B., Joewono, I. R. & Syam, A. F. A Quest for Survival: A Review of the Early  
495 Biomarkers of Pancreatic Cancer and the Most Effective Approaches at Present.  
496 *Biomolecules* **14**, 364 (2024).
- 497 4 Rashid, S. *et al.* Elucidating the Role of miRNA-326 Modulating Hedgehog Signaling in  
498 Pancreatic Carcinoma. *Pancreas* **53**, e42-e48 (2024).
- 499 5 Mok, E. T., Chitty, J. L. & Cox, T. R. miRNAs in pancreatic cancer progression and  
500 metastasis. *Clinical & Experimental Metastasis*, 1-24 (2024).
- 501 6 Min, L. *et al.* Evaluation of circulating small extracellular vesicles derived miRNAs as  
502 biomarkers of early colon cancer: a comparison with plasma total miRNAs. *Journal of*  
503 *extracellular vesicles* **8**, 1643670 (2019).
- 504 7 Tiwari, P. K. *et al.* Extracellular Vesicular miRNA in Pancreatic Cancer: From Lab to  
505 Therapy. *Cancers* **16**, 2179 (2024).
- 506 8 Shi, W. *et al.* Integrating a microRNA signature as a liquid biopsy-based tool for the early  
507 diagnosis and prediction of potential therapeutic targets in pancreatic cancer. *British*  
508 *Journal of Cancer* **130**, 125-134 (2024).
- 509 9 Chen, Z. *et al.* A machine learning model to predict the triple negative breast cancer  
510 immune subtype. *Frontiers in immunology* **12**, 749459 (2021).
- 511 10 Casà, C. *et al.* The impact of radiomics in diagnosis and staging of pancreatic cancer.  
512 *Therapeutic advances in gastrointestinal endoscopy* **15**, 26317745221081596 (2022).
- 513 11 Iwatate, Y. *et al.* Radiogenomics for predicting p53 status, PD-L1 expression, and  
514 prognosis with machine learning in pancreatic cancer. *British Journal of Cancer* **123**,  
515 1253-1261 (2020).
- 516 12 Han, Y. *et al.* Plasma extracellular vesicle messenger RNA profiling identifies prognostic  
517 EV signature for non-invasive risk stratification for survival prediction of patients with  
518 pancreatic ductal adenocarcinoma. *Journal of Hematology & Oncology* **16**, 7 (2023).
- 519 13 Halbrook, C. J., Lyssiotis, C. A., Pasca di Magliano, M. & Maitra, A. Pancreatic cancer:  
520 Advances and challenges. *Cell* **186**, 1729-1754 (2023).
- 521 14 Fahrmann, J. F. *et al.* Lead-Time Trajectory of CA19-9 as an Anchor Marker for  
522 Pancreatic Cancer Early Detection. *Gastroenterology* **160**, 1373-1383.e1376 (2021).
- 523 15 Wei, J., Yang, L., Wu, Y. N. & Xu, J. Serum miR-1290 and miR-1246 as Potential  
524 Diagnostic Biomarkers of Human Pancreatic Cancer. *J Cancer* **11**, 1325-1333 (2020).
- 525 16 Shams, R. *et al.* Identification of potential microRNA panels for pancreatic cancer  
526 diagnosis using microarray datasets and bioinformatics methods. *Sci Rep* **10**, 7559  
527 (2020).
- 528 17 Yang, J. *et al.* Early screening and diagnosis strategies of pancreatic cancer: a  
529 comprehensive review. *Cancer Commun (Lond)* **41**, 1257-1274 (2021).
- 530 18 Liu, J. *et al.* Combination of plasma microRNAs with serum CA19-9 for early detection of  
531 pancreatic cancer. *Int J Cancer* **131**, 683-691 (2012).
- 532 19 Wang, Z. Diagnostic performance for declined microRNA-133a in pancreatic cancer. *J*  
533 *Cell Biochem* **121**, 3882-3886 (2020).
- 534 20 Schultz, N. A. *et al.* MicroRNA biomarkers in whole blood for detection of pancreatic  
535 cancer. *Jama* **311**, 392-404 (2014).
- 536 21 Søkilde, R. *et al.* Refinement of breast cancer molecular classification by miRNA  
537 expression profiles. *BMC genomics* **20**, 1-12 (2019).

538 22 Watahiki, A. *et al.* MicroRNAs associated with metastatic prostate cancer. *PloS one* **6**,  
539 e24950 (2011).

540 23 Kim, D. H. *et al.* Exosomal miR-1260b derived from non-small cell lung cancer promotes  
541 tumor metastasis through the inhibition of HIPK2. *Cell Death Dis* **12**, 747 (2021).

542 24 Park, S. *et al.* Promotion of tumorigenesis by miR-1260b-targeting CASP8: Potential  
543 diagnostic and prognostic marker for breast cancer. *Cancer Sci* **113**, 2097-2108 (2022).

544 25 Morita, T. *et al.* Clinical relevance and functional significance of cell-free microRNA-  
545 1260b expression profiles in infiltrative myxofibrosarcoma. *Scientific reports* **10**, 9414  
546 (2020).

547 26 Hirata, H. *et al.* Genistein downregulates onco-miR-1260b and upregulates sFRP1 and  
548 Smad4 via demethylation and histone modification in prostate cancer cells. *Br J Cancer*  
549 **110**, 1645-1654 (2014).

