Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-23-301

Reviewer A

Thanks very much for the comments, which are very helpful in improving the quality of this article. We have revised the manuscript and paid much attention to your comments and suggestions.

Major comments

Comment 1: Page 10 Line 6 "These 5 genes were all considered as risk factors for survival of patients, and their high expressions may shorten the survival time of PCA patients [all P<0.05; hazard ratio (HR), 1.002–1.058]."
Although this is narrowly statistically significant, the effect size is quite small (0.2% to 6%). Why is there such a large difference from Figures 2K-2N? Is this just the difference between the training and validation set (in which case it is quite worrisome) or something else? It would probably make more sense to show the results from the validation set and not the training set in Figures 2K-2N.
Reply 1: In this study, univariate Cox regression analysis was used to screen the genes related to the prognosis of pancreatic cancer patients, and finally, five genes were obtained. Their expression levels were respectively related to the patients' OS. In addition, Kaplan-Meier survival analysis was performed to further investigate the differences in patient survival rates between the high gene expression group and the low gene expression group. Due to the different analytical methods used, it is understandable that there are differences in the results between these two locations. Thank you very much for the reviewer's suggestion that displaying the results of the validation set at Figures 2K–2N may be more meaningful. However, unfortunately, there was no difference in survival rates between high and low gene expression groups in the validation set, so it was not shown.

Comment 2: Figure 6A Is the difference between normoxic and hypoxic for TGM2 really significant (***)? Means look nearly identical. Please double check the significance for all proteins in this panel.
Reply 2: We have rechecked the processing and analysis of the expression data of four genes ((BHLHE40, ENO1, SDC4, and TGM2) in the panel and confirmed that through a paired t-test, the difference in expression data of the four genes under normoxic and hypoxic conditions was significant. In addition, we supplemented the statistical methods used here in the manuscript.
Changes in the text: We have modified our text (see Page 14, line 13).

Comment 3: Figure 4. How correlated is the risk score with grade? It is surprising the effect of grade on survival is so low, which only makes sense if the tumor grade effect is partly subsumed into the risk score. Is the risk score predicting the grade or is it predicting survival independent of grade? Can you run a model with and without the risk score and compare?
Reply 3: In response to your question, we analyzed the relationship between grade (grade 1,

grade 2, grade 3, and grade 4) and risk score through the Wilcoxon test, and the results showed that the risk score of grade 2 and grade 3 samples was higher than that of grade 1, respectively. However, there was no significant difference in the risk score of the samples between other different grades, which may be due to the small sample size. In this study, multivariate Cox regression analysis revealed that the risk score was an independent diagnostic factor, which means it can predict the prognosis of patients independently of other clinical features such as age, gender, grade, and stage. Our research focus in this section is to investigate whether the risk score is an independent prognostic factor, so we did not run a model without the risk score.

Comment 4: Methods section in abstract should be more concise and address only the specifics needed to understand the results. PubMed abstract length is set at 250 words, it is a good idea to avoid exceeding this.

Reply 4: It is an excellent suggestion to us. We have recognized the shortcomings of the abstract section and made modifications. Thank you very much for your proposal.

Changes in the text: We have modified our text as advised (see Page 2, line 2-14).

Comment 5: Does including the genes downstream of BHLHE40 improve the predictive power of the model?

Reply 5: We believe that the question you raised is highly thought-provoking. There are two reasons why the downstream genes of BHLHE40 were not included in our model. Firstly, the model we have established is based on hypoxia-related genes, while the relationship between downstream genes of BHLHE40 and hypoxia has not been confirmed. Secondly, the downstream genes of BHLHE40 have a certain expression correlation with BHLHE40. When constructing a model through LASSO regression analysis, by adding penalty terms and removing unnecessary variables from the model, the simplicity and stability of the model are ensured. Therefore, it is not necessary to include downstream genes of BHLHE40 in the model.

Minor comments

Comment 1: There are many points in the methods section where the default settings are used, which is acceptable but it assumes a familiarity with the specific package used. Please go through this section carefully and clarify any major choices in the parameters. For packages that are less standard, briefly explain the choices made and important parameters used. The papers associated with the packages should be cited where possible (i.e. [1]).

Some examples:

a. “The HRGs were obtained from the MSigDB, which contained 200 genes that were upregulated in response to low oxygen levels.

