A panel of 8 miRNAs as a novel diagnostic biomarker in pancreatic cancer

Benyuan Deng, MD<sup>a</sup>, Ming Wang, MD<sup>b</sup>, Zhongwu Liu, MM<sup>a</sup>.*

**Abstract**
Pancreatic cancer (PC) is one of the major causes of cancer mortality in developed countries. Therefore, there is an urgent need to derive biomarkers for early diagnosis of PC patients at high risk.

This study was designed to identify a panel of miRNAs that might serve as biomarkers for the early diagnosis of PC.

The data containing both PC and control samples were extracted from the Gene Expression Omnibus (GEO) database. EdgeR was applied to identify the differentially expressed miRNAs and genes between PC patients and healthy controls. Then a miRNA-mRNA network was constructed based on the differentially expressed miRNAs and genes. The miRNAs-based biomarker for PC was finally constructed by random forest. Finally, AUC was used to evaluate the performance of miRNAs to classify PC and control samples.

A total of 33 differentially expressed miRNAs, 753 differentially expressed genes, and 8 miRNAs (hsa-mir-139, hsa-mir-31, hsa-mir-196b, hsa-mir-221, hsa-mir-203b, hsa-mir-215, hsa-mir-144, and hsa-mir-4433b) that play important roles in PC were identified. The target genes of these miRNAs were found to be mainly enriched in negative regulation of acute inflammatory response cell-substrate responses, and O-glycan processing pathways. The constructed biomarkers based on these 8 miRNAs could distinguish samples coming from PC and healthy controls.

We identified a panel of eight-miRNAs that would serve as early diagnostic biomarkers for PC patients.

**Abbreviations:** GEO = Gene Expression Omnibus, miRNAs = microRNAs, PC = pancreatic cancer, ROC = receiver-operating characteristic, TCGA = The Cancer Genome Atlas.

**Keywords:** diagnostic biomarker, genes, miRNAs, pancreatic cancer

1. Introduction
Pancreatic cancer (PC) is reported to be the fourth major death cause of various malignancies around world especially in developed countries. PC is also known as a highly fatal cancer with an estimated 5-year survival rate <5%. Although many therapeutic improvements have been made in the past few years, yet most of the PC patients dyed <1 year after diagnosis, mainly due to its early metastatic nature,

surgical and medical intrusions inoperable, thus resulting in high mortality and poor prognosis. Therefore, there is an urgent need to understand the molecular mechanism involving the initiation, progression, and metastasis of PC to identify novel effective diagnostic or therapeutic markers to improve the management of PC.

Recently, new molecules are being actively investigated like microRNAs (miRNAs), which have emerged as one of the promising players in the prognosis, diagnosis, and therapy of cancer. miRNAs are noncoding 20 to 25 nucleotide endogenous RNA sequences which could regulate gene expression through base-pairing with target mRNA and regulate biological functions of many tumors in cell proliferation, apoptosis, invasion, and molecular pathways.

Several studies have confirmed that miRNAs played important roles during PC developments including tumor initiation, invasion, and metastasis. For example, miR-376 was reported to be overexpressed in human Panc-1 PC cells relative to other cancer cells. miR-21 was found to enhance PI3K-AKT signaling and activate the proliferation and invasiveness of PC cells. In addition, miR-21 could increase the invasion and metastasis of cancer cells by inducing tumor-activating environment. Let-7 family, the second discovered miRNA, was also found in several cancers including PC as a tumor suppressor. All the miRNAs mentioned above may be considered as the potential biomarkers for the diagnosis or therapy management of PC. However, these reported miRNAs may be not always sensitive or effective markers for PC, for so far there were still no dependable management strategies for
PC. Therefore, great efforts and studies were needed to explore more potential miRNAs that could be taken as reliable biomarkers in PC.

In this study, we constructed a miRNA-mRNA network based on the gene and miRNA expression profiles of 4 PC and 4 control samples and further identified the potential miRNAs associated with the development of PC. The miRNAs identified in this study could serve as early miRNAs-based diagnostic biomarkers for PC.

