Development of a prognostic risk model for colorectal cancer and association of the prognostic model with cancer stem cell and immune cell infiltration

Jian Zhang[1], Peter C. Ambe[2], Aasma Shaukat[3]

[1]Department of Clinical Laboratory, Benxi Iron and Steel General Hospital, Benxi, China; [2]Department of Surgery II, Witten/Herdecke University, Witten, Germany; [3]Division of Gastroenterology, NYU Grossman School of Medicine, New York, NY, USA

Contributions: (I) Conception and design: J Zhang; (II) Administrative support: J Zhang; (III) Provision of study materials or patients: J Zhang; (IV) Collection and assembly of data: J Zhang, PC Ambe; (V) Data analysis and interpretation: J Zhang, A Shaukat; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jian Zhang, MS. Department of Clinical Laboratory, Benxi Iron and Steel General Hospital, No. 29, Renmin Road, Benxi 117017, China. Email: zjbg163163@163.com.

Background: The development of a prognostic model for patients with colorectal cancer (CRC) can facilitate the assessment of patient survival and the effectiveness of clinical treatments. A reasonable prognostic model can provide a basis for individualized treatment, prognostic risk stratification, and subsequent therapy for CRC patients. The aim of our study was to construct a prognostic model for patients with CRC using sequencing data derived from The Cancer Genome Atlas (TCGA) database.

Methods: Sequencing data of paracancerous tissues (n=51) and CRC samples (n=647) were downloaded from the TCGA database. Least absolute shrinkage and selection operator (LASSO) and Cox regression analyses were employed to identify prognostic factors. A restricted cubic spline (RCS) model was used to assess the nonlinear relationship between risk score and poor overall survival (OS). The Genomics of Drug Sensitivity in Cancer (GDSC) database was accessed to evaluate the correlation between the prognostic model's risk score and drug sensitivity. The single-sample gene set enrichment analysis (ssGSEA), estimate, and CIBERSORT algorithms were applied to quantify the association between prognostic genes and immune cell infiltration in CRC.

Results: Our findings revealed that six genes, including Niemann-Pick C1-like 1 (*NPC1L1*) [hazard ratio (HR) =1.53; 95% confidence interval (CI): 1.08–2.17; P=0.02], glucagon-like peptide 2 receptor (*GLP2R*) (HR =0.68; 95% CI: 0.48–0.97; P=0.04), solute carrier family 8 member A3 (*SLC8A3*) (HR =0.67; 95% CI: 0.47–0.96; P=0.03), alpha-1-microglobulin/bikunin precursor (*AMBP*) (HR =0.64; 95% CI: 0.45–0.91; P=0.01), single-pass membrane protein with coiled-coil domains 2 (*SMCO2*) (HR =0.68; 95% CI: 0.48–0.97; P=0.03), and tetratricopeptide repeat domain 16 (*TTC16*) (HR =1.55; 95% CI: 1.09–2.20; P=0.02) function as independent prognostic factors for CRC. Based on these six genes, the developed prognostic assessment model identified a strong association between high risk score and poor OS (HR =2.43; 95% CI: 1.67–3.53; P<0.001) in patients with CRC. Furthermore, the analysis revealed a nonlinear relationship (P<0.001) between continuous variation in risk score and the risk of poor OS. Additionally, specific genes included in the prognostic model were found to be strongly associated with cancer stem cell and immune cell infiltration in CRC.

Conclusions: We developed a prognostic risk model incorporating a six-gene panel for patients with CRC. Our analysis revealed a nonlinear relationship between this prognostic model and OS in patients with CRC. A high risk score was associated with poor prognosis, indicating that the adverse outcomes observed in patients with CRC may be influenced by cancer stem cell and immune cell infiltration. Our model provides a promising predictive method for the prognosis of CRC patients, but it still needs to be validated in a larger sample size.

Introduction

Colorectal cancer (CRC) is the most prevalent malignancy of the digestive system and ranks third globally in terms of incidence, constituting 9.6% of all cancer cases (1). In China, the incidence of CRC in 2022 reached 517,000 cases, making it the second most common cancer following lung cancer (2,3). Among patients with CRC, 35% are diagnosed with metastatic disease, and up to 50% of those initially diagnosed with nonmetastatic CRC eventually develop metastases (4-6). Despite a gradual decline in CRC mortality rates, accurate prognostic prediction remains crucial in analyzing survival (5). Consequently, the development of an effective prognostic model is crucial in studying the survival duration of patients with CRC.

The prognosis of CRC is influenced by multiple factors, including age, gene mutations, aberrant gene transcription, microsatellite instability, cancer stem cells, drug sensitivity, gene sensitivity, and immune cell infiltration (7-12). Furthermore, genes associated with specific signaling pathways, such as ferroptosis,

pyroptosis, cuproptosis-related long non-coding RNAs, programmed cell death, lipid metabolism, mitosis, the cell cycle, lactation, methylation, hypoxia, immune cell-based prognosis model, and lactate metabolism have also been implicated in the prognosis of CRC (13-20). Despite the advancements in these models, limitations persist, such as the need for more precise and individualized predictions. Many models struggle with accuracy and may not fully capture the complexity of CRC progression. Furthermore, the integration of molecular markers and the development of time-dependent models are suggested to enhance prediction capabilities. Addressing these limitations is crucial to meet the clinical needs of personalized treatment plans and improved patient management in CRC. Given that the prognosis of CRC is a dynamic process, the rate of progression varies among patients and can also fluctuate within the same patient over time (21,22). To address this variability, some researchers have used serum markers such as carcinoembryonic antigen (CEA), carbohydrate antigen 19-9, and cancer antigen 125, along with perioperative longitudinal measurement data to construct dynamic prediction models, which can serve as personalized and adaptive prognostic tool for patients with CRC (23). Consensus molecular subtypes (CMSs) are employed in multiomics methodologies to classify patients with CRC (24). The prognostic significance of CMS has been validated in cases of metastasis, with CMS2 demonstrating the most favorable prognosis and CMS1 tumors being linked to an elevated risk of progression and mortality post chemotherapy (25,26). CMS4 is also associated with a poor prognosis (26). However, the clinical application of CMS is significantly constrained due to the extensive number of markers involved, the necessity for cross-platform analysis, the integration of multiomics data, and the intricate processes of data acquisition and analysis (25-27).

The CRC literature indicates that CRC-related prognostic models grounded in specific signaling pathways often suffer from the absence of critical indicators. Furthermore, the clinical applicability of CRC prognostic models derived from multiomics data is hindered by the challenges associated with multiplatform integration,

---

Highlight box

Key findings
- We developed a prognostic risk model incorporating a six-gene panel for patients with colorectal cancer (CRC).

What is known and what is new?
- Genes associated with specific signaling pathways, such as ferroptosis, pyroptosis, programmed cell death, lipid metabolism, mitosis, the cell cycle, lactation, hypoxia, and lactate metabolism, have been implicated in the prognosis of CRC.
- In our study, a higher risk score was associated with poorer prognosis, indicating that the adverse outcomes observed in patients with CRC may be influenced by cancer stem cell and immune cell infiltration. Our analysis revealed a nonlinear relationship between this prognostic model and poorer overall survival in patients with CRC.

What is the implication, and what should change now?
- We established a six-gene panel to complement CRC prognostic stratification and clinical treatment differentiation.

---

multi-index considerations, and complex algorithms. Using differentially expressed genes (DEGs) screening and prognostic index screening within The Cancer Genome Atlas (TCGA) database, we aimed to develop a relatively straightforward and broadly applicable prognostic model. Additionally, we analyzed the correlation between the proposed prognostic model and cancer stem cells, drug sensitivity, and immune cell infiltration in CRC. We present this article in accordance with the TRIPOD reporting checklist (available at https://jgo.amegroups.com/article/view/10.21037/jgo-2024-985/rc).

