An Integrative Data Mining for the Identification and Validation of Oncogenic Biomarkers in Pancreatic Carcinoma

Guihua Jin

First Affiliated Wenzhou Medical University

Qingqing Ruan

First Affiliated Hospital of Wenzhou University

Fugen Shangguan

First Affiliated Hospital of Wenzhou Medical University

Linhua Lan (✉ paullee@163.com)

First Affiliated Hospital of Wenzhou Medical University <https://orcid.org/0000-0003-4804-7767>

Primary research

Keywords: pancreatic carcinoma, bioinformatics analysis, gene expression, microRNA, biomarkers, LAMC2, RUNX2

Posted Date: July 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-40998/v1>

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

[Read Full License](#)

Abstract

Background: Pancreatic carcinoma (PC) is a severe disease associated with high mortality. Although strategies for cancer therapy made great progress, outcomes of pancreatic ductal adenocarcinoma patients remain extremely poor. Therefore, it is urgent to find novel biomarkers and therapeutic targets to improve outcomes of patients.

Methods: To identify general applicable targets for early diagnosis and therapy, we selected related microarray data which including three mRNA microarray datasets GSE62165, GSE15471, GSE32676 and two miRNA datasets GSE24279, GSE32678, and combinative analysis was performed by GEO2R. Functional and pathway enrichment analysis were performed using the DAVID database. MiRTarBase, miRWalk, Diana Tools and TBtools were used to get keys. TCGA database, HPA database and western blot experiments were used to verify diagnostic and prognostic value of key genes.

Results: By integrating mRNA and miRNA expression profiles, we identified 114 differentially expressed genes (DEGs) and 114 differentially expressed miRNAs (DEMs), respectively. Furthermore, three overlapping key genes, RUNX2, LAMC2 and FBXO32, were found by compared with DEMs target genes and DEGs. In detail, deregulation of 8 key miRNAs were closely related to poor outcomes and participated in crucial genes regulation which may contribute to build a miRNA biomarker panel for prognosis. Moreover, we confirmed that RUNX2 showed a potential property for distinguishing PC and normal people. We also demonstrated that aberrant over-expression of LAMC2 was associated with poor prognosis of PC patients as well as human tumor status and subtypes. The protein levels of RUNX2 and LAMC2 in PC patients were further verified by IHC from Human Protein Atlas and western blot experiments.

Conclusions: In summary, our current study identified that RUNX2 and LAMC2 may be promising targets for early diagnosis and therapy of PC patients.

Background

Pancreatic carcinoma (PC) is one of the most malignant cancer types in the world. Although considerable advancements in diagnostic and pharmacological therapy have greatly achieved, the 5-year overall survival (OS) rate of PC is still less than 9% ^[1]. Therefore, novel diagnostic and prognostic biomarkers as well as drug targets are urgently needed to improve individualized survival.

Microarray is a high-throughput platform for analysis of gene expression. It has revolutionized RNA and DNA studies and also been extensively conducted as an efficient tool for exploration of genes and biological processes of human diseases, which contribute to uncover the underlying biological mechanisms. Several studies reported the differentially expressed genes (DEGs) and differentially expressed microRNA (DEMs) of PC separately during the past years ^[2–4]. However, there remain questions about the interaction between DEGs and DEMs during the progression of the PC. Thus, it is necessary to realize these issues which could be helpful to find key genes.

In this study, we applied three mRNA profiling datasets (GSE62165 ^[5], GSE15471 ^[6] and GSE32676 ^[7]) and two microRNA (miRNA) profiling datasets (GSE24279 ^[8] and GSE32678 ^[7]), which were downloaded from the Gene Expression Omnibus (GEO) database, to obtain DEGs and DEMs between PC and normal tissue samples. Subsequently, DEMs targeted genes (DEMTGs) were predicted, and by overlapping analysis between DEMTGs and DEGs to filter potential genes and miRNAs. We further clarified differentially expression of key genes through The Cancer Genome Atlas (TCGA) database. Finally, the diagnostic and prognostic value of these key genes and miRNAs were evaluated and predicted by receiver operating characteristic (ROC) and survival analysis.

Methods

Data collection and DEGs/DEMs/tDEGs/tDEMs analysis

Gene expression profiles were downloaded from the GEO public database (<http://www.ncbi.nlm.nih.gov/geo/>) and the inclusion criteria were as follows: (1) pancreatic cancer, (2) January 1, 2009 to December 31, 2019, (3) homo sapiens, (4) tissue, (5) sample count exceeds 30, (6) mRNA or miRNA expression data available. As a result, a total of 5 datasets were selected, which included GSE62165, GSE15471, GSE24279, GSE32676 and GSE32678.