2. Material and methods

2.1. Data and pre-processing

In this study, 4 groups of data containing both PC and control samples were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) and the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/), in which, 2 groups were miRNA expression profile, whereas the other 2 were gene expression profile. The 4 groups of data were divided into train and test groups (Table 1). There were 4 PC samples and 4 control samples in TCGA-PC of train group, whereas 10 PC samples and 10 control samples in GSE119794 of test group, these samples contained both miRNA expression profile and gene expression profile data.

For the mRNA and miRNA sequencing data from TCGA and GSE119794, count data was used. The mature miRNA ID was converted into pre-miRNA ID according to the information in the MiRBase database (http://www.mirbase.org/) with the aim to unify the miRNA IDs in TCGA and GSE119794. In the study, only the overlapped miRNA and mRNA in TCGA and GSE119794 was analyzed to verify the accuracy of TCGA data in GSE119794.

2.2. Identification of the differentially expressed miRNAs and genes

EdgeR was used to identify the differentially regulated miRNAs or genes between normal and cancer in PC in the TCGA data. MiRNAs or genes with adjusted FDR <0.2 were screened out as differentially regulated miRNAs or genes. Paired Wilcoxon rank sum test was used to validate the hub miRNAs and genes in GSE119794.

2.3. miRNA-mRNA network

The target genes of miRNA were predicted through miRNA target gene information stored in miRTarBase database. As for all the obtained target genes, the differentially expressed genes in PC and control were screened out to reconstruct the miRNA-mRNA relationship profile. And a miRNA-mRNA network was finally constructed by putting the miRNA-mRNA relationship profile into cytoscape (www.cytoscape.org/) software.

2.4. GO-function enrichment analysis

The disturbed functional categories significantly enriched in the target genes of the hub miRNAs were identified by GO-function. And, the cumulative hypergeometric distribution model was also constructed by GO-function.

\[
P(X = k) = \frac{C_k C_{N-k}}{C_N}
\]

N presents the number of genes in the background; M presents the number of interested genes; n presents the number of genes in a pathway; X presents the number of interested genes in a pathway.

2.5. Area under the curve

In short, first, if the miRNA expression in the sample was lower than the given threshold value of PC, then it was classified as PC, with the expression range of miRNA in all samples ranged from 0 to the maximum; second, calculating the true-positive rate (sensitivity) and false-positive rate (specificity) of each threshold. The proportion that actual PC be classified as PC was measured by sensitivity, whereas the proportion that actual control be classified as control was measured by specificity. Third, receiver-operating characteristic (ROC) represented the curve of sensitivity versus specificity.

2.6. Random Forest analysis

The input samples were divided into two groups by random forest (RF) to create multiple decision trees in each iteration step, and 1 group was the samples randomly selected with replacement, whose sizes were the same as the input; the remaining samples were the other group, known as the out-of-bag samples, whose size were about one-third of the input at random. RF repeated the following process that contained 2 steps: constructing a decision tree with randomly selected samples and validating the tree with out-of-bag samples.

2.7. Ethical approval

The ethical approval was not necessary. This study downloaded the data from the Gene Expression Omnibus (GEO) database and the Cancer Genome Atlas (TCGA). Human or animal tissue was not involved in this study.

3. Results

3.1. miRNA-mRNA network

The differentially regulated miRNA or genes between 4 normal and 4 cancer samples in PC in the TCGA data were identified using EdgeR. Thiety-three differentially regulated miRNAs and 753 differentially regulated genes were identified with FDR values <0.2. A miRNA-mRNA network was constructed through the relationship between these differentially expressed genes and miRNA and their target genes. In short, for all the obtained target genes, the differentially expressed genes in PC relative to control were screened out, then a miRNA-mRNA profile (Fig. 1) was reconstructed by these differentially expressed genes and miRNAs. The miRNA target gene information was applied for the relationship between miRNA and its target genes.
in miRTarBase\textsuperscript{[24,25]} database was verified and showed high reliability.

