Development and Validation of a Novel Glycolysis-Related Risk Signature for Predicting Survival in Pancreatic Adenocarcinoma

Ang Li

The First Hospital of Jilin University

Sinan Hou

Lianyungang Traditional Chinese Medicine Hospital

Jian Chen

Zhongda Hospital affiliated to Southeast University

Huimin Hou

The First Hospital of Jilin University

Yanfang Jiang (✉ yanfangjiang1@hotmail.com)

The First Hospital of Jilin University <https://orcid.org/0000-0002-7403-2705>

Research

Keywords: Pancreatic Adenocarcinoma, Glycolysis, Risk Signature, Survival

Posted Date: May 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29859/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Clinica Chimica Acta on July 1st, 2021. See the published version at <https://doi.org/10.1016/j.cca.2021.03.020>.

Abstract

Background

Pancreatic adenocarcinoma (PAAD) is one of the leading causes of cancer death worldwide. Through data mining, an increasing number of biomarkers have been identified for predicting survival of PAAD. However, the ability of single gene biomarkers to predict patient survival is still insufficient. This study aimed to develop a novel risk signature for predicting survival of PAAD.

Methods

mRNA expression profiling was performed in a large PAAD cohort (N = 177) from The Cancer Genome Atlas (TCGA) database. Gene Set Enrichment Analysis (GSEA) was analyzed to detect whether the gene sets showed statistically differences between PAAD and adjacent normal tissues. Univariate Cox regression analysis was used to analyze and identify genes related to overall survival (OS), then subjected to multivariable Cox regression to further confirm the prognostic genes and obtain the coefficients. The expression level of selected genes weighted by their coefficients through linearly combining, we constructed a risk score formula for prognostic prediction. The three-mRNA signature for survival prediction is validated by Kaplan–Meier curve analysis.

Results

We demonstrated that a set of three genes (KIF20A, CHST2, and MET) were significantly associated with OS. Based on this three-gene signature, 177 PAAD patients were classified into high-risk groups and low-risk groups using the median risk score as cut off value. Additionally, multivariate Cox regression analysis revealed that the three-gene signature had independent prognostic value.

Conclusions

To our best knowledge, we first develop a glycolysis-related risk signature for predicting survival of pancreatic adenocarcinoma. The findings provide insight into identification of patients with poor prognosis in PAAD and improve novel therapy targets for this disease.

Introduction

Pancreatic adenocarcinoma (PAAD), the third leading cause of cancer death worldwide, is the most common type of pancreatic cancer (PC) [1]. Due to lack of specific clinical symptoms, PAAD is usually diagnosed at an advanced stage with invasive and metastatic characters [2, 3]. At present, surgery, radiotherapy, and chemotherapy are still the major treatment methods for PAAD. The therapeutic potential of these methods is limited, leading to a high mortality rate[1, 4]. Moreover, the prognosis of PAAD

patients is mainly evaluated through histopathology. However, patients with the same stage can show different prognoses and treatment responses [5]. Thus, it is crucial to develop reliable prognostic biomarkers and improve novel therapy targets for PAAD.

Based on mass data generated by next generation sequencing technology and abundant clinical data, a variety of genomic databases have been built up, giving us a deeper understanding of the relationship between genomic alteration and disease. Through data mining, we have identified many mRNAs, miRNAs and lncRNAs as prognostic biomarkers of patients with PAAD [6–8]. Whereas, the ability of single gene biomarkers to predict patient survival is still insufficient. Studies have shown that evaluating genetic traits involving multiple genes can improve prognosis prediction [9]. However, not all pathways have been studied to develop PAAD risk signature. Therefore, more efforts are needed to find more efficient biomarkers for PAAD.

