Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-23-833

Reviewer A

In their manuscript the authors addressed the predictive efficacy and prognostic performance of the clustering related to FAO metabolism using bulk RNA and scRNA sequencing analysis. They were able to isolate a Scoring System they named GFAO_Score that showed good evaluation of outcome (survival) and adjuvant therapies. This score will help determining better stratification in therapy decisions.

Minor:
The manuscript is well written and the data is clearly presented. If i may, I´d ask the authors to also discuss a bit more studies of CRC and FAO in further animal models, as for example, FAO has been shown to be essential for tumoral ISC in a multiplex fly model of CRC in vivo (Zipper et al, 2022, Metabolites) and further tumor entities.

Reply to Reviewer A's comments:

Thank you very much for your affirmation and constructive comment on our research. Animal model plays an important role in exploring the potential correlation between genomics and the development of malignancy. According to your advice, we have looked up some research related to animal models to discuss the influence of FAO on the biological progress of colorectal cancer.

Changes in the text:
We have made meticulous modifications to our text as advised and marked in blue. We have added the following content to Page 17, Paragraph 4, Lines 392-401 (section 4, Discussion), and added an abbreviation of "Intestinal Stem Cell" in Line 499

Animal models are indispensable work in the study of colorectal cancer, which helps explore the biological progress of tumor growth and subsequently develop anti-tumor drugs. The mouse colorectal cancer model has been widely used in various malignancy research in the world. Many studies have verified that the alteration of FAO affects the progress of colorectal cancer through experiments of the mouse model in vivo (32). In addition, the signaling pathways of intestinal development, regeneration, and disease between drosophila and mammals are highly conserved (33). The previous study has proved that the change of expression of FAO-related genes has a significant influence on intestinal stem cell (ISC) based on a new CRISPR-Cas9 fly model, in which ISC is the established cell of origin for colorectal cancer in the mammalian intestine (34).

32. Dai W, Xiang W, Han L, et al. PTPRO represses colorectal cancer tumorigenesis and

progression by reprogramming fatty acid metabolism. Cancer Commun (Lond) 2022;42:848-867.
33. Tafesh-Edwards G, Eleftherianos I. The role of Drosophila microbiota in gut homeostasis and immunity. Gut Microbes 2023;15:2208503.
34. Zipper L, Batchu S, Kaya NH, et al. The MicroRNA miR-277 Controls Physiology and Pathology of the Adult Drosophila Midgut by Regulating the Expression of Fatty Acid β-Oxidation-Related Genes in Intestinal Stem Cells. Metabolites 2022;12:315.

Reviewer B

In this study, Zou et al. investigate the fatty acid oxidation pathway using large publically available molecular and genomic datasets to establish a GFAO score to predict CRC patient outcomes. Three genes were used to develop the score with patients in the high scoring group displaying unfavorable survival outcomes and those in the low scoring group showing a better response to chemotherapy. Overall, this is a good study that has important implications for sub-classifying patients to personalize treatment options. Several points should be addressed to improve the manuscript.

1. The manuscript should be carefully checked for grammatical errors. There are several instances awkward phrasing including "The former research, line 49", "helps the option of, line 60", "It has (been?)", line 50, etc. Also, ATP, NADPH, and NADH should not be italicized, line 45; same for EMT, HIPPO, etc, line 391.

Replies to Reviewer B's comments:

Reply 1:

Thank you very much for your suggestions, which are valuable in improving the quality of our manuscript. It is a great pity for the grammatical errors and the improper italicized formats for the specific items. We invited one of our friends, who is a native English speaker, to go through the whole article and make the corresponding revisions.

Changes in the text:

In Line 74, we have replaced "The former research" with "The previous research". In Line 87, we have corrected "helps the options of following additional treatments" into "contributes to the following additional treatments". In Line 76, we have revised "It has illustrated" with "It has been illustrated". And we withdraw the incorrect italicized format mentioned in the comments:

Following changes:

L74: "The former research" → "The previous research"

L76: "It has " → "It has been"

L87: "helps the option of" → "contributes to the following additional treatments"

L70: "*ATP*, *NADPH*, *NADH,* and *FADH2*" → "ATP, NADPH, NADH, and FADH2"

L164: "*and*" → "and"

L655: "0.001", "0.0001" → "*0.001*", "*0.0001*"

L267: "*and*" → "and"

L669: "shrinkage, and selection" → "shrinkage and selection"

L384-385: "*HIPPO, MAPK, NOTCH, PI3K-AKT, RAS, TGFB*, and *WNT*" → "HIPPO, MAPK, NOTCH, PI3K-AKT, RAS, TGFB, and WNT"

L387: "*TGFβ1*", "*EMT*" → "TGFβ1", "EMT"

2. Pertaining to constructing the FAO score (section 2.2, line 87), more detail is needed to explain how you established the optimal cutoff value and how you combined the seven different FAO scores.

