ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Predictive value of an eight-mRNA signature in colon adenocarcinoma prognosis

Yuying Yang[1†], Cuiying Wang[2†], Hongqian Wei[1], Bing Zhou[1], Songtao Hou[1], Xiaochen Pang[1], Wenhai Dong[1], and Zhongqiu Chai[1*]

[1]Department of Anal Medicine, Binhai New Area Hospital of Traditional Chinese Medicine, Fourth Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, China

[2]Department of Geriatrics, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China

Abstract

Colorectal cancer is a prevalent malignancy, with colon adenocarcinoma as the most common type. Early diagnosis biomarkers and effective risk stratification are crucial for optimal treatment. In this study, gene expression data from the Cancer Genome Atlas and Gene Expression Omnibus (GEO) were analyzed to identify relevant genes for colon adenocarcinoma. These datasets were standardized and subjected to weighted gene co-expression network analysis and differentially expressed gene analysis. Univariate Cox regression and least absolute shrinkage and selection operator Cox regression analyses were performed to generate a risk profile and identify prognosis-related genes. Receiver operating characteristic (ROC) analysis, Kaplan–Meier (KM) curve, and Cox analyses validated the risk signature. Immune cell infiltration patterns and immunological activities in high- and low-risk groups were assessed using single-sample gene set enrichment analysis (ssGSEA). GSEA was used to investigate the signaling pathways associated with low-risk and high-risk groups, whereas ssGSEA was used to analyze those associated with high-risk groups. A line graph was created to predict the overall survival (OS) of patients. Quantitative real-time polymerase chain reaction confirmed differential gene expression between normal and cancerous colon tissues. The eight genes identified – *ACOX1*, *ATP8B1*, *CHGA*, *NAT2*, *PKIB*, *SLC39A8*, *TINAG*, and *VEGFA* – correlated with tumor immunity and clinical outcomes. This eight-gene risk profile can accurately stratify risk and predict OS based on KM curves, ROC analysis, and regression models. GSEA analysis revealed calcium ion metabolism as the top pathway in the GEO dataset. These findings provide a foundation for prognostic evaluation and may guide therapeutic decision-making in colon adenocarcinoma.

Keywords: Colon adenocarcinoma; Weighted gene co-expression network analysis; Gene set enrichment analysis; Prognosis

†*These authors contributed equally to this work*

***Corresponding author:**
Zhongqiu Chai
(chaizhongqiu@126.com)

Citation: Yang Y, Wang C, Wei H, *et al*. Predictive value of an eight-mRNA signature in colon adenocarcinoma prognosis. *Eurasian J Med Oncol.* 2025;9(2):234-249. doi: 10.36922/EJMO025060024

Received: February 5, 2025

Revised: April 1, 2025

Accepted: April 8, 2025

Published online: May 9, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

According to the American Cancer Society, colorectal cancer accounted for over one in 10 cancer cases and fatalities in 2020, with over 1.9 million new diagnoses (including anal) and 935,000 deaths globally. In general, colorectal cancer ranks second in terms of cancer-associated mortality and third in terms of incidence.[1] The most prevalent

pathological form of colon cancer is colon adenocarcinoma (COAD).[2] Oncogene activation and inactivation are two examples of the multistage, multigene processes involved in the formation of COAD.[2,3] Treatment of tumors requires early detection and timely aggressive excision. Patients with colorectal cancer have a 5-year survival rate of almost 90% if they receive an early diagnosis, compared to only 12% for those who have distant metastases.[4] A 2020 study on colorectal cancer incidence revealed a declining trend in highly developed nations, possibly due to improved disease screening.[1,5] Identifying suitable biomarkers is crucial for risk assessment, early diagnosis, and treatment outcome prediction. Risk stratification in COAD can be enhanced using gene expression profiles.[6] However, many studies have not thoroughly examined the genes associated with the clinical prognosis of COAD patients. Prognostic biomarkers can significantly influence the risk classification of COAD patients,[7] with high-risk groups benefiting from more intensive treatment to prevent undertreatment. In contrast, low-risk groups should receive low-intensity treatment regimens to avoid overtreatment. Hence, to assess the clinical prognosis of COAD patients, we evaluated gene modules and identified potential biomarkers.