As shown in Figure 1, some critical miRNAs (hsa-mir-139, hsa-mir-31, hsa-mir-196b, hsa-mir-221, hsa-mir-203b, hsa-mir-215, hsa-mir-144, and hsa-mir-4433b) were screened out. The mature miRNA ID and pre-miRNAs ID information of the 8 miRNAs were shown in Table 2. The key target genes were also identified (HOXA10, BCL2, ITGA3, MET, SH2B3, and ERO1A).

### 3.2. Validation of the 8 miRNAs in independent data

Furthermore, the expression patterns of the above 8 miRNAs (hsa-mir-139, hsa-mir-31, hsa-mir-196b, hsa-mir-221, hsa-mir-203b, hsa-mir-215, hsa-mir-144, and hsa-mir-4433b) were verified in the test data composed of 10 PC samples and 10 control samples from GSE119794 data. From Figure 2, we could see that hsa-mir-139 still maintained a significantly low expression pattern (Paired Wilcoxon rank sum test, \(P < .002\), Fig. 2A) of PC relative to control in the independent large sample verification set.

The same results were observed in hsa-mir-196b (Paired Wilcoxon rank sum test, \(P < .033\), Fig. 2C), hsa-mir-221 (Paired Wilcoxon rank sum test, \(P < .002\), Fig. 2D), hsa-mir-203b (Paired Wilcoxon rank sum test, \(P < .014\), Fig. 2E), and hsa-mir-215 (Paired Wilcoxon rank sum test, \(P < .028\), Fig. 2F), whereas the significance of hsa-mir-31 (Paired Wilcoxon rank sum test, \(P < .131\), Fig. 2B) and hsa-mir-144 (Paired Wilcoxon rank sum test, \(P < .064\), Fig. 2G) were near the significance threshold. The test value of hsa-mir-4433b with more 0 value may lead to high \(P\) value (Paired Wilcoxon rank sum test, \(P < .354\), Fig. 2H).

### 3.3. Validation of the 6 genes in independent data

Similarly, the expression patterns of the 6 genes (HOXA10, BCL2, ITGA3, MET, SH2B3, and ERO1A) were verified in test data from 10 PC and 10 control samples from GSE119794. Figure 3 demonstrated that HOXA10 maintained a significantly high expression pattern of PC relative to control in the independent large sample verification set (Paired Wilcoxon rank sum test, \(P < .004\), Fig. 3A).
Figure 2. The expression levels of 8 hub miRNAs in validation data. The expression values of (A) hsa-mir-139, (B) hsa-mir-31, (C) hsa-mir-196b, (D) hsa-mir-221, (E) hsa-mir-203b, (F) hsa-mir-215, (G) hsa-mir-144 and (H) hsa-mir-4433b in 4 pancreatic cancer patients and 4 controls from test dataset GSE119794.
Figure 3. The expression levels of six hub genes in validation data. The expression values of (A) HOXA10, (B) BCL2, (C) ITGA3, (D) MET, (E) SH2B3 and (F) ERO1A in 4 pancreatic cancer patients and 4 controls from test dataset GSE119794.
sum test, \( P < .002, \text{Fig. 3A} \)). The same results were observed in ITGA3 (Paired Wilcoxon rank sum test, \( P < .004, \text{Fig. 3C} \)), MET (Paired Wilcoxon rank sum test, \( P < .006, \text{Fig. 3D} \)), and ERO1A (Paired Wilcoxon rank sum test, \( P < .002, \text{Fig. 3F} \)).