Methods

Data collection

The DEGs between paracarcinoma tissues (n=51) and CRC tumor tissues (n=647) were identified and retrieved from TCGA database (https://portal.gdc.cancer.gov). The selection of DEGs was based on the criteria of $|\log_2$ fold change$| \geq 2$ and an adjusted P value (P.adj) <0.05. Additionally, prognostic factors associated with CRC (n=644) were identified using the criterion of P.adj <0.05. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Prognostic factor screening

We identified the intersection of DEGs and potential prognostic factors. Nonzero coefficients were subsequently screened using the least absolute shrinkage and selection operator (LASSO). These selected nonzero coefficients were then used in a Cox proportional hazards model, ultimately identifying six genes as independent prognostic factors for patients with CRC. These six genes were subsequently employed to construct both a nomogram model and a risk scale model. The purpose of this study is to preliminarily explore a prognostic risk model for colorectal cancer. We may focus more on internal data analysis rather than conducting external validation. In addition, due to resource limitations, we were unable to collect enough samples, and therefore external validation was not included in the study.

Survival analysis in patients with CRC

The risk scores based on six independent prognostic genes were used to evaluate the proportional risk hypotheses for overall survival (OS) and disease-specific survival (DSS)

via the "survival" package (version 3.3.1) in R (The R Foundation of Statistical Computing, Vienna, Austria) and to conduct fitted survival regression analyses. The outcomes of these analyses were subsequently visualized with the "survminer" (version 0.4.9) and "ggplot2" (version 3.3.6) R packages. The survival analysis was stratified into two cohorts according to the expression levels of the six genes and the corresponding risk scores, with the median expression value serving as the threshold to delineate high-expression (top 50%) and low-expression (bottom 50%) groups.

Restricted cubic spline (RCS)

A line graph was employed to illustrate the relationship between the hazard ratio (HR) and the variables within the proportional hazards model, specifically the Cox model, with the x-axis denoting the variables and the y-axis denoting the HR derived from the Cox regression analysis. The "survival" R package (version 3.3.1) was used for conducting the Cox regression analysis, the "rms" R package (version 6.4.0) was employed to construct an RCS model and perform correlation analysis, and the "ggplot2" R package (version 3.3.6) was used for data visualization. The RCS model was evaluated with the "plotRCS" R package (version 0.1.3).

Drug sensitivity analysis

The Genomics of Drug Sensitivity in Cancer (GDSC; https://www.cancerrxgene.org/) database was used to identify drugs and cell lines, facilitating the integration of transcriptome data from patients with CRC in TCGA to calculate spectral data and upload drug sensitivity ratings for the samples. By integrating the core algorithms of oncoPredict (version 0.2) and pRRophetic (version 0.5) R package [28,29], we predicted the specific drug sensitivity of the samples, thereby estimating the tolerance levels of the analyzed samples to the particular drugs. The sensitivity of the sample to the specific drug increases as the corresponding value decreases. Furthermore, Spearman correlation analysis was conducted to examine the relationship between the risk score and the drug sensitivity score.

Cancer stem cells

RNA-sequencing (RNA-seq) data and corresponding

80

Zhang et al. A prognostic model for CRC

clinical information for 620 CRC samples were obtained from TCGA dataset. The messenger RNA (mRNA) stem cell index (mRNAsi) was calculated using the one-class logistic regression (OCLR) algorithm developed by Malta *et al.* (30). All analyses and computations were conducted using R software version 4.0.3.

Immune cell infiltration

Single-sample gene set enrichment analysis (ssGSEA) and the ESTIMATE and CIBERSORT algorithms were used to measure the correlation of prognostic genes with immune cell infiltration in CRC. The Spearman correlation between prognostic molecules and immune infiltration was assessed, with the results being visualized through heatmaps. Utilizing the ssGSEA algorithm available in GSVA R package (version 1.46.0), we analyzed 24 immune cell markers to quantify the immune infiltration within the corresponding data from patients with CRC in TCGA database (31,32). The stromal score, immune score, and ESTIMATES score for the CRC patient data within TCGA were computed via the "estimate" R package (version 1.0.13) (31). The CIBERSORT algorithm, accessible via the CIBERSORTx platform (https://cibersortx.stanford.edu/), was employed to quantify immune cell infiltration via markers representative of 22 distinct immune cell types in accordance with previously established methodologies (33,34).

Statistical analysis

The data were expressed as the mean ± standard deviation. LASSO and Cox proportional hazards regression analyses were conducted to identify prognostic indicators via the "glmnet" package (version 4.1.7), "survival" package (version 3.3.1), and "rms" package (version 6.3-0) in R software version 4.2.1. Spearman correlation coefficient analysis was employed to evaluate the relationship between risk score and drug susceptibility. The DEGs in CRC were analyzed using the R packages "DESeq2" (version 1.36.0) and "edgeR" (version 3.38.2). Prognostic factors in CRC were identified with the "survival" R package (version 3.3.1). The Wilcoxon rank-sum test was employed to assess differences between the two groups via the "ggplot2" (version 3.3.6), "stats" (version 4.2.1), and "car" (version 3.1-0) R packages. Kaplan-Meier survival analysis was employed to assess the association of prognostic indicators with OS and DSS via the "survival" (version 3.3.1), "survminer" (version 0.4.9), and "ggplot2" (version 3.3.6) R packages. Nomogram and calibration

models were constructed and visualized using the R "rms" package (version 6.3-0), while the RCS model was evaluated with the "plotRC" R package (version 0.1.3). The selection of prognostic biomarkers was conducted using the Cox proportional hazards model; when the HR is greater than 1, it indicates that the variable is associated with an increased risk of the event occurring. When the HR is less than 1, it indicates that the variable is associated with a decreased risk of the event occurring. A P value with two-sided test of less than 0.05 was considered indicative of statistical significance.

Results

Screening of potential prognostic factors for CRC

To identify the potential prognostic genes for CRC, we conducted a differential expression analysis using paracancerous tissues (n=51) and CRC samples (n=647) from the TCGA database. This analysis yielded 2,154 DEGs, selected based on the criteria of $|\log_2$ fold change$|$ ≥2 and P.adj <0.05. Additionally, we identified 1,705 CRC-related prognostic factors, with the criterion of P.adj <0.05 being applied. From these, 152 potential prognostic genes were selected for subsequent studies (*Figure 1A*). We then employed LASSO regression analysis using cross-validation curves (*Figure 1B*) and LASSO coefficient path plots (*Figure 1C*) to refine the selection of nonzero coefficients. This analysis identified 39 nonzero coefficients. Further examination using Cox regression analysis revealed that six genes, including Niemann-Pick C1-like 1 (*NPC1L1*), glucagon-like peptide 2 receptor (*GLP2R*), solute carrier family 8 member A3 (*SLC8A3*), alpha-1-microglobulin/bikunin precursor (*AMBP*), single-pass membrane protein with coiled-coil domains 2 (*SMCO2*), and tetratricopeptide repeat domain 16 (*TTC16*) function as independent prognostic factors for CRC (*Figure 1D*). Simultaneously, a volcano plot was employed to illustrate the expression levels of these six prognostic genes. Compared to those of the paracarcinoma tissues, the expressions of *NPC1L1*, *AMBP*, *SMCO2*, and *TTC16* were markedly upregulated, whereas the expression levels of *GLP2R* and *SLC8A3* were significantly downregulated in CRC tumor tissues (*Figure 1E*).