Reply 2:

Thank you very much for your valuable comments. We have added more details of the optimal value determination and combination of the FAO scores in the text according to your advice.

Changes in the text:

We have made meticulous modifications to our text, please see Pages 7, section 2.2, Lines 112-124, for the details as follows:

We determined the optimal cutpoints of seven FAO scores according to *survminer* package. Based on the optimal cutpoints, we stratified patients into high and low FAO score groups seven times and performed K-M survival analyses to explore the difference in OS each time.

Unsupervised clustering algorithm was applied to cluster analysis of FAO scores in 511 colorectal cancer samples. FAO scores consist of seven different FAO scores (GO_FAAO, GO_FABO, GO_FAOO, FAO, WP_FABO, WP_FAOO, and REACTOME_FAO scores) originated from seven gene sets related to fatty acid oxidation. We used the *ConsensusClusterPlus* package for the above steps and conducted 1000 repetitions to ensure the stability of the classification. Lastly, the patients were divided into two subtypes, Cluster 1 and Cluster 2, according to the parameter kappa 2. Taking the *P-value* less than 0.05 as statistically significant, the OS and progression-free survival (PFS) analyses of the two subtypes were subsequently performed.

3. There is some discrepancy regarding explanation of development and evaluation of GFAO score (section 2.4). On line 128, you state that only LASSO regression analysis was performed. Later in the text, you state that univariant Cox regression and LASSO regression was used (line 252). Please make the corrections to ensure consistency.

Reply 3:

Thank you very much for your meticulous reminders. We are very sorry for the discrepancy in descriptions in the development and evaluation of GFAO_Score (section 2.4). Based on your comments, we have made the corrections to ensure consistency. We used three different methods to identify the DEGs. A Venn diagram was applied to determine the intersection between C0 and C1 that met the screening parameters (log2 fold change, |logFC| > 1, and false discovery rate (FDR) < 0.05). Further, DEGs from the above-mentioned intersection were introduced into a univariate Cox regression along with survival information of patients to identify genes related to prognosis, which were identified as the hub genes. LASSO regression was subsequently performed on hub genes to avoid multicollinearity and overfitting as well as develop the final model.

Changes in the text:

We have revised the manuscript based on your comments. The details and changes are marked in blue, please see Pages 9, section 2.4, Lines 166-172:

Venn diagrams were used to determine the intersection of DEGs generated from three methods. DEGs in the above intersection were introduced into a univariate Cox regression along with survival information of patients with CRC to identify differentially expressed prognostic genes, which were defined as hub genes. Using the *glmnet* R package, least absolute shrinkage and selection operator (LASSO) regression analysis was further performed on hub genes in order to avoid multicollinearity and overfitting as well as develop the final model.