The relationship between COAD and the genome was evaluated using weighted gene co-expression network analysis.[8] Furthermore, the least absolute shrinkage and selection operator (LASSO) Cox regression and univariate proportional hazards analyses were used to identify an mRNA signature closely linked to the prognosis of COAD, surpassing clinical criteria. Ultimately, the gene signature and clinical characteristics were combined to create a nomogram to predict prognosis.

2. Materials and methods

2.1. Data retrieval process

All bioinformatics and statistical analyses were performed using R software (version 4.2.0). Raw data were obtained from the Cancer Genome Atlas Program (TCGA) (https://portal.gdc.cancer.gov/), which included 647 colorectal cancer patients with complete survival information. From the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database, the GSE39582 dataset[9] was used for validation.[10] The TCGA dataset contained 647 tumor samples and 51 normal samples, whereas the GSE39582 dataset included 556 colorectal cancer samples.

2.2. Co-expression network construction

In this study, intramodular hub genes were selected using the criterion of gene significance>0.2 and module membership>0.8. Weighted gene co-expression networks were developed using weighted gene co-expression network analysis, and network topology was analyzed with a soft-threshold power ranging from 1 to 30. The relational matrix was then converted into the adjacent matrix and transformed into a topological overlap matrix for average link hierarchy clustering to classify related modules. The correlation between modules and clinical traits was assessed using Pearson methods.

2.3. Modeling-associated genes identification and prognostic signature construction

The LASSO regression analysis was performed on the survival-related genes to avoid the collinearity of high-dimensional transcriptome data to identify modeling genes. The value of this mRNA signature for the prognosis of colorectal patients was verified using the Kaplan–Meier (KM) curve and log-rank test. The modeling data set's core genes were identified using univariate Cox regression analysis with $p<0.05$ as the threshold. Multi-factor Cox regression analysis was used to analyze the risk score, which was then used to classify patients into high- and low-risk groups.

2.4. Predictive accuracy of the risk signature

The eight-mRNA signature's predictive significance was evaluated through internal and external validation.

2.5. Protein expression levels and genetic alterations of prognostic genes validation

The cBioPortal database (https://www.cbioportal.org/) was used to investigate the mutation status of key genes in the mRNA signature, whereas the Human Protein Atlas database[11] (https://www.proteinatlas.org/) was utilized to verify the protein expression levels of these genes through immunohistochemistry.

2.6. Immune landscape differences of colon adenocarcinoma patients across risk strata

The single-sample gene set enrichment analysis (ssGSEA) algorithm was employed to ascertain the impact of high- and low-risk groups on immune cell infiltration and immune function in the TCGA and GEO cohorts.

2.7. Univariate and multi-factor Cox regression analysis

Cox regression analyses, both univariate and multivariate, were conducted on age, sex, stage, TNM staging, and other clinical factors in the TCGA and GEO cohorts. This analysis was conducted to confirm the mRNA risk signature's independent predictive value. Univariate and multivariate Cox regression, KM curves, and log-rank tests were performed using the R package survival (version 3.3 1) and "survminer" (version 0.4.9).

2.8. Association between the risk signature, pathological clinical features, and chemotherapeutic responses

We examined whether the mRNA risk characteristics could predict clinical responses of colorectal cancer patients to commonly used chemotherapeutic agents by analyzing the risk differences associated with clinical information and various pathological characteristics in the TCGA and GEO datasets.

2.9. Nomogram development and validation

Multivariable survival analysis and nomograms for predicting the overall survival (OS) of every patient were established using the Cox regression model and created using the R package "rms" (version 6.3-0).

2.10. Cell culture

We obtained normal human colonic epithelial cells, HCT-116 cell lines, and LoVo cell lines from the Peking Union Cell Resource Center (China). All cells were cultivated in whole culture with 10% fetal bovine serum (ExCell, South America) and 1% double antibody at 37°C in a humidified incubator with 5% carbon dioxide. Human colonic epithelial cells were cultured in Dulbecco's Modified Eagle Medium, a high-sugar medium (Gibco, Thermo Fisher Scientific Biochemical Products, China), whereas LoVo and HCT-116 cells were cultured in Roswell Park Memorial Institute-1640 media.