Correlation of six prognostic genes with clinical parameters

We further used TCGA database to examine the associations between *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* with pathological stages I–IV (*Figure 2A*), T stage
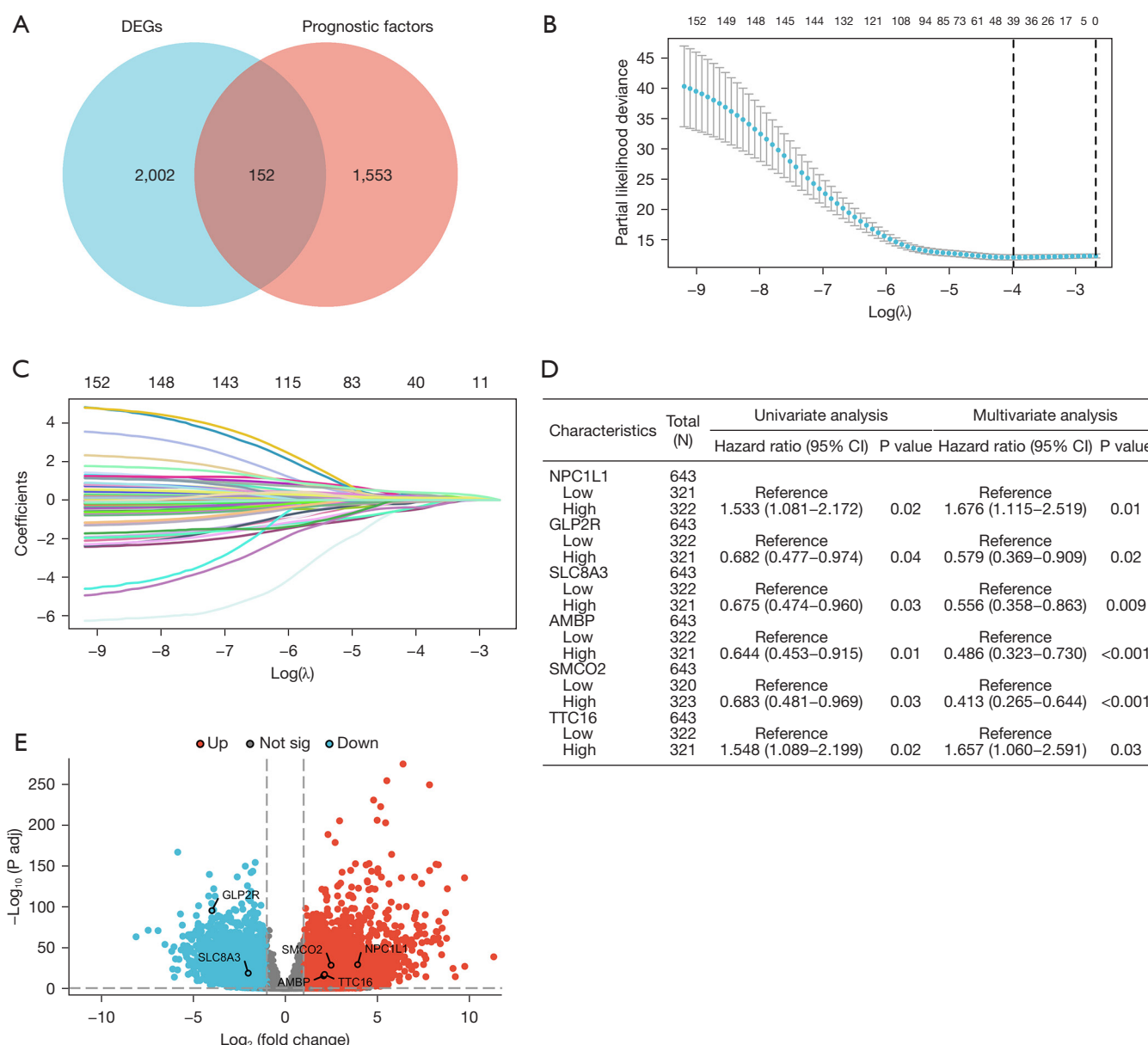
Figure 1 Screening of potential prognostic factors for CRC. Potential prognostic genes were selected using TCGA database (A). LASSO cross-validation curves (B), LASSO coefficient path plots (C), and Cox regression analysis (D) were used to select the independent prognostic genes for predicting the OS of patients with CRC. A volcano plot was employed to illustrate the expression levels of the six prognostic genes (E). DEG, differentially expressed gene; CI, confidence interval; sig, significance; CRC, colorectal cancer; TCGA, The Cancer Genome Atlas; LASSO, least absolute shrinkage and selection operator; OS, overall survival.

(*Figure 2B*), N stage (*Figure 2C*), M stage (*Figure 2D*), perineural invasion (*Figure 2E*), and CEA (*Figure 2F*). Our findings indicate that the majority of clinical parameters did not exhibit significant associations with the expression levels of these six prognostic genes, with only a few parameters showing significant differences in patients with CRC.

Correlation of six prognostic genes with OS and DSS in patients with CRC

As illustrated in *Figure 3A*, expression levels of *NPC1L1* and *TTC16* were significantly correlated with reduced OS as compared to the low expression of these genes. Conversely,
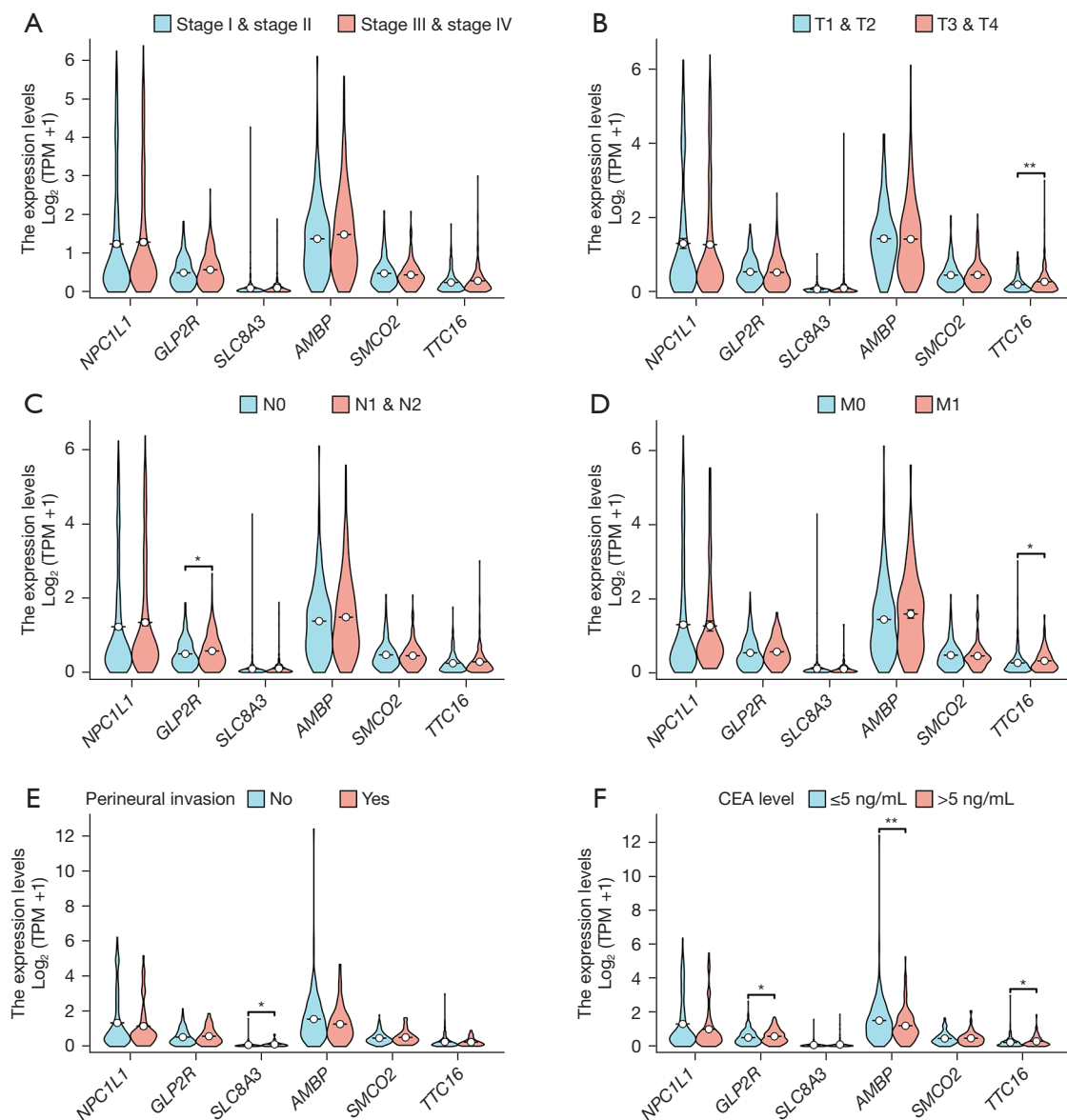
Figure 2 Correlation of six prognostic genes with clinical parameters. TCGA database was used to examine the associations between *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* with pathological stages I–IV (A), T stage (B), N stage (C), M stage (D), perineural invasion (E), and CEA (F) in patients with CRC. *, P<0.05; **, P<0.01. TPM, transcripts per million; *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; T, tumor; N, node; M, metastasis; CEA, carcinoembryonic antigen; TCGA, The Cancer Genome Atlas; CRC, colorectal cancer.