Accordingly, in the present study, we aimed to develop a novel risk signature for predicting survival of pancreatic adenocarcinoma. We have drawn hallmark gene sets from 177 PAAD patients with whole genome expression profiles from the TCGA database. We identified 199 mRNAs were significantly associated with glycolysis using Gene Set Enrichment Analysis (GSEA) and built a three-gene risk signature that can accurately predict patient prognosis. This risk signature can successfully predict PAAD patients who are in the high-risk group with poor prognosis. Additionally, based on Cox multivariate regression analyses, we proved that this three-gene risk signature had a better prognostic value than other clinical information of patients with PAAD.

Materials And Methods

Clinical data and mRNA expression dataset

The clinical data and mRNA expression profiles of PAAD patients were extracted from the TCGA database (<https://cancergenome.nih.gov/>). Additionally, the clinical data patients age, gender, grade, TNM stage, T stage, N stage, and M stage was included (Table 1).

Functional enrichment analyses

GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) was used to explore whether the identified gene sets showed significant differences between the groups. The algorithm can present an overall trend of the raw data and does not need to specify differential gene threshold. Next, we analyzed the expression levels of mRNAs in PAAD samples and adjacent noncancerous tissues. Finally, we determined gene set for further analysis by normalized p values ($p < 0.05$) and normalized enrichment score.

Data processing and risk-parameter calculation

Univariate Cox regression analysis was used to analyze and identify genes related to overall survival (OS), then subjected to multivariable Cox regression to further confirm the prognostic genes and obtain the coefficients. The selected mRNAs were divided into the risk (hazard ratio: $HR > 1$) gene and protective

($0 < HR < 1$) gene. The expression level of selected genes weighted by their coefficients through linearly combining, we constructed a risk score formula as follows:

$$\text{Risk score} = \sum (\beta_i \times \text{Exp}_i)$$

β_i representing coefficients and Exp_i indicating the expression level of each selected mRNAs. The median risk score was used as cut-off value, the 177 patients were classified into high-risk and low-risk groups.

Statistical analysis

Kaplan–Meier survival curves and the log-rank method were used to validate the prognostic significance of the risk score. The survival function of the glycolysis based prognostic risk model was analyzed by receiver operating characteristic (ROC) curve. Moreover, we performed univariate Cox, multivariate Cox regression analysis, and data stratification analysis to test whether the risk score was independent of the clinical features, including age, grade, and stage. $P < 0.05$ was considered as statistically significant. Statistical analyses were performed using SPSS 20.0 software (SPSS, Inc., Chicago, IL, USA).

Results

Initial screening of genes using GSEA

We obtained clinical data from 177 patients with PAAD, along with an expression data for 56753 mRNAs, from the TCGA database. The hallmark gene sets from the Molecular Signatures Database (MSigDB) were selected to represent well-defined biological process or courses, which contained 50 specific gene sets. GSEA was analyzed using the above database to detect whether the gene sets showed statistically differences between PAAD and adjacent normal tissues. A mount of 36 gene sets were upregulated in PAAD. Among them, HALLMARK_GLYCOLYSIS, and HALLMARK_ESTROGEN_RESPONSE_LATE gene sets were significantly enriched (Figure1). We then selected GLYCOLYSIS ($P = 0.028$, $NES = 1.54$) for further analysis, which has the top-ranking function and contained 199 genes.

Identification of survival-associating glycolysis-related mRNAs

In order to identify novel genetic biomarkers associated with the survival of patients with PAAD, we applied univariate Cox regression of 199 genes that were enriched via glycolysis. A total of 7 genes were significantly associated with OS ($P < 0.05$) and then entered them into multivariate Cox regression analysis. Three independent genes (KIF20A, CHST2, and MET) were selected via multivariate COX regression analysis (Table 2). With $HR > 1$ associated with poorer survival (KIF20A and MET), with $HR < 1$ associated with better survival (CHST2).

