Peer Review File
Article information: https://dx.doi.org/10.21037/tcr-24-319

Reviewer A

The authors present a manuscript identifying a signature of genes associated with lactate metabolism that could serve as prognostic predictors in ovarian cancer following a comprehensive bioinformatics analysis. The proposal is novel and effectively contextualizes the main findings. However, this reviewer believes that several aspects require clarification.

Major Comments:

1. All analyses are based on GDC TCGA-OV data. Nevertheless, the authors do not explain the filters used to select the samples (e.g., primary site, association with metastasis, stage, or histological type). It is not advisable to include all samples indiscriminately, as tumors could be associated with differential biological events.
Response: Thanks very much for the reviewer's suggestion. The filters used to select the samples as followed: a. we select the samples in primary site; b. we select the samples at stage of Grade 3; c. we select the samples at stage of II-IV. (line 104-105)
Changes in the text: The filters used to select the samples as followed: a. primary site; b. Grade 3; c. stage of II-IV.
2. The authors do not specify the data structure used (i.e., raw counting, FPKM, FPKM-UQ, or TPM). This reviewer recommends using different data structures depending on the bioinformatics tool (e.g., for analyses in DESeq2, normalized raw counting is recommended, whereas TPM is advised for immune estimation analyses). I urgently suggest clarifying this aspect.
Response: Thanks very much for the reviewer's suggestion. TCGA data is downloaded from TCGA counts data, converted to TPM format based on the counts data, and then normalized to $\log_2$ (TPM+1). (line 105-106)
Changes in the text: TCGA data is downloaded from TCGA counts data, converted to TPM format based on the counts data, and then normalized to $\log_2$ (TPM+1).

3. Towards the end of the discussion, the authors state that the main limitation lies in the lack of in vitro and in vivo models that confirm their findings. However, given the limitations of the bioinformatics scope, it seems more concerning that the authors did not validate their results with other ovarian cancer cohorts. In the same Xena repository, cohorts are available to confirm the findings in TCGA.
Response: Thanks very much for the reviewer's suggestion. This is a limitation of the

nature analysis article, which is a characteristic of itself and not a unique feature of this article. Just like retrospective analysis, the precision of grasping this data may be flawed, which is the limitation of such research.

Changes in the text: Some limitations to this study should be mentioned. It is still necessary to confirm these findings in vivo and in vitro by examining these eight LRGs. Furthermore, future research should also consider the related molecular mechanisms.

4. In "The immune activity of two lactylation-related clusters in OC 'Immunoeconomics,'" the authors contrast the molecular signature with immune checkpoints and assume they could represent the degree of infiltration or the tumor's immunosuppressive profile. However, to reach these conclusions, it is necessary to incorporate other bioinformatics tools; for example, I advise conducting the analysis based on CIBERSORT or MIXTURE (https://github.com/elmerfer/MIXTURE/). Additionally, it would be more convenient and valuable to evaluate the Immune Checkpoint Inhibitor (ICI) score through an analysis that encompasses them, drawing more robust conclusions. This could be achieved through Gene Set Variation Analysis (GSVA) (https://bioconductor.org/packages/release/bioc/html/GSVA.html).

Response: Thanks very much for the reviewer's suggestion. We have tried many times but have not yet succeeded in mastering the method. Considering the issue of revision time, we are ready to submit the revision comments first. We are still trying.

Minor Comments:

1. Figure 1E displays the correlations of highlighted genes. However, it is not specified whether these correlations were statistically significant. Please clarify whether all of them were significant or just some.

Response: Thanks very much for the reviewer's suggestion. The absolute value of the correlation coefficient represents different correlation strengths within different ranges: below 0.3: indicates no correlation; 0.3 to 0.5: indicates low correlation; 0.5 to 0.9: indicates significant correlation; 0.9 to 1.0: indicates extremely high correlation. (line 120-123)

Changes in the text: The absolute value of the correlation coefficient represents different correlation strengths within different ranges: below 0.3: indicates no correlation; 0.3 to 0.5: indicates low correlation; 0.5 to 0.9: indicates significant correlation; 0.9 to 1.0: indicates extremely high correlation.

2. The PCA plot in Figure 2 indicates that the two principal components represent less than 50% of the data variability (which might be related to the unclear heatmap in the preceding figure). Could you provide the results for components 3 and 4? Please ensure

that at least 60% of the variance is explained.