higher expression levels of *GLP2R*, *SLC8A3*, *AMBP*, and *SMCO2* were significantly associated with prolonged OS. *Figure 3B* shows that high expression of *NPC1L1* and *TTC16* was also significantly correlated with shorter DSS in patients with CRC. However, the expression levels of *GLP2R*, *SLC8A3*, *AMBP*, and *SMCO2* did not exhibit a

significant correlation with DSS.

A prognostic nomogram model was established to predict OS in patients with CRC

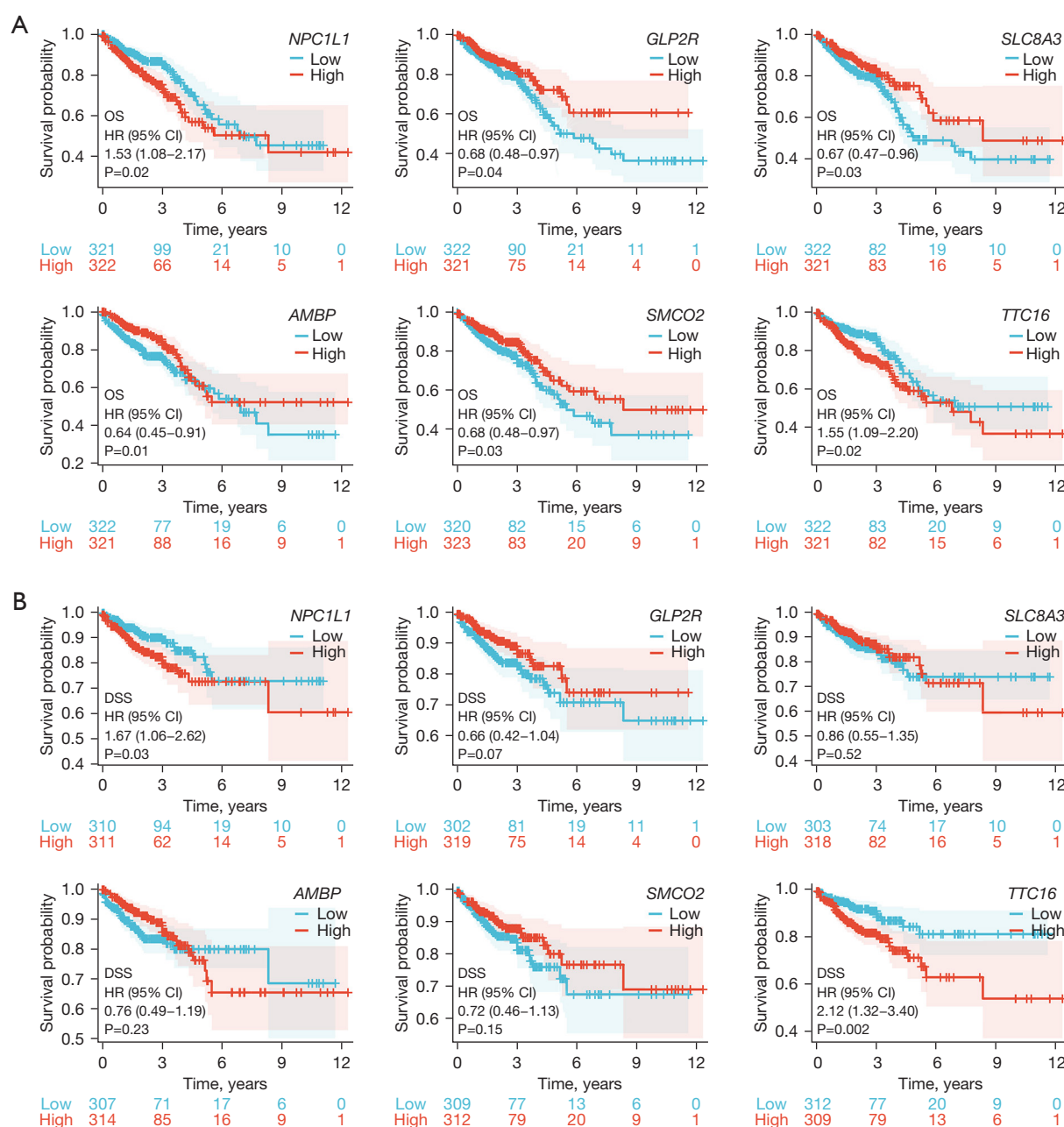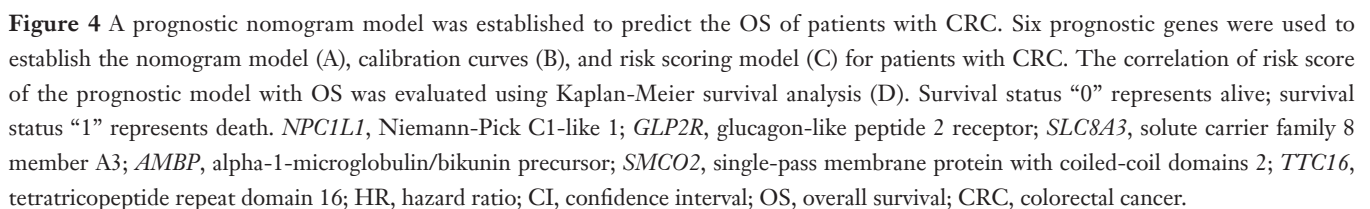We used six genes to construct a prognostic nomogram

Figure 3 Correlation of six prognostic genes with OS and DSS in patients with CRC. The correlation of six prognostic genes with OS (A) and DSS (B) was evaluated using Kaplan-Meier survival analysis. HR, hazard ratio; CI, confidence interval; *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; OS, overall survival; DSS, disease-specific survival; CRC, colorectal cancer.