### 3.4. Correlations between miRNAs and genes

Sixteen miRNA and gene correlations between the 8 hub miRNAs and 6 hub genes were identified (shown in Fig. 4). From the 16 paired miRNAs and genes, 8 pairs were significantly

---

**Figure 4.** Correlation between miRNAs and genes. (A–P) presents correlation between (A) HOXA10 and hsa-mir-196b, (B) HOXA10 and hsa-mir-215, (C) HOXA10 and hsa-mir-144, (D) BCL2 and hsa-mir-196b, (E) BCL2 and hsa-mir-144, (F) BCL2 and hsa-mir-196b, (G) ITGA3 and hsa-mir-215, (H) ITGA3 and hsa-mir-144, (I) MET and hsa-mir-139, (J) MET and hsa-mir-31, (K) SH2B3 and hsa-mir-215, (L) SH2B3 and hsa-mir-4433b, (M) ERO1A and hsa-mir-221, (N) ERO1A and hsa-mir-203b, (O) ERO1A and hsa-mir-4433b.
correlated, which indicated that these miRNA and genes may play important roles in PC through the mutual regulation.

3.5. Functional analysis of the target genes

The target genes of these microRNAs were subjected to do GO functional enrichment analysis\(^{[26,27]}\) so as to analyze the possible roles of the 8 miRNAs played during the development of PC (Fig. 5). It was found these target genes with FDR <0.2 were mainly enriched in negative regulation of acute inflammatory response pathway, cell-substrate response pathway, o-glycan processing pathway, and so on.

3.6. ROC curve analysis of the 8 miRNAs in PC

ROC curve and the area under the ROC curves (AUC) were performed in GSE119794 test data to investigate the characteristics of the 8 miRNAs as biomarkers of PC; the test data included 10 PC patients and 10 control samples. The ROC curves illustrated strong separation between the tumor tissues and control group, with an AUC of 0.77 (95% confidence interval [CI]: 0.53–1) for hsa-mir-139 (Fig. 6A), hsa-mir-196b (AUC = 0.72, 95% CI: 0.48–0.96, Fig. 6C), hsa-mir-211 (AUC = 0.79, 95% CI: 0.58–1, Fig. 6D), hsa-mir-203b (AUC = 0.74, 95% CI: 0.51–0.97, Fig. 6E), and hsa-mir-144 (AUC = 0.79, 95% CI: 0.58–1, Fig. 6G) were similar with that of hsa-mir-139.

---

**Figure 5.** GO enrichment analysis of target genes of the eight hub miRNAs. The blue node represents the -log10 (P value) of each GO pathway.
Figure 6. The receiving-operating characteristic (ROC) curve of mRNA expression for distinguishing control and pancreatic cancer (PC) samples. ROC analysis of the eight miRNAs (A) hsa-mir-139, (B) hsa-mir-31, (C) hsa-mir-196b, (D) hsa-mir-221, (E) hsa-mir-203b, (F) hsa-mir-215, (G) hsa-mir-144, (H) hsa-mir-4433b, and (I) signature used to predict 10 PC patients from GSE119794. AUC = area under the curve, CI = confidence interval.
Furthermore, a miRNA-based biomarker for PC was constructed through the 8 miRNAs integrated by random forest algorithm. As shown in Figure 6f, this biomarker could accurately distinguish all PC samples from the control samples.

4. Discussions

In the past few years, numerous studies have shown that dysregulation of miRNAs contributes to the pathogenesis, diagnosis, and treatment of various human cancers including PC, which might provide a novel and promising field in the treatment of cancers.

