Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-24-818

Reviewer A

In this article, the authors utilized TCGA and two GEO datasets to identify amino acid metabolism-related genes that are significantly associated with ESCC patients' survival. The authors identified two potentially important genes. While this manuscript presents some interesting findings, further clarifications are required.

Major Comments:

1. Please elaborate on why did you perform Cox regression analysis on AAMRGs before you performed DEGs? You identified 18 AAMRGs that can predict the survival of ESCC patients but only two of them are differently expressed between cancer and normal tissues?

Answer: Thank you for your careful comment. We have identified 629 AAMRGs from the GeneCards database, Molecular Signatures Database (MSigDB), and PubMed database. Among these 629 AAMRGs, the key genes closely related to the survival prognosis of ESCC patients, termed prognostic-related genes, were of great importance. Therefore, we conducted a single-factor Cox regression analysis to screen for these prognostic-related genes. Differentially expressed genes (DEGs) also play a crucial role in tumorigenesis. We filtered out 2 DEGs with consistent expression trends from the TCGA_GTEx-ESCC, ESCC, and GEO datasets for further analysis.

2. Please elaborate on why you performed analysis to predict RNA-miRNA and mRNA-TF? This part of the analysis does not contribute to your overall findings and you didn't address the rationale for performing these analyses in the manuscript.

Answer: Thank you for your suggestive comment. mRNA-miRNA analysis can uncover disease-related regulatory networks, aiding in the identification of potential

biomarkers and therapeutic targets. mRNA-TF analysis helps in understanding how transcription factors orchestrate gene expression and maintain cellular homeostasis. Although these analyses are not the main focus of this article, they help elucidate the function of the 18 AAMRGs. We have added the principles of mRNA-miRNA analysis in lines 143 to 146 and mRNA-TF analysis in lines 150 to 153.

3. Please provide detailed method on how you performed differential gene expression analysis.

Answer: Thank you for your suggestive comment. We have add detailed information as "Genes with logFC >0 and P.value < 0.05 are up-regulated differentially expressed genes (up regulated genes), and genes with logFC < 0 and P.value < 0.05 are down-regulated differentially expressed genes (down regulated genes) in lines163 to 166."

4. L141: you mentioned that the number of clusters was set to 2 and 8. Please elaborate on why you picked these two numbers specifically and you did not include results where the number of clusters was set to 8. In addition, do these two subtypes/clusters have biological/clinical meaning (i.e. subtypes corresponding to ESCC subtypes in the clinic)?

Answer: Thank you for your constructive comment. In this study, we employed a consensus clustering approach to determine the optimal number of clusters for ESCC disease subtypes. We selected 2 and 8 clusters for experimentation and assessed their stability and consistency through multiple experiments and analyses. The reasons for choosing these two numbers of clusters are as follows: First, the criteria for selecting the number of clusters mainly included cluster stability and Delta curve evaluation. The results showed that when the number of clusters was 2, the CDF curve exhibited the least variation, indicating the best cluster stability at this point. We found that when the number of clusters was set to 2, there was a significant difference between ESCC patient samples in the two subtypes (cluster 1 and cluster 2). Specifically, the KM survival curve showed that the high-risk tumor stages were more concentrated in cluster 1, and the prognosis for cluster 1 was worse than that for cluster 2 (as shown in Figure

6D). Second, we also analyzed the situation with 8 clusters and found its complexity to be high. No significant new clinical features were discovered when correlating with clinical data. Therefore, we ultimately chose to present the research results with 2 clusters.

These two subtypes are classified based on data characteristics and are related to specific biological and clinical significance. By comparing clinical data, we found that cluster 1 mainly included high-risk advanced tumors, while cluster 2 was primarily concentrated in early or mid-stage tumors. Additionally, the KM survival curve clearly showed that the prognosis for cluster 1 was poorer, with significantly lower survival times compared to cluster 2. Based on these findings, we concluded that these two subtypes not only have significant differentiation in data characteristics but are also closely related to clinical features, holding potential clinical application value. We have described this information in the result section in lines 288 to 292.

5. You performed survival analysis to look at 1- and 2- year progression-free survival analysis. What about longer time intervals, such as 5-year?

Answer: Thank you for your constructive suggestion. We currently primarily present the 1-year and 2-year prognosis analysis results instead of 5-year prognosis analysis as you mentioned about. This is because the data at these time points are more abundant, and the model's predictive performance is better at these two time points.

As seen in Figure 10, the 1-year calibration curve shows that the model's predictive performance is quite close to the ideal line. Although the 2-year predictive performance is slightly worse, it still has good clinical utility. Based on the good performance of the 1-year and 2-year prognosis models, we infer that the model could also be reasonably applied to longer-term survival prediction to some extent.

However, we fully agree with your viewpoint and a 5-year survival analysis is critical for clinical application which need to be supplemented in future on the premise of sufficient ESCC cases. And we have mentioned this limitation in the manuscript's conclusion section in lines 426 to 428.

6. A qPCR was performed to validate the findings. Please include patient information (e.g. age, gender, stage) as a separate table and a detailed method describing how and where these patient samples were acquired, including IRB approval number.

Answer: Thank you for your careful comment. The clinical information of ESCC patients enrolled in qPCR validation was provided in Table S2. The collection time of ESCC tissue and IRB approval number were described in lines 206 to 208.

The authors claimed that amino acid metabolism is an important cellular process in ESCC. However, the pathway analysis the authors performed failed to identify any major amino acid metabolism processes, or metabolism processes in general. Please elaborate on this disagreement between your claims and your data.

