Peer Review File

Reviewer A

Comment 1: This paper studied the prognostic significance of ARGs in colorectal cancer. The study used data from public databases for statistical experiments and wet lab experiements on tumor and normal tissues from patients as I understood. The authors concluded they had developed a prognostic model based on ARG signatures.

Reply 1: We would like to thank you for your careful reading, helpful comments, and constructive suggestions, which has significantly improved the presentation of our manuscript. We have carefully considered all comments from the reviewer and revised our manuscript accordingly. The manuscript has also been double-checked, and the typos and grammar errors we found have been corrected. In the following section, we summarize our responses to each comment from the reviewer. We hope that our responses have well addressed your concerns. Finally, thank you again for your advice.

Comment 2: The manuscript was difficult to read and understand and not only because most statistical analysis used were not familiar to me. Analyses used data from common databases as well as from tissues - however, in the methods section there is not information about these tissues and how they were recruited, while data collection from websites were described.

Reply 2: We are sorry for the mistakes in this manuscript and inconvenience they caused in your reading. The manuscript has been thoroughly revised and edited by a native speaker. We've revised the text description in Method 2.1 and added a table (Table 2) to make it clearer. In fact, there are two main sources of data analyzed in this paper, one from database data and another from tumor tissue stored in our laboratory. The database data includes the TCGA and GEO databases, from which we downloaded the transcriptome expression profiles of cancer patients and some clinically relevant data for analysis. The description of this part can be found in section 2.1 of the methodology. Tumor tissue samples were used to verify the reliability of the analytical data. We apologize that this part of the information is indeed missing, and for this reason we have added part 2.7 to the methods section, which will hopefully solve your problem. Changes in the text: The Methods 2.1 section of the article was revised and a Table 1 was included to provide a clearer representation of our data sources. In addition to this, we have added a new section 2.7 to provide more details on the tumor tissues that we used in our experiments.

Comment 3: The cluster analysis was not describe so I could understand how the data (and what data) was divided.

Reply 3: We're sorry for the lack of clarity in this section, so we've revised the text description in Method 2.3. Clustering is achieved by the statistical analysis method K-means clustering. We have provided several literature notes on this. In short, subtype classification based on gene expression cluster analysis can classify cancers into different subtypes based on similar expression patterns.

Changes in the text: The Methods 2.3 section of the article was rewritten and literature support was provided for the statistical data handling method.

Comment 4: Tumor specimen are mentioned under 2.4 how they were divided, it is unclear if this data were from a website or from a wet lab procedure.

Reply 4: I'm sorry for confusing you because the formulation is not clear. We've revised Section 2.4 of the Methods section to make it clear to you. In a nutshell, the tumor samples mentioned here are made from the TCGA database and the GEO database integrated together.

Changes in the text: Line 120-129, the description has been modified.

Comment 5: The overall number of analyses were many, too many, and it was difficult to follow. Perhaps could be good with a flow-chart to describe data in and various analyses?

Comment 6: As a whole I did not understand how the data was handled or the results and the manuscript could probably be better to describe this. There were so much data being used for analysis and unclear where the data came from.

Reply 5-6: Thank you for your suggestion, based on which we have inserted a flowchart throughout the article so that readers can better understand our research ideas.

Changes in the text: We have inserted a flowchart that represents the design and ideas of the article, as shown in Figure 1.

Comment 7: Most important - the relation to prognosis - where does this important piece of information come from? Own tumors? Or was this information not related to patients but through associations with various results?

Reply 7: We totally understand the reviewer's concern. We have added more details of the relation of gene expression and prognosis. On the one hand, tumor grade and TNM stage are commonly used to evaluate tumor prognosis. Second, the five-year survival rate can also be used as an indicator of cancer prognosis. In our study, we focused on the relationship between these indicators and the expression of apoptosis genes. We constructed a model that calculates the risk score of this patient by detecting the expression of ten genes in the tumor tissue, and based on the risk score we can predict the patient's 1-, 3-, and 5-year survival rates, which was validated in the test set. Survival analysis shows that our model has a good prediction effect. However, this study may still have a long way to go before it can be used in the clinic, and at this point we can only categorize high and low risk patient sets in specific datasets. In the clinic, however, each patient is individually specific, and this is where our weakness in this study lies.

The article is generally fluent. Please note abbreviations have to be defined in both the Abstract and the Main Text. Please check on the comments and suggestions below for improving the presentation.

Some Suggestions / Comments:

we apologize for the language problems in the original manuscript.

Line 45: "47" => "Forty-seven" (better to spell out number <10 for formal writing; please check across the whole paper)

Reply 1: Thank you very much for your suggestion. We've checked the full text and changed a

lot of numbers, e.g. 47 in the abstract has been changed to forty-seven.

Changes in the text: Line 45 has been modified. After checking we also modified Line 48: "10" => "ten"

Line 50: "indicated that the dissimilarities…" (simple past tense for Methods and Results presentation)

Reply 2: Thank you very much for your suggestion. Based on your suggestion, we've modified the voice issue here and examined the entire methodology and results section.

Line 60: last point, please answer or delete the point

Reply 3: Thanks for your suggestion. We have deleted the point.

Line 88: "downloaded" => "download" (noun usage)

Reply 4: Thank you very much for your suggestion. Based on your suggestion, we've modified the voice issue here and examined the entire methodology and results section.

Line 126, 127: please first define "ROC", "AUC" (abbreviations have to be defined in both the Abstract and the Main Text.)

Reply 5: Thank you very much for your suggestion. Based on your suggestion, we have checked the full text and added definitions where abbreviations appear for the first time.

Line 187: please first define "ECM" (abbreviations have to be defined in both the Abstract and the Main Text.)