4. Results described in Figure legend 1g (comparison of K-M curves) are not depicted in the Figure.

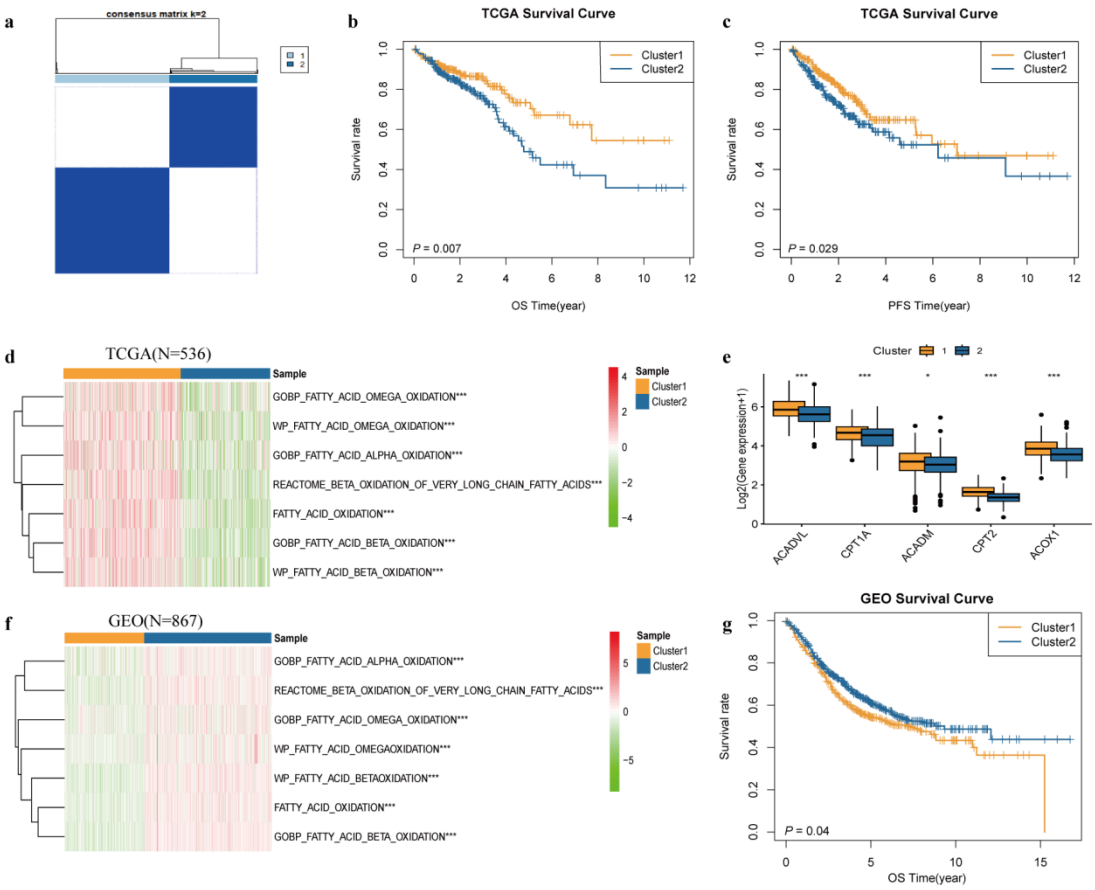
Reply 4:

We feel great thanks for your careful examination of the figures and pointing out our mistakes. It is a great pit that we left out of the picture, Figure 1g. We have made a deep reflection on this situation. We have re-uploaded the correct Figure 1g and modified the description of Figure 1 in the Result and Figure Legend parts in the text. In addition, we have gone through all of the figures and the corresponding legends in the submitted manuscripts. There was a small episode in a one-to-one correspondence between the diagrams and legends in Figure 3, whereas no such mistakes in the Result of the text. We strictly proofread and re-correspond to the legends in Figure 3, which re-corresponded Fig. 3b, Fig. 3f, and Fig. 3h, and modified these parts in the Legends.

We tried our best to improve the manuscript and made some changes marked in blue in the revised paper which will not influence the content and framework of the text. We appreciate your warm work earnestly and hope the correction will meet with approval.