Response: Thanks very much for the reviewer's suggestion. This is a limitation of this analysis. Similar results have also been obtained in other study [1], and we are attempting to re cluster, but the workload is relatively large and we are still in the process.

[1] Zhou Y, Wei X, Li W, Zhang S, Zhao Y. Comprehensive analysis of mitophagy-related subtypes of breast cancer and the association with immune related characteristics. Heliyon. 2023 Dec 3;9(12):e23267.

Changes in the text: A consistency cluster and principal component analysis (PCA) showed that TCGA-OC patients were well stratified into two clusters when clustering variable (k) was 2 (Figure 2A-D and Table S2). Similar results have also been obtained in other study (7).

3. In Figure 2F, the survival analysis spans an extended period of 15 years. It would be more appropriate to limit it to 5 years since, based on the median, differences in survival could be overestimated due to varying participant sizes. Additionally, I suggest reanalyzing the data by stratifying expression quartiles or randomly selecting a fixed number of subjects and comparing both groups.

Response: Thanks very much for the reviewer's suggestion. We hope to monitor patients for a longer survival time, so we have designed a time change.

Changes in the text: We hope to monitor patients for a longer survival time. Overall survival was statistically significantly longer in the low-risk group 1 than in high-risk group 2 (HR: 0.622, 95% CI: 0.460–0.841; P=0.00207; Figure 2F and Table S2).

4. Please clarify how the immune cell values estimated by EPIC were determined (means, medians?).

Response: Thanks very much for the reviewer's suggestion. The data is sourced from TIMER's website (https://cistrome.shinyapps.io/timer/), which uses deconvolution algorithms to infer the abundance of tumor infiltrating immune cells (TIICs) from gene expression profiles.

Changes in the text: The data is sourced from TIMER's website (https://cistrome.shinyapps.io/timer/), which uses deconvolution algorithms to infer the abundance of tumor infiltrating immune cells (TIICs) from gene expression profiles.

5. Please specify the basis for determining the weight coefficients for each marker in developing the risk score equation.

Response: Thanks very much for your question. Lasso regression is an extension of linear regression that achieves feature selection and complexity adjustment by

introducing an L1 penalty term. With the "glmnet" R package, Cox regression analysis was also applied to assess the prognostic value of the LRGs in the TCGA-OC cohort. Based on the minimum criteria, the λ condition was determined for variables with nonzero coefficients. The risk score was calculated using the following formula: risk score = sum (expression level of each gene × corresponding coefficient).

Changes in the text: With the "glmnet" R package, Cox regression analysis was also applied to assess the prognostic value of the LRGs in the TCGA-OC cohort. Based on the minimum criteria, the λ condition was determined for variables with nonzero coefficients. The risk score was calculated using the following formula: risk score = sum (expression level of each gene × corresponding coefficient).

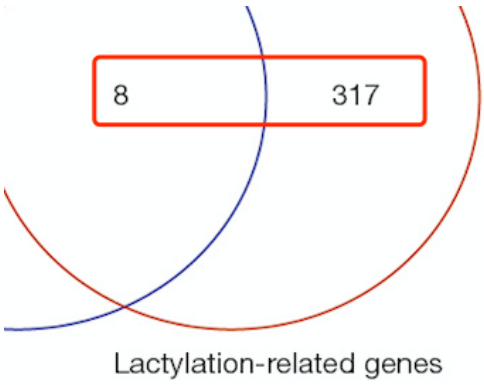
Reviewer B

1. Figure 1

Please check whether "332 lactylation-related genes" is correct, as it is "8+317" (=325) in figure 1A.

Response: Thanks very much. We have changed in figure 1A.

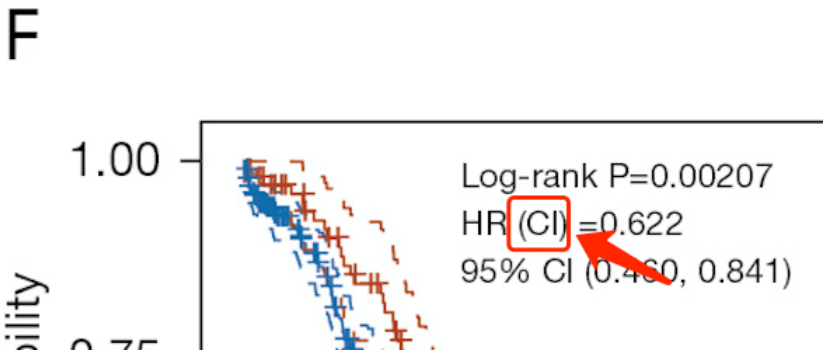
94 **Figure 1 Identification of key lactylation-related genes in OC.** (A) The prognostic
95 lactylation-related genes were screened out using a Venn diagram. Univariate Cox analysis
96 identified 707 prognostic genes, and 332 lactylation-related genes are listed in this study; (B)