550 27 Permeth-Wey, J. *et al.* Plasma MicroRNAs as Novel Biomarkers for Patients with  
551 Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Cancer Prev Res (Phila)*.  
552 2015; 8 (9): 826–34. In: (Epub 2015/09/01. <https://doi.org/10.1158/1940-6207.CAPR-15-0094> PMID: 26314797).

554 28 Zhao, K. *et al.* Exosomal hsa-miR-151a-3p and hsa-miR-877-5p are potential novel  
555 biomarkers for predicting bone metastasis in lung cancer. *Aging (Albany NY)* **15**, 14864-  
556 14888 (2023).

557 29 Yang, Z. *et al.* Contribution of a Circulating 2'-O-methylated MicroRNA Panel to the  
558 Diagnosis of Pancreatic Ductal Adenocarcinoma. *Journal of Cancer* **15**, 1583 (2024).

559 30 Neoptolemos, J. P. *et al.* Therapeutic developments in pancreatic cancer: current and  
560 future perspectives. *Nat Rev Gastroenterol Hepatol* **15**, 333-348 (2018).

561 31 Ko, S. Y. *et al.* The glycoprotein CD147 defines miRNA - enriched extracellular vesicles  
562 that derive from cancer cells. *Journal of Extracellular Vesicles* **12**, 12318 (2023).

563

564

565 **Figure legends**

566 Figure 1: Schematic presentation of the workflow of this study. The use radiomics to aid finding  
567 novel EV charged miRNAs to allow PDAC diagnostics.

568 Figure 2: The base line information of clinical parameters of patients enrolled from four centers  
569 in this multi-center trial.

570 Figure 3: Propensity Score Matching (PSM) allows matching data of benign pancreas lesions  
571 and PDAC patients from DUH to UMMD & JHC patients according to the age factor, all of DUH  
572 patients successful matched similar patients (A). The different radiomic features between the  
573 benign lesions and PDAC patients CT images(B). 12 most important radiomic features  
574 differentiating between the benign pancreatic lesion and PDAC patients' CT images identified by  
575 the Boruta algorithms (C) Four radiomic features were selected by Lasso Regression to build  
576 model signature (D&E). Applying the four radiomic features related signature in image analysis  
577 show high accuracy in predicting the PDAC manifestation in the WUH test dataset(F).

578 Figure 4: EV miRNA presenting the risk group stratification based on radiomics signature by  
579 WGCNA analysis featuring green mode discovering our key module for further analysis ( $r=0.21$ ,  
580  $p=0.047$ ) (A). The number of low abundance miRNA in the entire EVseq dataset cohort is  $n=$   
581 295 (B). Out of those low abundance miRNAs,  $n=12$  present matching candidates differentially  
582 expressed in high risk group patients. Alignment to our radiomics feature parameters identified  
583 three core miRNAs (hsa-miR-1260b, hsa-miR-151a-3p and hsa-miR-5695) (C). The three key  
584 miRNAs show significantly different expression levels in tumor condition, both for serum (D-F)  
585 and tissue (G-I).

586 Figure 5: Ten machine learning algorithms demonstrate that three key miRNAs show a high  
587 accuracy to diagnosis PDAC in early stage, no matter in training or test datasets. The best  
588 machine learning algorithms is GBM (cutoff:0.75) (A). Three miRNAs prediction ability of training  
589 dataset (GSE10106817) in GBM model is 0.978(B). Three miRNAs prediction ability of test  
590 dataset (GSE113486 and GSE112264) in GBM model is 0.919, and 0.857, respectively(C&D).  
591 Data distribution before removal of batch effect of our center data and GSE109319 dataset (E).  
592 Data distribution after remove batch effect of our center and GSE109319 dataset (F). Three EV  
593 miRNAs prediction ability to identify cancer of our center data and GSE109319 in GBM model is  
594 0.897(G).

595 Figure 6: Stratification of abundancy levels of shared mRNAs by Non-negative matrix  
596 factorization method allows the clustering of patients into two subtypes (C1 and C2) (A). Patients



597 of the C1 with a poor outcome in OS and DFS(B&C). C1 subtype patients are characterized  
598 with older age and bigger average tumor size, higher number of tumor cell positive lymph nodes  
599 as compared to C2 patients (D-F). C1 patients are predominantly female, have tumors with  
600 pathological classification marks of perineural invasion and advanced tumor stage (G-I).

601 Figure 7: C2 subtype is positively associated with elevated levels of transcripts regulating gene  
602 pathways encoding for CD8 T cells, cytotoxic lymphocytes, NK cells(A&B). Commonly known  
603 immune checkpoints also higher expressed in C2 subtypes(D). Aligning in vitro drug sensitivity  
604 and expression data from the GDSC database to subtype signature, predicts that tumors of C1  
605 patients might be more sensitive to AKT INHIBITOR VIII, Bleomycin, Dasatinib, GNF-2, PF -  
606 562271, Refametinib, BMS-509744 as C2 subtypes.

607 Figure 8: GO functional enrichment analysis indicated that the C1 subtype is enriched for  
608 intermediate filament organization, CC ribosome, as well as MF symporter activity(A-C).  
609 Pathway enrichment analysis showed that the C1 subtype was activated with the Reactome fatty  
610 acids, Reactome diseases of metabolism, as well as Reactome biological oxidations pathways,  
611 and may be inhibited with Reactome apoptosis, Reactome DNA repair, and Reactome signaling  
612 by Hippo pathways (D-I).

613

#### 614 **Data availability**

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

617

#### 618 **Funding**

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

620

#### 621 **Authors contribution**

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

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

624 Data interpretation, writing and review manuscript: all authors

625