The GEO2R web tool was used to screen for DEGs and DEMs between the PC and normal tissue samples in each dataset. The log2FoldChange (logFC) was calculated and the P-values were adjusted to correct for the occurrence of false-positive results by using the default method. Then, $P\text{-value} < 0.05$ and $|\log_2FC| > 1.5$ as the cut-off criteria for significant DEGs, and $P\text{-value} < 0.05$ and $|\log_2FC| > 0.5$ for significant DEMs was settled. Subsequently, volcano plots of DEGs and DEMs were used to quickly identify differences and Venn analysis was performed to get overlapping up or down regulated DEGs/DEMs in all mRNA/miRNA datasets, respectively ^[9]. By adding common shared up regulated DEGs/DEMs with down regulated DEGs/DEMs, obtained datasets were named as tDEGs and tDEMs in current study.

Enrichment analysis and key genes acquisition

Target genes of tDEMs (tDEMTGs) were predicted by the datasets which downloaded from miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>), miRWalk (<http://mirwalk.umm.uni-heidelberg.de/>) and Diana Tools (<http://diana.imis.athena-innovation.gr/DianaTools/index.php>) ^[10–12]. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were applied for the potential relevant functional annotation and pathway enrichment analysis of tDEMTGs and tDEGs were analyzed by the Database for Annotation Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) ^[13]. Then, tDEMTGs were aligned with tDEGs to obtain the intersection termed "key genes" for further analysis. Additionally, tDEMs which were relevant to key genes regulation termed as "candidate miRNAs".

Survival analysis

To evaluate the prognostic value of key genes and miRNAs, the overall survival (OS) of PC patients was investigated based on the online database, Kaplan-Meier Plotter (<http://kmplot.com/analysis/index.php>)^[14]. High and low level cohorts of indicated genes were set according to the median value of gene expressions. The OS plot was obtained with the hazard ratio (HR), the 95% confidence interval (CI) and logrank P-value information.

Key genes expression analysis in TCGA and Genotype-Tissue Expression (GTEx) tissues

Gene Expression Profiling Interactive Analysis (GEPIA^[15], <http://gepia.cancer-pku.cn/>) is an interactive web server for estimating the RNA-Seq expression data from the TCGA (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>) and GTEx (<https://www.gtexportal.org/>) databases. The expression of key genes was analyzed in various tumor and non-tumor tissues using GEPIA, and the comparison of the key genes expression at different stages of PC was also performed. The clinical information of patients was also downloaded from TCGA for further data validation.

Data validation

Blood and saliva GEO data verification: We investigated a validation by comparing the key genes mRNA standardized values in five independent GEO datasets: four of which were from blood samples (GSE74629 [16], GSE49641 [17], GSE49515 [18] and GSE15932), and the last one was from saliva samples (GSE14245 [19]).

Immunohistochemical (IHC) verification: The IHC staining images of PC and normal tissue were obtained from the Human Protein Atlas [20] (HPA, www.proteinatlas.org) to validate expression of key genes. The staining, intensity, quantity and location of IHC images present in HPA was calculated to each gene.

Western Blot (WB) verification: A total of 6 pancreatic tissue samples of PC patients were obtained by surgical resection and further divided into the tumor group and the adjacent non-cancerous group (normal). The study was approved by the Institutional Ethics Committee of Wenzhou Medical University and written informed consent was obtained from each patient before their enrollment. WB protocol was performed according to our previously described procedures [21]. The following primary antibodies were used: RUNX2 (ABclonal, cat. no. A2851), LAMC2 (ABclonal, cat. no. A1869) and β -Actin (Abmart, cat. no. P30002). The protein expression levels in samples were quantified by Image J software.

Statistical Analysis

The discriminative ability of key genes and miRNAs in the GEO datasets was calculated by ROC analysis, the pROC^[22] R package was performed in R 3.6.2 (<http://www.R-project.org/>). The area under the curve (AUC), 95% CI and P-value for assessing the predictive accuracy and specificity of ROC were calculated using SPSS version 23.0 (IBM, Chicago, IL, USA). All scatter and bar plots were generated by Graph Pad Prism 7 Software (GraphPad Software, Inc.). Comparisons between means were performed by ANOVA. P-value < 0.05 was considered statistically significant.

Results

Initial identification of tDEGs, tDEMs and tDEMTGs in pancreatic cancer

To find key genes which expressed differentially between PC patients and healthy people, we decided to filter them out in two sides: mRNA and miRNA targeted method. Firstly, we selected three expression array profiling datasets (GSE62165, GSE15471 and GSE32676) and two non-coding RNA array profiling datasets (GSE24279 and GSE32678) from GEO database. As for mRNA, 1398, 607 and 1157 DEGs were extracted from GSE62165, GSE15471 and GSE32676. Among them, there were 997, 562, 669 (103 shared genes) up-regulated and 401, 45, 488 (11 shared genes) down-regulated DEGs were identified. Taking advantage of Venn analysis in up and down regulated DEGs respectively, we captured 114 total shared DEGs, which we termed as tDEGs in this study, where 103 were high-expression and 11 were low-expression genes (Fig. 1A&B). As for miRNA, the same operations were also performed. 417 and 569 DEMs were extracted from GSE24279 and GSE32678. Two miRNA datasets totally shared 66 DEMs, which we termed as tDEMs, including 24 were up-regulated and 42 were down-regulated miRNAs (Fig. 1C&D).