2. Figure 2

This "CI" is not needed, please delete it.

Response: Thanks very much. We have deleted it.



3. Figure 3

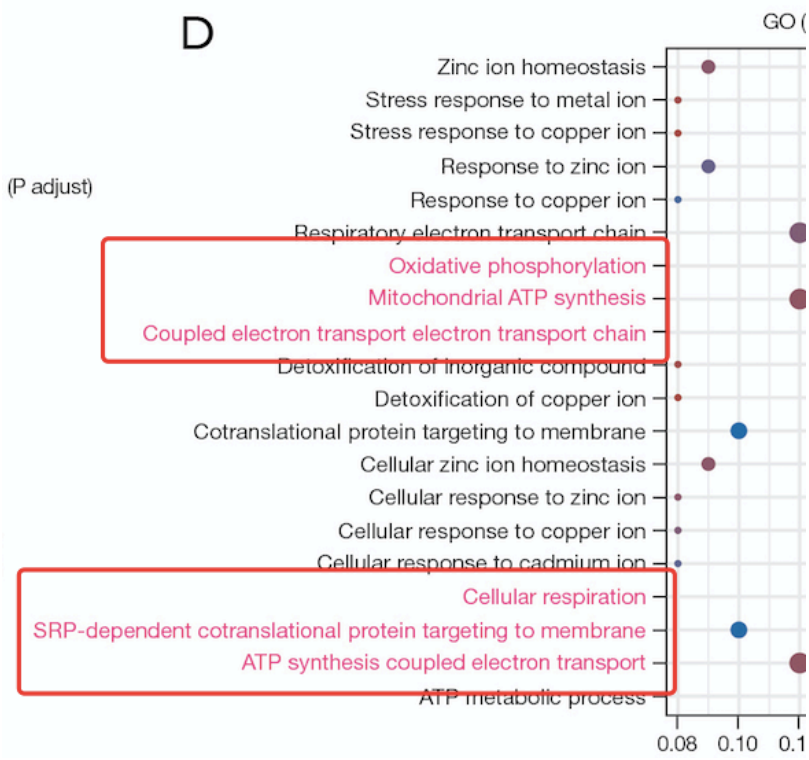
a. A summarised legend for a figure with different parts should be provided, followed by legends for each part. However, the legend of figure 3B is missing. Please provide.

410 **Figure 3 Identification of underlying mechanisms between the two OC groups.** (A)
411 Volcano plots of DEGs in the G1 OC samples compared to G2 samples. Red indicates an
412 upregulated gene, while blue indicates a downregulated gene; (C) KEGG pathways enriched
413 for upregulated and downregulated DEGs; (D) The enriched biological processes for
414 upregulated and downregulated DEGs. OC, ovarian cancer; GO, Gene Ontology; DEGs,
415 differentially expressed genes; KEGG, kyoto encyclopedia of genes and genomes; G1,

Response: Thanks very much. We have revised it.

b. Some items are in pink. Are they have special meaning? If so, please explain this in the figure legend.

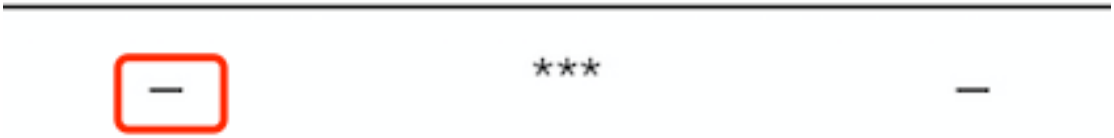
Response: Thanks very much. We have revised it.



4. Figure 4

Please explain this symbol "-" in the figure legend.

