Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-2158

Reviewer 1#

Comment 1: The manuscript is not well written, and the methodology is not clear. I don't know very well why it comes from those genes. The results are neither clear nor consistent. The authors should rewrite the paper making it clearer, more concise and with more precise data.

Response:

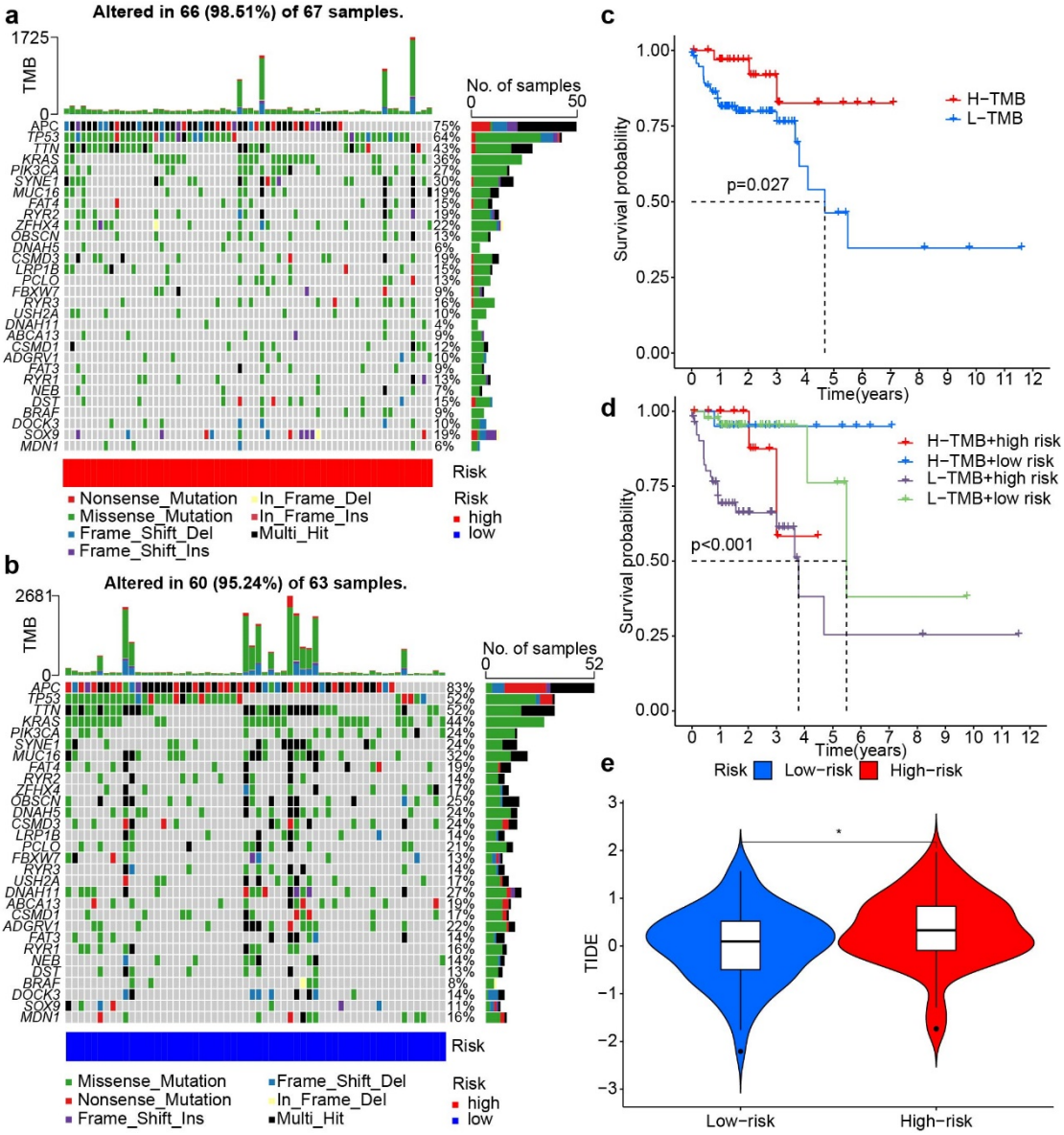
We thank the reviewer for this insightful comment. The methodology and result sections of the original draft did not make the details clear. Considering the Reviewer's suggestion, we have strengthened these sections. We tried our best to make some changes to improve the manuscript using concise and precise data. These changes will not influence the content and framework of the paper. However, we do invite a friend of ours who is a native English speaker from the USA to help polish our article. We hope the revised manuscript could be acceptable to you. And here we did not list the changes but marked them in yellow in the revised paper. We appreciate your warm work earnestly and hope that the correction will meet with approval.