aimed at predicting the OS of patients with CRC. By assigning weighted risk scores to these six genes, we calculated total points to estimate the 1-, 3-, and 5-year survival probabilities (*Figure 4A*). Prognostic calibration curves demonstrated the divergence between the predicted probabilities and the observed probabilities of the model at 1-, 3-, and 5-year intervals for the OS of patients with CRC. Statistical analyses indicated that our prognostic

Figure 4 A prognostic nomogram model was established to predict the OS of patients with CRC. Six prognostic genes were used to establish the nomogram model (A), calibration curves (B), and risk scoring model (C) for patients with CRC. The correlation of risk score of the prognostic model with OS was evaluated using Kaplan-Meier survival analysis (D). Survival status "0" represents alive; survival status "1" represents death. *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; HR, hazard ratio; CI, confidence interval; OS, overall survival; CRC, colorectal cancer.

model exhibited a good fit [concordance index =0.671; 95% confidence interval (CI): 0.645–0.698; likelihood ratio test =40.25, P<0.001; Wald test =40.69, P<0.001] (*Figure 4B*). Finally, we developed a prognostic risk scoring model incorporating six predictive genes (*Figure 4C*) and found that a lower risk score correlated with prolonged survival trend (HR =2.43; 95% CI: 1.67–3.53; P<0.001) (*Figure 4D*).

Nonlinear relationship between prognostic genes with the HR of OS

If a majority of binary variables are employed, the continuous variation between gene expression and HR of OS cannot be adequately captured. To address this limitation, we used the RCS model, which offers a more flexible fit across a spectrum of independent variables, thereby better representing the nonlinear relationship between gene expression and the HR for OS. Our findings indicate that *AMBP* and *TTC16* but not *NPC1L1*, *GLP2R*, *SLC8A3*, or *SMCO2* exhibited a nonlinear association with HR for OS. The elevated expression of *AMBP* was correlated with an increased risk of mortality, while the higher expression of *TTC16* was similarly correlated with an elevated risk of death in patients with CRC (*Figure 5*). Furthermore, our findings revealed a nonlinear relationship between the risk score derived from these six genes and OS. Specifically, an elevated risk score correlated significantly with an increased risk of mortality (*Figure 5*). According to the RCS curve, the risk score is categorized into three groups for selecting the best cut-off point. Ultimately, the optimal critical values for the risk score are –0.721 and –0.119. Compared to the group with a risk score ≤–0.721, the HR values for the groups with –0.721< risk score ≤–0.119 and risk score >–0.119 are 1.90 (95% CI: 1.13–3.18; P=0.02) and 3.22 (95% CI: 1.98–5.23; P<0.001), respectively.

Estimated correlation of prognostic genes with drug sensitivity in CRC

We used a prognostic risk score derived from six prognostic genes in conjunction with data from the GDSC database to evaluate the sensitivity of therapeutic agents associated with CRC. Our findings indicated a significant inverse association between elevated prognostic risk scores and sensitivity to various CRC treatment agents. Among these, uprosertib exhibited the most pronounced correlation (r=–0.462; P<0.001; *Figure 6*).

Correlation of prognostic gene expression with cancer stem cells in CRC

We employed a logistic regression machine learning algorithm (OCLR) to compute the mRNAsi. Our findings indicated that mRNAsi levels were significantly reduced in the high-expression groups of *GLP2R* and *SLC8A3* compared to the low-expression groups. Conversely, the expression levels of *NPC1L1*, *SMCO2*, and *TTC16* did not exhibit a correlation with mRNAsi (*Figure 7*). However, mRNAsi levels were significantly elevated in the high-expression group of *AMBP* compared to the low-expression group (*Figure 7*).

Correlation of prognostic genes with immune cell infiltration in CRC

To comprehensively investigate the correlation between NPC1L1, GLP2R, SLC8A3, AMBP, SMCO2, and *TTC16* and immune cell infiltration, we used the ssGSEA, ESTIMATE, and CIBERSORT algorithms to calculate their respective correlations. Specifically, the relationship between these prognostic genes and 24 types of immune cell enrichment scores was assessed using the ssGSEA algorithm. The findings revealed that *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* significantly correlated with the enrichment of various immune cells. Notably, *SLC8A3* exhibited a particularly strong association with multiple immune cell types (*Figure 8A* and Table S1). The ESTIMATE algorithm demonstrated comparable results, revealing a significant positive correlation between SLC8A3 and stromal score, immune score, and ESTIMATE score (*Figure 8B* and Table S2). Conversely, the CIBERSORT algorithm indicated a noticeable reduction in the correlation between *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* and the immune cell enrichment score (*Figure 8C* and Table S3).

Discussion

The prognosis of CRC involves a complex interaction of a variety of mechanisms. Currently, the CMS system for CRC uses extensive sequencing data and network cluster analysis to categorize CRC into five distinct CMS molecular subtypes. Patients with the CMS1 immunotype are particularly challenging to treat, exhibit a poor prognosis, and are frequently associated with *BRAF* gene mutations (35,36). Additionally, the molecular typing conducted by Nunes *et al.* (7), which integrates whole-genome deep
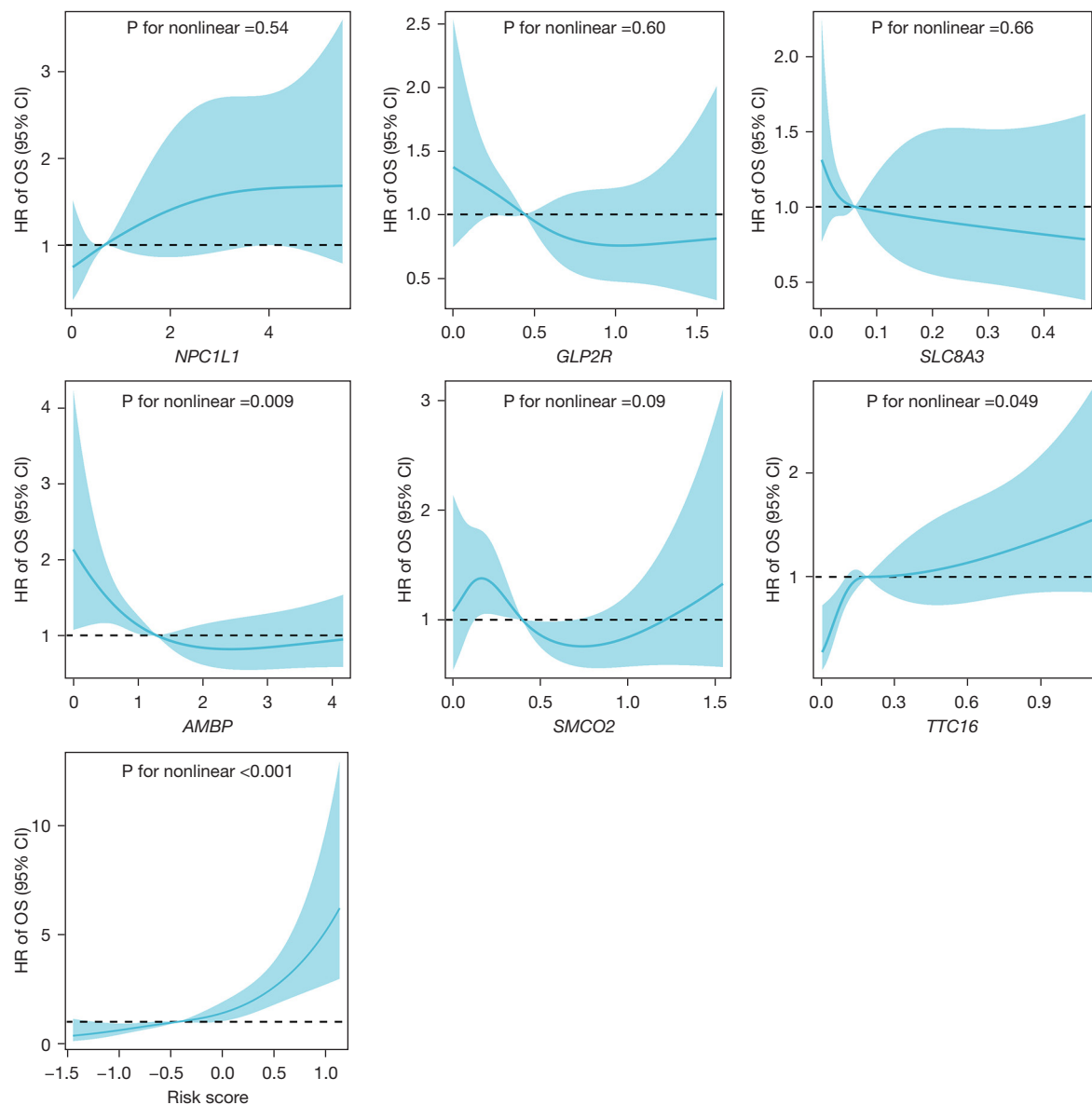
Figure 5 Nonlinear relationship between prognostic genes with the HR for OS. An RCS model was employed to evaluate the nonlinear relationship between prognostic genes (*NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16*) with the HR for OS. HR, hazard ratio; OS, overall survival; CI, confidence interval; *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; RCS, restricted cubic spline.