2.11. RNA extraction and quantitative real-time polymerase chain reaction analysis

Total RNA was extracted from LoVo cells, HCT-116 cells, and human colonic epithelial cells using the MolPure® Cell/Tissue Total RNA Kit (Yisheng Biotechnology Co., China), following the manufacturer's instructions. Reverse transcription of isolated RNA to complementary DNA (cDNA) was performed using the Hifair® III 1st Strand cDNA Synthesis Kit (Yisheng Biotechnology Co., China). The SYBR Green polymerase chain reaction kit was used to amplify the resulting cDNA, and *ACTB* was used as an internal reference gene. Relative mRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The primer sequences are shown in Table 1.

3. Results

3.1. Preprocessing of the dataset

The Surrogate variable analysis package was used to convert the microarray data into an expression matrix and perform batch corrections. Using the classification and regression training package, 647 patients with full survival data from the TCGA dataset were divided into two groups in a 6:4 ratio: 372 cases for the modeling set and 246 cases for the internal validation set. Furthermore, the GSE39582 dataset, consisting of 556 individuals with complete survival information, was used as an external validation set.

3.2. Construction of the weighted gene co-expression network

The TCGA dataset consisted of 647 tumor samples and 51 normal samples, with clinical data used in constructing a weighted gene co-expression network. No samples were excluded, and sample clustering was satisfactory (Figure 1A). The optimal soft threshold was found to be seven using topological computation (Figure 1B). Gene modules were categorized using the topological overlap matrix approach, with each module containing at least 50 genes. The gene module height for this analysis was set to 0.75 (Figure 1C). Pearson's correlation was used to assess the association between clinical characteristics and modules. Among the five identified modules, the blue module (r=0.84/−0.84, $p<0.05$) was the core module (Figure 1D). This module comprised 349 genes strongly associated with colorectal cancer progression.

Table 1. Primer sequences of eight genes

Gene	Forward primer (5'–3')	Reverse (5'–3')
ATP8B1	AGAACCACCACACTCAATGAAC	GAGCAAGAAGAAGAACTGTCGTA
SLC39A8	CCAGAGATGAATGATATGCTGAGAG	AATGAGTAGAATGGCTGTGAATCC
TINAG	GAATGCGTTGTGCTACTGTGA	GCTGTCCATCCATAGTCTCCTT
CHGA	CCAAGACCTCGCTCTCCAA	CTGGCTGCTCTGGTTCTCA
NAT2	GGTGTCTCCAGGTCAATCAAC	GAACCATGCCAGTGCTGTATT
VEGFA	GCGGATCAAACCTCACCAAG	GCTCCAGGGCATTAGACAGC
ACOX1	AAATTTTGTGCACCGAGGGC	CTGTCTGGGCATAAGTGCCA
ACTB	AGCGAGCATCCCCCAAAGTT	GGGCACGAAGGCTCATCATT
PKIB	CATCTTCAGCAAGGGCAG	CATCTTCCTTCACGGAGAG

3.3. Identification of patient risk signatures in the modeling sets

Through univariate Cox regression analysis, 12 genes from the 349 genes in the blue module (Figure A1A) were found to be significantly associated with survival (p<0.05) in the modeling dataset. These genes were subjected to LASSO regression analysis (Figure A1B and C) to determine the final eight genes involved in modeling. Multivariate Cox regression analysis (Figure A1D) was performed for the above eight genes, in which the coefficient value of each gene was involved in the construction of the risk score (Table 2). The formula for risk score is given in Equation I.