Specify this is from the Hallmark genes data set and cite [2]:

“The HRGs were obtained from hallmark gene set collection of the MSigDB, which contains 200 genes that are commonly upregulated in human mammalian epithelial cells and a variety of cancers in response to low oxygen levels.[2]”

b. CIBERSORT is very sensitive to the reference cell set. Specify the signature matrix used.

c. Page 6 Line 11 “We calculated the area under the curve (AUC) of the receiver operating characteristic (ROC) curve to assess prognostic ability”

Please be more specific. Are you calculating the probability a patient will survive to a specific

time i.e Survival (1 year, high risk)?
d. Specify and describe the clustering metric used in 1A and 1D
Reply 1: (a) we have modified our text according to your instruction. (b) we have specified the signature matrix used. (c) The value of AUC is in the range of 0.5 to 1.0. In the case of AUC > 0.5, the closer the value of AUC to 1, the higher the predictive power of the model. When the value of AUC is greater than 0.7, the model has good predictive accuracy. Considering that the predictive accuracy of HRG-based risk score for predicting the OS is not high, with AUC values being lower than 0.7, even one of them lower than 0.5, we believe that it is better to change the focus to the prognostic role of HRG-based risk score, not the predictive accuracy, and to remove the ROC curve and related content from the manuscript. (d) We have supplemented the content of consensus clustering in the methods section of the text.
Changes in the text: (a) We have modified our text as advised (see Page 6, line 15-19). (b) We have modified our text as advised (see Page 7, line 26). (c) we have deleted methods, results, and figures that include ROC curves in the text (see Page 7, line 6-7; Page 12, line30-33; Page 13, line7-8; Page 29, line7-16;). (d) We have modified our text as advised (see Page 5, line 25-33).

Comment 2: It is unclear what the x-axis in Figure 1E is
Reply 2: We have added annotations for the x-axis in Figure 1E.

Comment 3: Is there any previous evidence for any of the 5 HRGs (BHLHE40, ENO1, PLAUR, SDC4, and TGM2) being significantly associated with OS? If so please cite.
Reply 3: According to previous studies, ENO1, SDC4, and TGM2 expression is related to the prognosis of pancreatic cancer patients. We have added these contents to the discussion and cited corresponding references.
Changes in the text: We have modified our text as advised (see Page 19, line 1-3; Page 19, line 11-13; Page 19, line 17-18;).

Comment 4: Figure 2F provide a legend
Reply 4: We have modified the legend of Figure 2F
Changes in the text: We have modified our text as advised (see Page 28, line 26).

Comment 5: Page 10 Line 11 “Finally, 4 key HRGs, including BHLHE40, ENO1, SDC4, and TGM2, were screened out.”
Should be “LASSO regression analysis established 4 key HRGs, BHLHE40, ENO1, SDC4, and TGM2, were associated with patient survival. “Screened out” suggests they were excluded from further analysis. Same thing on Page 13, Line 23
Reply 5: Thank you very much for your correction. We have made the modifications according to your proposal.
Changes in the text: We have modified our text as advised (see Page 12, line 3-6; Page 15, line 20).

Comment 6: Page 10 line 19 “It was found that patients with higher expressions of BHLHE40, SDC4, and TGM2 had lower survival rates (Figure 2K,2M,2N).”

This is a key result. How much lower?

Reply 6: The reduction in survival rate is not easy to calculate. We have revised the text to "had worse prognosis" and added a P-value.

Changes in the text: we have modified our text (see Page 12, line 14-18).

Comment 7: "A hypoxia-related prognostic model was constructed to evaluate the prognosis of each patient as follows: Risk score = (0.0018 × BHLHE40 expression) + (0.0020 × ENO1 expression) + (0.0031 × SDC4 expression) + (0.0002 × TGM2 expression)."

I assume the coefficients are from the LASSO model? Should ENO1 be used here?

Reply 7: Thank you for the reviewer's question. Although the K-M curve results showed no significant difference in prognosis between patients with high and low expression of ENO1, we found a correlation between ENO1 and patient OS through previous univariate COX regression. Therefore, we included ENO1 in the construction of the model. The successful construction of the model also demonstrates the importance of including this gene in the model.

Reviewer B

Thanks very much for the comments, which are very helpful in improving the quality of this article. We have revised the manuscript and paid much attention to your comments and suggestions.

