

LETTER TO EDITOR

Analysis of tumor microenvironment-related key mRNAs and construction of a prognosis signature in colon cancer

Dear Editor,

Tumor microenvironment (TME) serves as a crucial factor during the process of colon cancer, which can provide valuable information for prognosis of colon cancer patients.¹ We used bioinformatics analysis flow (Figure S1) to integrated analyze the prognostic TME-related mRNAs and construct a prognostic model to predict the overall survival in colon cancer patients.

In total, 2831 mRNAs were identified as prognostic mRNAs and were combined with the results of ESTIMATE analysis for further analysis via WGCNA (weight gene co-expression network analysis).^{2,3} The process of WGCNA was shown in Figures S2A, S2B, and 1A(a), eight gene modules were identified (Figure 1A[b]), cells that were correlated with any one of the three scores were screened out (the criterion was correlation value >0.4 and P -value $< .05$), and the mRNAs involved in these cells were identified as TME-related key mRNAs (the criterion was corresponding |Gene Significance (GS)| > 0.2 ; |Module Membership (MM)| > 0.7 ; P -value of GS and MM $< .05$). Finally, 317 mRNAs were identified as TME-related key mRNAs. GO and KEGG analysis indicated that TME-related key mRNAs mainly enriched in the cell cycle (Figures S3A and S3B). It revealed that cell cycle also plays a vital role in regulating the TME of colon cancer. The PPI network was used to identify the 20 hub mRNAs and to assess the interactions (Figures S2C and 2CD): the cluster that owned the highest score contained 833 interactions but only 43 mRNAs, and the number of interactions was much larger compared to the number of mRNAs involved in the PPI network indicating that the mechanism of TME in colon cancer was complex.

In the internal training set, least absolute shrinkage and selection operator (LASSO) was performed based on the TME-related key mRNAs.⁴ We performed cross-validation, and the $\log(\lambda)$ that owned the lowest deviance was selected to screen the mRNAs with the coefficients that were not 0 (Figure S4). In total, 21 mRNAs were identified and fur-

ther included to construct the prognostic signature to predict overall survival of colon cancer patients via multivariate Cox regression analysis. The signature was as follows: risk score = (exp level of APOL3 * -0.473) + (exp level of FIGNL1 * -0.188) + (exp level of HLX * 0.496) + (expression level of MCM6 * -0.251) + (exp level of RNF114 * -0.645) + (exp level of TIMM13 * -0.692) + (exp level of TNFRSF9 * -0.491). The efficacy of the signature was assessed and validated to be great in the internal training set, the internal validation set, and the external validation (Figure 1B). In the entire set, the area under the curve (AUC) at 1, 3, and 5 years was 0.745, 0.673, and 0.651, respectively. The AUC at 1 year ranged from 0.729 to 0.766 in different sets indicating the great efficiency at 1 year, though decreased to about 0.65 at 3 and 5 year, and the signature still remained good stability and efficiency. Univariate and multivariate analyses proved the risk score derived from the signature to be an independent prognostic factor (Figure 1C). The entire set was classified into different subgroups based on several clinical features; the AUC of different subgroups at 1, 3, and 5 years did not alter much indicating that our signature was stable in different situations (Figure 2). We further checked if our signature and mRNA involved in the signature could have a great correlation with the TME, which was estimated via TME scores and infiltrating immune cells abundance (assessed by ESTIMATE and CIBERSORT algorithm). The results were shown in Figure S5 indicating that our signature and mRNA involved in the signature had a great correlation with TME.

In conclusion, we comprehensively analyzed mRNA that might involve in mediating of TME in colon cancer, we identified TME-related key mRNAs, and mining the potential interactions and the functional enrichment of these mRNAs, moreover, we constructed a robust and stable signature in order to predict the overall survival of colon cancer patients. Our analysis could help to strengthen the knowledge of specific mRNAs in the regulation of TME in colon cancer and the signature could further assist to