Reply 6: Thank you very much for your suggestion. Based on your suggestion, we have checked the full text and added definitions where abbreviations appear for the first time.

Line 252: "has" => "had" (simple past tense for Methods and Results presentation)

Reply 7: Thank you very much for your suggestion. Based on your suggestion, we've modified the voice issue here and examined the entire methodology and results section.

Line 258: "and with a poor prognosis…"

Reply 8: Thank you very much for your suggestion. It's true that we've omitted a preposition here.

Line 258: "indicated" (simple past tense for Methods and Results presentation)
Line 278: "showed that … had" (simple past tense for Methods and Results presentation)
Line 287: "aligned…exhibited" (simple past tense for Methods and Results presentation)

Reply 9: Thank you for pointing out this problem in manuscript. Based on your suggestion, we've modified the voice issue here and examined the entire methodology and results section.

Line 308: please first define "IHC" (abbreviations have to be defined in both the Abstract and the Main Text.)

Reply 10: Thank you so much for your careful check. Based on your suggestion, we have checked the full text and added definitions where abbreviations appear for the first time.

1. Figures
    (1) Figure 1 has no subparts A and B. please check and revise the text.
        Reply: We are very sorry for our negligence of Figure 1. We rechecked and revised the text to match the figures.
    (2) Please check if Figure 2A-2C in the text should be Figure 3A-3C, and 3A-3E should be 4A-4E. Please check through **all figure citations** and make sure they are correct and matchable.
    (3) The citation of Figure 11 is missing in the main text. Please check and revise.
        Reply (2-3): We apologize for any inconvenience this may have caused. We have completed a one-to-one comparison and revision of all the charts. In the newly revised manuscript, we made sure they matched perfectly.
    (4) It is suggested to add the website link instead of its logo to each database in Figure 1.
        Reply: Thank you for your rigorous consideration. According with your advice, we have replaced all the icons with links to websites in the latest Figure 1 file available, which is indeed more intuitive and convenient.
    (5) Please add a label to indicate the meaning of different color dots in Figure 2A.
        Reply: Thank you for the above suggestions. We have added a label to the Figure 2A diagram to indicate the meaning of different color dots.
    (6) Figure 2B: Please revise the words to "P value" and "Hazard ratio (95% CI)".

        Reply: As shown in Figure 2B, we have changed the column names according to your instructions. Your suggestions are very helpful in improving the quality of our articles. Thank you very much.
    (7) Figure 3A: there is a typo.

        Reply: We apologize for the typo in figure 3A. We revised the artwork and re-uploaded it. Beyond that, we scrutinized the text in the other drawings. Anyway, thank you very much for your careful attention to detail.
    (8) Please check the legend of Figure 3A, which cannot match with its figure.
        - **Legend:** Red represents risk factors, and green dots represent favorable factors.
        - **Figure 3A:**

- Risk factors
- Favorable factors

    -    Please also indicate the meaning of purple dots.

Reply: We have made correction according to the Reviewer's comments. Red represents anoikis genes, and purple represent risk factors

(9) Please indicate the meaning of asterisks in the legend of Figure 4B, 5B and 9A.

Reply: We are very sorry for our negligence. We've redone the Figure legend for each of these graphs.

(10) Figure 5A: the word should be "unknown".
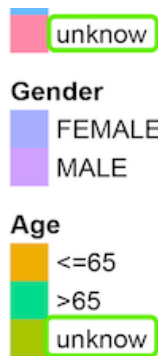
Reply: We have made correction according to the Reviewer's comments.

(11) The data are overlapping in Figure 7B, please check and revise.

Reply: Thank you so much for your careful check. We avoided overlapping data by making the fonts smaller and switching positions to stagger them.

(12) Please revise "1 years" to "1 year" in Figure 7D.

Reply: Thank you so much for your careful check.

(13) There are some overlapping in Figure 9A, please check and revise.

Reply: Thank you for pointing out this problem in manuscript. We processed Figure 9A somewhat to avoid overlapping areas as much as possible.

(14) Please add a unit to futime in Figure 9A.

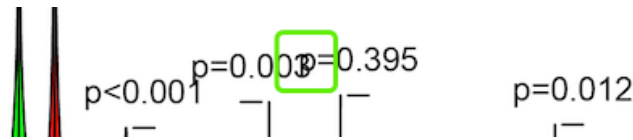Reply: Thank you so much for your careful check. We have added a unit to

(15) Please delete (%) in the axes of Figure 9B as their rates are 0-1.

Reply: Thank you for pointing out this problem in manuscript. We have deleted (%) in the axes of Figure 9B

Please add a unit to time in Figure 9C.

Reply: Thank you so much for your careful check. We have added a unit to time in Figure 9C.

(16) Figure 10C: The data are overlapping. Please modify.

p<0.001 p=0.003 p=0.395 p=0.012

Reply: As shown in Figure 10C, we modified the overlap in Fig.

(17) Plesae indicate the meaning of **** in Figure 11 legend.

Reply: Thank you so much for your careful check. We have indicated the meaning of **** in Figure 11 legend.

(18) Please provide legend for Supplementary Figure 1.

Reply: Thank you so much for your careful check. We have provided legend for Supplementary Figure 1 (Line 545).

2. Table

(1) The citation of Table 1 is missing in the main text. Please check and revise.

(2) Please add a unit to Age in Table 1.

Reply (1): Thank you so much for your careful check. We checked and revised the main text (Line 128).

(2) We have added a unit to Age in Table 1.

3. The citation of Supplementary Appendix is missing in the text. Please check and revise.

Reply: Thank you so much for your careful check. We checked and cited the Supplementary Appendix in the text (Line 161).