Comment 1: Hypoxia has been found to be one of the factors contributing to chemoresistance in pancreatic cancer, but also a major driver of the formation of the tumor immunosuppressive microenvironment. What methods can be used to identify the degree of hypoxia in the tumor microenvironment? What help can the model in this study provide? Suggest adding relevant content to the discussion.

Reply 1: We fully agree with your proposal and have added relevant content to the discussion

Changes in the text: We have modified our text as advised (see Page 17, line 9-12; Page 17, line 17-22).

Comment 2: The abstract is not sufficient and needs further modification. The research background did not indicate the clinical needs of the research focus.

Reply 2: We have modified the abstract section and indicated the clinical needs of the research focus.

Changes in the text: We have modified our text as advised (see Page 1, line 31-33).

Comment 3: What is the greatest advantage of the prognostic model in this study? What is the biggest problem we are facing? It is suggested to add relevant content to the discussion.

Reply 3: Our model is accurate and reliable in predicting the prognosis of pancreatic cancer patients, which is confirmed by the retrospective data of public databases. However, more prospective data are needed to verify its clinical applicability. We had explained it in the second to last paragraph of the discussion.

Comment 4: What is the spatial distribution of hypoxia-related heterogeneity in pancreatic

cancer? Is it necessary to use the new technology of spatial transcriptomics? How to plan and discuss the next steps based on the content of this study?

Reply 4: What we know is that pancreatic cancer tissue has fewer central blood vessels, leading to microenvironment hypoxia, but there are few studies to explore the spatial distribution of hypoxia-related heterogeneity in pancreatic cancer cells. Spatial transcriptome is a new technology that may be of significance for exploring the spatial distribution of hypoxia-related heterogeneity in pancreatic cancer cells. However, our study focuses on building a hypoxia prognosis model and does not explore the heterogeneity of its spatial distribution. The next step of this study had already explained in the second-to-last paragraph of the discussion. Our next research will mainly focus on collecting more clinical samples and using prospective real-world data to confirm the clinical applicability of the model. In addition, further exploration is needed for the cellular signal transduction between BHLHE40 and downstream genes.

Comment 5: What is the role of the hypoxic microenvironment in tumor progression? It is suggested to add relevant content to the discussion.

Reply 5: It is an excellent suggestion to us. We have added relevant content to the discussion.

Changes in the text: We have modified our text as advised (see Page 17, line 3-5).

Comment 6: The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as "The contributions of hypoxia to poor outcomes in pancreatic cancer, PMID: 36523300". It is recommended to quote the article.

Reply 6: We have quoted the article.

Changes in the text: We have modified our text as advised (see Page 3, line 18).

Comment 7: What is the biological significance of the BHLHE40/TLR3 axis in the proliferation of pancreatic cancer cells and the synthesis of genetic material? It is suggested to add relevant contents.

Reply 7: Thank you very much for your proposal. According to previous studies, the main impact of hypoxia microenvironment on the biological behavior of tumor cells is to promote invasion and migration. Therefore, this study focuses on the impact of the hypoxia-related gene BHLHE40 and its downstream gene TLR3 on the invasion and migration of pancreatic cancer cells, without exploring its biological significance in cell proliferation and genetic material synthesis. Your proposal is very meaningful and may be considered for further in-depth research in the future.

Reviewer C

Thank you for your comments on our article. According to your suggestion, we have revised our previous manuscript, and the detailed point-by-point answers are listed below.

Comment 1: First of all, my major concern for this study is the poor predictive accuracy of HRG-based risk score for predicting the OS, with AUC values being lower than 0.7, even one of them lower than 0.5. So, this is a failed study of the predictive accuracy of HRG-based

prognosis prediction model. Because of this, the authors' conclusion on the prediction model is also misleading. The authors need to consider whether the current research focus is appropriate for the data and consider to include other clinical factors and biomarkers as additional predictors to improve the predictive accuracy. Otherwise, please change the focus to the prognostic role of HRG-based risk score, not the predictive accuracy.

Reply 1: Thank you for pointing this out. We have carefully considered your question and changed the focus to the prognostic role of HRG based risk score, rather than predictive accuracy. We have deleted the ROC curve and related content from the manuscript.

Changes in the text: We have modified our text as advised (see Page 7, line 6-7; Page 12, line30-33; Page 13, line7-8; Page 29, line7-16;).

