Received: 21 December 2023 | Revised: 6 March 2024 | Accepted: 26 March 2024

DOI: 10.1002/ags3.12802

AGSurg Annals of Gastroenterological Surgery Open Access WILEY

ORIGINAL ARTICLE

A new TGF-β risk score predicts clinical and immune landscape in colorectal cancer patients

Bing Tang | **Binggang Liu** | **Zhiyao Zeng**

Department of Gastrointestinal Surgery, Central Hospital of Yongzhou, Yongzhou, Hunan, China

Correspondence
Bing Tang, Department of Gastrointestinal Surgery, Central Hospital of Yongzhou, 396 Yiyun Road, Lengshuitan, Yongzhou, Hunan 425000, China.
Email: 13874615070@163.com

Abstract

Background: Aberrant TGF-β signaling pathway can lead to invasive phenotype of colorectal cancer (CRC), resulting in poor prognosis. It is pivotal to develop an effective prognostic factor on the basis of TGF-β-related genes to accurately identify risk of CRC patients.

Methods: We performed differential analysis of TGF-β-related genes in CRC patients from databases and previous literature to obtain TGF-β-related differentially expressed genes (TRDEGs). LASSO-Cox regression was utilized to build a CRC prognostic feature model based on TRDEGs. The model was validated using two GEO validation sets. Wilcoxon rank-sum test was utilized to test correlation of model with clinical factors. ESTIMATE algorithm and ssGSEA and tumor mutation burden (TMB) analysis were used to analyze immune landscape and mutation burden of high-risk (HR) and low-risk (LR) groups. CellMiner database was utilized to identify therapeutic drugs with high sensitivity to the feature genes.

Results: We established a six-gene risk prognostic model with good predictive accuracy, which independently predicted CRC patients' prognoses. The HR group was more likely to experience immunotherapy benefits due to higher immune infiltration and TMB. The feature gene TGFB2 could inhibit the efficacy of drugs such as XAV-939, Staurosporine, and Dasatinib, but promote the efficacy of drugs such as CUDC-305 and by-product of CUDC-305. Similarly, RBL1 could inhibit the drug action of Fluphenazine and Imiquimod but promote that of Irofulven.

Conclusion: A CRC risk prognostic signature was developed on basis of TGF-β-related genes, which provides a reference for risk and further therapeutic selection of CRC patients.

KEYWORDS
colorectal cancer, drugs, immune, prognosis, TGF-β

1 | INTRODUCTION

The second predominant cause of cancer-related death worldwide is colorectal cancer (CRC), one of the most prevalent malignant tumors.[1] With the increase in therapeutic options, the progression of CRC has been inhibited, and overall survival (OS) has been prolonged.[2,3] In addition, population-based CRC screening like fecal occult blood tests and colonoscopy, has led to a consistent decrease in incidence and mortality of CRC.[1,4,5] In clinical settings, TNM classification is frequently used for risk assessment and therapeutic decision-making.[6] But due to high molecular heterogeneity, even among patients with similar clinicopathological features, the risk of recurrence and death can vary greatly.[6] To accurately identify the risk of CRC patients, fresh prognostic factors are urgently required.

TGF-β signaling can inhibit epithelial cell growth in normal tissue and promote the progression of tumor cells in late-stage cancer tissue.[7] In late-stage CRC, abnormal TGF-β signaling pathway can lead to the invasive phenotype of CRC, resulting in poor prognosis. For example, loss of SMAD4 expression is an independent adverse prognosticator for stage II/III CRC and liver metastasis after curative surgery.[8] TGF-β can also inhibit intestinal immune cells and induce immune tolerance, which plays a protective role against luminal bacterial antigens.[9,10] TGF-β signaling disruption in the colon promotes cancer progression through epithelial cell transformation or tumor-matrix interaction.[11,12] Repression of TGF-β signaling is effective in the preclinical and clinical treatment of CRC.[13] TGF-β-related prognostic risk models have been developed in other cancers such as bladder cancer, and their establishment provides important reference for individualized precision therapy for bladder cancer patients.[14] Therefore, the development of TGF-β-related prognostic models has important clinical significance. However, there has been little research on the development of TGF-β-related prognostic risk models and their association with tumor microenvironment (TME) phenotypes in CRC.

In this study, we integrated CRC-related genes and clinical profiles from TCGA and GEO to establish a novel TGF-β-related prognostic signature and conducted immune analysis for specific patient populations. Finally, drug sensitivity prediction was performed based on the prognostic feature genes. This model provided some reference for the analysis of prognostic outcomes of cancer patients.

2 | MATERIALS AND METHODS

2.1 | Data collection

RNA sequencing transcriptome data and clinical information of CRC were collected from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/), including 701 samples (normal: 51, tumor: 650). Gene expression profiles of TCGA-COAD/READ were obtained in COUNT format from TCGA, and the array dataset GSE29621 (tumor: 65) and GSE39582 (tumor: 579) was from GEO (https://www.ncbi.nlm.nih.gov/geo/). TGF-β-related genes were searched in published literature.[14]

2.2 | TGF-β gene analysis

The differential analysis of transcriptome sequencing data between tumor and normal groups was performed using the edgeR package[15] to obtain differentially expressed genes (DEGs) with threshold $|\log FC| > 0.585$ and $FDR < 0.05$. Finally, TGF-β-related DEGs (TRDEGs) were obtained. STRING website (https://string-db.org/) was utilized to generate a PPI network to visualize association between genes. The integrated analysis of CNV differential data and TGF-β gene data from TCGA was performed to determine the key CNV-driven TRDEGs in CRC. The RCircos package[16] was used to plot the specific location of TGF-β genes on chromosomes. GO and KEGG enrichment analyses were conducted for relevant genes.

2.3 | Selection of prognosis-related feature genes for building prognostic models

We selected tumor patient samples with survival time greater than 30 days and merged these clinical data with the screened DEGs. Then, univariate Cox regression analysis assessed association of DEGs with prognosis, and genes with $p < 0.05$ were selected as candidate feature genes. To prevent overfitting, LASSO regression analysis was utilized to expel redundant variables, and multivariate Cox regression was finally conducted to screen for feature genes for building prognostic models. The formula for RiskScore is as follows: exp represents the gene expression level of the patient, and Coef values are regression coefficients obtained through multivariate Cox regression analysis.

$$\text{Risk Score (patient)} = \sum_{i=1}^{n} \text{exp}_{\text{gene}} * \text{Coef}_{\text{gene}}$$

Patient risk scores were computed according to expression level and hazard ratio of each gene, and samples were assigned into high-risk (HR) and low-risk (LR) groups on the basis of median risk score. The timeROC package[17] was used to draw ROC curves and calculate AUC values for 1-, 3-, and 5-year survival. The survminer R package[18] was used to draw K-M curves to evaluate OS differences among varying groups. Prediction performance was evaluated by K-M survival curves and time-dependent ROC, and then the score distribution and survival status distribution of HR and LR groups were plotted. Finally, the reliability of the model was verified by using GSE29621 as a validation set and drawing the score distribution, survival status distribution, model survival curve, and ROC curve of the HR and LR groups.

2.4 | GSEA enrichment

The c2.cp.kegg.v2022.1.Hs.symbols.gmt gene set database was used, and the JAVA application (http://software.broadinstitute.org/gsea/index.jsp) was utilized to complete GSEA on mRNAs in different groups analyzed in 1000 random sample permutations.[19]

2.5 | Independent prognostic analysis

To determine whether the signature had independent prognostic ability, univariate and multivariate Cox regression analyses were done. After multivariate Cox regression analysis, a nomogram was plotted using the rms package,[20] and a calibration curve was developed to verify nomogram accuracy.

2.6 | Immune analysis

We utilized ESTIMATE method[21] to compute ESTIMATE, immune, and stromal scores to forecast levels of infiltrating immune cells and stromal cells, which constituted basis for inferring tumor purity. In addition, ssGSEA[22] was performed on the basis of levels of 29 immune-related feature genes using the R package GSEAbase.[23]

2.7 | Mutation analysis

The corresponding gene mutation data of TCGA-COAD/READ cohort were obtained from TCGA. Tumor mutation burden (TMB) was an indicator used to reflect overall degree of gene mutations in tumor cells, usually expressed as total number of somatic mutations per megabase of the tumor genome region. Tumors with high TMB levels may have more tumor neoantigens being identified by immune system. The number of non-synonymous mutations in each patient sample was obtained by maftools R package.[22,24] Mutation data of top 30 genes in each risk group were organized, and a waterfall plot was drawn using the GenVisR package.[15]

2.8 | Drug sensitivity prediction

CellMiner (https://discover.nci.nih.gov/cellminer/) was utilized to screen for anti-tumor drugs that were significantly correlated with TGF-β feature genes in terms of sensitivity.

3 | RESULTS

3.1 | Expression and functional analysis of TGF-β genes in CRC tumors

To properly comprehend expression patterns of TRDEGs implicated in carcinogenesis, we analyzed samples from TCGA dataset and identified 8866 DEGs with significant expression differences between tumor and normal groups, with 5156 upregulated and 3710 downregulated in tumor group (Figure 1A). The intersection of these 8866 genes with TGF-β genes yielded 97 TRDEGs with differential expression between normal and CRC tumor tissues (Figure 1B). PPI network analysis revealed interaction relationships among TRDEGs (interaction score>0.4), with 92 nodes and 938 edges (Figure 1C). Finally, the positions of these TRDEGs on chromosomes were drawn (Figure 1D). GO enrichment analysis revealed that TRDEGs were mainly enriched in biological functions like TGFR signaling pathway, transmembrane receptor protein serine/threonine kinase signaling pathway, and BMP signaling pathway (Figure 1E). KEGG unveiled that related genes were enriched in cytokine-cytokine receptor interaction, signaling pathways regulating pluripotency of stem cells, cellular senescence, cell cycle, and transcriptional misregulation in cancer (Figure 1F).