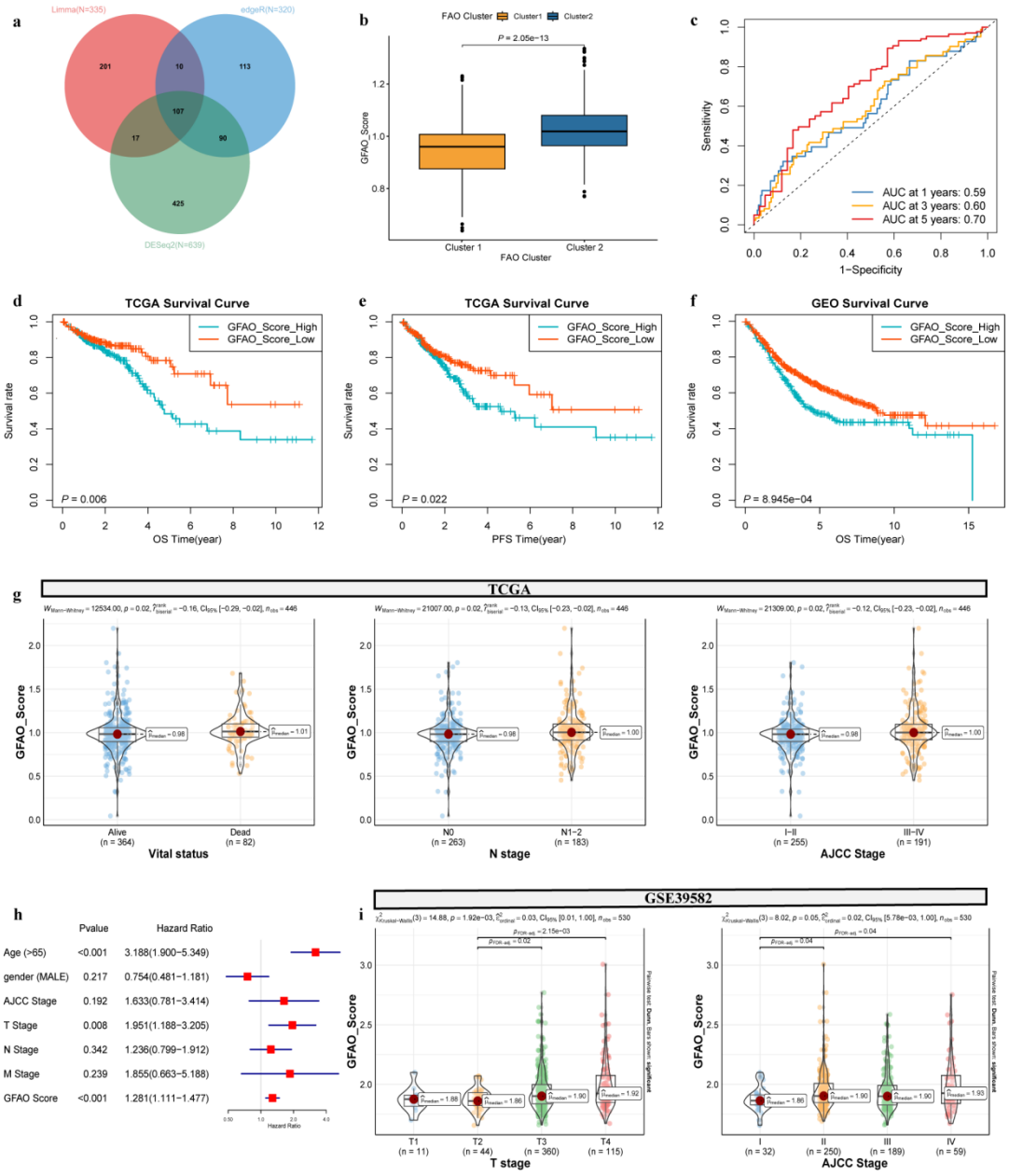
Changes in the text:

We have re-uploaded the Fig.1 (including the Fig.1g) and revised the part of the corresponding description:



-Revision of Figure 1. Please see Page 11.

We have re-uploaded the Fig.1g in Figure 1. We have added "(Fig. 1g)" in Line 244, and corrected "Fig. 1f-g" into "Fig. 1f" in Line 242. In addition, we have revised the sequencing of the description of Fig.1, please check Lines 240-244.



-Revision of the legends of Figure 3 (please see Page 27 to Page 28)

In Lines 283-284, we have added "In the GEO validation cohort, patients with high GFAO_Score also had the unfavorable prognosis (Fig 3f)." the description of Fig 3f.

In Lines 659-661, the legends of Figure 3, we have corrected "(b) Cross-validation of selection

of tuning parameters in LASSO regression analysis" into "(b) Comparison of GFAO_Score between Cluster 1 and Cluster 2. The Wilcoxon test was used to determine the statistical significance of the difference, and $P < 0.05$ was considered statistically significant".

In Lines 663-664, we corrected "(f) Forest plots of multivariate Cox regression analyses, including GFAO_Score, age, gender, TNM, and clinical stage in TCGA" into "(f) Comparison of OS between high and low GFAO_Score in the GEO validation cohort. Log-rank test *P-value* is shown".

In Lines 665-667, we have revised "(h) Comparison of OS between high and low GFAO_Score in the GEO validation cohort. Log-rank test *P-value* is shown" into "(h) Forest plots of multivariate Cox regression analyses, including GFAO_Score, age, gender, TNM, and clinical stage in TCGA".