Comment 2: Second, the title needs to indicate the research design of this study, i.e., a bioinformatics analysis and an experimental validation of the risk score model.

Reply 2: It is an excellent suggestion to us. We have changed the title to "a bioinformatics analysis and an experimental validation of the hypoxia-related prognostic model"

Changes in the text: We have modified our text as advised (see Page 1, line 3-4).

Comment 3: Third, the abstract needs some revisions. The background did not indicate the knowledge gap and potential clinical significance of this research focus. The methods did not report the prognosis outcomes in the databases used and how the prediction model was developed and validated. The results need to quantify the findings by reporting statistics, such as HR values, expression levels, AUC values, and accurate P values. As I commented above, the authors should revise the conclusion accordingly.

Reply 3: It is an excellent suggestion to us. We have added the knowledge gap and potential clinical significance of this research focus in the background of the abstract. We have modified the methods and reported the prognosis outcomes in the databases used and how the prediction model was developed and validated. Due to the large number of research results and the limitation of the number of abstract words, we did not add any content to the results section of the abstract.

Changes in the text: We have modified our text as advised (see Page 1, line 31-33; Page 2, line 2-29).

Comment 4: Fourth, in the introduction of the main text, the authors need to briefly review what has been known on the prognostic biomarkers in PCA and the prognosis predictive models based these biomarkers and have comments on their limitations and knowledge gap to indicate the needs for the current research focus.

Reply 4: We sincerely appreciate the valuable comments. we have reviewed the on the prognostic biomarkers in PCA, explained the limitations of these biomarkers and the significance of the model we constructed

Changes in the text: We have modified our text as advised (see Page 4, line 11-16).

Comment 5: Fifth, in the methodology of the main text, please have an overview of the research and experimental procedures of this study including the questions to be answered by these procedures. Please describe the databases used including the clinical factors and prognosis

outcomes. In statistics, please describe the details for ascertaining the independent prognostic role of the risk score. It is also necessary to report the threshold AUC values for a good predictive model. P value for statistical significance is also needed.

Reply 5: Your suggestion really means a lot to us. We have added the research procedures in the methods section of the manuscript. Detailed information about the datasets is provided in Table 1 and the clinical factors and prognosis outcomes were included.

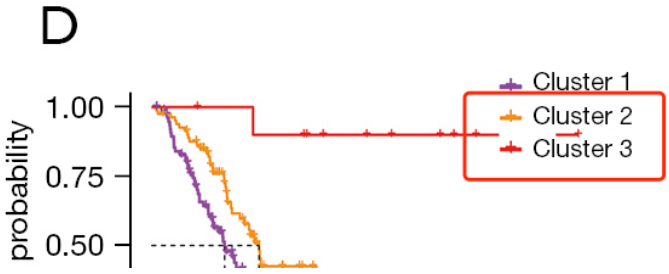
Changes in the text: We have modified our text as advised (see Page 10, line 12-32).

Reviewer D

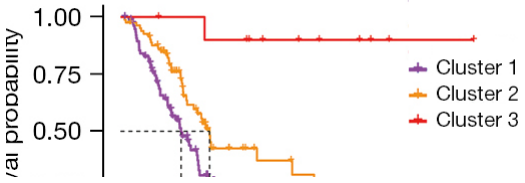
Comment 1: Figure 1

a) Please explain CDF in the legend.

b) The line of cluster 2 was covered, please adjust.



Reply: a) Thank you for your reminder. We have explained CDF in the legend. b) We have adjusted Figure 1D.



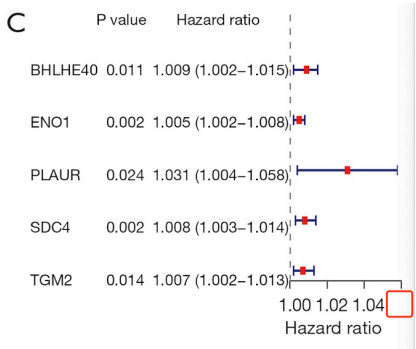
Changes in the text: a) We have modified our text as advised (see Page 29, line 3).

Comment 2: Figure 2

a) Please explain FC and FDR in the legend.