The target genes of tDEMs were predicted by miRTarBase, miRWalk and Diana Tools databases. And the 114 consistent expression genes, which termed as tDEMTGs, were found by Venn analysis (Fig. 1E).

GO and pathway enrichment analysis

The GO term and KEGG pathway analysis were performed via DAVID. Firstly, the results of GO term enrichment analysis varied a lot between tDEGs and tDEMTGs (Fig. 2A&B). Biological process (BP) analysis of GO showed tDEGs were significantly enriched in extracellular matrix associated part, such as cell adhesion, extracellular matrix organization and disassembly, collagen catabolism and organization. As for tDEMTGs, top 2 clusters of genes were enriched in the regulation of transcription and proliferation, which showed a similar result in tDEGs, but other genes were mainly responsible for the regulation of cell death associated part. For GO cell component (CC) analysis, the tDEGs were significantly enriched in extracellular part, such as extracellular region, extracellular exosome, extracellular space, extracellular matrix and proteinaceous extracellular matrix. However, the results of tDEMTGs were focused on plasma membrane part, cytosol, organelle lumen, membrane-enclosed lumen and intracellular organelle lumen. The same differences also showed in molecular function (MF) analysis, tDEGs were mainly enriched in protein binding, but tDEMTGs were enriched in nucleotide binding, transcription regulator activity, ribonucleotide binding, purine ribonucleotide binding and purine nucleotide binding.

Additionally, the differences between tDEGs and tDEMTGs were also confirmed KEGG pathway analysis. The result disclosed that tDEGs were involved in PI3K-Akt signaling pathway, focal adhesion, ECM-receptor interaction, amoebiasis and pathways in cancer (Fig. 2C). However, tDEMTGs were significantly enriched in pathways in cancer and partially enriched in pancreatic cancer (Fig. 2D).

Identification of key genes and miRNAs

To identify potential key genes, tDEMTGs were aligned with tDEGs to obtain the intersection part, which we termed as key genes for further analysis (Fig. 1F). Importantly, the result showed runt related transcription factor 2 (RUNX2), laminin subunit gamma-2 (LAMC2) and F-box protein 32 (FBXO32) were commonly shared and listed in Table 1. Subsequently, we screened candidate miRNAs involved in the regulation of key genes from 66 tDEMs, and found 16 up-regulated and 19 down-regulated key miRNAs (Table 2).

Table 1
The differential expression of key genes in different GEO datasets.

Gene	Datasets	ID	LogFC	P.Val	Expression
RUNX2	GSE62165	11725180_a_at	3.7	3.10E-29	UP
	GSE15471	232231_at	2.93	9.52E-11	UP
	GSE32688(mRNA)	232231_at	1.69	0.1046045	UP
LAMC2	GSE62165	11737325_a_at	3.49	4.80E-25	UP
		11743514_a_at	4.09	7.56E-26	UP
		11743516_s_at	4.29	1.49E-26	UP
		11743515_s_at	4.35	1.06E-28	UP
		11747923_s_at	4.37	7.93E-31	UP
	GSE15471	207517_at	1.60	3.83E-09	UP
		202267_at	2.76	1.76E-13	UP
	GSE32688(mRNA)	202267_at	4.93	0.0002728	UP
		207517_at	2.25	0.0038994	UP
FBXO32	GSE62165	11719394_a_at	2.88	3.83E-22	UP
	GSE15471	225803_at	1.72	9.65E-14	UP
		241762_at	1.83	2.64E-11	UP
		225328_at	1.91	5.28E-12	UP
	GSE32688(mRNA)	241763_s_at	2.44	0.0095169	UP
		241762_at	2.32	0.0215962	UP
		225803_at	1.88	0.0176426	UP
		225345_s_at	1.56	0.0702839	UP

Table 2
The differential expression of candidate miRNAs in different GEO datasets.

Target gene	miRNA	Datasets	LogFC	Expression
RUNX2	hsa-miR-519c-3p	GSE32678	0.531577	UP
		GSE24279	3.2867924	
	hsa-miR-876-3p	GSE32678	0.755198	UP
		GSE24279	9.9730133	
	hsa-miR-1265	GSE32678	0.883218	UP
		GSE24279	0.6795887	
	hsa-miR-92a-2-5p	GSE32678	-1.231335	DOWN
		GSE24279	-0.7331099	
	hsa-miR-193b-5p	GSE32678	0.6020578	DOWN
		GSE24279	-1.218209	
	hsa-miR-517-5p	GSE32678	-0.606971	DOWN
		GSE24279	-1.2242733	
	hsa-miR-488-3p	GSE32678	-1.145003	DOWN
		GSE24279	-1.145003	
RUNX2/FBXO32	hsa-miR-23b-3p	GSE32678	1.214636	UP
		GSE24279	0.7358448	
	hsa-miR-221-5p	GSE32678	1.277623	UP
		GSE24279	1.8034373	
	hsa-miR-199a-5p	GSE32678	1.240204	UP
		GSE24279	1.6889355	
	hsa-miR-376a-5p	GSE32678	0.546421	UP
		GSE24279	2.3186641	
	hsa-miR-1825	GSE32678	1.225039	UP
		GSE24279	4.0684489	
	hsa-miR-1227-3p	GSE32678	0.687478	UP
		GSE24279	3.0619077	
	hsa-miR-198	GSE32678	-1.004687	DOWN