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics

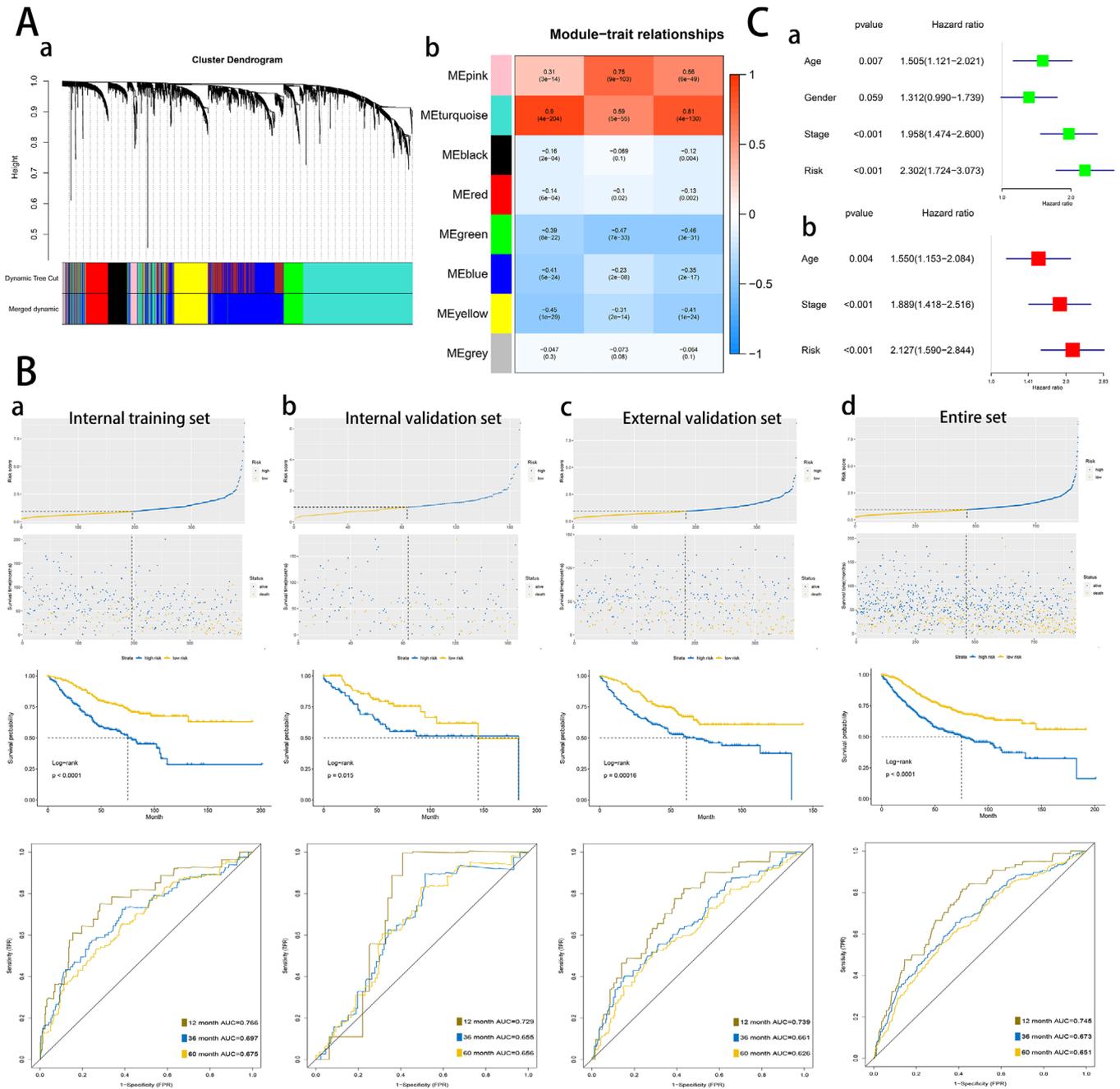


FIGURE 1 WGCNA and construction and validation of the prognostic signature. **A**, WGCNA based on overall survival-related mRNAs: (a) hierarchical cluster analysis dendrogram to identify co-expression models along with relevant color assignments; (b) correlation analysis between the gene modules and TME-related scores including immune score, stromal score, and ESTIMATE score. Each cell contained a relevant correlation and *P*-value. The cells were colored by correlation, which decreased in degree from red to blue. Gene Significance (GS) and Module Membership (MM) were computed to assess the relationship between gene expression and ESTIMATE scores (immune score, stromal score, and estimate score) and individual module. **B**, Construction and validation of prognostic signature. Efficacy of the prognostic signature in the internal training set (a), the internal validation set (b), the external validation set (c), and the entire set (d). The upper panel presented the risk score analysis of the signature, the middle panel presented the Kaplan-Meier survival analysis of the signature, and the cutoff of high-risk and low-risk groups was the median risk score in the internal training set. The lower panel presented the ROC curves of the signature. **C**, Univariate and multivariate Cox analysis in the entire set. (a) Forest plot depicting the results of univariate Cox analysis. (b) Forest plot depicting the results of multivariate Cox analysis (The risk factors that were identified as statistically significant in univariate Cox regression were further analyzed by multivariate Cox regression). *P*-value < .05 was identified as statistically significant