3.2 | Construction of a CRC prognostic risk model

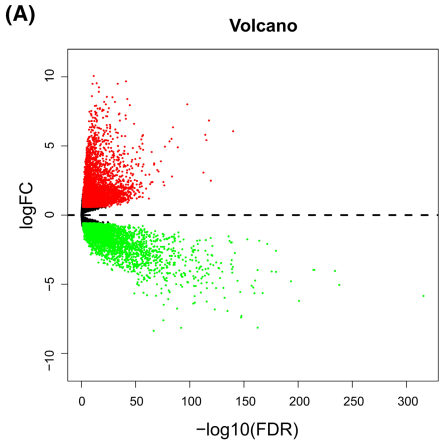
Here, we built a prognostic risk model to estimate prognosis of each CRC patient. Ten candidate genes significantly correlated with CRC patient survival were identified through univariate regression analysis in training set (Table S1). LASSO analysis was then performed and screened out eight feature genes after removing collinear genes (Figure 2A,B; Table S2). Multivariate regression analysis on the eight feature genes resulted in six genes used to construct the risk model: TGFB2, RBL1, BMP5, INHBB, NODAL, and PPP2CB (Table S3; Figure 2C). We computed risk score for each patient using following formula: Risk score=(0.162 * TGFB2 expression)+(−0.250 * RBL1 expression)+(−0.052 * BMP5 expression)+(0.219 * INHBB expression)+(−0.127 * NODAL expression)+(−0.629 * PPP2CB expression).

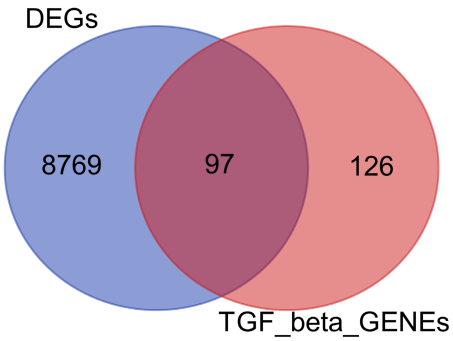
3.3 | Validation of the prognostic risk model

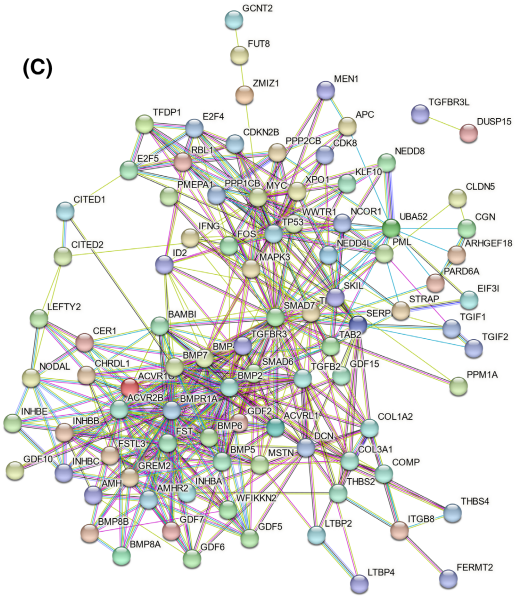
According to the median risk score of patient population in TCGA cohort, the patients were assigned to HR and LR groups. Patients in HR group had poorer survival outcomes (Figure 3A,B). Among the prognostic feature genes, TGFB2 and INHBB were upregulated in HR group, while RBL1, BMP5, NODAL, and PPP2CB were lowly expressed in HR group (Figure 3C). Survival curve reported that OS rate of LR group was higher than HR group ($p<0.001$, Figure 3D). Notably, our TRDEG prognostic model performed well in predicting 1-, 3-, and 5-year OS of these CRC patients, with AUC values all greater than 0.7 (Figure 3E). Moreover, the model also showed good validation performance in validation set GSE29621 (Figure 3F–J). Similarly, the validation results of the model in the other validation set were also good. Patients in the high-risk group had worse vital status and a worse survival rate (Figure S1A–E).

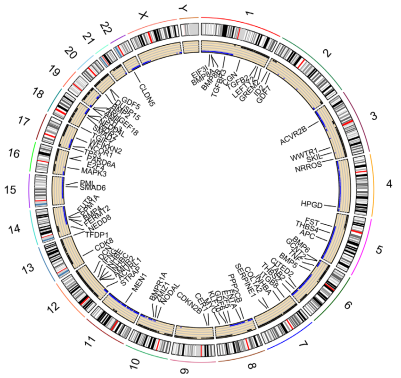
3.4 | GSEA enrichment analysis

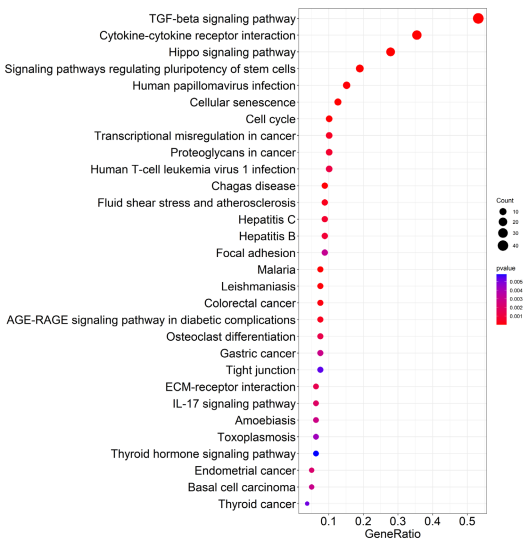
GSEA was utilized to analyze genes associated with HR and LR groups of CRC patients in TCGA set to perform KEGG analysis and assess possible biological functions. The results showed that BASAL CELL CARCINOMA, NOTCH SIGNALING PATHWAY, and ECM RECEPTOR INTERACTION pathways exhibited significant enrichment in HR group, while BASE EXCISION REPAIR, CELL CYCLE, and MISMATCH REPAIR pathways in LR group (Figure 4).


(A)
Volcano
logFC
10
5
0
−5
−10
0
50
100
150
200
250
300
−log10(FDR)


(B)
DEGs
8769
97
126
TGF_beta_GENEs


(C)
GCNT2
FUT8
ZMIZ1
MEN1
TGFBR3L
DUSP15
TFDP1
E2F4
APC
CDKN2B
PPP2CB
CDK8
NEDD8
E2F5
RBL1
KLF10
PMEPA1
PPP1CB
MYC
XPO1
CLDN5
CITED1
WWTR1
NCOR1
UBA52
CGN
IFNG
FOS
TP53
CITED2
NEDD4L
PML
ARHGEF18
ID2
MAPK3
PARD6A
SKIL
EIF3I
LEFTY2
SMAD7
STRAP
BAMBI
SERP
TGIF1
CER1
TGFBR3
TAB2
TGIF2
BMP7
BMP
NODAL
CHRDL1
SMAD6
TGFB2
GDF15
PPM1A
ACVR1
BMP2
ACVR2B
BMPR1A
INHBE
GDF2
ACVRL1
COL1A2
INHBB
BMP6
FST
DCN
FSTL3
BMP5
MSTN
COL3A1
GDF10
INHBC
GREM2
COMP
AMH
INHBA
THBS2
AMHR2
WFIKKN2
THBS4
BMP8B
GDF7
GDF5
LTBP2
ITGB8
BMP8A
GDF6
LTBP4
FERMT2


(D)
X
Y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22


(E)
TGF-beta signaling pathway
Cytokine-cytokine receptor interaction
Hippo signaling pathway
Signaling pathways regulating pluripotency of stem cells
Human papillomavirus infection
Cellular senescence
Cell cycle
Transcriptional misregulation in cancer
Proteoglycans in cancer
Human T-cell leukemia virus 1 infection
Chagas disease
Fluid shear stress and atherosclerosis
Hepatitis C
Hepatitis B
Focal adhesion
Malaria
Leishmaniasis
Colorectal cancer
AGE-RAGE signaling pathway in diabetic complications
Osteoclast differentiation
Gastric cancer
Tight junction
ECM-receptor interaction
IL-17 signaling pathway
Amoebiasis
Toxoplasmosis
Thyroid hormone signaling pathway
Endometrial cancer
Basal cell carcinoma
Thyroid cancer
0.1
0.2
0.3
0.4
0.5
GeneRatio
Count
pvalue

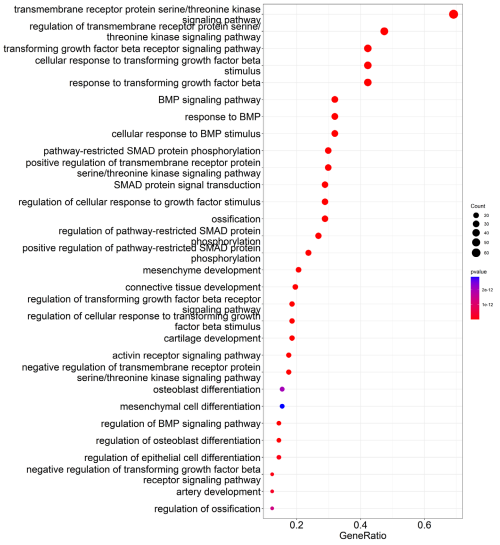

(F)
transmembrane receptor protein serine/threonine kinase signaling pathway
regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
transforming growth factor beta receptor signaling pathway
cellular response to transforming growth factor beta stimulus
response to transforming growth factor beta
BMP signaling pathway
response to BMP
cellular response to BMP stimulus
pathway-restricted SMAD protein phosphorylation
positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
SMAD protein signal transduction
regulation of cellular response to growth factor stimulus
ossification
regulation of pathway-restricted SMAD protein phosphorylation
positive regulation of pathway-restricted SMAD protein phosphorylation
mesenchyme development
connective tissue development
regulation of transforming growth factor beta receptor signaling pathway
regulation of cellular response to transforming growth factor beta stimulus
cartilage development
activin receptor signaling pathway
negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
osteoblast differentiation
mesenchymal cell differentiation
regulation of BMP signaling pathway
regulation of osteoblast differentiation
regulation of epithelial cell differentiation
negative regulation of transforming growth factor beta receptor signaling pathway
artery development
regulation of ossification
0.2
0.4
0.6
GeneRatio
Count
pvalue

FIGURE 1 The expression and functional role of TGF-β genes in CRC tumors. (A) Screening of DEGs associated with tumors. Red indicates upregulated genes, while green indicates downregulated genes. (B) Acquisition of TRDEGs in CRC. (C) Analysis of the interaction between TRDEGs displayed as a PPI network. Nodes represent proteins, and the connections between nodes represent interactions between two proteins. Different colors correspond to different types of interactions. (D) Distribution of 15 TRDEGs on human chromosomes. The outer circle represents a schematic diagram of the chromosomes, while the inner circle represents the corresponding mutation information at specific positions. (E, F) Enrichment maps for GO (E) and KEGG (F) of TRDEGs. The redder the color, the more significant the enrichment. The larger the circle, the greater the number of genes enriched in the corresponding pathway.