-Additional revisions in the manuscript:

In Line 68, we have changed "The abnormal alteration of fatty acid β-oxidation (FAO)" into "The abnormal alteration of fatty acid β-oxidation (FABO)"

In Lines 95-96, we have changed "The Cancer Genome Atlas (TCGA) Pan-Cancer Clinical Data Resource and TCGA-GDC (https://cancergenome.nih.gov/)" into "TCGA-GDC (https://cancergenome.nih.gov/) and The Cancer Genome Atlas (TCGA) Pan-Cancer Clinical Data Resource".

In Line 235, we have corrected "including EMT" to "including CELL CYCLE"

In Line 240, we have corrected "the ssGSEA method" into "the heatmap".

In Line 384, we have changed "EMT" into "CELL CYCLE".

5. In the tSNE plots depicted in Fig. 2, is it possible to further refine to display particular epithelial subtypes (enterocytes vs stem cells, for example)?

Reply 5:

Great thanks for your valuable opinions which have found a very angle worthy of in-depth exploring. We have indeed neglected to conduct the analysis of subtypes of epithelial cells. Our original intention focused on the relationship between the FAO enrichment among subtypes of

epithelial cells and the common cancer pathways after re-clustering in epithelial cells. We innovatively verified that higher levels of oxidative enrichment status of fatty acids may be associated with poor prognosis based on gene set enrichment scoring at the single cell level. It is pretty meaningful and constructive that you suggested further refining to display particular epithelial subtypes, which will increase the whole quality of our research. We hope we can collect more colorectal single-cell data, accurately name epithelial subtypes, and explore different transcriptome profiles and regular expressions among subtypes in the future. At the same time, we noticed that a previous study (Joanito I et al, 2022, Nat Genet) analyzed a large single-cell data set of CRC composed of 63 patients, identified two epithelial subsets of iCMS2 and iCMS3 in CRC, and discussed the insight the relationship between various subtypes and common cancer pathways.

1. Joanito I, Wirapati P, Zhao N, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat Genet. 2022;54(7):963-975. doi:10.1038/s41588-022-01100-4

6. In addition to AKR1B10, ZFHX4 and AQP8 should also be included in the discussion.

Reply 6:

We sincerely appreciate the valuable comments. We have checked the literature carefully and added more discussion on *AKR1B10*, *ZFHX4*, and *AQP8* in the revised manuscript.

Changes in the text:

We have added the discussion about *AKR1B10*, *ZFHX4*, and *AQP8* on Page 19, Paragraph 10, section 4 (Discussion), Line 448-460:

ZFHX4 is one of the DNA repair pathway genes, and somatic mutations in this gene are associated with poor survival in esophageal squamous cell carcinoma and multiple myeloma (38-39). Moreover, *ZFHX4* encodes a zinc finger protein and is associated with the maintenance of tumor-initiating cells in glioblastoma (40). However, the role of this gene in CRC is not clear. Based on the derivation of GFAO_Score, patients with overexpression of *ZFHX4* are more likely to be stratified into the high-risk group. It will be a valuable research direction to explore the relationship between *ZFHX4* mutation and the prognosis of CRC. Besides, the aquaporins are a family of small membrane transport proteins, and *AQP8* functions as water-selective transporters (41). The dysregulation of *AQP8* has been proven to be involved in tumorigenesis. For example, *AQP8* increases viability, inhibits apoptosis, and facilitates metastasis in cervical cancer cells (42). In CRC, *AQP8* represses progression by regulating PI3K/AKT signaling and PCDH7 expression (43). However, the association between *AQP8* expression and fatty acid metabolism pathways in CRC remains unexplored.

38. Qing T, Zhu S, Suo C, et al. Somatic mutations in ZFHX4 gene are associated with poor overall survival of Chinese esophageal squamous cell carcinoma patients. Sci Rep 2017;7:4951.
39. Pawlyn C, Davies FE. Toward personalized treatment in multiple myeloma based on molecular characteristics. Blood 2019;133:660-675.
40. Chudnovsky Y, Kim D, Zheng S, et al. ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. Cell Rep 2014;6:313-24.
41. Walz T, Fujiyoshi Y, Engel A. The AQP structure and functional implications. Handb Exp Pharmacol 2009;190:31-56.
42. Li W, Song Y, Pan C, et al. Aquaporin-8 is a novel marker for progression of human cervical cancer cells. Cancer Biomark 2021;32:391-400.
43. Wu Q, Yang ZF, Wang KJ, et al. AQP8 inhibits colorectal cancer growth and metastasis by down-regulating PI3K/AKT signaling and PCDH7 expression. Am J Cancer Res 2018;8:266-279.