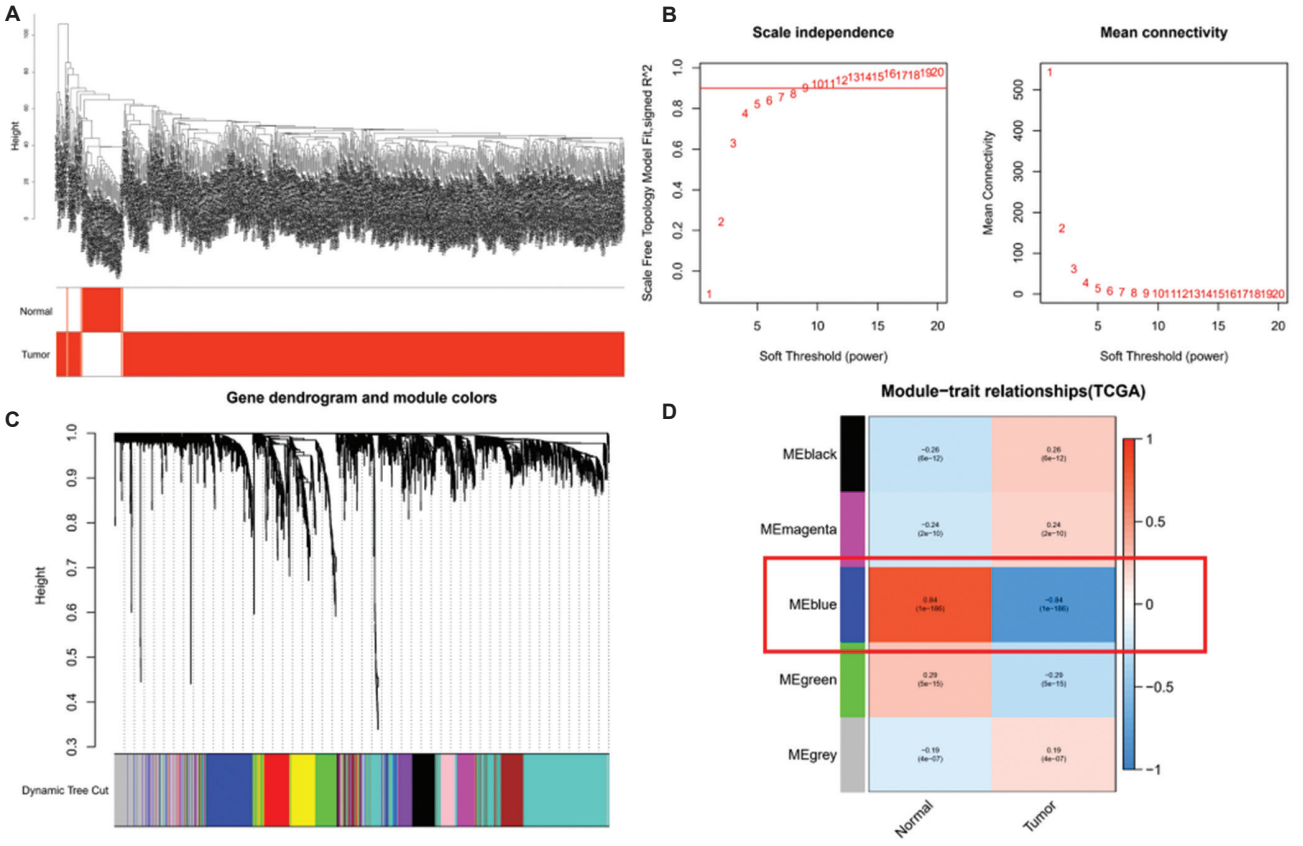


Figure 1. COAD candidate gene identification. (A) COAD and conventional control clustering dendrogram. (B) Soft thresholds in topological computations identified the optimal Soft Threshold of seven. (C) Gene clustering on a dissimilarity measure in a dendrogram. (D) A heat map showing the connections between traits and modules.
Abbreviations: COAD: Colon adenocarcinoma; ME: Module eigengene; TCGA: The cancer genome atlas program.

Table 2. Multifactorial Cox regression analysis of mRNAs

Gene	Coefficient	Hazard ratio	Hazard ratio (95% lower confidence interval)	Hazard ratio (95% higher confidence interval)	p-value
ACOX1	−0.0628	0.9391	0.8589	1.0268	0.1679
ATP8B1	−0.0384	0.9624	0.9324	0.9933	0.0174
CHGA	0.0079	1.0079	1.0026	1.0132	0.0032
NAT2	0.0100	1.0101	0.9118	1.1189	0.8480
PKIB	0.1356	1.1453	1.0902	1.2031	0.0000
SLC39A8	−0.0359	0.9647	0.9218	1.0096	0.1214
TINAG	−0.1228	0.8845	0.7763	1.0078	0.0652
VEGFA	0.0558	1.0574	1.0187	1.0974	0.0033

Risk score = (−0.0628×*ACOX1*) + (−0.0384×*ATP8B1*) + (0.0079×*CHGA*) + (0.0100×*NAT2*) + (0.1356×*PKIB*) + (−0.0359×*SLC39A8*) + (−0.1227×*TINAG*) + (0.0558×*VEGFA*) (I)

