Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-4541

Reviewer A

Comment 1. Data analysis was performed on 159 tumour samples with 10 non-tumour samples for comparison. More non-tumour samples are needed to draw a comparison between the two groups.

Reply 1: Thank you for your suggestion. By now, the TCGA-ESCA mRNA-Seq database only included 169 samples for esophageal cancer (159 tumor + 10 non-tumor samples, after filtering out outliers), which pass stringent quality control. So, we have included all the qualified data in our research. As what pointed out by the reviewer, we acknowledge that more non-tumor samples can help us to detect finer differences between tumor and tumor-adjacent tissues, generally the more the samples included in genomic research, the better for detecting genomic changes. Although, it's possible to include some more tumor-adjacent tissues data from other projects or databases in our study, batch effects could skew the differences between tumor and tumor-adjacent tissues, even making it difficult to distinguish true differences from noise. In TCGA dataset, the number of non-tumor samples are small, this is one of limitations in our present study. We have clarified it in our study's limitations. And we will collect more esophageal cancer tumor and non-tumor samples to further research in future.

Changes in the text: We have clarified this as a limitaion of our study (see Page 13, line 250-252).

Comment 2. How were the tumour samples and non-tumour samples selected for the study ie was it a random selection or was there a bias?

Reply 2: Thank you for your suggestion. All the sample data were from TCGA database. TCGA utilizes a strict set of criteria for inclusion into the study due to the rigorous and comprehensive nature of the work being performed. For details,

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v11.p8, could see "Study Inclusion/Exclusion Criteria". Because of small number of samples, our present study inevitably has the problem of bias. We have clarified it in our study's limitations.
Changes in the text: See Page 13, line 252.

Comment 3. The 4 ARG's that were used to create the prognostic signature need to be listed in the Abstract as the prognostic signature generated is based on these 4 genes.

Reply 3: As Reviewer's suggested that we have added the information.
Changes in the text: We have added the information in our text as advised (see Page 2, line 34-35).

Comment 4. A number of epidemiological risk factors have been identified for esophageal adenocarcinoma and squamous cell carcinoma. Some of these factors include age, gender, race, obesity, reflux, smoking, alcohol consumption and diet. The population demographics of the tissue samples needs to be included as OC risk and prevalence is dependent on this.

Reply 4: It is really true as reviewer suggested. Some factors have been identified as esophageal cancer risk factors, such as age, gender, race, alcohol consumption and reflux. We have supplied a supplementary file with the demographics of the tissue sample, please see Table S3
Changes in the text: See Page 8, line 152-153.

Comment 5. The supplementary section needs further details of the analysis and the raw data

Reply 5: Thank you for your suggestion. We have uploaded the traning set as Table S1 and the testing set as Table S2 (including complete survival information and the expression of 7 ARGs, significantly associated with overall survival).
Changes in the text: See page 8, line 150-151.

Comment 6. A total of 7 ARGs were found to be significantly associated with OS ($P<0.05$) although only 4 of these genes were selected for further Cox regression analysis to construct the OS prediction model. Why were the full complement of 7 ARGs found not used to construct this model?

Reply 6: Thank you for your suggestion. In the study, univariate and multivariate cox regression analysis methods were applied. Firstly, seven ARGs were detected by univariate Cox regression analysis. Then 4 ARGs were selected from the 7 ARGs by multivariate Cox regression, including *DNAJB1, BNIP1, VAMP7* and *TBK1,* which were used to construct the OS prediction model. We have modified the expression in the text.
Changes in the text: See page8-9, line 157-160

Comment 7. The Cox analysis should also be performed using adenocarcinoma vs squamous cell carcinoma as co-variables.

Reply 7: As Reviewer's suggested we have added the tumor type (adenocarcinoma vs squamous cell carcinoma) as co-variables and re-analyze this part of data.
Changes in the text: See page 10, line 198, Figure 6B,C and Figure 8A,B.

Comment 8. Introduction, page 4, sentence on lines 58-60 needs a reference.

Reply 8: Thank you for your suggestion. We have added the reference.
Changes in the text: See page 4, line 60 and page 15,line 293-294.

Reviewer B

Comment 1. The information on this test of the predictor on the entire dataset should be removed from the paper.

Reply 1: Thank you for your suggestion. We have removed the data about the entire dataset.

Changes in the text: See page 2, line 39; page 6, line 94; page 10, line 186,187,190,192,193,201; page 11, line 224; page 13, line 253; page 19, line 373-376,379 and Figure 7 and Figure 8.

Comment 2. If they are going to look at clinical variable contribution to overall survival along with their ARG signature they need to consider tumor type, squamous cell carcinoma versus adenocarcinoma in addition to other clinical variables. This is crucial.

Reply 2: It is really true as reviewer suggested we have added the tumor type (adenocarcinoma vs squamous cell carcinoma) as co-variables and re-analyze this part of data.

Changes in the text: See page 10, line 198， Figure 6B,C and Figure 8A,B

Comment 3. The authors may want to consider using these to verify that their classifier for survival is somewhat accurate. The Aoyagi dataset of esophageal tumor samples is small but has survival data.

Reply 3: Thank you for your suggestion. Aoyagi dataset of esophageal tumor samples included twenty paired esophageal carcinoma samples, one biopsy and one surgical specimen from each patient. We get the data link from ocomine and search it in GEO database (GSE32701). However, there is no complete overall survival information can be found. In the present study, we randomly separated the TCGA-ESCA dataset into a training and a testing dataset to establish and validate the prognostic signature. We will colletct esophageal cancer tumor samples and complete follow up infomation to further validate in independent ESCA cohort in future. We have clarified this as a limitaion of the present study.

Changes in the text: See page 13, line 252-254.