Construction of a three-mRNA signature to predict patient outcomes

Based on integrating the expression level and corresponding regression coefficients derived from multivariate Cox regression analysis, a prognostic risk score model was established as follows:

Risk score = $0.1755 * \text{expression of KIF20A} + (-0.1400) * \text{expression of CHST2} + 0.0214 * \text{expression of MET}$

We then ranked the risk score in ascending order. According to the risk score formula, 177 PAAD patients were classified into high-risk groups (n=88) and low-risk groups (n=89), using the median risk score as cut off value (Figure 2A). The survival time of the patients are shown in Figure 2B. Patients with the low-risk group had lower mortality rates, whereas patients in high-risk group showed poorer survival. Additionally, a heatmap displayed the expression profiles of 3 mRNAs (Figure 2C). Compared with the low-risk group, the high-risk mRNAs (KIF20A and MET) was upregulated in high-risk group, whereas the expression of the protective type of mRNAs (CHST2) was downregulated in high-risk group. The sensitivity and specificity of the three-mRNA signature by area under the receiver operating characteristic curve (AUC) value in the ROC curve at 1-year survival was calculated. The ROC curve analysis score was 0.718 (Figure 3), indicating the good sensitivity and specificity of the three-mRNA signature in predicting survival of PAAD patients.

Risk score generated from the signature as an independent prognostic indicator

To verify whether the risk score is independent of other clinical features, we performed univariate and multivariate Cox regression analyses to evaluate the importance of these indicators in PAAD patients, which included risk score, age, gender, grade, and TNM stage as covariables (Figure 4). We found that the risk score (HR: 1.242; 95% CI: 1.157-1.334; $P < 0.001$) age (HR: 1.028; 95% CI: 1.006-1.050; $P = 0.012$), and grade (HR: 1.377; 95% CI: 1.020-1.859; $P = 0.037$) were associated with patient survival in the univariate analysis. We identified risk score and age had independent prognostic value, as these factors showed significant differences in both univariate and multivariate analyses. All these indicated that the three-gene signature had competitive prognostic value for survival prediction.

Validation of three-mRNA signature for survival prediction by Kaplan–Meier curve analysis

Compared high-risk group and low-risk group, the Kaplan-Meier analysis showed a significant difference in the survival of the patients ($P < 0.001$; Figure 5A). Additionally, the risk score was a stable prognostic marker for patients with PAAD stratified by age (< 65 or ≥ 65), TNM stage (stage I or stage II-IV), T stage (stage 1-2 or stage 3-4), and N stage (Figure 5B, 5C, and 5D). The patients in the high-risk group had significantly shorter OS than those in the low-risk group in M0 subgroup (Figure 5E). Since there were only 4 patients in M1 subgroup, it is regrettable that the analysis of M1 subgroup has not been completed. All these results indicated that the risk score was robust in predicting the prognosis of patients with PAAD.

Discussion

Currently, the study of energy metabolism has attracted people's attention. Human malignancy is inextricably linked to energy metabolism. Tumor cells have the ability to proliferate indefinitely and are prone to distant metastasis, which requires enough biosynthetic precursors and energy to promote cell division, migration, and invasion. Hypovascularization in PAAD reduce the transfer of biosynthetic

precursors to cancer cells, leading to an energy crisis [10]. However, tumor cells have “metabolic reprogramming” ability to change their energy metabolism [11]. Then, the major pathway of glucose metabolism is transformed from oxidative phosphorylation to anaerobic glycolysis to meet cell proliferation needs [12]. Biomarkers have been thought of in terms of cancer and energy metabolism. Although many studies have been conducted on PAAD and glycolysis [13–15], the research that involved in prognostic biomarkers of cancer related to glycolysis is limited. In this study, we attempted to identify a glycolysis-related prognostic biomarker for PAAD patients.