Target gene	miRNA	Datasets	LogFC	Expression
		GSE24279	-0.6024842	
	hsa-miR-937-3p	GSE32678	-0.768939	DOWN
		GSE24279	-1.2399531	
FBXO32	hsa-miR-345-5p	GSE32678	0.821379	UP
		GSE24279	7.9474776	
	hsa-miR-1233-3p	GSE32678	0.528843	UP
		GSE24279	1.0034567	
	hsa-miR-1289	GSE32678	0.583983	UP
		GSE24279	0.5373088	
	hsa-miR-621	GSE32678	0.962951	UP
		GSE24279	0.9769538	
	hsa-miR-1248	GSE32678	1.377747	UP
		GSE24279	1.0123159	
	hsa-miR-30b-3p	GSE32678	-0.705179	DOWN
		GSE24279	-0.7012921	
	hsa-miR-373-5p	GSE32678	-0.943119	DOWN
		GSE24279	-1.5962925	
	hsa-miR-492	GSE32678	-0.665303	DOWN
		GSE24279	-1.1011978	
	hsa-miR-510-5p	GSE32678	-1.069006	DOWN
		GSE24279	-2.499187	
	hsa-miR-575	GSE32678	-0.77259	DOWN
		GSE24279	-0.5620953	
	hsa-miR-604	GSE32678	-0.508662	DOWN
		GSE24279	-1.2474484	
	hsa-miR-611	GSE32678	-1.159661	DOWN
		GSE24279	-1.6342346	
	hsa-miR-617	GSE32678	-0.816815	DOWN

Target gene	miRNA	Datasets	LogFC	Expression
		GSE24279	-1.5017728	
	hsa-miR-921	GSE32678	-0.90506	DOWN
		GSE24279	-1.9104782	
	hsa-miR-520d-5p	GSE32678	-0.660262	DOWN
		GSE24279	-0.6523876	
FBX32/LAMC2	hsa-miR-199b-5p	GSE32678	1.544751	UP
		GSE24279	1.0106128	
LAMC2	hsa-miR-588	GSE32678	0.526075	UP
		GSE24279	4.3557666	
	hsa-miR-891a-5p	GSE32678	-0.713461	DOWN
		GSE24279	-1.3790634	
	hsa-miR-622	GSE32678	-0.961764	DOWN
		GSE24279	-1.1362947	
LAMC2/RUNX2	hsa-miR-628-3p	GSE32678	-0.64033	DOWN
		GSE24279	-0.9205919	

Diagnostic and prognostic value of candidate miRNAs

To test the clinical applicability of the screening results, we explored the diagnostic significance of candidate miRNAs by ROC curve analysis (Fig. 3A). However, the mean AUC values of them were all lower than 0.85 (Fig. 3B). Then, to further evaluate the prognostic value, survival analysis was applied and the results showed 10 up-regulated key miRNAs indicated a significantly lower OS rate: hsa-miR-519c-3p, hsa-miR-1265, hsa-miR-1825, hsa-miR-1227-3p, hsa-miR-1233-3p, hsa-miR-1289, hsa-miR-621, hsa-miR-1248, hsa-miR-199b-5p and hsa-miR-588 (Fig. 4A). Correspondingly, three down-regulated candidate miRNAs were positively related with OS rate: hsa-miR-488-3p, hsa-miR-30b-3p and hsa-miR-628-3p (Fig. 4B). These data indicated that these 13 candidate miRNAs could be potential for clinical application as prognostic markers.

Diagnostic and prognostic value of key genes

We then compared the transcriptional levels of key genes in multiple cancers with those in normal samples by using TCGA database via GEPIA (Fig. 5A&B&C). The mRNA expression levels of RUNX2, LAMC2 and FBX32 showed significant differences between cancer and adjacent non-cancerous tissues especially in pancreatic adenocarcinoma (PAAD, Fig. 5.D). ROC and survival analysis were also

performed to ensure the application prospect of key genes. All the three key genes indicated a good discriminating ability, of which the mean AUC values of each key gene in three GEO datasets were all greater than 0.85 (Fig. 5E). Furthermore, the mean AUC values of LAMC2 and FBXO32 both exceeded 0.95 which indicated that these two molecules might be promising biomarkers to PC diagnosis. The results of survival analysis showed only LAMC2 expression level affected OS rate significantly (HR = 3.06, P = 0.00011, Fig. 5F). It suggested that LAMC2 may be an important molecule which participate in the development of PC and also act as a potential prognostic biomarker.

Verification of key genes

In addition to high accuracy and specificity, the diagnostic methods should also be painless and woundless. Blood and saliva are easily obtained samples that we can get with little harm for the body. Thus, we verified the diagnostic possibility of key genes in 4 blood cells GEO datasets (GSE74629, GSE49641, GSE49515 and GSE15932) and 1 saliva cells GEO dataset (GSE14245). Interestingly, we found only RUNX2 had significant reduction compared with normal controls in saliva and blood cells (GSE14245 and GSE49641) which suggested that RUNX2 could be considered as one of promising candidate biomarkers of PC. However, LAMC2 and FBXO32 were not significantly different between the control and PC samples.