3.5 | Relationship between the prognostic model and clinical features and construction of nomogram

By verifying the correlation between stage, T, N, M staging, and risk score in the TCGA dataset, an association of higher risk score with poorer clinical stage was unraveled (Figure 5A). Furthermore, based on univariate and multivariate Cox regression analyses, it was found that compared with age, grade, and stage, risk score could be used as an independent prognostic indicator for CRC patients ($p < 0.001$) (Figure 5B,C). According to correlation of clinicopathological features with risk scores, a nomogram was developed to predict patient OS (Figure 5D). Calibration curves for 1-, 3-, and 5-year survival presented good consistency between predicted and actual survival results (Figure 5E–G). Overall, our prognostic model for CRC based on TRDEG was reliable.

3.6 | TME and immune cell infiltration in HR and LR groups

TME characteristics were inherently related to the effectiveness of cancer immunotherapy, and TME characteristics significantly affected the progression and metastasis of cancer. Therefore, we analyzed the TME of CRC patients. First, infiltration level of immune microenvironments in HR and LR groups was analyzed using various algorithms. The ssGSEA method was utilized to measure enrichment scores of varying immune cell subtypes, related activities or pathways, to study relationship of risk scores with immune cells and functions. HR group showed higher levels of immune-related functions and immune cell infiltration (Figure 6A). Immune score, stromal score, and ESTIMATE score of HR group were significantly higher than LR group, while the tumor purity score was substantially lower (Figure 6B). Immune checkpoint analysis revealed that most immune checkpoint molecule levels in HR group were significantly higher than LR group (Figure 6C). From the above results, it could be concluded that our prognostic model linked to TGF-β-related genes was related to TME and could indicate the immune status of CRC patients.

3.7 | Mutation analysis

Other evidence suggests that due to the higher antigen count, high TMB patients may benefit from immunotherapy.[25] We performed statistical analysis of all mutated genes in the HR and LR groups. The results showed that TMB rate in HR group was notably higher than LR group (Figure 7A). We selected the top 30 mutation load genes from the SNP dataset and plotted waterfall plots to delineate specific gene mutation spectra of CRC populations in two groups (Figure 7B,C). This further supported the notion that patients in HR group were more suited to immunotherapy.

3.8 | Correlation study between prognostic features gene and drug sensitivity

To explore the drugs related to feature genes, we used the CellMiner database to predict relationship of risk model genes with drug sensitivity. TGFB2 was positively correlated with XAV-939, Staurosporine, and Dasatinib (0.652, 0.587, 0.481), while RBL1 was positively correlated with Fluphenazine and Imiquimod (0.626, 0.490), and TGFB2 was negatively correlated with CUDC-305 and by-product of CUDC-305 (−0.494, −0.475), and RBL1 was negatively correlated with Irofulven (−0.493) (Figure 8). Positive correlation represented that the gene could inhibit the efficacy of the drug, while negative correlation indicated that the gene could promote the efficacy of the drug.

4 | DISCUSSION

In this study, we first used the TCGA-COAD/READ dataset combined with the TGF-β gene set to obtain 97 TRDEGs. LASSO-Cox regression was then done to obtain a six-gene risk prognosis model. Among these six featured genes, TGFB2 and INHBB were highly expressed in the HR group, while RBL1, BMP5, NODAL, and PPP2CB were lowly expressed in the HR group. A study showed that high expression of miR-149-5p represses CRC cell migration and invasion, and that TGFB2 is a direct target of miR-149-5p in CRC and is negatively linked to miR-149-5p expression,[26] consistent with our conclusion that TGFB2 was highly expressed in the HR group. In Yuan et al.'s study, increased INHBB is implicated in poor OS and DFS in CRC patients, making it a novel biomarker for promoting CRC progression.[27] BMP5 is a tumor suppressor in CRC and can signal through Jak–Stat pathway to inhibit activation of the oncogene EPSTI1.[28] NODAL can enhance the proliferation rate, motility, invasiveness, and EMT ability of CRC cells, and knocking down its expression leads to an increase in lipid peroxidation in CRC cells.[29] In this study, the gene showed significant differences between the HR and LR groups, indicating that NODAL could be a potential therapeutic target. RBL1 is considered to regulate the cell cycle and its expression level increases

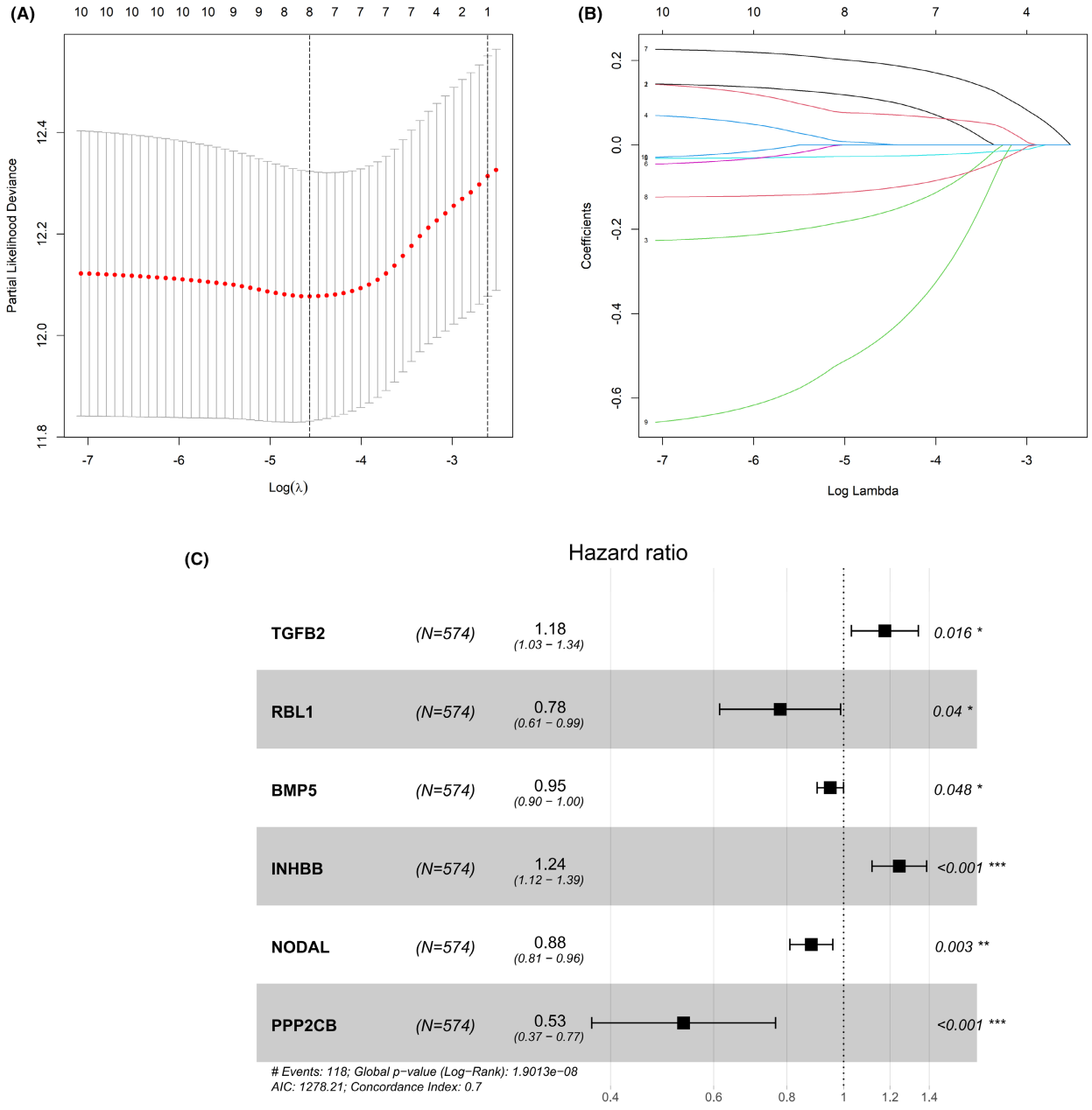
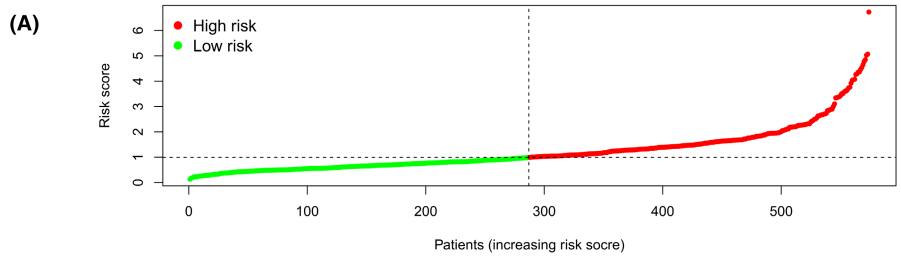
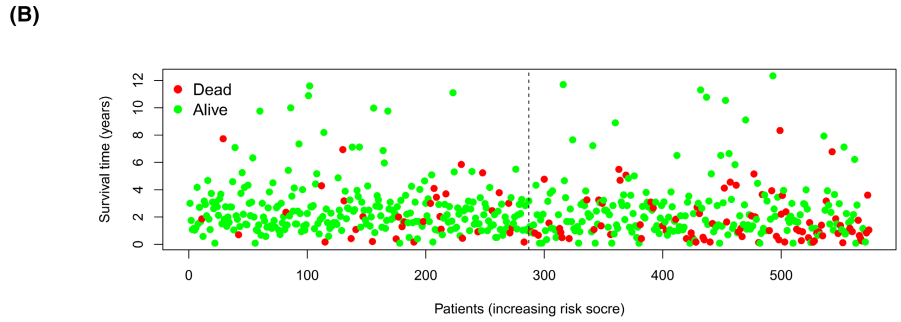
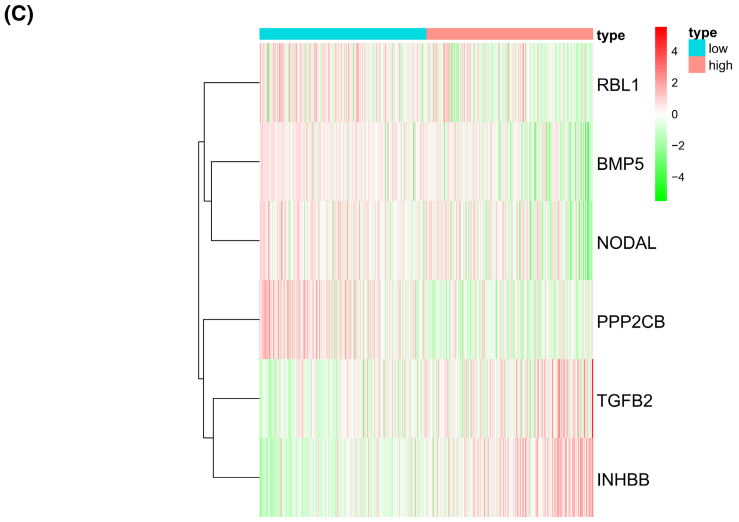