Here, we would like to briefly summarize the entire article so that the reviewer can understand it better. The innovativeness of this paper is that we first constructed a ZNF protein gene signature to predict the prognosis and guide the treatment of COAD patients. There are three parts in our article: development and validation of the ZNF protein genes signature in COAD patients, functional exploration behind the signature, and clinical associations with the signature. In the first part, we extracted the prognosis-related ZNF protein genes in the TCGA-COAD profile by the univariate Cox analysis. Then we used Lasso analysis and Multivariate Cox analysis to develop a ZNF protein gene signature in the training cohort and verified this signature in the entire TCGA-COAD cohort and GSE17537. In the second part, we employed GO analysis, KEGG analysis, GSEA analysis, and ssGSEA analysis to explore the differences between the high- and low-risk groups in the functional enrichment, TIME, immune cells, and immune functions among the TCGA-COAD cohort. Additionally, on the one hand, IHC was performed to validate the expression of these model genes. On the other hand, mutation data analysis, TIDE, and drug sensitivity analysis were used to compare the distinctions between the high- and low-risk groups. In the last part, the univariate Cox analysis and the Multivariate Cox analysis were performed to explore whether the risk score is an independent prognosis factor compared with other clinical factors. Then we constructed a nomogram consisting of the risk scores and other clinical factors to guide the clinical therapies.

Comment 2: Currently, the treatment of colorectal cancer is dependent on the mutational status of the RAS, BRAF and PIK3CA genes. In this regard, it would be very interesting to classify the expression based on the mutational status of these genes.

Response: We think this is an excellent suggestion. In this article, we employed mutation analysis to explore the differences in mutational status between the two risk groups. The results were that the mutational frequency of KRAS in the low-risk group was more than that in the high-risk group (44% VS 36%), and the mutational frequency of BRAF and PIK3CA was more in the high-risk group than that in the low-risk group (9% VS 8%, 27% VS 24%) (Figure 9a and 9b). These results indicated that COAD patients with high-risk scores, BRAF mutation, and PIK3CA mutation were more likely to have a worse prognosis than COAD patients with

low-risk scores and KRAS mutation. However, these discrepancies between the high- and low-risk groups were not too evident and the potential mechanism needs to be studied deeply. These questions really could be the key points in future research and we have added this part to the limitation of this article.



Reviewer 2#

Comment 1: What is the advantage of the current gene-expression based prognostic model compared to other gene-expression based models in CRC such as consensus molecular subtyping (CMS, https://doi.org/10.1038/nm.3967), CINSARC (doi: 10.1038/s41379-022-01166-9) or a 15-gene expression signature (https://doi.org/10.1158/2767-9764.CRC-22-0489). The authors should comment on this.

Response: Thanks for your consideration for the advantages of this article. After we read these articles, it can be summarized that all these models show great value for colon cancer patients. However, there is also a set of discrepancies among the three models. The CMS model had a great integration of mutation, copy number, methylation, microRNA, and proteomics factors. This model could translate molecular subtypes to the clinic comprehensively. However, the CMS model requires sufficient tumor issues and a large cost for the technology of genome-wide expression analysis, which is so difficult to employ in practice. Nowadays, how to identify the standardization of the CMS classification is also a serious problem. Additionally, the CINSARC signature was a robust tool for stage II-III CRC patients. This signature was able to reduce the risk of the under- or treatment of stage II-III CRC patients, but there was no evident survival difference in the non-treated group. Moreover, the CINSARC signature containing 67 genes could be used in clinical practice at a high cost. Lastly, a 15-gene expression signature had the high ability to stratify stage II CRC patients to guide adjuvant chemotherapy. Despite the ability, this signature combined pT would perform better accuracy of relapse prediction for patients. Compared with these gene signatures, our gene model, consisting of 7 ZNF protein genes, can easily be put into clinical practice and is more economical for patients. Furthermore, it is also an independent factor for predicting the prognosis of COAD patients.