Then, to further clarify the expression of RUNX2 and LAMC2 in PC tissue, we studied the IHC experiments of these two genes in HPA (Fig. 6B, Supplementary Fig. 1). First, the result of LAMC2 was consistent with previous analysis. The staining of anti-LAMC2 antibodies were all above medium in PC patients, which were hardly detected in normal pancreas tissues. Although staining quantity were varying (5.5/12 were > 75%, 2.5/12 were 75%-25% and 4/12 were < 25% stained), almost all of their staining intensity were strong. Besides, LAMC2 were visualized in cytoplasmic/membranous while RUNX2 were not corroborating with predicted. Though a slight up-regulation of RUNX2 could be found in some samples (17%), most of them which just like normal samples could not be detected. Consistently, immune blot was performed to detect the level of RUNX2 which showed a similar result with IHC (Fig. 6C). These data indicated that LAMC2 may play a crucial role in the PC therapy.

In a second validation study, we examined RUNX2 and LAMC2 performance in different clinical stage, sex, age, tumor subtype and personal tumor status groups (Fig. 7). We found RUNX2 was only significantly up-regulated in different stage groups. However, LAMC2 were significantly high-expressed not only in advanced stage PC patients, but also in ductal type PC group (P = 0.03) and PC patients (P = 0.02).

According to above evidences, we further found LAMC2 might be the main factor in influencing median survival time of PC patients by combination analysis of RUNX2 and LAMC2 (Fig. 8).

Discussion

Recently, increased morbidity and mortality of PC were concerned as a severe challenge for human health. Despite great advantages in past years, the early diagnosis and efficient therapeutic strategies of PC are still huge problems owing to lacking of deep understanding about PC carcinogenesis. Currently, due to development of bioinformatics technology, it is much easier to identify the general genetic alterations in diseases which may shed light on determination of hub targets for clinical utility. Studies of miRNA biology have expanded considerably since first discovery, and have been considered as attractive targets because of their crucial roles in modulation of gene expression under healthy, inflamed and carcinogenic pathological state. Therefore, identifying general diagnostic and prognostic biomarkers in cancer is vital to analyze the changes of gene expression in combination with miRNAs.

In the present study, we initially screened out 114 tDEGs and 114 tDEMs from 3 and 2 GEO datasets respectively. By GO function and KEGG analysis, we obtained an in-depth understanding of these genes attached to PC initiation and progression. Further, intending to investigate the general profiles of molecular alterations in PC, we then identified 3 key genes and 35 important miRNAs by aligning tDEMTGs with tDEGs. We investigated diagnostic and prognostic value of them by ROC and survival tests. And we also confirmed protein expression with IHC and WB assays. According to the data of current study, we suggested that LAMC2 and RUNX2 could serve as potential candidates for the clinical application of PC in the future.

MicroRNAs (miRNAs) are a class of non-coding RNAs which can bond to 3'-untranslated region (3'-UTR) of targeted mRNA to regulate their proteins expression levels [23]. Multiple studies have demonstrated that miRNAs participate in the management of all cancer hallmarks. It is commonly considered that miRNAs could be important molecular tools for noninvasive diagnosis and prognosis of cancer. Therefore, it is of great importance to screen most commonly suitable DEGs which could be used as treatment targets or diagnostic markers by discussing the interaction of miRNAs and mRNAs. In this study, we finally obtained 8 key miRNAs from candidate miRNAs by further screening through ROC and survival analysis, which enriched miRNA profiles of PC and may help to highlight the diagnostic and therapeutic potential of miRNAs cluster in PC (Fig. 9).

RUNX2 belongs to RUNX family, and is known as one of the major determinants of osteoblast differentiation and bone formation. Recent studies found RUNX2 was overexpressed in several tumor tissues and may play a vital role in tumor initiation, progression, invasion and metastasis [24–28]. Our results demonstrated that although overexpression of RUNX2 was related to malignant behaviours, it didn't significantly affect the survival time of the PC patients which might be a bit inconsistent with previous research. Intriguingly, we found that RUNX2 presented a high diagnostic ability in PC. Histological analysis is an invasive examination that is not suitable for the early diagnosis of PDAC, in terms of safety. Therefore, blood or body fluid testing may provide a good alternative method to assist clinicians with diagnosis. Unlike high-expression of RUNX2 in pancreatic tissue, it was down-regulated in blood datasets, and its expression in peripheral blood mononuclear cell (PBMC, GSE49641) was significantly lower in PC group than in healthy control samples. Saliva samples (GSE14245) showed a similar result which suggesting its application possibility as a noninvasive diagnostic biomarker.