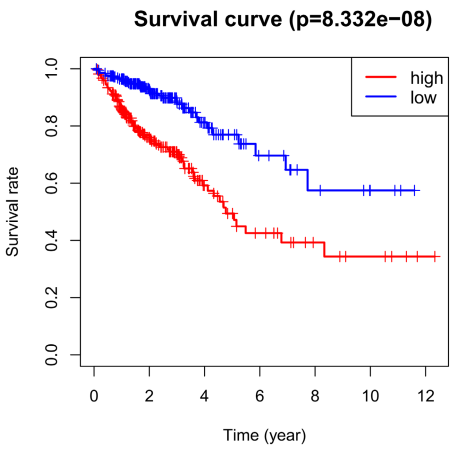
FIGURE 2 Development of a prognostic risk model in CRC. (A) Using partial-likelihood deviance to select the optimal λ value. lambda.min (left vertical line): the λ value when MSE (Mean Squared Error) is minimized; lambda.1st (right vertical line): the λ value for the simplest model within a range of variance from lambda.min. lambda.min is typically chosen. (B) Distribution plot of LASSO coefficients. Each colored line represents the variation of the regression coefficient β for a variable. The numbers below the x-axis indicate the penalty values (tuning coefficients), and the numbers above the x-axis represent the remaining number of variables at that value. As λ increases, the regression coefficients for various variables decrease, and some may become zero, indicating minimal contribution of those variables to the model, making them removable. (C) Forest plot of model feature genes. HR>1: risk factors. HR<1: protective factors.

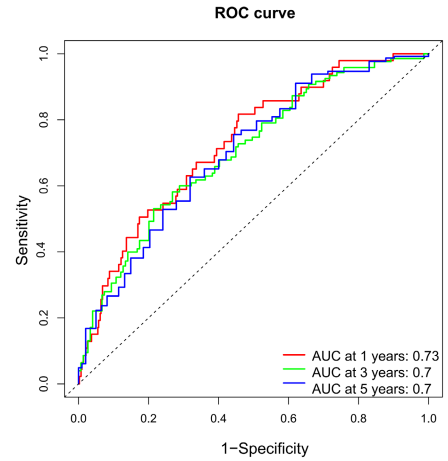
FIGURE 3 Validation of the risk prognostic model. (A–E) Analysis of CRC patients in the training set (TCGA): (A) Distribution of patient sample risk scores. Red color represents high-risk patients, while green color represents low-risk patients. The dashed line represents the median value of the risk score. (B) Survival status of each sample. Red color represents patient's deceased status, while green color represents patient's survival status. (C) Expression levels of model feature genes in TCGA patient population. Red color represents upregulation, while green color represents downregulation. (D) Survival analysis plot. (E) AUC values for patient's 1-year, 3-year, and 5-year risk scores. The closer the AUC (area under the curve) is to 1.0, the higher the accuracy of the detection method. (F-J) Analysis of CRC patients in the validation set (GSE29621): (F) Distribution of patient sample risk scores. (G) Relationship between sample risk scores and survival status. (H) Expression levels of model feature genes in the GSE29621 dataset. (I) Survival analysis plot. (J) AUC values for patient 1-year, 3-year, and 5-year risk scores.

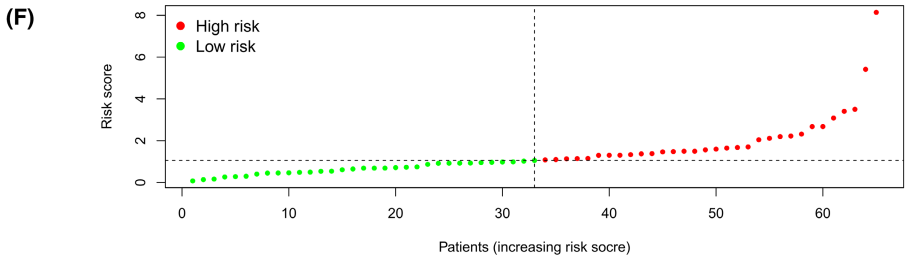

(A)
High risk
Low risk
Risk score
Patients (increasing risk socre)

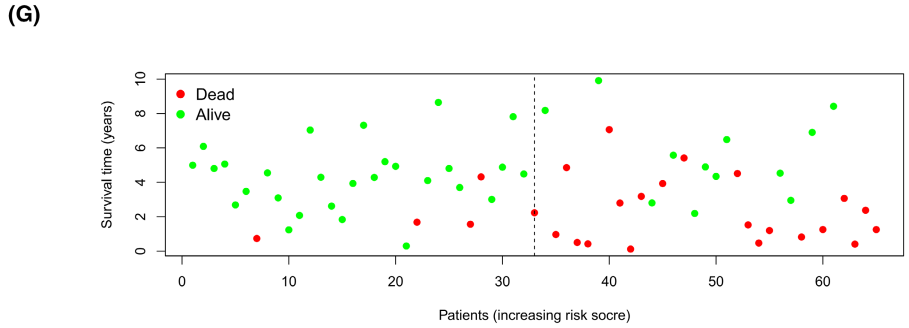

(B)
Dead
Alive
Survival time (years)
Patients (increasing risk socre)

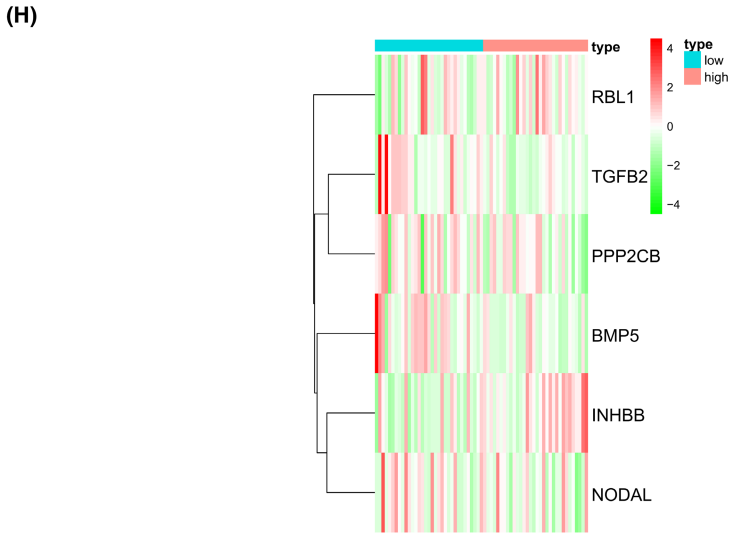

(C)
type
low
high
RBL1
BMP5
NODAL
PPP2CB
TGFB2
INHBB

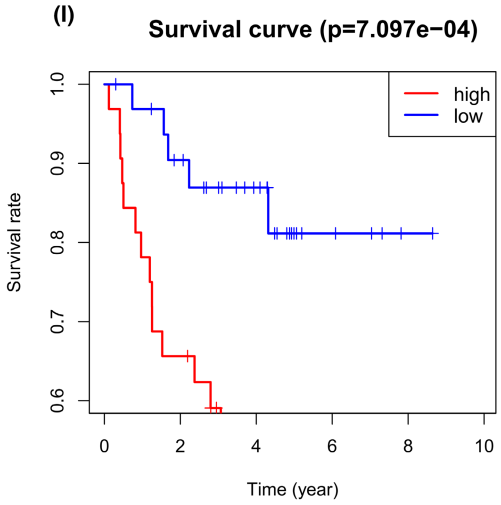

(D)
Survival curve (p=8.332e−08)
high
low
Survival rate
Time (year)