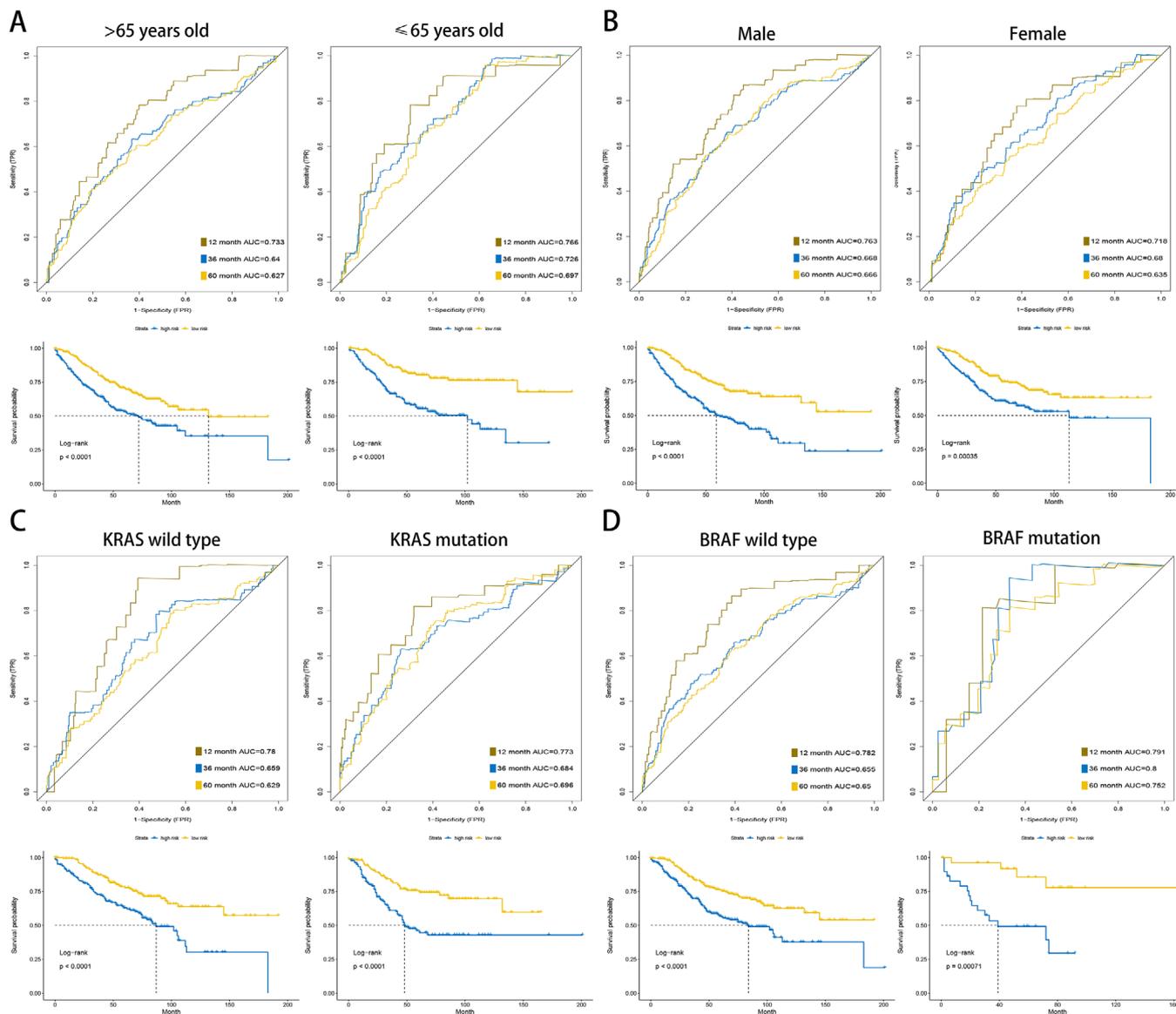


FIGURE 2 Evaluation of the signature in different subgroups classified by different clinical features. The entire set was divided into diverse subgroups according to age (A), gender (B), KRAS status (C), and BRAF status (D); the efficacy was evaluated in diverse groups. The upper panel presented the ROC curves and the lower panel presented the Kaplan-Meier survival analysis of the signature in corresponding groups; the cutoff of high risk and low risk groups was the median risk score in the internal training set

better predict the prognostic outcomes of colon cancer patients.

ACKNOWLEDGMENTS

We want to thank the GEO database for providing the raw data. The [National Natural Science Foundation of China](#) (No. 81702302 and 81472228) granted funding support for our research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AVAILABILITY OF DATA AND MATERIALS

Data and materials are all available.

CONSENT FOR PUBLICATION

All authors consent to publish.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This is a study based on data from public datasets, so we want to waive the ethics consent.

AUTHOR CONTRIBUTION

ZYZ performed and designed the study, and wrote this letter. JMX, GDH, and QYF assisted to improve the design. YL, YHM, YQX, and PZ helped to analyze the data and write the article.

Zhiyuan Zhang 

Guodong He
Qingyang Feng
Peng Zheng
Yang Lv
Yihao Mao
Yuqiu Xu
Jianmin Xu

*Department of General Surgery, Zhongshan Hospital,
Fudan University, Shanghai, China*

Correspondence

Jianmin Xu, Department of General Surgery, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai 200032, China.
Email: xujmin@aliyun.com

Zhiyuan Zhang, Guodong He, and Qingyang Feng contributed equally to this work.

ORCID

Zhiyuan Zhang  <https://orcid.org/0000-0003-0336-7783>

REFERENCES

1. Feng Q, Chang W, Mao Y, et al. Tumor-associated macrophages as prognostic and predictive biomarkers for postoperative adjuvant chemotherapy in patients with stage II colon cancer. *Clin Cancer Res.* 2019;25(13):3896-3907.
2. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
3. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
4. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997;16(4):385-395.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.