However, there were no significant differences detected in peripheral blood (GSE74629 and GSE15932). In addition to these, RUNX2 had been described that could regulate multiple carcinogenesis genes and molecules in PC, including VEGF (vascular endothelial growth factor), matrix metalloproteinases and survivin [29]. Combined with the results of transcription factors prediction, we believed that the role of RUNX2 in PC progression might concentrate more on regulating the target genes, such as LAMC2.

LAMC2 is a member of laminins family which are the main structural component of basement membranes and participate in a wide variety of biological processes including cell adhesion, differentiation, migration and metastasis [30, 31]. In this study, results of the enrichment analysis indicate that LAMC2 are closely related with the process of cell adhesion, extracellular matrix organization/disassembly and positive regulation of cell proliferation. We found LAMC2 expression was elevated in PC tumor tissues through GEO database, IHC and WB analysis. Additionally, PC patients who harbored high expression of LAMC2 had a poor prognosis. According to these data, we hypothesized that LAMC2 expression may lead to an increased risk of tumor recurrence. Pancreatic ductal adenocarcinoma (PDAC) accounts for most human PC cases (more than 95%) [32], we found that the expression of LAMC2 is mainly up-regulated in PDAC, which indicates that it is not only suitable for most PC patients, but it may also help to identify the subtypes of PC. Because carbohydrate antigen 19 – 9 (CA19-9) which acts as the most commonly used PC marker is not accurate for the diagnosis of PDAC [33], combining with the good discriminating ability that LAMC2 showed in the AUC curve, we suggest that it may be a potential indicator in the auxiliary diagnosis of PC.

Conclusions

In summary, we confirmed that RUNX2 and LAMC2 are key genes that facilitate the progression of PC through bioinformatics and experimental analysis. Eight key miRNAs (miR-588, miR-199b, miR-1227, miR-628, miR-488, miR-1825, miR-519 and miR1265) which participated in the regulation of key genes expression might enrich the specific miRNA profiles of PC. These findings provide a new perspective on the underlying molecular mechanism of PC, suggesting that LAMC2 and RUNX2 may be valuable biomarkers and therapeutic targets for PC patients and may also offer powerful evidence and clues for the future genomic individualized treatment of PC.

- **List of abbreviations**

Area under the curve, AUC; biological process, BP; carbohydrate antigen 19 – 9, CA19-9; cell component, CC; confidence interval, CI; Database for Annotation Visualization and Integrated Discovery, DAVID; DEMs target genes, DEMTGs; differentially expressed genes, DEGs; differentially expressed miRNAs, DEMs; F-box protein 32, FBXO32; Genotype-Tissue Expression, GTEx; Gene Expression Profiling Interactive Analysis, GEPIA; gene ontology, GO; pancreatic carcinoma, PC; Gene Expression Omnibus, GEO; hazard ratio, HR; Human Protein Atlas, HPA; immunohistochemical, IHC; Kyoto Encyclopedia of Genes and Genomes, KEGG; laminin subunit gamma-2, LAMC2; log2FoldChange, logFC; molecular function, MF; overall survival, OS; pancreatic ductal adenocarcinoma, PDAC; pancreatic adenocarcinoma, PAAD;

microRNA, miRNA; receiver operating characteristic, ROC; runt related transcription factor 2, RUNX2; The Cancer Genome Atlas, TCGA; target genes of tDEMs, tDEMTGs; VEGF, vascular endothelial growth factor.

Abbreviations

Area under the curve, AUC; biological process, BP; carbohydrate antigen 19-9, CA19-9; cell component, CC; confidence interval, CI; Database for Annotation Visualization and Integrated Discovery, DAVID; DEMs target genes, DEMTGs; differentially expressed genes, DEGs; differentially expressed miRNAs, DEMs; F-box protein 32, FBXO32; Genotype-Tissue Expression, GTEx; Gene Expression Profiling Interactive Analysis, GEPIA; gene ontology, GO; pancreatic carcinoma, PC; Gene Expression Omnibus, GEO; hazard ratio, HR; Human Protein Atlas, HPA; immunohistochemical, IHC; Kyoto Encyclopedia of Genes and Genomes, KEGG; laminin subunit gamma-2, LAMC2; log2FoldChange, logFC; molecular function, MF; overall survival, OS; pancreatic ductal adenocarcinoma, PDAC; pancreatic adenocarcinoma, PAAD; microRNA, miRNA; receiver operating characteristic, ROC; runt related transcription factor 2, RUNX2; The Cancer Genome Atlas, TCGA; target genes of tDEMs, tDEMTGs; VEGF, vascular endothelial growth factor.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Ethics Committee of Wenzhou Medical University and written informed consent was obtained from each patient before their enrollment.

Consent for publication

Not applicable.

Availability of data and material

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

The Authors declare that there is no conflict of interest.

Funding

This work was partially supported by Natural Science Foundation Youth Fund Project of China (No. 81902803). And it was also supported by the incubation project of the First Affiliated Hospital of Wenzhou Medical University.