(E)
ROC curve
AUC at 1 years: 0.73
AUC at 3 years: 0.7
AUC at 5 years: 0.7
Sensitivity
1−Specificity









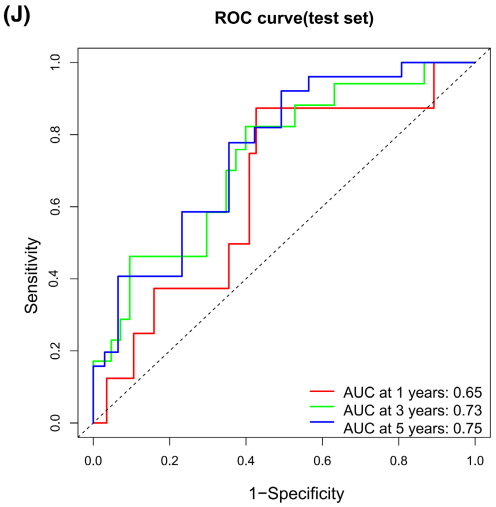


FIGURE 3 (Continued)

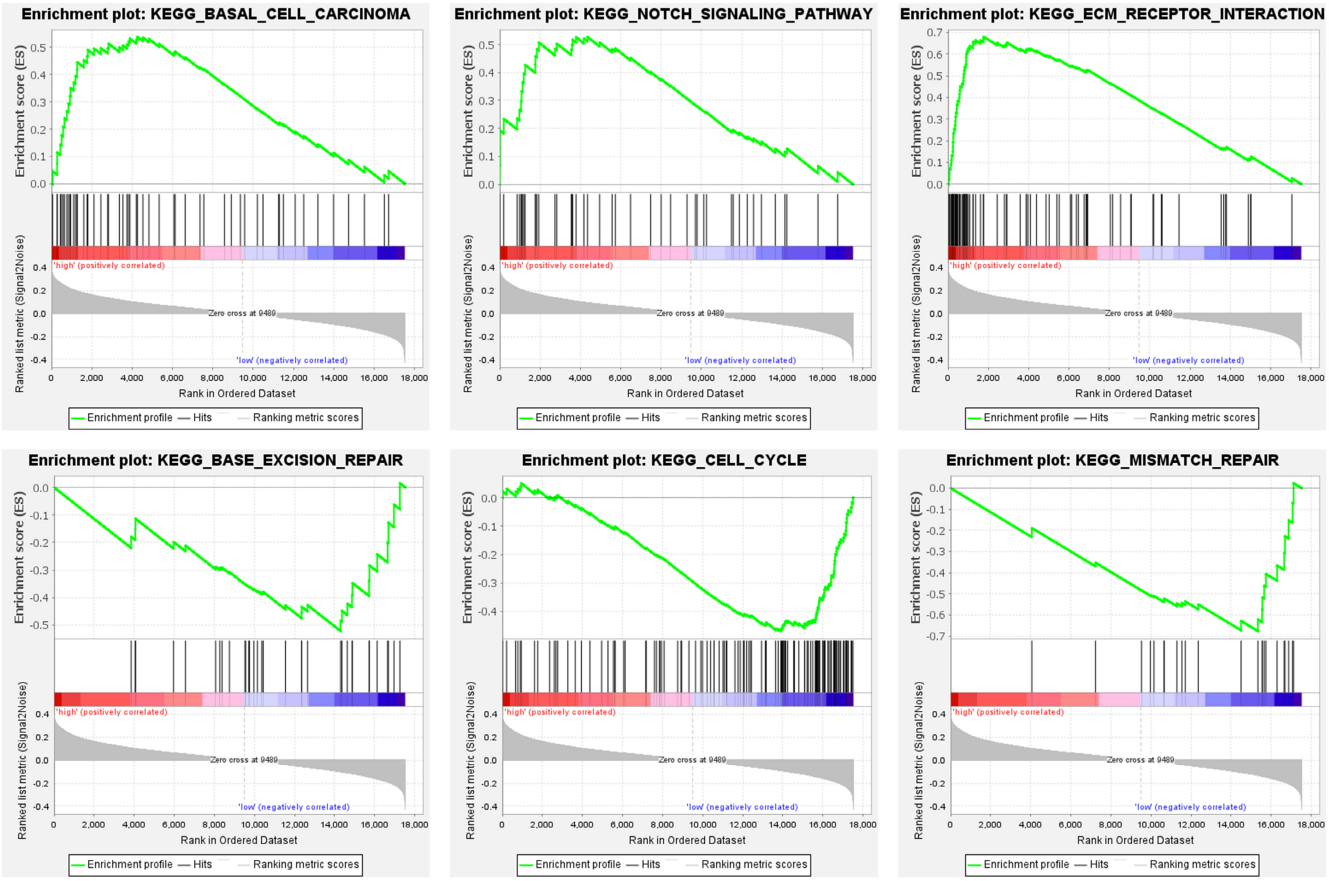


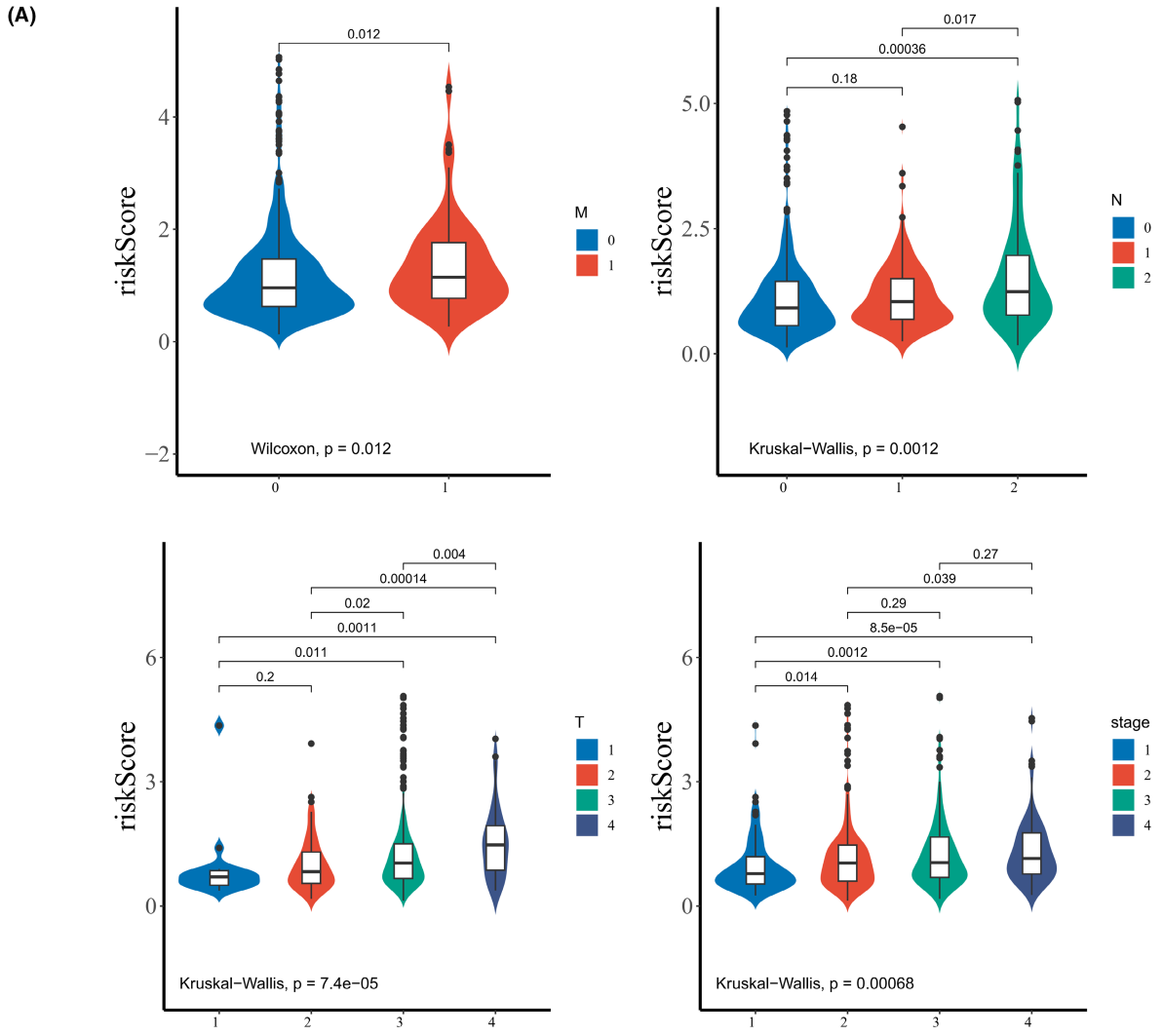
FIGURE 4 GSEA pathway enrichment analysis. ES>0 indicates high expression of the pathway in the high-risk group. ES<0 indicates high expression of the pathway in the low-risk group.

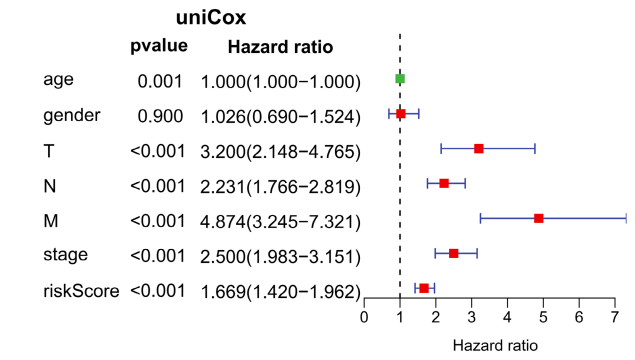
during the G1-S transition.[30] In bladder cancer, PPP2CB is identified as one of key genes in Wnt signaling, which is linked to cancer cell EMT progression and immune cell infiltration as a prognosticator.[31] Hence, genes we obtained participated in cancer progression. The TGF-β-related featured genes in our model were reliable and could possibly impact the development of CRC. In training and validation sets, we divided CRC patients into HR and LR groups on the basis of risk score. Prognosis of LR group patients was significantly better than that of HR group patients. Multivariate Cox regression revealed that the developed model was an independent prognostic predictor for CRC. To expand value of risk score in clinical practice, we constructed a nomogram with clinical indicators to offer a personalized prognosis risk assessment system for clinicians and individualized treatment options for patients.