Answer: Thank you for your good comment. We first explored the relationship between genes and biological processes through differential expressed genes and GSEA enrichment analysis. In our analysis, the primary amino acid metabolic processes were indeed not directly enrich in major amino acid metabolism processes. However, these does not completely negate the importance of amino acid metabolism in ESCC.

In fact, the impact of amino acid metabolism may be more complex and deeply involved in other cellular processes or signaling pathways. GSEA analysis focuses on the overall enrichment trend of gene sets in a specific biological process rather than the significance of a single process. Therefore, other pathways may become more significant leading to the amino acid metabolism pathways less prominent in the GSEA analysis.

We also conducted univariate and multivariate Cox regression analyses and found that the expression of two amino acid metabolism-related genes (BCAT1 and MMACHC) significantly affects ESCC prognosis. Subsequently, we further analyzed the correlation between the relative expression levels of these amino acid metabolism-related genes and clinical variables (such as pathological stage and clinical T/N/M stage) to gain a deeper understanding of their role in the biological processes of ESCC. In the

future, we plan to use more specialized research methods to further clarify the specific role and impact mechanisms of amino acid metabolism in ESCC. Thank you for your valuable comment.

Minor Comments:

1. L213: please report P-values in a separate table.

Answer: Thank you for your careful comment. In L213, the PCA analysis for removing batch effects does not involve p-values. Therefore, we have not provided the p-values in our manuscript.

2. All figure legends need detailed figure titles and descriptions.

Answer: Thank you for your good suggestion. We have provided bold titles and descriptions of figures in the revised manuscript.

Reviewer B

1. Tables:

1) The P values in the main text are inconsistent with Table 4.

PI3K/AKT signaling pathway (P<0.001), pre-Notch expression and processing (P=0.004), the TGF-β signaling pathway (P=0.005), the Hippo signaling regulation pathway (P=0.03), the MAPK family signaling cascades (P=0.03), and other pathways (*Figure 7B-7G* and *Table 4*).

Reply: In Table 4 of the manuscript, two decimal places were retained, so there are some differences between the table and the main text. To avoid any misunderstandings, we have revised Table 4.

2) Table 4-6: It is suggested to revise all P=0 to "P<0.001".

	P value
	0
	0
	0

Reply: We have made Table 4-6 the necessary modifications.

3) Please define “KEGG” in Table 4 foot, “DN” in Table 5 foot, “ECM” in Table 7 foot and “ESCC” in Table 9 foot.

Reply: We have provided the full name in lines 738, 734, 747, and 758.

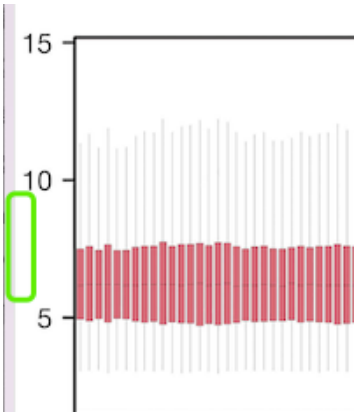
4) It should be “Table 9”.

328 multi-factor Cox regression model was constructed (*Table 8*), a nomogram analysis of

Reply: We sincerely apologize for our mistake. We have revised table 8 with table 9 in lines 338.

2 Figures:

1) Figure 2: Please add the description of y-axis in Figure 2A-B.



Reply: We sincerely apologize for our mistake. We have revised Figure 2A-B and submitted the updated version.

2) The P value in the main text is inconsistent with Figure 5E.

positively associated with the survival time of the ESCC patients (P=0.01). A high expression of *BCAT1* appeared to be associated with longer OS in the ESCC patients although this was not statistically significant (P=0.17) (*Figure 5D,5E*).

Reply: Two decimal places were retained, so there are some differences between the table and the main text. We sincerely apologize for this misunderstanding. To avoid any misunderstandings, we have revised P=0.01 with P=0.015 in lines 283.

3) Please unify “MMACHC”, “MMAHC”, “MMAHCC” in your whole manuscript and Figures 3-

6, 10B, 11B.

To explore the expression differences of the two AAMRGs (i.e., *BCAT1* and *MMACHC*) follow-up analysis. In the survival analysis, the expression of *MMAHC* was found to be positively associated with the survival time of the ESCC patients (P=0.01). A high groups of the *BCAT1* (D) and *MMAHCC* (E) genes in TCGA-ESCC data set. ns, no

Reply: Two decimal places were retained, so there are some differences between the table and the main text. We sincerely apologize for these misunderstandings. We have revised with the correct format of MMACHC in lines 278, 279, 282, and 653.

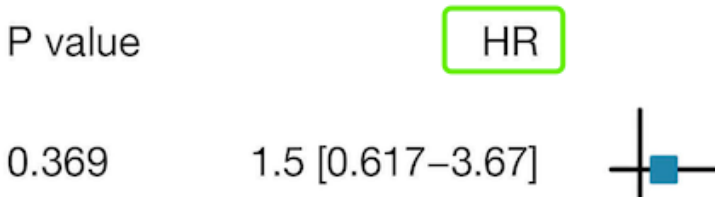
4) Please define "UV", "DN" in Figure 8 legend.

Reply: We have revised with the full name of UV and DN.

5) Figure 9: Please indicate the meaning of colorful bars (red/blue/grey).

Reply: We have added the description in lines 701-704.

6) Figure 10A: Please revise "HR" to "HR [95% CI]".



Reply: We have revised Figure 10A and submitted the updated version.

7) Figure 10B: Please add unit for age.

Reply: We have revised Figure 10B and submitted the updated version.

8) Figure 10C-D: Please remove "%" from the x/y-axis.

Reply: We have revised Figure 10C-D and submitted the updated version.