Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-22-2036

Reviewer A

Colon cancer (CC) is one of the most common cancers and the second leading cause of cancer-related mortality worldwide. Therefore, it is crucial to identify new prognostic biomarkers for cancer patients.

In this study, Chai et al. showed that the extracellular matrix (ECM) related genes (CXCL13, CXCL14, SFRP5, AND THBS4) are useful genes for the discrimination of colon cancer patients into low-risk and high-risk groups.

While the authors' results may be important for developing new diagnostic tools for colon cancer patients, several issues need to be addressed.

Major point

Although examining the expression of some ECM-related genes may be useful for the discrimination of colon cancer patients, it has already been reported that the patients who have high expression of the genes related to mesenchymal cells, including ECM-related genes, have a poor prognosis (PMID: 26457759, 26457759). The authors need to discuss the relationship between the author's study and these studies. In addition, the authors should show the relationship between the high-risk group from the analysis of ECM-related genes and CMS4 patients.

Thank you for your thoughtful comments and helpful suggestions.

Recently, the consensus molecular subtypes (CMSs) groups developed by Guinney and colleagues were considered the most reliable classification system available for CC. This system divides CC into four subtypes (CMS1-CMS4) with distinguishing features. Among the four CMSs, CMS4, the mesenchymal type, shows a poor prognosis. It is characterized by the activation of several critical signaling pathways including transforming growth factor-β (TGF β) signaling, angiogenesis, and ECM remodeling pathways (1). Although this classification system has been recognized as a critical step forward in distinguishing subtypes of CC, the utilization of this approach for individual patient prognostication has been hampered as analysis of thousands of genes is required. Therefore, simpler approaches such as gene signature-based prognostic risk models are urgently needed to aid in clinical decision making. In view of the association between ECM remodeling pathway activation and the poor outcomes of CMS4 subtype in CC patients, herein, we focused our efforts on developing an ECM-based prognostic signature for patient prognosis.

In the revised manuscript, we carefully analyzed and discussed the association between the ECM-related risk score and CMSs. We divided the TCGA samples into CMS4 and non-CMS4 groups. (1) We found that the ECM-based risk scores were significantly higher in CMS4 group than non-CMS4 group (Figure 6G). (2) The ECM-related genes signature remained effective at discriminating survival after adjusting to CMS (Figure 6U and 6V). These new results, to some extent, suggest our ECM-related genes signature is not only able to reflect individual risk classification, but also identify CMS4

subtype.

Please also refer to Line 34-37, Line 90-102, and Line 317-321 in the revised manuscript.

Reviewer B

The study provides an ECM-based signature risk model which estimates individual risk classification of colon cancer. The topic is of high impact (molecular mechanisms of CC are still not understood in many details) and the approach is properly chosen (tumor microenvironmental factors are increasingly recognized as pivotal for affecting prognosis and treatment options). The methodical approach is not novel, I have seen dozens of virtually identical applications to different cancer entities (differential expression analysis, Cox survival regression), but nevertheless this approach is justified and is varied in a specific way by considering ECM genes. The analysis is sound and the results of interest for scientists interested in CC-markers.

Thank you for these thoughtful and enthusiastic comments.

I suggest a few of points for major review to possibly improve the manuscript.

1. L 149: Cibersort is not novel. The first version was published in 2015.

Thank you for the comment. We have deleted “novel” in the revised manuscript.

2. L108: It is not clear how ECM genes were selected. Please describe the criteria and provide the list of ECM genes in the supplement.

Thank you for the comment and suggestion. We have added a brief description of how ECM genes were selected and downloaded in the revised manuscript, we also provided a list in the supplement (Supplement table 1).

3. L163: Please provide a list of differentially expressed genes in the supplement.

Thank you for the suggestion. We have provided a list of differentially expressed ECM genes in the supplement (Supplemental table 2).

4. Please decipher what T, N and M stages mean.

Thank you for the suggestion. In the introduction of the revised manuscript (Line 68-69), we mentioned Tumor, Nodal Involvement, Metastasis (TNM) Stages.

5. The authors completely ignore previous subtyping schemes of CC, such as CMS (consensus molecular subtypes of CRC; CMS1-CMS4) and CRIS (CRC intrinsic subtypes). Assignments are given for TCGA samples in the literature. I strongly suggest to associate them with the risk score (eg in the survival curves in Fig.6, and/or Fig. 7, as color bar in Fig.8 A,B, or so).

Thank you for the comment and suggestions.

In the revised manuscript, we carefully analyzed the association between the ECM-related risk score and consensus molecular subtypes (CMSs). In our analysis, we excluded samples with unknown AJCC stage or less than 1 month of survival time, so we had 366 samples in total. Interestingly, we found that only 333 samples were assigned with CMS in the CRC Subtyping Consortium (CRCSC) literature (1). We then divided them into CMS4 and non-CMS4 groups. (1) We found that the risk

scores were significantly higher in CMS4 group than non-CMS4 group (Figure 6G). (2) The ECM genes signature remained effective at discriminating survival after adjusting to CMSs (Figure 6U and V). (3) We also performed Cox regression analysis on the 333 samples with CMSs subtype information. We found that our ECM signature still showed significant association with overall survival. However, CMSs were not significantly associated with overall survival, which might be due to the small number of CMS4 samples (Author response figure 1). (4) Since there were 33 samples with unknown CMSs label, we couldn't provide a new heatmap such as Figure 4, 8A, and 8B. However, these news results suggest our ECM-related genes signature is not only able to reflect individual risk classification, but also identify CMS4 subtype to some extent.