Immune cell infiltration is one of the TME key immunological characteristics, and it is essential for tumor cells to evade the immune system and the emergence of an inflammatory environment.[32] We looked at the connection between TRDEGs and the level of immune cell infiltration in CRC. HR group patients showed higher immune, stromal, and ESTIMATE scores. Maintaining self-immune tolerance and regulating the intensity and duration of immunological responses in peripheral tissues are made possible by immune checkpoints, which act as immune system regulators.[33,34] This study found that many immune checkpoints were noticeably upregulated in HR group patients. HR group patients had substantially higher TMB values than LR group patients. A higher TMB represents a stronger degree of tumor mutations, leading to stronger immunogenicity and triggering T cell responses.[35] These findings suggested that HR group patients may benefit more from immune checkpoint blockade and other immunotherapies.

Finally, we used the CellMiner database to predict relationship of risk model genes with drug sensitivity. TGFB2 repressed efficacy of XAV-939, Staurosporine, and Dasatinib drugs but promoted the efficacy of CUDC-305 and by-product of CUDC-305 drugs. Similarly, RBL1 could inhibit the effect of Fluphenazine and Imiquimod drugs but promote the effect of Irofulven. Therefore, our study provided a new perspective for clinicians to implement clinical anti-tumor drugs based on patient characteristics to maximize the efficacy of the drugs.

The pivotal role of TGF-β in tumor evolution has emerged, and its related bioinformatics analysis has always been a focus of researchers. We built a TGF-β-related gene CRC prognosis model by referencing databases and literature information. The model could predict CRC patients' prognoses accurately and was a unique prognostic biomarker and potential therapeutic target for CRC patients. HR group patients were more suitable for immune therapy and may





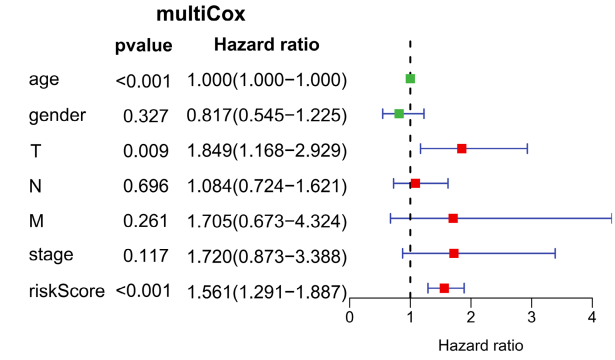
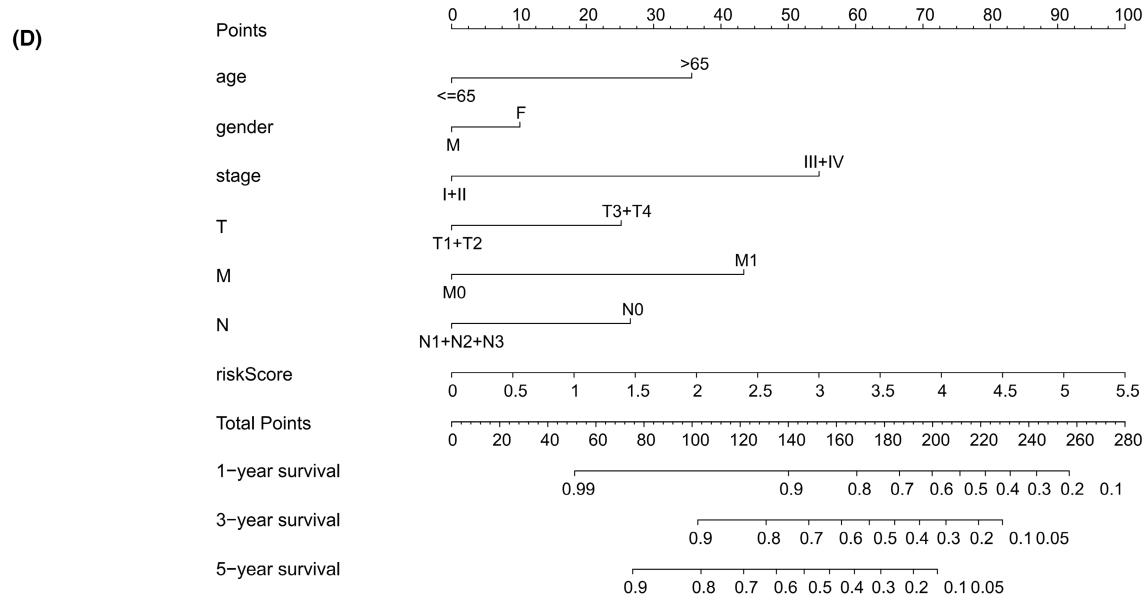
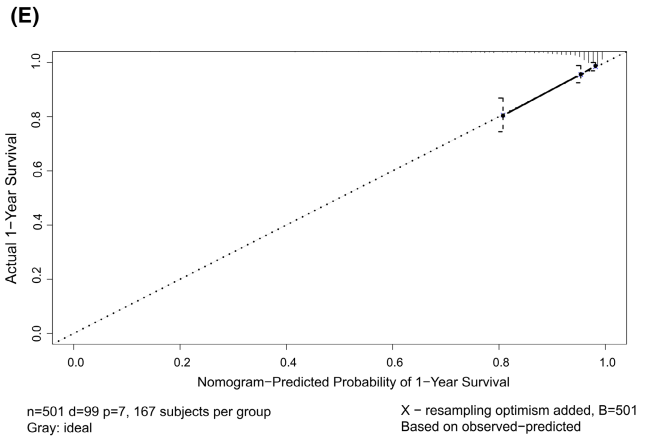
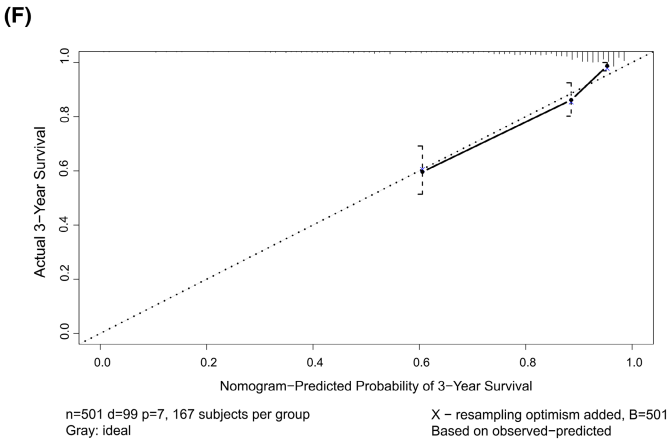


FIGURE 5 Validation of the prognostic features of TGF-β-related genes and generation of a nomogram. (A) Correlation between risk score and patient tumor grade and TNM stage. Univariate (B) and multivariate (C) Cox regression analysis of clinicopathological features and risk scores in the TCGA training set. HR>1: Risk factor. HR<1: Protective factor. (D) Predictive nomogram for the 1-year, 3-year, and 5-year survival of CRC patients, combining age, grade, and stage. (E-G) Calibration curves evaluate the accuracy of the predicted results for patient 1-year, 3-year, and 5-year outcomes. The x-axis represents the predicted event rate (predicted risk), and the y-axis represents the observed event rate (observed risk), both ranging from 0 to 1. The dashed line on the diagonal is the reference line, indicating the scenario where the predicted value equals the observed value. If the predicted value matches the observed value, the red line coincides perfectly with the reference line. If the predicted value is greater than the observed value, indicating an overestimation of risk, the curve will be below the reference line. If the predicted value is less than the observed value, indicating an underestimation of risk, the curve will be above the reference line.







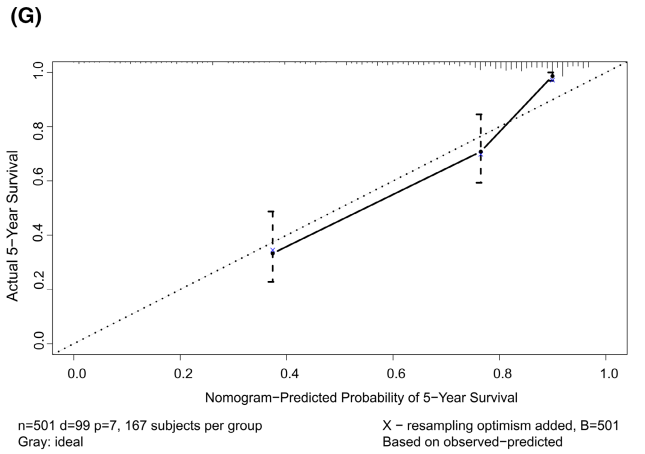


FIGURE 5 (Continued)

benefit more from this therapy. However, the limitations of this study still need to be continuously optimized in subsequent analyses and problem-solving. Firstly, the dataset used to validate the model was limited, and further clinical validation is needed. Secondly, validation analysis of in vitro experiments is also crucial. In conclusion, the model constructed in this study provides important reference value for prognosis prediction and clinical therapeutic selection of CRC patients.