Recent studies showed that clinicopathological features such as age, sex, and metastatic diagnosis are not sufficient to precisely predict the outcome of patients with cancer [16]. The ideal biomarker is readily available, cost-effective, reliably measurable, and highly accurate. Thus, a growing number of mRNAs have been identified as biomarkers for predicting survival of PAAD. For example, Strippoli et al. report that the overexpression of ERCC1 showed to be significantly associated with shorter survival and poor disease control [17]. Giovannetti et.al reported that hENT1 is predictive factor of response to gemcitabine, the overexpression is linked to a longer overall survival in PAAD [18]. Additionally, miRNA-21 and miRNA-155 have been considered as novel biomarkers shown prognostic and predictive value [19, 20]. However, these biomarkers cannot serve as independent prognostic indicator for predicting patient prognosis. Particularly, single gene expression levels interfered by many factors, preventing these markers from providing powerful predictive effects. Therefore, we constructed a gene signature containing multiple related genes by statistical model, and improved the prediction efficiency by combining the prediction effect of each component gene. In our study, we reported a glycolysis-related risk signature (KIF20A, CHST2, and MET) and then proved the prognostic value in PAAD.

Conclusions

In conclusion, we firstly develop a glycolysis-related risk signature for predicting survival of pancreatic adenocarcinoma. The three-gene signature had competitive prognostic value for survival prediction which could help clinicians to stratify pancreatic adenocarcinoma patients and improve their prognosis. Additionally, it is of great significance to improve novel therapy targets for PAAD.

Abbreviations

PAAD: Pancreatic adenocarcinoma; PC: Pancreatic cancer; TCGA: The Cancer Genome Atlas; GSEA: Gene Set Enrichment Analysis; OS: Overall survival; CI: Confidence interval; HR: hazard ratio; ROC: Receiver operating characteristic; AUC: Area under curve; MSigDB: Molecular Signatures Database.

Declarations

Acknowledgements

None.

Author contributions

YFJ conceived and designed the study. AL collected the data and wrote the manuscript. SNH and HMH summarized the data, AL and JC revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant nos. 30972610, 81273240, 91742107, and 81570002), National Key Research and Development Program (Grant nos. 2017YFC0910000 and 2017YFD0501300), Jilin Province Science and Technology Agency (Grant nos. 20190101022JH, 2019J026, 20170622009JC, 2017C021, 2017J039, SXGJXX2017-8, JJKH20180197KJ, DBXM154-2018, and 2018SCZWSZX-015), The “521” Foundation of Lianyungang (Grant nos. LYG52105-2018082).

Availability of data and materials

The current study was based on the results of relevant published studies.

Ethics approval and consent to participate

Not applicable

Informed consent

For this type of study formal consent is not required.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics 2017 CA Cancer J Clin. 2017;67(1):7–30.
2. Rahib L, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014;74(11):2913–21.
3. Ferri MJ, et al. Improved Pancreatic Adenocarcinoma Diagnosis in Jaundiced and Non-Jaundiced Pancreatic Adenocarcinoma Patients through the Combination of Routine Clinical Markers Associated to Pancreatic Adenocarcinoma Pathophysiology. PLoS One. 2016;11(1):e0147214.
4. Kamisawa T, et al. Pancreatic cancer. Lancet. 2016;388(10039):73–85.