Comment 2: Even though gene-expression profiling is nowadays more easily available, clinical implementation is lacking. H&E histopathology is still the cornerstone of CRC diagnosis. Are the authors capable of assessing known histopathologic features with regards to their score (e.g., comparing the top and bottom expressers in TCGA-CRC based on HE morphology, whole slide images are here available, see https://portal.gdc.cancer.gov/, with regards to differentiation and other markers like tumor budding (and also see the further references: https://doi.org/10.1038/s41416-020-01222-8,https://doi.org/10.1038/s41417-023-00695-y).

Response: Regarding the assessment of the correlation of morphological features of H&E with genetic models, we must candidly state that we were not able to implement this analysis in the current study due to the lack of dedicated histopathology expertise in our current research team. We recognize this as a limitation of our study and thank you for pointing out this important aspect. We will address this issue in the future through skills training or seeking collaborations.

Comment 3： Taking morphologic features into account, should definitely be discussed as some histopathologic biomarkers are also associated with a different immune response, e.g., reduced NK cells, such as the zinc finger risk score proposed in this study (10.3390/cancers15030994). This is quite an intriguing finding.

Response: Thank you for your interest in the relationship between morphologic features and ZNF protein gene signature. We have read this paper. As the comment described, the author of this paper reported that colorectal cancer patients with stroma areactive invasion front areas (SARIFA), an independent prognostic predictor for a poorer outcome, have a great reduction of NK cells in the peripheral blood (PB). Similarly, the score of NK cells in the high-risk COAD patients was significantly lower than that in the low-risk COAD patients. What an intriguing finding! In our paper, immune cells such as CD8+ T cells, NK cells, macrophages and so on owned higher scores in the high-risk subset than it in the low-risk subset, which indicated that low-risk patients had better immune environments. So we employed TIDE analysis and found that patients in the low-risk group were less likely to experience immune dysfunction and rejection experience. This comment about the possible relationship between the reduction of NK calls and ZNF protein gene signature has great

implications for our future research. Despite this, owing to the limitation of funding and time, we could not experiment to verify if cell content changes in peripheral blood are consistent with our results. We will explore in depth the specific roles of these immune cells and their impact on disease progression, especially taking morphological changes into account to more fully understand their clinical significance in the future research.

Comment 4: As histopathologic markers are used in routine clinical practice, these findings could have immediate impact.

Response: We gratefully appreciate for your valuable suggestion. Microsatellite instability attaches great importance to the treatment of CRC. The differently expressed genes between MSI-H and MSS of GO enrichment analysis included ZNF813, ZNF426, ZNF611, ZNF320 and ZNF573, despite the limited number of DEGs of MSI-L and MSS, MSI-H, and MSI-L(doi: 10.3892/mmr.2019.9849). The result indicated that ZNF protein genes are possible to be associated with MSI status. We will explore the potential combination of ZNF protein genes and MSI status in the future and have added it to our limitation.

Comment 5: Why is only colon cancer investigated? Most classification systems (like WHO) and most treatment guidelines always look at colo-rectal carcinoma (CRC) – even though, of course, there are quite some differences? Are the results translatable to rectal cancer?

Response: Thank you for your rigorous consideration. It is really true as the Reviewer suggested that most classification systems and treatment guidelines always look at colorectal carcinoma (CRC). First of all, colon cancer and rectal cancer essentially belong to two different anatomic divergences, although there are many similarities. Additionally, the two types of cancer also differ in their embryological origin, drug targets, and metastatic patterns (Rectal and colon cancer: Not just a different anatomic site. Cancer Treat Rev. 2015 Sep;41(8):671-9. doi: 10.1016/j.ctrv.2015.06.00). At the end, we applied this model in rectal cancer and it was unable to be suitable for predicting the prognosis and guiding the treatment of rectal cancer. If possible, we would like to research such a new model which can be applied to rectal cancer likewise.

Comment 6： Publicly available datasets and resources need to be cited correctly, like: Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.; Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22:bbab260. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. Cell. 2016;166:740–54.

Response: Thank you for pointing out this problem in the manuscript. We are very sorry for our negligence of citations. And now we have added the correct citations to the manuscript following the corresponding text (Page 7, lines 288-289; Page 11, line 381).