In this study, we aimed to construct miRNA-based markers for early diagnosis of PC. Firstly, we constructed a miRNA-mRNA network based on the differentially expressed miRNAs and genes of PC relative to controls. Subsequently, 8 miRNAs that may be related to the occurrence and development of PC were found by the miRNA-mRNA network. Among which, Hsa-mir-139 had been proven to be down-expressed in various types of cancer, and it was found to be a kind of tumor-suppressing miRNA and played a key role in tumorigenesis of lung cancer. A research by Kanno et al. revealed that hsa-mir-196b played a carcinogenic role in PC and its high expression, showing a significant association with the poor prognosis of PC. Moreover, transfection of miR-196b inhibitor showed an anti-tumor effect in a PC cell line, which may indicate miR-196b as a potential diagnostic and therapeutic biomarker in PC. It was also reported that has-mir-221 could inhibit apoptosis and promote the proliferation and invasion of PC cells by upregulation of its expression. These findings also suggest that has-mir-221 may be a promising target in PC treatment. Whereas in another study, it was demonstrated that lower expression of hsa-mir-221 in PC patients may contribute to their longer survival time comparing to those of PC patients with higher expression of has-mir-221. In other words, downregulation of hsa-mir-221 in PC cells could suppress cell proliferation, and inhibit PC progression, which also verified the potential value of has-mir-221 as a prognostic marker for the poor survival of PC patients. Another miRNA, hsa-mir-144, a well-known onco-miRNA, has been found to be deregulated in several types of cancer including gastric cancer, breast cancer, colorectal cancer, and PC. And in PC, hsa-mir-144 was demonstrated to be able to induce cell cycle arrest and apoptosis through targeting proline-rich protein 11 (PRR11). Therefore, hsa-mir-144 may also be considered as a promising therapeutic target in PC treatment.

All the findings listed above also verified the accuracy of our results in this study, and further studies were needed to explore the roles of these miRNAs played during the processes of PC development. The functional enrichment analysis of these 8 miRNAs showed the target genes of these 8 miRNAs were mainly enriched in negative regulation of acute inflammatory response, cell-substrate adhesion, and O-glycan processing pathways, which were all critical players in PC. For example, inflammation always played an important role in the development of PC. Moreover, relative research also found that isoagglutinins were able to bind differentially expressed O-glycan-derived PC proteins and the cell-substrate adhesion and O-glycan processing pathway may play important roles in the pathogenesis and progression of PC. All this further indicated that the 8 miRNAs obtained may be related to the occurrence and development of PC.

In addition, the obtained 6 genes including HOXA10, BCL2, ITGA3, MET, SH2B3, and ERO1A were also verified for their expression patterns. And previous studies have showed that HOXA10 was closely associated with tumor progression in many human cancers and HOXA10 could promote cell invasion and MMP-3 expression of PC cells via TGFb2-p38 MAPK pathway. Knockdown of HOXA10 downregulated the expression of MMP-3 in Panc-1 and BxPC-3 cells, whereas overexpression of HOXA10 upregulated MMP-3 expression in MIA PaCa-2 cells, which confirmed that HOXA10 could increase MMP-3 expression in pancreatic cancer cells. Furthermore, knockdown of HOXA10 decreased the expression of TGFb2 in Panc-1 and BxPC-3 cells, whereas the overexpression of HOXA10 increased the level of TGFb2 in MIA PaCa-2 cells, which indicated that HOXA10 could regulate the expression of TGFb2 in pancreatic cancer cells. Jiao et al. examined the relationship between pancreatic ITGA3 expression with the clinical and pathological characteristics of PC patients, and their results indicated that ITGA3 expression could be served as a prognostic or diagnostic target in PC. They found that high pancreatic expression of ITGA3 was associated with the histological type, histological grade, stage, T classification, vital status, and relapse, which may be achieved through ITGA3-activation of the PI3K-Akt signaling pathway. ITGA3 also influenced glycolysis, notch signaling, P53 signaling, TGF-β signaling, the mitotic spindle, interferon alpha response, and so on. Additionally, ITGA3 promoted the epithelial-mesenchymal transition, cell migration, and invasion processes that were the main biology underlies relapse and metastasis and the most serious problems in pancreatic cancer. mTOR signaling endoplasmic reticulum oxidoareductase 1 alpha (ERO1L), was also proved to regulate the progression of various human cancers including PC. In addition, the PC patients with high ERO1L expression showed shorter survival time, which indicated ERO1L as an independent prognosis factor for PC patients’ prognosis. ERO1L was found to be upregulated in pancreatic cancer tissues and cells, and patients with high ERO1L expression showed high metastasis ability, whereas patients with low ERO1L expression had low metastasis ability, which revealed that ERO1L could promote pancreatic cancer metastasis. Besides, ERO1L significantly increased the Wnt/β-catenin pathway activity through regulating the downstream genes of Wnt/β-catenin including MYC, CD44, RUNX2, and SNAI1. Previous studies have suggested that the serum metadherin mRNA and serum HtrA2 mRNA could be served as potential diagnosis biomarkers for colorectal cancer and breast cancer, respectively.