sequencing and transcriptome data, has further refined the prognostic classification of CRC into five prognostic subtypes. Our study introduced a prognostic model based on the TCGA database, incorporating six specific genes: *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* to develop a risk score model for CRC. Our study revealed a significant inverse correlation between the risk score and

OS in patients with CRC, indicating that higher risk scores are associated with poorer OS. Additionally, we identified associations between these genes and both the abundance of cancer stem cells and immune cell infiltration in CRC.

Cancer stem cells constitute a distinct subpopulation within tumors and are characterized by their capacity for self-renewal and their critical contribution to driving
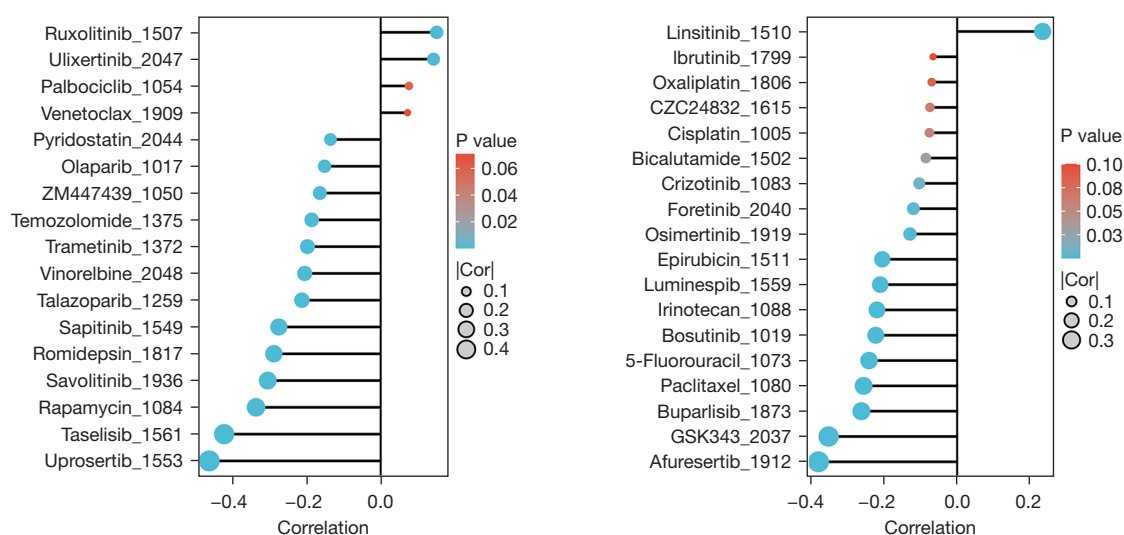
Figure 6 Estimated correlation of prognostic genes with drug sensitivity in CRC. The GDSC database was used to evaluate the correlation of the sensitivity of 35 drugs related to CRC treatment with prognostic risk scores. CRC, colorectal cancer; GDSC, Genomics of Drug Sensitivity in Cancer.
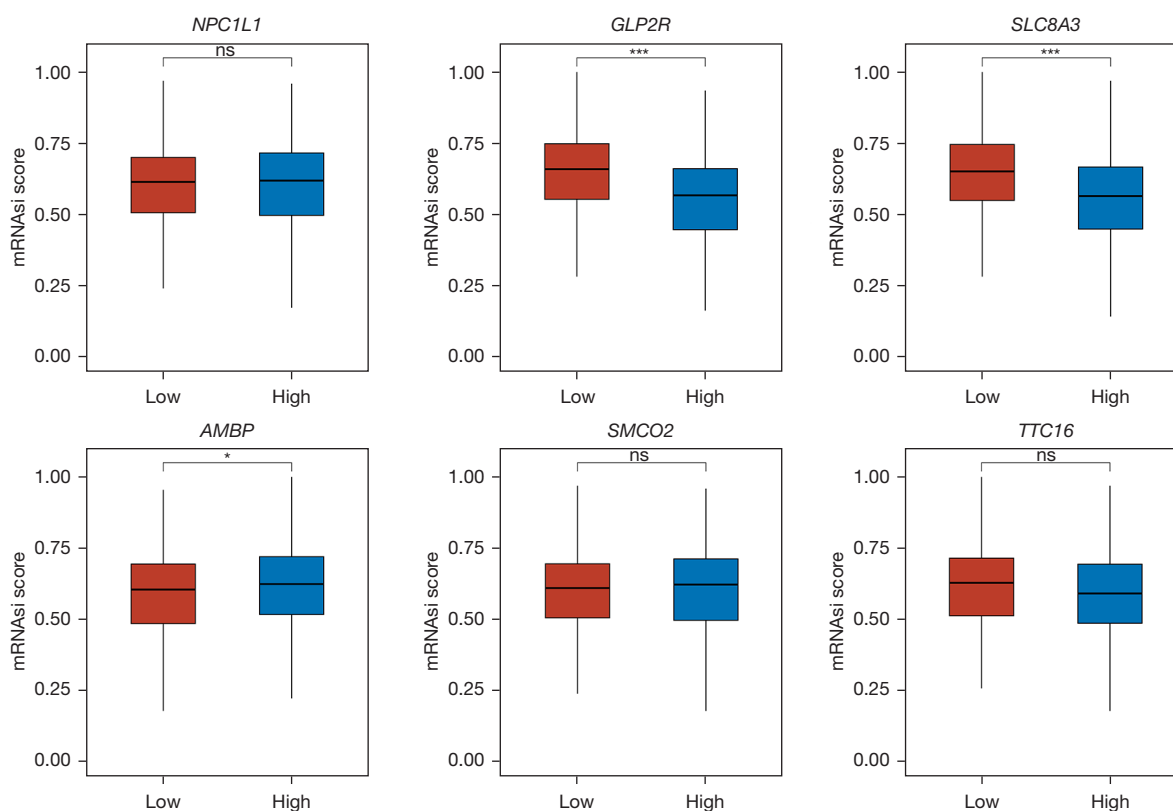

Figure 7 Correlation of prognostic gene expression with cancer stem cell abundance in CRC. OCLR was employed to evaluate the correlation of *NPC1L1*, *GLP2R*, *SLC8A3*, *AMBP*, *SMCO2*, and *TTC16* with cancer stem cell abundance in CRC. *, $P<0.05$; ***, $P<0.001$; ns, not significant. *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; CRC, colorectal cancer; OCLR, one-class logistic regression.