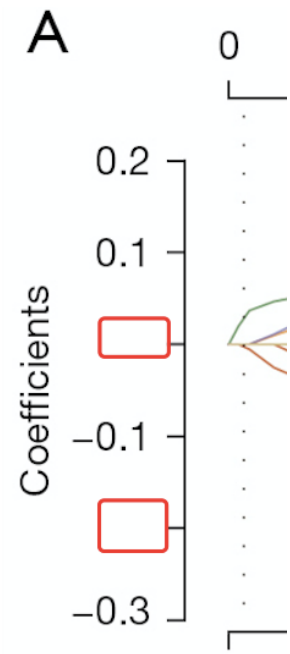
Response: Thanks very much. We have added the information.



5. Figure 6

a. Two numbers are missing. Please revise.

Response: Thanks very much. We have revised the information.



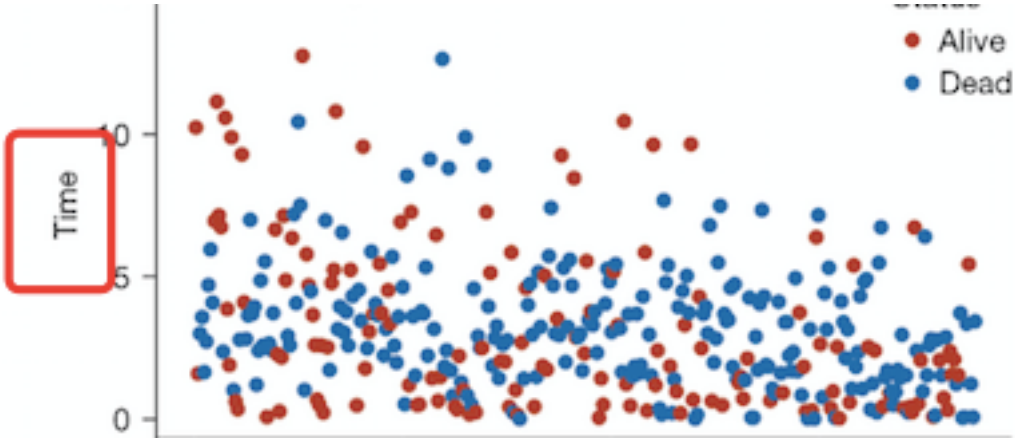
b. This term is incomplete, please revise.

Response: Thanks very much. We have revised the information.

c. Some terms are overlapped. Please revise.

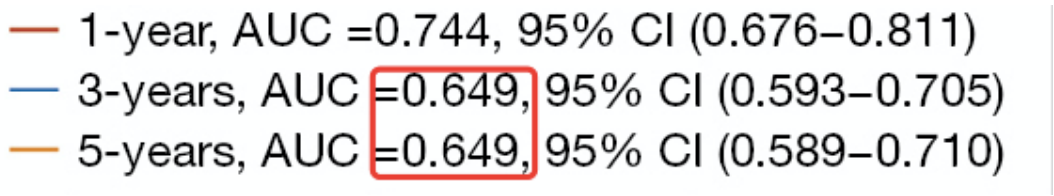
Response: Thanks very much. We have revised the information.

d. Please add the unit of time in figure 6C.
Response: Thanks very much. We have added the information.



e. Please check whether these two numbers are correct.
Response: Thanks very much. No problem.

261 characteristic (ROC) curve (AUC). The AUCs for 1 year, 3 years, and 5 years were 0.744, 0.649,
262 and 0.649, respectively (Figure 6E and Table S6), indicating that this model was a good
263 prognosticator.



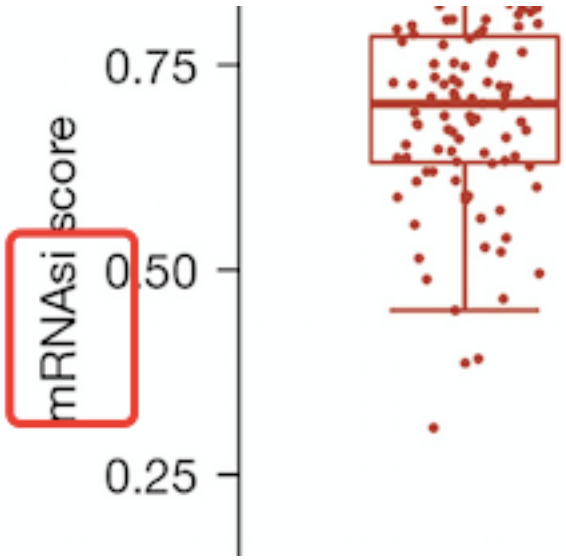
6. Figure 7

It is "CSC" in figure legend and main text, but it is "mRNAsi" in figure 7B. Please revise.