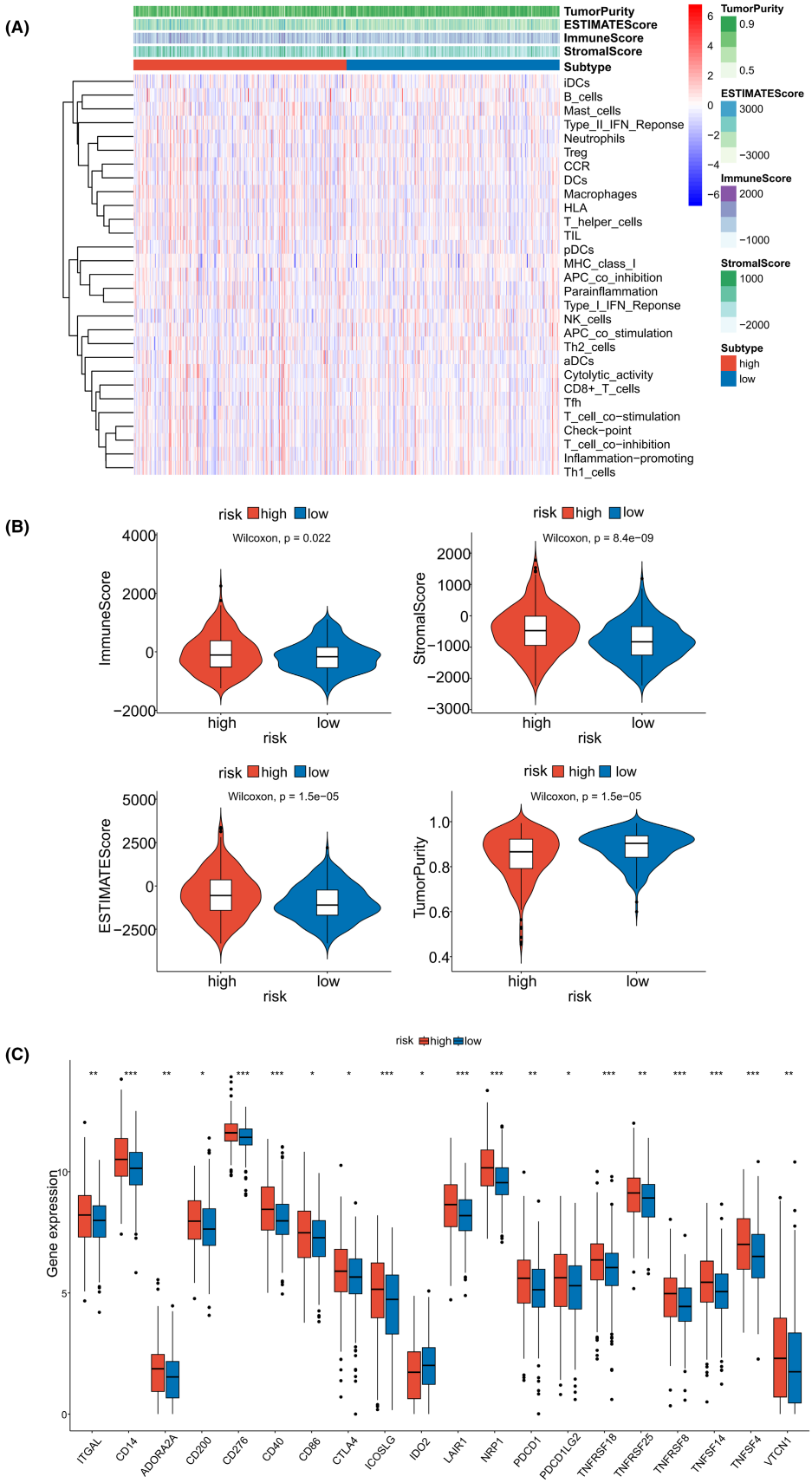


FIGURE 6 TME and immune cell infiltration in the HR and LR groups. (A) Heatmap of immune cell infiltration in the TME. (B) Violin plots of immune scores, stromal scores, ESTIMATE scores, and tumor purity scores. (C) Expression levels of immune checkpoint genes. Different colors represent different scores.

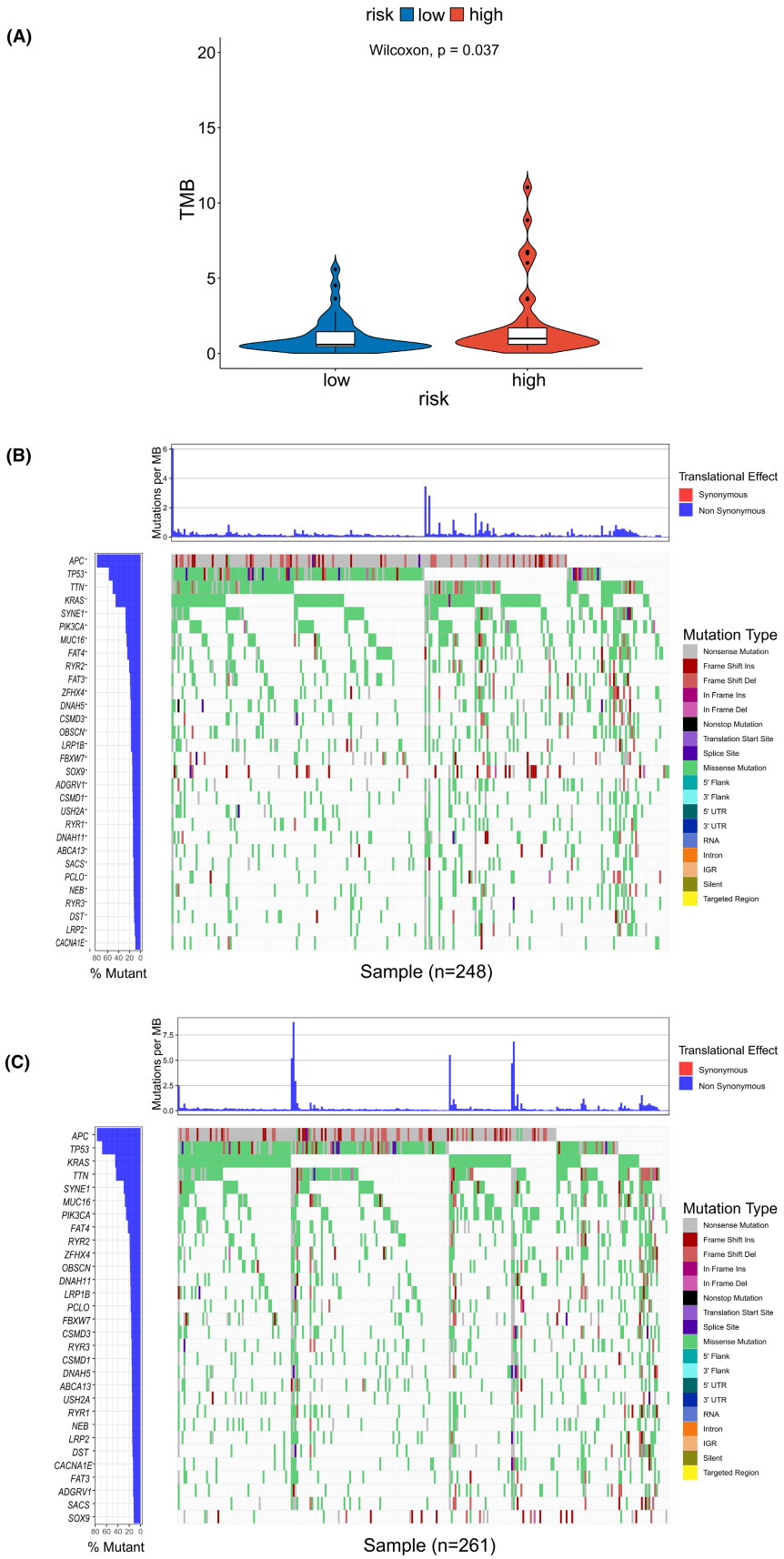


FIGURE 7 TMB analysis. (A) Comparison of TMB between the HR and LR groups. (B) Specific mutation status of the top 30 mutated genes in the LR group. (C) Specific mutation status of the top 30 mutated genes in the HR group. The x-axis represents samples, and the y-axis represents genes. The colored blocks in the central section represent gene mutations in the samples, where different colors represent different mutation types. The topmost histogram represents the number of mutations for each sample. The leftmost bar chart represents the proportion of mutations for each gene across all samples.