5. Sarantis P, et al. Pancreatic ductal adenocarcinoma: Treatment hurdles, tumor microenvironment and immunotherapy. *World J Gastrointest Oncol.* 2020;12(2):173–81.
6. Meijer LL, et al., Plasma miR-181a-5p Downregulation Predicts Response and Improved Survival After FOLFIRINOX in Pancreatic Ductal Adenocarcinoma. *Ann Surg*, 2018.
7. Ye Y, et al. High expression of AFAP1-AS1 is associated with poor survival and short-term recurrence in pancreatic ductal adenocarcinoma. *J Transl Med.* 2015;13:137.
8. Dell'Aquila E, et al. Prognostic and predictive factors in pancreatic cancer. *Oncotarget.* 2020;11(10):924–41.
9. Wang QW, et al. Prognostic Correlation of Autophagy-Related Gene Expression-Based Risk Signature in Patients with Glioblastoma. *Onco Targets Ther.* 2020;13:95–107.
10. Yang J, et al. The enhancement of glycolysis regulates pancreatic cancer metastasis. *Cell Mol Life Sci.* 2020;77(2):305–21.
11. DeBerardinis RJ, et al. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* 2008;7(1):11–20.
12. Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. *Cell.* 2008;134(5):703–7.
13. Baek G, et al. MCT4 defines a glycolytic subtype of pancreatic cancer with poor prognosis and unique metabolic dependencies. *Cell Rep.* 2014;9(6):2233–49.
14. Dovmark TH, et al. Connexin-43 channels are a pathway for discharging lactate from glycolytic pancreatic ductal adenocarcinoma cells. *Oncogene.* 2017;36(32):4538–50.
15. Kumari S, et al. MUC13 contributes to rewiring of glucose metabolism in pancreatic cancer. *Oncogenesis.* 2018;7(2):19.
16. Zhao L, et al. Identification of a novel cell cycle-related gene signature predicting survival in patients with gastric cancer. *J Cell Physiol.* 2019;234(5):6350–60.
17. Strippoli A, et al. ERCC1 expression affects outcome in metastatic pancreatic carcinoma treated with FOLFIRINOX: A single institution analysis. *Oncotarget.* 2016;7(23):35159–68.
18. Giovannetti E, et al. Transcription analysis of human equilibrative nucleoside transporter-1 predicts survival in pancreas cancer patients treated with gemcitabine. *Cancer Res.* 2006;66(7):3928–35.
19. Song WF, et al. MiR-21 upregulation induced by promoter zone histone acetylation is associated with chemoresistance to gemcitabine and enhanced malignancy of pancreatic cancer cells. *Asian Pac J Cancer Prev.* 2013;14(12):7529–36.
20. Mikamori M, et al. MicroRNA-155 Controls Exosome Synthesis and Promotes Gemcitabine Resistance in Pancreatic Ductal Adenocarcinoma. *Sci Rep.* 2017;7:42339.

Tables

Table 1 Clinical data of patients with pancreatic adenocarcinoma in this research

Variables	N	%
Age		
≤65	81	46
≥65	96	54
Gender		
Male	97	55
Female	80	45
Grade classification		
Grade I	31	18
Grade II-IV	144	82
Stage classification		
Stage I	21	12
Stage II-IV	153	88
T classification		
T1-2	31	18
T3-4	144	82
N classification		
N0	49	28
N1-3	123	72
M classification		
M0	79	95
M1	4	5

Table 2 Overall information of three mRNAs constructing the prognostic signatures

Gene	Ensemble ID	Hazard Ratio	Coefficient
KIF20A	ENSG00000112984	1.1918	0.1755
CHST2	ENSG00000175040	0.8694	-0.1400
MET	ENSG00000105976	1.0217	0.0214

Figures

Figure 1

Enrichment plots of two gene sets which had significant difference between PAAD and adjacent normal tissues by performing GSEA.

Figure 2

The three-mRNA signature predicts outcomes of PAAD patients. A. mRNA risk score distribution in PAAD patients. B. Survival time of PAAD patients in ascending order of risk score. C. A heatmap of expression profiles of 3 mRNAs.

ROC curve (AUC = 0.718)

Figure 3

The AUC for 1-year OS. OS, overall survival. AUC, area under the receiver operating characteristic curve.

Figure 4

Univariable and multivariable analyses for each clinical feature. A. Univariable analyses for each clinical feature. B. Multivariable analyses for each clinical feature.

Figure 5

Kaplan–Meier survival analysis for the patients with PAAD. A. The Kaplan–Meier curve for patients divided into high-risk and low-risk. B-E. Kaplan–Meier curves for prognostic value of risk-score signature for the patients divided by clinical feature.