Authors' contributions

Guihua Jin have drafted the work. Qingqing Ruan and Fugen Shangguan have substantively revised it. Linhua Lan have designed of this work. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank the GEO database, TCGA group, miRbase database, miRWalk database, DIANA TOOLS, HPA, GEPIA, Kaplan Meier plotter and TBtools for providing relevant data and tools that is publicly accessible.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
2. Almeida PP, Cardoso CP, de Freitas LM. PDAC-ANN: an artificial neural network to predict pancreatic ductal adenocarcinoma based on gene expression. *BMC Cancer*. 2020;20(1):82.
3. Yang C, Liu Z, Zeng X, Wu Q, Liao X, Wang X, Han C, Yu T, Zhu G, Qin W, et al. Evaluation of the diagnostic ability of laminin gene family for pancreatic ductal adenocarcinoma. *Aging*. 2019;11(11):3679–703.
4. Long NP, Jung KH, Anh NH, Yan HH, Nghi TD, Park S, Yoon SJ, Min JE, Kim HM, Lim JH, et al: An Integrative Data Mining and Omics-Based Translational Model for the Identification and Validation of Oncogenic Biomarkers of Pancreatic Cancer. *Cancers (Basel)* 2019, 11(2).
5. Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, Roskams T, Aerts S, Topal B. Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. *BMC Cancer*. 2016;16:632.
6. Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology*. 2008;55(88):2016–27.
7. Donahue TR, Tran LM, Hill R, Li Y, Kovochich A, Calvopina JH, Patel SG, Wu N, Hindoyan A, Farrell JJ, et al. Integrative survival-based molecular profiling of human pancreatic cancer. *Clin Cancer Res*. 2012;18(5):1352–63.
8. Bauer AS, Keller A, Costello E, Greenhalf W, Bier M, Borries A, Beier M, Neoptolemos J, Buchler M, Werner J, et al. Diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis by measurement of microRNA abundance in blood and tissue. *PLoS One*. 2012;7(4):e34151.
9. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* 2018:289660.
10. Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M, Kourtis K, Koziris N, Dalamagas T, Hatzigeorgiou AG. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res* 2011, 39(Web Server issue):W145-148.

11. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018;46(D1):D296–302.
12. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One.* 2018;13(10):e0206239.
13. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics.* 2012;28(13):1805–6.
14. Nagy A, Lanczky A, Menyhart O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep.* 2018;8(1):9227.
15. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47(W1):W556–60.
16. Irigoyen A, Jimenez-Luna C, Benavides M, Caba O, Gallego J, Ortuno FM, Guillen-Ponce C, Rojas I, Aranda E, Torres C, et al. Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers. *PLoS One.* 2018;13(4):e0194844.
17. Caba O, Prados J, Ortiz R, Jimenez-Luna C, Melguizo C, Alvarez PJ, Delgado JR, Irigoyen A, Rojas I, Perez-Florido J, et al. Transcriptional profiling of peripheral blood in pancreatic adenocarcinoma patients identifies diagnostic biomarkers. *Dig Dis Sci.* 2014;59(11):2714–20.
18. Shi M, Chen MS, Sekar K, Tan CK, Ooi LL, Hui KM. A blood-based three-gene signature for the non-invasive detection of early human hepatocellular carcinoma. *Eur J Cancer.* 2014;50(5):928–36.
19. Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park NH, Chia D, Wong DT. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology.* 2010;138(3):949–57. e941-947.
20. Uhlen M, Bjorling E, Agaton C, Szigartyo CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics.* 2005;4(12):1920–32.
21. Y LLWWYZLNXCDH. G, S M, J L, M G et al: Deferoxamine suppresses esophageal squamous cell carcinoma cell growth via ERK1/2 mediated mitochondrial dysfunction. *Cancer letters.* 2018;432:132–43.
22. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
23. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat Rev Drug Discov.* 2017;16(3):203–22.
24. Li H, Zhou RJ, Zhang GQ, Xu JP. Clinical significance of RUNX2 expression in patients with nonsmall cell lung cancer: a 5-year follow-up study. *Tumour Biol.* 2013;34(3):1807–12.
25. Wang ZQ, Keita M, Bachvarova M, Gobeil S, Morin C, Plante M, Gregoire J, Renaud MC, Sebastianelli A, Trinh XB, et al. Inhibition of RUNX2 transcriptional activity blocks the proliferation, migration and invasion of epithelial ovarian carcinoma cells. *PLoS One.* 2013;8(10):e74384.