Patients with colorectal cancer in the modeling set were categorized into two groups: High-risk (n=186) and low-risk (n=186), based on the median risk score of 0.9629. The high-risk group's survival time was significantly shorter than the low-risk group, according to the KM curve and log-rank test (p<0.001; Figure 2A). Using this risk score, the areas under the curve of the modeling group that were used to estimate the likelihood of survival after 1, 3, and 5 years were 0.708, 0.720, and 0.684, respectively (Figure 2B). Furthermore, among these eight mRNAs, *ACOX1*, *APT8B1*, *NAT2*, *SLC39A8*, and *TINAG* were overexpressed in low-risk patients (p<0.05), whereas *VEGFA* was overexpressed in high-risk patients (Figure 2C).

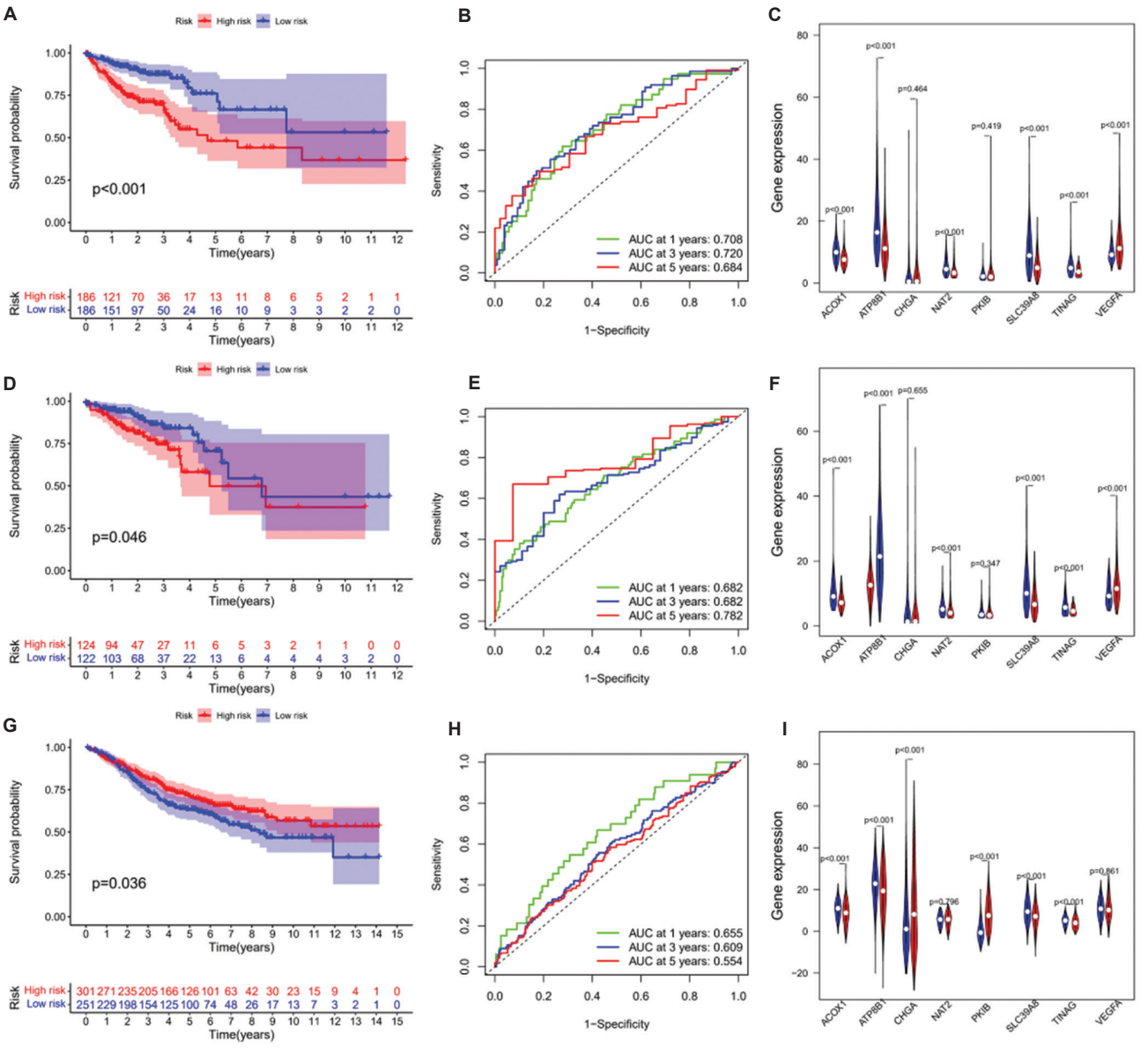


Figure 2. The predictive and prognostic value of the mRNA signature. (A) Survival curves for colon adenocarcinoma patients in the internal test group, (D) external test cohort, and (G) Kaplan–Meier training group. The eight-mRNA signature's time-dependent receiver operating characteristic curves for (B) training, (E) testing, and (H) the entire cohort were 1, 3, and 5 years. (C, F, and I) Violin plots of *ACOX1*, *ATP8B1*, *CHGA*, *NAT2*, *PKIB*, *SLC39A8*, *TINAG*, and *VEGFA* gene expression levels in various risk categories.
Abbreviation: AUC: Area under the curve.

3.4. Validation of the prognostic and predictive accuracy of the eight-mRNA risk signature