88

Zhang et al. A prognostic model for CRC
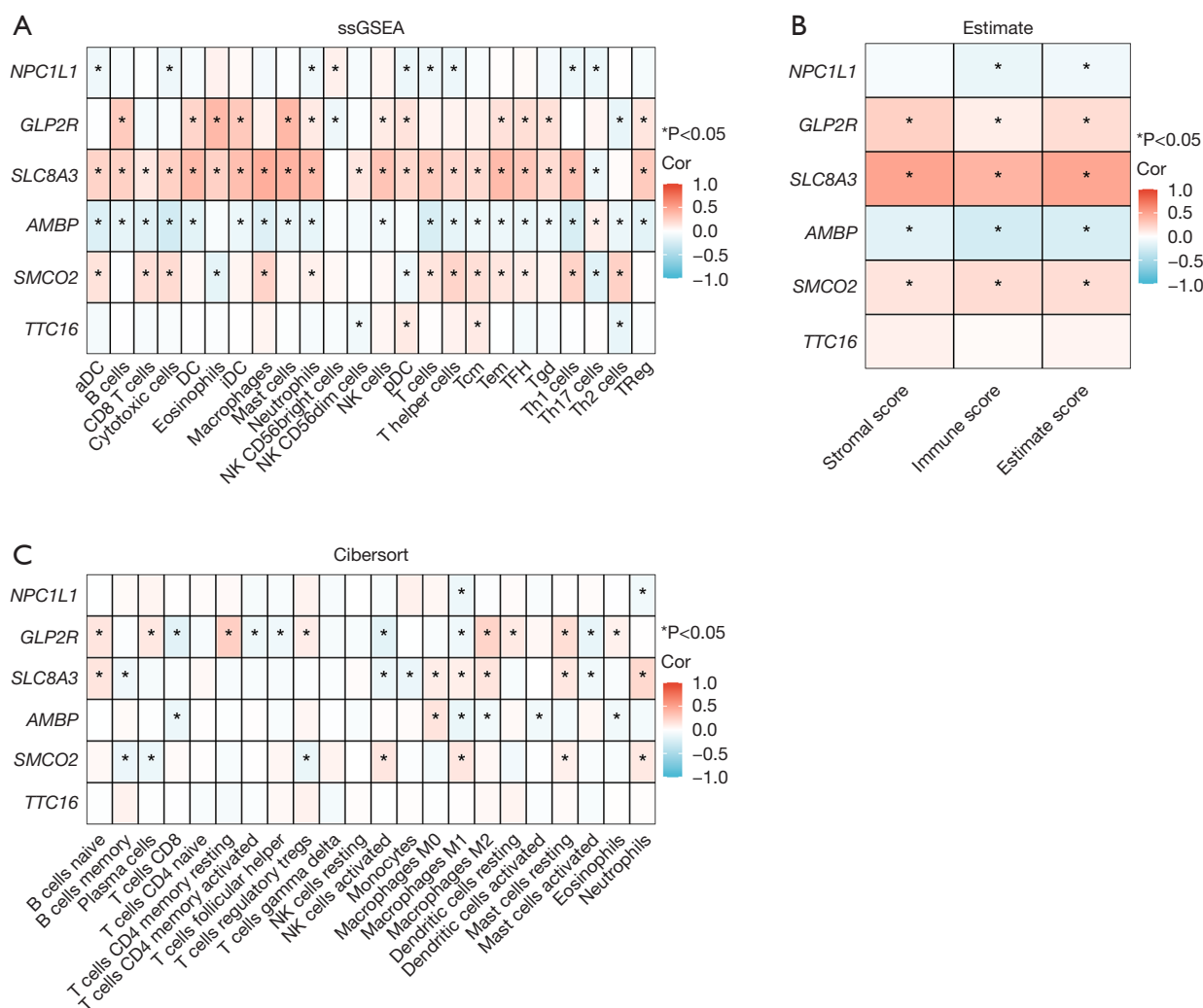
Figure 8 Correlation of prognostic genes with immune cell infiltration in CRC. ssGSEA (A) and the ESTIMATE (B) and CIBERSORT (C) algorithms were used to measure the correlation of prognostic genes with immune cell infiltration in CRC. ssGSEA, single-sample gene set enrichment analysis; Cor, correlation; aDC, activated dendritic cell; DC, dendritic cell; iDC, immature dendritic cell; NK, natural killer; pDC, plasmacytoid dendritic cell; Tcm, T central memory; Tem, T effector memory; TFH, T follicular helper; Tgd, T gamma delta; *NPC1L1*, Niemann-Pick C1-like 1; *GLP2R*, glucagon-like peptide 2 receptor; *SLC8A3*, solute carrier family 8 member A3; *AMBP*, alpha-1-microglobulin/bikunin precursor; *SMCO2*, single-pass membrane protein with coiled-coil domains 2; *TTC16*, tetratricopeptide repeat domain 16; CRC, colorectal cancer.

tumor heterogeneity and recurrence (37,38). The cancer-stem-cell state exemplifies the mimicking behavior of cancer cells, culminating in phenotypic plasticity. This mimicry enables cancer stem cells to emulate immune cells, vascular endothelial cells, or lymphangiogenic cells, thereby facilitating tumor progression (37). Furthermore, cancer stem cells induce the adoption of pro-tumor phenotypes by tumor-associated immune cells and stromal cells through paracrine and juxtacrine signaling mechanisms, ultimately

contributing to therapeutic resistance (37). We found that elevated expression levels of *GLP2R* and *SLC8A3* in patients with CRC correlated with a significantly reduced mRNAsi score. Furthermore, high expression of *GLP2R* and *SLC8A3* was associated with better OS in patients with CRC. These observations imply that the improved prognosis associated with *GLP2R* and *SLC8A3* expression may be linked to a lower mRNAsi score in patients with CRC. Jin *et al.* (39) identified the *SLC8A3* gene as being implicated in calcium

extrusion—from which a prognostic model for colon adenocarcinoma was developed—with the overexpression of SLC8A3 markedly suppressing the proliferation and migration of CRC rhabdomyosarcoma kidney origin (RKO) cells. These findings indicate that elevated expression of *SLC8A3* is associated with a more favorable prognosis and that *SLC8A3* exhibits antitumor properties. This finding was confirmed in our investigation, revealing a significant positive correlation between *SLC8A3* expression and various immune cell populations, particularly macrophages (r=0.439; P<0.001). More so, *SLC8A3* exhibited a statistically significant correlation with the stromal score (r=0.498; P<0.001), immune score (r=0.484; P<0.001), and ESTIMATE score (r=0.408; P<0.001) in patients with CRC.

Most existing sequence-based prognostic models for CRC employ a binary framework (40,41), whereby an individual's predicted risk remains static once the model is established. This static nature constitutes a relevant limitation as such models fail to account for the relationship between continuous variations in risk scores and prognosis. This flaw was considered in our study, where an RCS model was used to evaluate the significant nonlinear relationship between continuous changes in risk scores derived from six prognostic genes, and the adverse outcomes in patients with CRC. Our findings indicate that an elevated risk score is associated with an increased likelihood of poor prognosis in patients with CRC.