Comment 7: Some of the immunohistochemical stains show a strong background reactivity. Are the images representative? If yes, dilution etc. need to be adapted. Also more images of the immunohistochemical stains could be provided in the supplement.

Response: Thank you very much for pointing out this important issue. We agree with your opinion. These images are representative. Background staining may be due to prolonged DAB staining, a high secondary antibody or primary antibody concentration, etc. (DOI:

10.1002/0471142735.im2104s103). These problems could be solved by applying diluted polymer-conjugated secondary, bellowing uses biotinylation of the primary antibody before using the reagent or setting up a positive control group. We observed the slides and computed scores of every tissue at that time, and some representative slides were listed in the article by the data information technology due to the limited funding. Considering the Reviewer's suggestion, we have checked these slides again. Unfortunately, because the vitro experiment was conducted 8 months ago, some slides of COAD and normal tissues are more or less discolored, which is not as typical as before. We sincerely guarantee the authenticity of this experiment. Due to the limited time and funding, we could not supplement experimental validation again. Thanks for your understanding! The IHC staining score for each sample was equal to the product of the Positive Staining Cell Score and the Staining Intensity Score, where Staining intensity ranged from 0-3 and positive Staining Cell Score: 1 portrays 0-25%; 2 portrays 26-50%; 3 portrays 51-75%; and 4 portrays 76-100%. Staining intensity ranges from 0-3. As a result, background reactivity didn't exert influence on the IHC scores.

Comment 8: What about microsatellite/mismatch repair status? As this is of outmost importance in colon cancer, the authors should comment on this.

Response: It is really true as the Reviewer suggested that microsatellite or mismatch repair status attaches much importance to the treatment of colon cancer. However, the purpose of this article is to construct a new signature consisting of ZNF protein genes to better predict the prognosis and guide the clinical treatment of COAD patients. Whether the signature can combine microsatellite or mismatch repair status to reach a better role of guidance or the signature is associated with microsatellite or mismatch repair status may be an intriguing research direction. We have added it to the limitation of this article.

Comment 9: Are the comparisons for the 16 immune cells and 14 immune-related pathways corrected for multiple testing? If not, this should be done, or at least stated in the methods.

Response: We are so sorry for describing this part unclearly. Multiple tests have been performed during the comparisons for the 16 immune cells and 14 immune-related pathways. And we have explained it in the method.

Comment 10: Lines 244-246. When talking about KM curves, p-values of log rank tests and/or HR with 95% CI should be stated. Also e.g. median months of survival (is it overall survival? Cancer-specific survival? …) should be mentioned.

Response: We are very sorry for our negligence of these details. The statement of KM curves has been described in detail. The primitive format is "The Kaplan-Meier survival analysis revealed that patients in low-riskiness subset had a significantly greater survival probability than patients in high-riskiness subset (Figure 3c)", and the current format is "The Kaplan-Meier survival analysis of OS revealed that patients in the low-risk subset had a significantly greater survival probability than patients in the high-risk subset (Figure 3c, $p<0.001$)". Moreover, we added the definition of OS to the methodology.

Comment 11: Lines 268-273. Why secretion? Expression?

Response: We feel sorry for the inconvenience brought to the reviewer. "Expression" could be more appropriate.

Comment 12: Which were the settings for oncoPredict, e.g. batchCorrect or remove LowVaryingGenes?

Response: In our article, the oncoPredict was employed to perform drug response prediction.

Comment 13: Spelling/Grammar/Style: Professional Editing is highly recommended! • Introduction need to be introduced within the text at first mentioning them, like colon adenocarcinoma (COAD)…

• Reconsider alternative title such as "Gene Expression of Zinc Finger Proteins Enables Patient Stratification in Colon Cancer"

• L. 31 "GSE17537)"

• L. 33 "opposite risk-status"

• High-risk vs low-risk patients in whole manuscript may be better

• L. 47-49: rephrase

• Key findings: without "to what we know", also the definition of high-risk is that patients survive worse, so second sentence is redundant, last bullet point: better "to tailor treatment decisions"

• L. 76 "The five-year survival rate of colon cancer patients"

Response: We feel sorry for our carelessness. Corrections above all have been made. We do invite a friend of ours who is a native English speaker from the USA to help polish our article and make the corrections marked in yellow. We hope the revised manuscript will be acceptable to you.