Out of the 246 patients, 124 were placed in the high-risk group and 122 in the low-risk group based on the model set's median risk score. The KM curve showed that the high-risk group's survival time was significantly shorter than the low-risk group (p<0.05; Figure 2D). Using this risk score, the areas under the curve for estimating the likelihood of survival at 1, 3, and 5 years were 0.682, 0.682, and 0.782, respectively (Figure 2E). With *VEGFA* overexpression in high-risk patients and *ACOX1*, *APT8B1*, *NAT2*, *SLC39A8*, and *TINAG* overexpression in low-risk patients, the internal validation set in the eight-mRNA expression studies resembled the modeling set (Figure 2F). Furthermore, it is important to note that, out of the 556 patients in the external validation cohort, 274 were high-risk and 282 were low-risk, according to the model set's median risk score. Based on the KM curve, it was observed that the high-risk group's survival time was significantly shorter than the low-risk group (p<0.05; Figure 2G). The areas under the curve for estimating the probability of survival after 1, 3, and 5 years were 0.655, 0.609, and 0.554, respectively (Figure 2H), after employing the risk score. *PKIB* and *CHGA* were overexpressed in high-risk patients in the external validation queue, whereas *VEGFA*, *TINAG*, *SLC38A8*, *ATP8B1*, and *ACOX1* were overexpressed in low-risk patients (Figure 2I). In addition, the effect of the expression of these eight mRNAs on the total TCGA cohort's prognosis for patients with colorectal cancer was analyzed. The cut-off value for these mRNAs was their median expression level. While *PKIB8* (Figure 3E), *VEGFA* (Figure 3F), and *TINAG*