Our study was subject to some limitations. First, data collection was restricted to sequencing data from TCGA. No external datasets [such as Gene Expression Omnibus (GEO), International Cancer Genome Consortium (ICGC)] were employed for supplementary validation, which does not entirely guarantee the stability and reliability of the model. In addition, and we will further validate the reliability of the model in future research through single-center and multi-center sample sizes. Exclusion of other multiomics datasets may have limited the comprehensiveness of marker selection. Second, the functional roles of the genes identified in our study have not been validated. Therefore, the findings from our study should be interpreted with caution. Finally, the reliance on a single sequencing dataset for model prediction hinders the applicability of our model to dynamically forecast the prognosis of all patients with CRC.

Conclusions

Overall, we have developed a prognostic assessment model

for patients with CRC based on six genes. A higher risk score was significantly correlated with poorer OS. In addition, the analysis revealed a nonlinear relationship between the continuous variation in risk score and the risk of poor OS, indicating a continuous dose-effect relationship between risk score and OS. Additionally, specific genes included in the prognostic model were found to be strongly associated with cancer stem cell and immune cell infiltration in CRC, suggesting that the poor prognosis observed in patients with CRC may be influenced by these factors.

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74:229-63.

2. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 2024;4:47-53.

3. Mi X, Yao H, Lu Y, et al. Leptin increases chemosensitivity by inhibiting CPT1B in colorectal cancer cells. J Gastrointest Oncol 2024;15:2507-20.

4. Cervantes A, Adam R, Roselló S, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2023;34:10-32.

5. Eng C, Yoshino T, Ruíz-García E, et al. Colorectal cancer. Lancet 2024;404:294-310.

6. Jácome AA, Mathias-Machado MC, Gil M, et al. Later lines of systemic therapy in patients with metastatic colorectal cancer: real-world data from a setting with barriers to access cancer therapies. J Gastrointest Oncol 2024;15:2543-51.

7. Nunes L, Li F, Wu M, et al. Prognostic genome and transcriptome signatures in colorectal cancers. Nature 2024;633:137-46.

8. Xu Y, Liu K, Li C, et al. Microsatellite instability in mismatch repair proficient colorectal cancer: clinical features and underlying molecular mechanisms. EBioMedicine 2024;103:105142.

9. Heiser CN, Simmons AJ, Revetta F, et al. Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors. Cell 2023;186:5620-5637.e16.

10. Wu J, Li W, Su J, et al. Integration of single-cell sequencing and bulk RNA-seq to identify and develop a prognostic signature related to colorectal cancer stem cells. Sci Rep 2024;14:12270.

11. Glisan A, Nielsen E, Billion T, et al. Regional trends in colorectal cancer mortality in people aged 45-84 years in the US, 1999-2022. J Gastrointest Oncol 2024;15:2533-42.

12. Luo B, Liao M, Nie B, et al. Genomic profiles and their associations with microsatellite instability status, tumor mutational burden, and programmed death ligand 1 expression in Chinese patients with colorectal cancer. J Gastrointest Oncol 2024;15:2460-72.

13. Zhu J, Zhang J, Lou Y, et al. Developing a machine learning-based prognosis and immunotherapeutic response signature in colorectal cancer: insights from ferroptosis, fatty acid dynamics, and the tumor microenvironment. Front Immunol 2024;15:1416443.

14. Liu Y, Zhao Y, Zhang S, et al. Developing a prognosis and chemotherapy evaluating model for colon adenocarcinoma based on mitotic catastrophe-related genes. Sci Rep 2024;14:1655.

15. Li L, Li Y, Lin J, et al. A Pyroptosis-Related Gene Signature Predicts Prognosis and Tumor Immune Microenvironment in Colorectal Cancer. Technol Cancer Res Treat 2024;23:15330338241277584.

16. Huang H, Chen K, Zhu Y, et al. A multi-dimensional approach to unravel the intricacies of lactylation related signature for prognostic and therapeutic insight in colorectal cancer. J Transl Med 2024;22:211.

17. Huang A, Sun Z, Hong H, et al. Novel hypoxia- and lactate metabolism-related molecular subtyping and prognostic signature for colorectal cancer. J Transl Med 2024;22:587.

18. Xu Y, Qu H, Liang R, et al. A multi-gene blood-based methylation assay for early diagnosis of colorectal cancer. Transl Cancer Res 2024;13:6699-708.

19. Deng S, Zhu Q, Chen H, et al. Screening of prognosis-related Immune cells and prognostic predictors in Colorectal Cancer Patients. BMC Cancer 2023;23:195.

20. Pang L, Wang Q, Wang L, et al. Development and validation of cuproptosis-related lncRNA signatures for prognosis prediction in colorectal cancer. BMC Med Genomics 2023;16:58.

21. Li Z, Li C, Pu H, et al. Trajectories of perioperative serum carcinoembryonic antigen and colorectal cancer outcome: A retrospective, multicenter longitudinal cohort study. Clin Transl Med 2021;11:e293.

22. Li C, Zhang D, Pang X, et al. Trajectories of Perioperative Serum Tumor Markers and Colorectal Cancer Outcomes: A Retrospective, Multicenter Longitudinal Cohort Study. EBioMedicine 2021;74:103706.

23. Li C, Zhao K, Zhang D, et al. Prediction models of colorectal cancer prognosis incorporating perioperative longitudinal serum tumor markers: a retrospective longitudinal cohort study. BMC Med 2023;21:63.

24. Takei S, Tanaka Y, Lin YT, et al. Multiomic molecular characterization of the response to combination immunotherapy in MSS/pMMR metastatic colorectal cancer. J Immunother Cancer 2024;12:e008210.

25. van de Weerd S, Torang A, Zwager LW, et al. Consensus molecular subtype transition during progression of colorectal cancer. J Pathol 2023;261:298-308.

26. Wang W, Kandimalla R, Huang H, et al. Molecular

subtyping of colorectal cancer: Recent progress, new challenges and emerging opportunities. Semin Cancer Biol 2019;55:37-52.

27. Valdeolivas A, Amberg B, Giroud N, et al. Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics. NPJ Precis Oncol 2024;8:10.

28. 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.

29. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol 2014;15:R47.

30. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 2018;173:338-354.e15.

31. 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.

32. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 2013;39:782-95.

33. Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 2018;1711:243-59.

34. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.

35. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350-6.

36. Rejali L, Seifollahi Asl R, Sanjabi F, et al. Principles of Molecular Utility for CMS Classification in Colorectal Cancer Management. Cancers (Basel) 2023;15:2746.

37. Saw PE, Liu Q, Wong PP, et al. Cancer stem cell mimicry for immune evasion and therapeutic resistance. Cell Stem Cell 2024;31:1101-12.

38. Sarabia-Sánchez MA, Tinajero-Rodríguez JM, Ortiz-Sánchez E, et al. Cancer Stem Cell markers: Symphonic masters of chemoresistance and immune evasion. Life Sci 2024;355:123015.

39. Jin M, Yin C, Yang J, et al. Identification and validation of calcium extrusion-related genes prognostic signature in colon adenocarcinoma. PeerJ 2024;12:e17582.

40. Chen L, Ge M, Mo S, et al. Construction of a New Ferroptosis-related Prognosis Model for Survival Prediction in Colorectal Cancer. Curr Med Chem 2024. [Epub ahead of print]. doi: 10.2174/0109298673296767240116215814.

41. Li Q, Liu H, Jin Y, et al. Analysis of a new therapeutic target and construction of a prognostic model for breast cancer based on ferroptosis genes. Comput Biol Med 2023;165:107370.

(English Language Editor: J. Gray)