26. Zhou S, Zhang S, Wang L, Huang S, Yuan Y, Yang J, Wang H, Li X, Wang P, Zhou L, et al. BET protein inhibitor JQ1 downregulates chromatin accessibility and suppresses metastasis of gastric cancer via inactivating RUNX2/NID1 signaling. *Oncogenesis*. 2020;9(3):33.
27. Fritz AJ, Hong D, Boyd J, Kost J, Finstaad KH, Fitzgerald MP, Hanna S, Abuarqoub AH, Malik M, Bushweller J, et al. RUNX1 and RUNX2 transcription factors function in opposing roles to regulate breast cancer stem cells. *J Cell Physiol* 2020.
28. Boregowda RK, Olabisi OO, Abushahba W, Jeong BS, Haenssen KK, Chen W, Chekmareva M, Lasfar A, Foran DJ, Goydos JS, et al. RUNX2 is overexpressed in melanoma cells and mediates their migration and invasion. *Cancer Lett*. 2014;348(1–2):61–70.
29. Akech J, Wixted JJ, Bedard K, van der Deen M, Hussain S, Guise TA, van Wijnen AJ, Stein JL, Languino LR, Altieri DC, et al. Runx2 association with progression of prostate cancer in patients: mechanisms mediating bone osteolysis and osteoblastic metastatic lesions. *Oncogene*. 2010;29(6):811–21.
30. Garg M, Braunstein G, Koeffler HP. LAMC2 as a therapeutic target for cancers. *Expert Opin Ther Targets*. 2014;18(9):979–82.
31. Kosanam H, Prassas I, Chrystoja CC, Soleas I, Chan A, Dimitromanolakis A, Blasutig IM, Ruckert F, Gruetzmänn R, Pilarsky C, et al. Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues. *Mol Cell Proteomics*. 2013;12(10):2820–32.
32. Tanaka S. Molecular Pathogenesis and Targeted Therapy of Pancreatic Cancer. *Ann Surg Oncol*. 2016;23(Suppl 2):197–205.
33. Molina V, Visa L, Conill C, Navarro S, Escudero JM, Auge JM, Filella X, Lopez-Boado MA, Ferrer J, Fernandez-Cruz L, et al. CA 19 – 9 in pancreatic cancer: retrospective evaluation of patients with suspicion of pancreatic cancer. *Tumour Biol*. 2012;33(3):799–807.

Figures

Figure 1

Identification of DEGs, DEMs and key genes. A. The volcano plot of mRNA GEO datasets GSE62165, GSE15471 and GSE32688 respectively; B. Venn diagram of up-regulated DEGs and down-regulated DEGs set from GSE62165, GSE15471 and GSE32688 datasets; C. The volcano plot of miRNA GEO datasets GSE24279 and GSE32688 respectively; D. Venn diagram of up-regulated DEMs and down-regulated DEMs set from GSE24279 and GSE32688 datasets; E. The target genes, tDEMTGs, of 66 consistent

DEMs (tDEMs) were predicted by miRTarBase, miRWalk and Diana Tools databases. F. Venn diagram of tDEMTGs and tDEGs to get key genes.

Figure 2

Identification of DEGs, DEMs and key genes. A. The volcano plot of mRNA GEO datasets GSE62165, GSE15471 and GSE32688 respectively; B. Venn diagram of up-regulated DEGs and down-regulated DEGs set from GSE62165, GSE15471 and GSE32688 datasets; C. The volcano plot of miRNA GEO datasets

GSE24279 and GSE32688 respectively; D. Venn diagram of up-regulated DEMs and down-regulated DEMs set from GSE24279 and GSE32688 datasets; E. The target genes, tDEMTGs, of 66 consistent DEMs (tDEMs) were predicted by miRTarBase, miRWalk and Diana Tools databases. F. Venn diagram of tDEMTGs and tDEGs to get key genes.

Figure 3

ROC analysis of key genes related DEMs in GSE24279 and GSE32688. A. ROC curve for 35 key genes related DEMs in GSE24279 and GSE32688. B. Mean value of AUC results of 35 DEMs.

Figure 4

Survival analysis of 13 key DEMs expression in TCGA PAAD cohort. Green color indicates that the miRNA is downregulated, whereas red color indicates upregulation.

Figure 5

Expression, ROC and survival analysis of key genes. (A-C) Differential expression of key genes (RUNX2, LAMC2 and FBXO32) in various tumor and non-tumor tissues. D. Differential expression of key genes in PC patients and normal controls. E. ROC curve of key genes in GSE62165, GSE15471 and GSE32688. F. Survival analysis of key genes in TCGA PAAD cohort.

Figure 6

Verification of key genes. A. RUNX2, LAMC2 and FBXO32 expression levels (Log2FoldChange value) in GSE14245, GSE74629, GSE49641, GSE49515 and GSE15932. B. Representative immunohistochemistry staining of RUNX2 and LAMC2 in pancreatic ductal adenocarcinoma (PC tissue) and control pancreatic tissue (Normal tissue) in Human Protein Atlas (HPA). Scales bars represent 200 μ m. C. Western blot analysis for RUNX2 and LAMC2 expression proteins in 6 paired samples from PC patients.

Figure 7

The correlation of RUNX2 and LAMC2 expression with the clinico-pathological characteristics of patients. A. Expression levels of RUNX2 and LAMC2 in pancreatic cancer (PC) patients of different clinical stages. B. Expression levels of RUNX2 and LAMC2 in different sex groups of PC patients. C. Expression levels of RUNX2 and LAMC2 in different age groups of PC patients. D. Expression levels of RUNX2 and LAMC2 in different tumor subtypes of PC patients. E. Expression levels of RUNX2 and LAMC2 in different tumor status groups of PC patients.

Figure 8

Survival analysis of combined RUNX2 and LAMC2 in TCGA PAAD cohort.

Figure 9

The key genes and key miRNA interaction network.

Supplementary Files

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

- [sm.docx](#)