Response: Thanks very much. We have revised the information.

430 **Figure 7 Comparison of the tumor stemness between the OC groups.** (A) Tumor imr
431 dysfunction and exclusion (TIDE) scores and (B) cancer stem cell score of both OC gr
432 OC, ovarian cancer; *, P<0.05; ****, P<0.0001; G1, Group 1; G2, Group 2.

272 and G1 TCGA-OC patients in terms of CSC scores (Figure 7B and Table S7). According



7. When using **abbreviations** in table/figure or table/figure description, please mention the entire expression in a footnote below the corresponding table/figure. Please check and revise. Such as:

Figure 3: IL-17, EGFR, ECM, AGE-RAGE, ATP.

Figure 4: ICIs.

Figure 5: IC.

Here is an example:

Response: Thanks very much. We have added the information.

8. Reference

The authors mentioned "studies...", while only one reference was cited. Change "Studies" to "A study" or add more citations. Please revise. Please number references consecutively in the order in which they are first mentioned in the text.

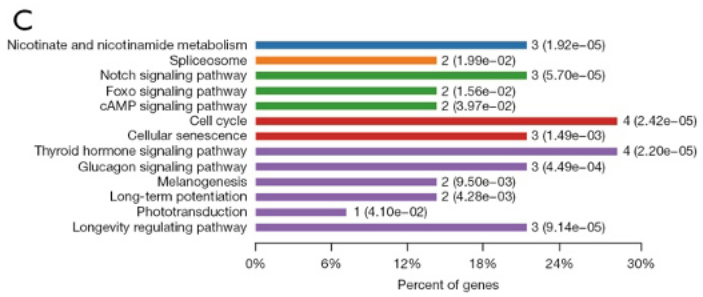
As previous studies indicated lactase and delactylases to be the most important regulatory genes, one lactase and six delactylases were included as a key LRGs……

Response: Thanks very much. We have revised the information.

9. Figure 1

Please explain the meaning of different colors.

Thank you. We have revised the Figure 1C.

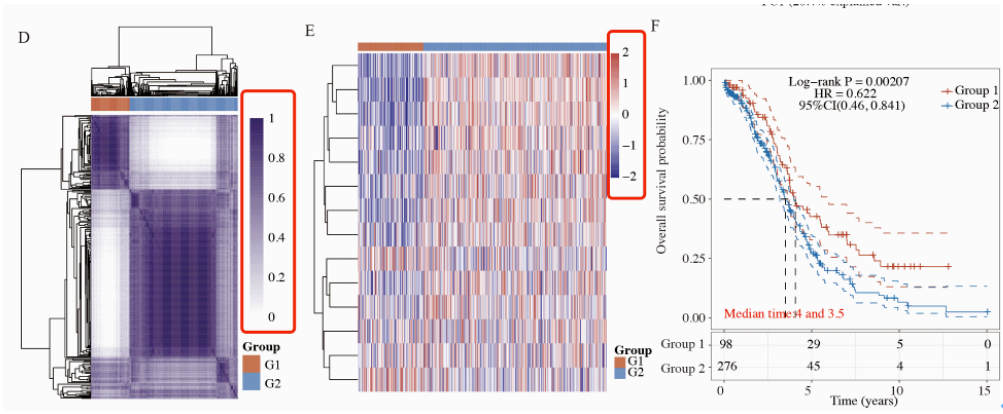


10. Figure 2D, 2E

Please explain the meaning of the pointed bar (red box).

Figure 2D: A value close to 1 indicates that the two samples are consistently assigned to the same cluster in most or all clustering processes, indicating that the clustering result is highly consistent and reliable. A value close to 0 indicates that the two samples are rarely or never assigned to the same cluster, indicating that they belong to different groups.

Figure 2E: In different subgroups, the heat maps of gene expression are shown. Red represents high expression, and blue represents low expression.



11. Figure 3B

Please explain the meaning of the pointed bar (red box).

Figure 3B: In different subgroups, the heat maps of gene expression are shown. Red represents high expression, and blue represents low expression.