To conclude, in this study, we identified 8 miRNAs and 6 genes that were closely associated with PC development. And the biomarkers constructed based on the 8 miRNAs could completely distinguish PC and control. Therefore, these miRNAs could be considered as biomarkers in the diagnosis or prognosis of PC. However, much more efforts and studies are still needed to make the miRNAs as useful tools in the future treatment management of PC.

Author contributions

BY.D, M.W, and ZW.L conceived this study; BY.D, M.W. and ZW.L performed the experiments; BY.D, M.W, and ZW.L prepared the manuscript. All authors approved the final version of this manuscript.
References

[1] Giovannetti E, Funel N, Peters GJ, et al. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer Res 2010;70:4528–38.

[2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.

[3] Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359–86.

[4] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65:5–29.

[5] Sultana A, Tudur Smith C, Cunningham D, et al. Meta-analyses of chemotherapy for locally advanced and metastatic pancreatic cancer: results of secondary end points analyses. Br J Cancer 2008;99:6–13.

[6] Baradaran B, Shahbazi R, Khordadmehr M. Dysregulation of key microRNAs in pancreatic cancer development. Biomed Pharmacother 2019;109:1008–15.

[7] Schouwen K, Le N, von Arnim U, et al. Circulating and fecal microRNAs as biomarkers for inflammatory bowel diseases. Inflamm Bowel Dis 2018;24:1547–57.

[8] Hussain SP. Pancreatic cancer: current progress and future challenges. Int J Cancer 2015;136:2558–69.

[9] Chitkara D, Mittal A, Mahato RI. miRNAs in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects of p53. J Biol Sci 2016;12:270.

[10] Ying H, Elpek KG, Vinjamoori A, et al. PTEN is a major tumor suppressor in pancreatic ductal adenocarcinoma. Mol Cell Proteomics 2016;15:2396.

[11] Sarker S, Dubaybo H, Ali S, et al. Down-regulation of miR-221 inhibits proliferation of pancreatic cancer cells through up-regulation of PTEN, p53, and p21(WAF1). J Cancer Res 2013;5:465–77.

[12] Kanno S, Nosho K, Ishigami K, et al. MicroRNA-196b is an independent prognostic biomarker in patients with pancreatic cancer. Carcinogenesis 2017;38:425–31.

[13] Berindan-Neagoe I, Montoug Padel C, Pascalli B, et al. MicroRNAome genome: a treasure for cancer diagnosis and therapy. CA Cancer J Clin 2014;64:311–36.

[14] Sun C, Sang M, Li S, et al. Hsa-miR-139-5p inhibits proliferation and causes apoptosis associated with down-regulation of c-Met. Oncotarget 2015;6:39756–92.

[15] Kanno S, Nosho K, Ishigami K, et al. MicroRNA-196b is an independent prognostic biomarker in patients with pancreatic cancer. Carcinogenesis 2017;38:425–31.

[16] Xu Q, Li P, Chen X, et al. miR-221/222 induces pancreatic cancer progression through the regulation of matrix metalloproteinases. Oncotarget 2015;6:14153–64.

[17] Hsu SD, Lin FM, Wu WY, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 2014;42:D78–85.

[18] Hsu SD, Lin FM, Wu WY, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res 2011;39:D163–9.

[19] Deng et al. Medicine (2020) 99:38

[20] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.

[21] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets. Nucleic Acids Res 2012;40:D13–17.

[22] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[23] Wang J, Zhou X, Zhu J, et al. GO-function: deriving biologically relevant functions from statistically significant functions. Brief Bioinform 2012;13:216–27.