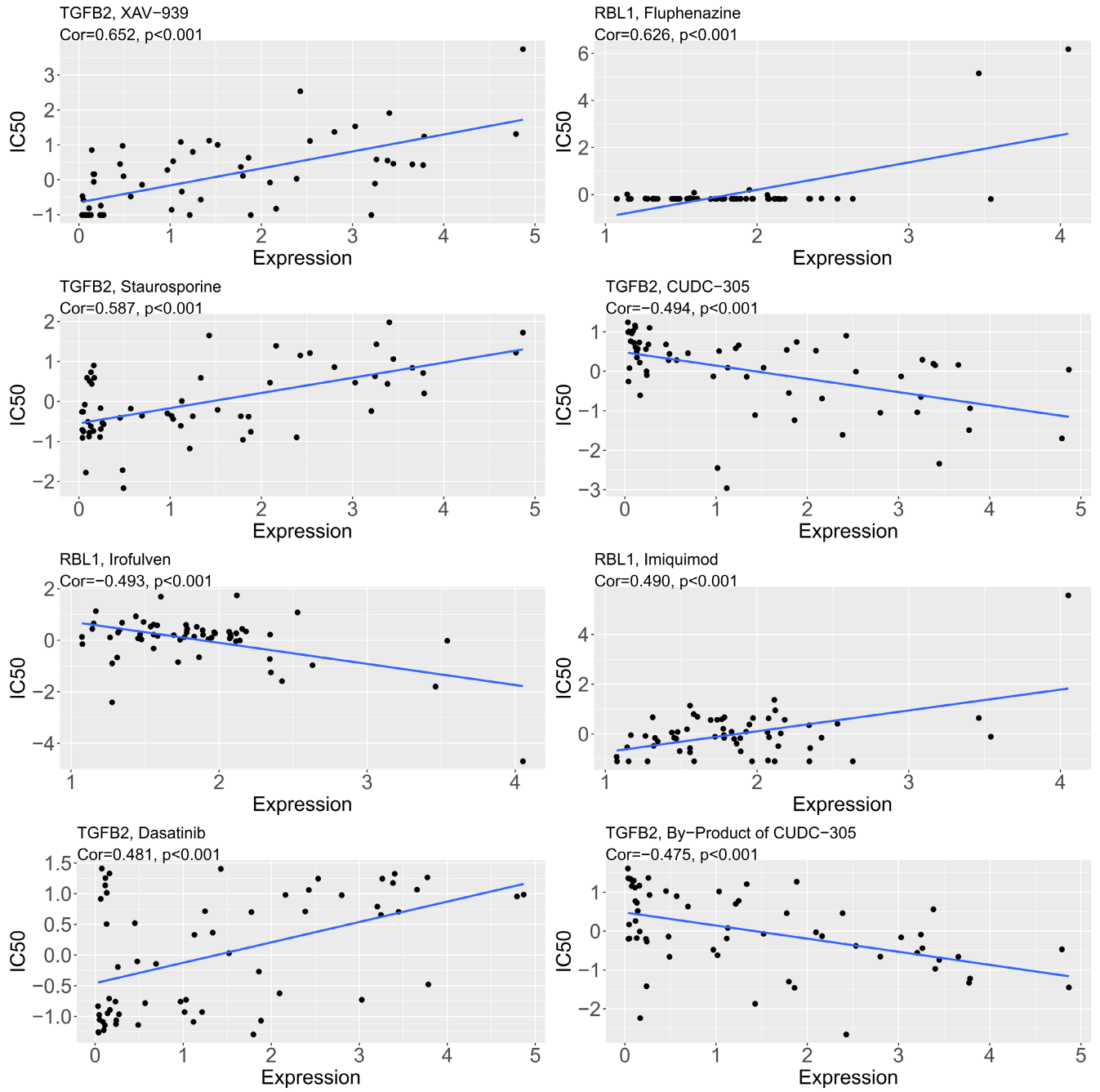


FIGURE 8 Correlation between prognostic feature genes and drug sensitivity. Positive correlation indicates that the gene can inhibit the effectiveness of the drug, while negative correlation suggests that the gene can promote the effectiveness of the drug.

AUTHOR CONTRIBUTIONS

Bing Tang contributed to the study design and wrote the manuscript. Binggang Liu acquired the data and performed data analysis. Zhiyao Zeng revised the article and gave the final approval of the version to be submitted. All authors read and approved the final manuscript.

FUNDING INFORMATION

Not applicable.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available from the corresponding author upon request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Approval of the research protocol: N/A.

Informed Consent: N/A.

Registry and the Registration No. of the study/trial: N/A.

Animal Studies: N/A.

ORCID

Zhiyao Zeng https://orcid.org/0009-0000-7730-1926

REFERENCES

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70:145–64.
2. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394:1467–80.
3. Guren MG. The global challenge of colorectal cancer. Lancet Gastroenterol Hepatol. 2019;4:894–5.
4. Hissong E, Pittman ME. Colorectal carcinoma screening: established methods and emerging technology. Crit Rev Clin Lab Sci. 2020;57:22–36.
5. Bui NC, Cho HN, Lee YY, Suh M, Park B, Jun JK, et al. Stages of adoption for fecal occult blood test and colonoscopy tests for colorectal cancer screening in Korea. Cancer Res Treat. 2018;50:416–27.
6. Weiser MR, Gonen M, Chou JF, Kattan MW, Schrag D. Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. J Clin Oncol. 2011;29:4796–802.
7. Principe DR, Doll JA, Bauer J, Jung B, Munshi HG, Bartholin L, et al. TGF-beta: duality of function between tumor prevention and carcinogenesis. J Natl Cancer Inst. 2014;106:djt369.
8. Mizuno T, Cloyd JM, Vicente D, Omichi K, Chun YS, Kopetz SE, et al. SMAD4 gene mutation predicts poor prognosis in patients undergoing resection for colorectal liver metastases. Eur J Surg Oncol. 2018;44:684–92.
9. Ihara S, Hirata Y, Koike K. TGF-beta in inflammatory bowel disease: a key regulator of immune cells, epithelium, and the intestinal microbiota. J Gastroenterol. 2017;52:777–87.
10. Batlle R, Andres E, Gonzalez L, Llonch E, Igea A, Gutierrez-Prat N, et al. Regulation of tumor angiogenesis and mesenchymal-endothelial transition by p38alpha through TGF-beta and JNK signaling. Nat Commun. 2019;10:3071.
11. Inamoto S, Itatani Y, Yamamoto T, Minamiguchi S, Hirai H, Iwamoto M, et al. Loss of SMAD4 promotes colorectal cancer progression by accumulation of myeloid-derived suppressor cells through the CCL15-CCR1 chemokine Axis. Clin Cancer Res. 2016;22:492–501.
12. Yamamoto T, Kawada K, Itatani Y, Inamoto S, Okamura R, Iwamoto M, et al. Loss of SMAD4 promotes lung metastasis of colorectal cancer by accumulation of CCR1+ tumor-associated neutrophils through CCL15-CCR1 Axis. Clin Cancer Res. 2017;23:833–44.
13. Villalba M, Evans SR, Vidal-Vanaclocha F, Calvo A. Role of TGF-beta in metastatic colon cancer: it is finally time for targeted therapy. Cell Tissue Res. 2017;370:29–39.
14. Liu Z, Qi T, Li X, Yao Y, Othmane B, Chen J, et al. A novel TGF-beta risk score predicts the clinical outcomes and tumour microenvironment phenotypes in bladder cancer. Front Immunol. 2021;12:791924.
15. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: genomic visualizations in R. Bioinformatics. 2016;32:3012–4.
16. Naquin D, d'Aubenton-Carafa Y, Thermes C, Silvain M. CIRCUS: a package for Circos display of structural genome variations from paired-end and mate-pair sequencing data. BMC Bioinformatics. 2014;15:198.
17. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med. 2013;32:5381–97.
18. Therneau TM, Grambsch PM. Modeling survival data: extending the cox model. New York: Springer-Verlag; 2000.
19. Yao B, Wang L, Wang H, Bao J, Li Q, Yu F, et al. Seven interferon gamma response genes serve as a prognostic risk signature that correlates with immune infiltration in lung adenocarcinoma. Aging (Albany NY). 2021;13:11381–410.
20. Huang C, Liu Z, Xiao L, Xia Y, Huang J, Luo H, et al. Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination. Front Oncol. 2019;9:1159.
21. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
22. Xu Q, Chen S, Hu Y, Huang W. Landscape of immune microenvironment under immune cell infiltration pattern in breast cancer. Front Immunol. 2021;12:711433.
23. Yin Y, Tian Y, Ren X, Wang J, Li X, Zeng X. Qualification of necroptosis-related lncRNA to forecast the treatment outcome, immune response, and therapeutic effect of kidney renal clear cell carcinoma. J Oncol. 2022;2022:3283343.
24. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.
25. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol. 2019;30:44–56.
26. Qu L, Chen Y, Zhang F, He L. The lncRNA DLGAP1-AS1/miR-149-5p/TGFB2 axis contributes to colorectal cancer progression and 5-FU resistance by regulating smad2 pathway. Mol Ther Oncolytics. 2021;20:607–24.
27. Yuan J, Xie A, Cao Q, Li X, Chen J. INHBB is a novel prognostic biomarker associated with cancer-promoting pathways in colorectal cancer. Biomed Res Int. 2020;2020:6909672.
28. Chen E, Yang F, He H, Li Q, Zhang W, Xing J, et al. Alteration of tumor suppressor BMP5 in sporadic colorectal cancer: a genomic and transcriptomic profiling based study. Mol Cancer. 2018;17:176.
29. Wu T, Wan J, Qu X, Xia K, Wang F, Zhang Z, et al. Nodal promotes colorectal cancer survival and metastasis through regulating SCD1-mediated ferroptosis resistance. Cell Death Dis. 2023;14:229.
30. Ventura E, Iannuzzi CA, Pentimalli F, Giordano A, Morrione A. RBL1/p107 expression levels are modulated by multiple signaling pathways. Cancers (Basel). 2021;13:13.
31. Sun S, Wang Y, Wang J, Bi J. Wnt pathway-related three-mRNA clinical outcome signature in bladder urothelial carcinoma: computational biology and experimental analyses. J Transl Med. 2021;19:409.
32. Lei X, Lei Y, Li JK, Du WX, Li RG, Yang J, et al. Immune cells within the tumor microenvironment: biological functions and roles in cancer immunotherapy. Cancer Lett. 2020;470:126–33.
33. Okazaki T, Honjo T. The PD-1-PD-L pathway in immunological tolerance. Trends Immunol. 2006;27:195–201.
34. Lai G, Li K, Deng J, Liu H, Xie B, Zhong X. Identification and validation of a gene signature for lower-grade gliomas based on Pyroptosis-related genes to predict survival and response to immune checkpoint inhibitors. J Healthc Eng. 2022;2022:8704127.
35. Addeo A, Friedlaender A, Banna GL, Weiss GJ. TMB or not TMB as a biomarker: that is the question. Crit Rev Oncol Hematol. 2021;163:103374.

SUPPORTING INFORMATION

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

How to cite this article: Tang B, Liu B, Zeng Z. A new TGF-β risk score predicts clinical and immune landscape in colorectal cancer patients. Ann Gastroenterol Surg. 2024;8:927–941. https://doi.org/10.1002/ags3.12802