Please also see our response to reviewer A.

6. Fig. 2B: gene names are too small; provide as table in the supplement.

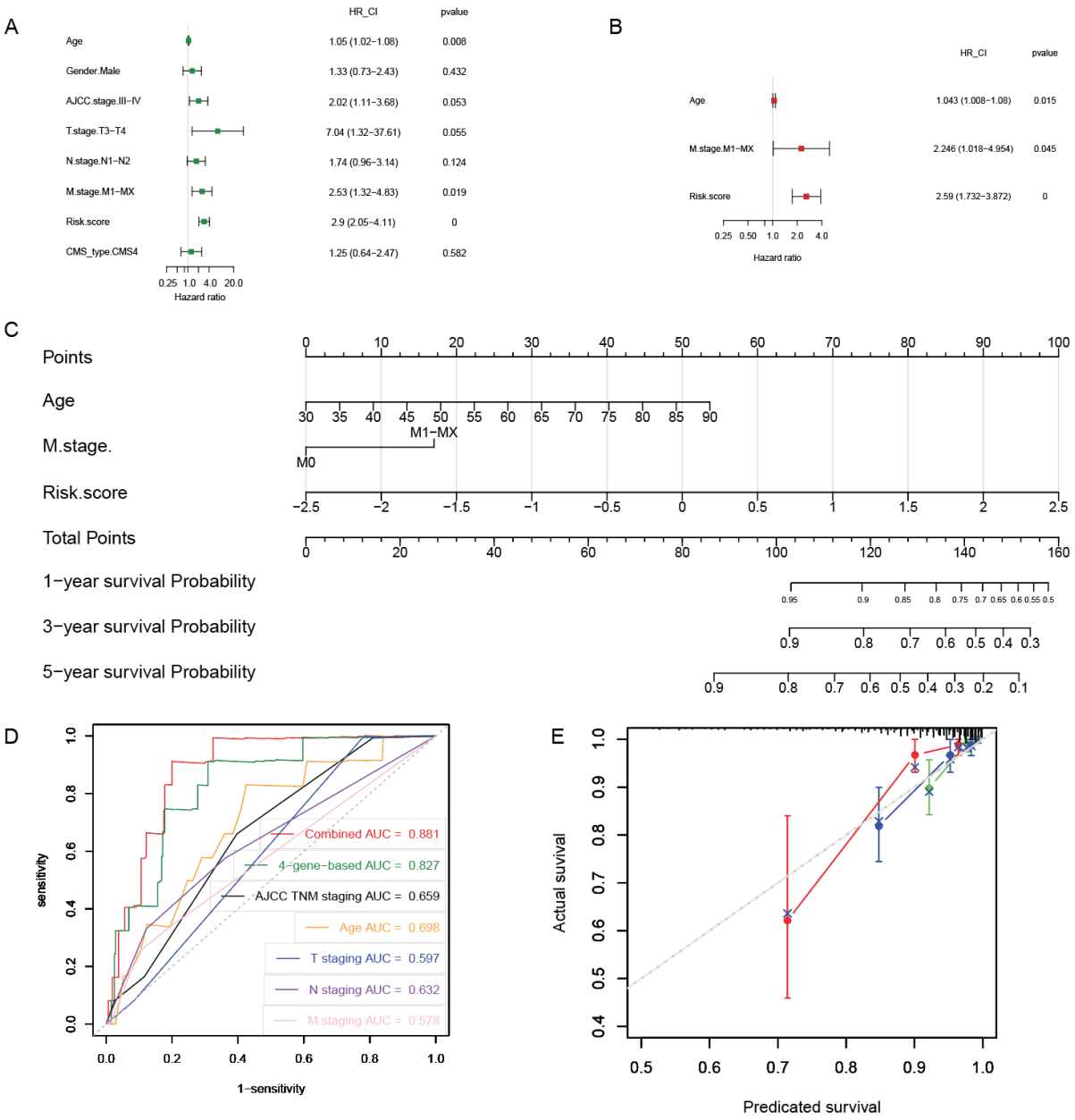
Thank you for the comment and suggestion. We provided a list of differentially expressed ECM genes in the supplement (Supplemental table 2).

7. Fig. 2C-F: Might you annotate the gene signature categories (eg BP, MF, CC etc…) also in the figure.

Thank you for the suggestion. In the revised manuscript Figure 2, we added the categories.

8. Fig 3 B, C show technical parameters. I suggest shifting these parts into the supplement.

Thank you for the suggestions. We have shifted these parts into the supplement (Supplemental figure 1) and added the technical parameters in the figure.



Author response figure 1. Construction and verification of nomogram using 333 samples with CMSs information. (A) Univariate Cox regression analysis in training cohort. (B) Multivariate Cox regression analysis in training cohort. (C) The prognostic nomogram constructed based on the risk score of ECM-based signature and clinical factors predicted the overall survival rate of COAD patients at 1, 3, and 5 years. (D) Time-dependent ROC curves for the prognostic performance of the nomogram in COAD cohort. (E) Time-dependent calibration curves show the concordance between predicted and observed 1-, 3-, and 5-year survival rates. ECM, extracellular matrix; COAD, Colon Adenocarcinoma; ROC, receiver-operating characteristic.