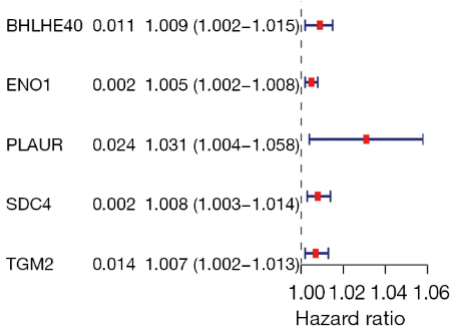
b) As there are no symbols "*, **," in the figure, please delete the explanations in the legend.

c) Please provide the scale bar of the x-axis.



Reply: a) Thank you for your reminder. We have explained FC and FDR in the legend. b) We

have deleted the explanations of "*, **," in the legend. c) We have provided the scale bar of the x-axis.



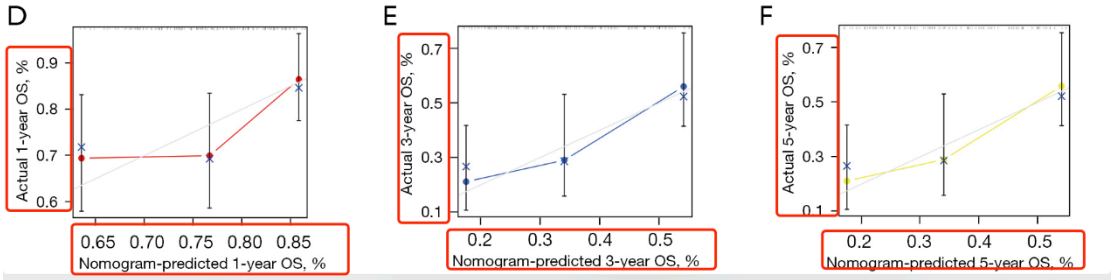
Changes in the text: a) We have modified our text as advised (see Page 29, line 16-17). b) We have deleted the explanations of "*, **," in the legend (see Page 29, line 15-16).

Comment 3: Figure 4

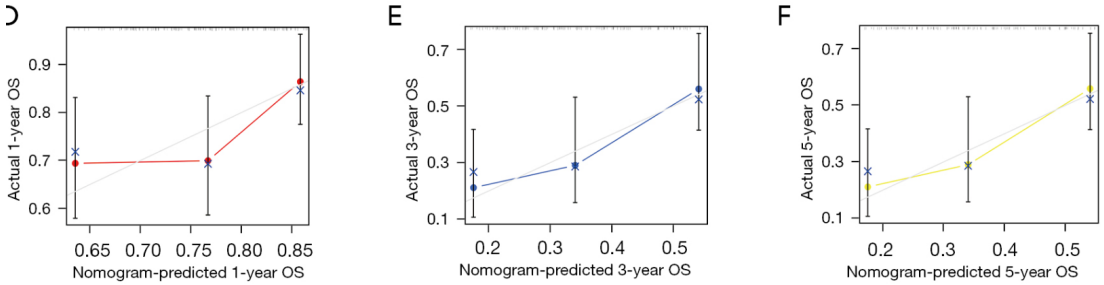
The correct format for the y-axis should be one of the following, please revise.

a) If the description is "Percent OS", the numbers should be 0-100.

b) If the description is "OS", the numbers should be 0-1.0.

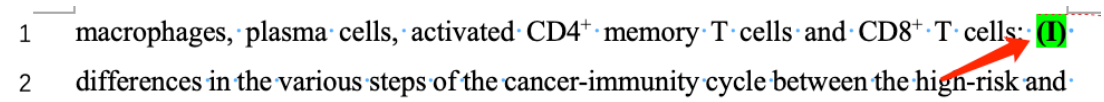


Reply: Thank you for your reminder. We have revised the description to "OS" and deleted the "%".



Comment 4: Figure 5

Please check if here should be E.



Reply: Thank you for your reminder. We have revised "I" to "E".

Changes in the text: We have modified our text as advised (see Page 30, line 17).

Comment 5: Figure 6

Since the figure 6B was obtained from the HPA dataset, please follow the policy from the HPA database (https://www.proteinatlas.org/about/licence).
Otherwise, we suggest removing the figure 6B.
Reply: We have followed the policy from the HPA database and indicated the website in the figure legends that directly links to this figure.
Changes in the text: We have modified our text as advised (see Page 30, line 26-28).

Comment 6: Figure 7
Please explain FDR in the legend.
Reply: We have explained FDR and FC in the legend.
Changes in the text: We have modified our text as advised (see Page 31, line 15-16).