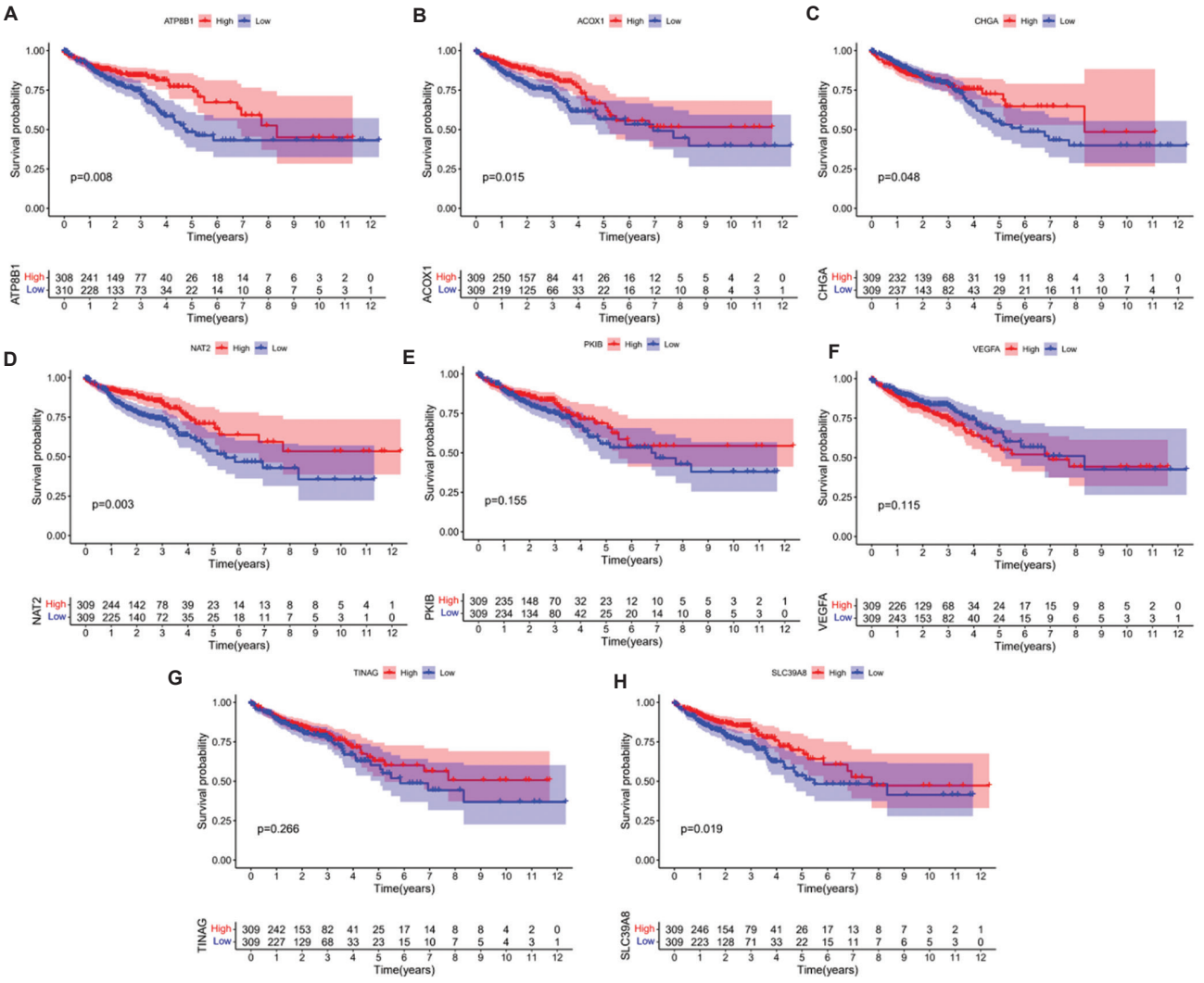


Figure 3. Survival analysis of the overall cohorts of the cancer genome atlas. (A) *ATP8B1*, (B) *ACOX1*, (C) *CHGA*, (D) *NAT2*, (E) *PKIB*, (F) *VEGFA*, (G) *TINAG*, (H) *SLC39A8*.

(Figure 3G) were not associated with survival outcomes, survival analysis revealed that high expression of *AT8B1* (Figure 3A), *ACOX1* (Figure 3B), *CHGA* (Figure 3C), *NAT2* (Figure 3D), and *SL39AB* (Figure 3H) were protective prognostic variables.

3.5. Protein expression levels and external validation of genetic changes

Representative patient samples were taken from the Human Protein Atlas database to examine protein expression levels. It was found that *PKIB* (Figure 4E) and *VEGFA* (Figure 4H) were moderately to highly positive in immunohistochemical staining. However, *ACOX1* (Figure 4A), *ATP8B1* (Figure 4B), *CHGA* (Figure 4C), *NAT2* (Figure 4D), *SLC39A8* (Figure 4F), and *TINAG* (Figure 4G) showed weak positive or negative staining. Protein expression levels were consistent with gene expression levels, which supported the preliminary results in the TCGA cohort: Low expression levels of *ACOX1*, *APT8B1*, *SLC39A8*, and *TINAG*, and high expression levels of *VEGFA* are risk factors for survival in colorectal cancer patients. In addition, we conducted a mutation analysis of all colorectal cancer patients in the TCGA cohort through the cBioPortal cancer Genomics database and found that the mutation frequency of *NAT2* in the eight-mRNA was 8%, whereas *PKIB* had no mutations (Figure 4I).

3.6. Immune cell distribution across risk categories

Using the ssGSEA algorithm, we analyzed the TCGA and GEO queues and discovered that, compared to the GEO queues, high- and low-risk groups in TCGA had a significantly higher level of immune cell infiltration ($p<0.05$; Figure 5A). This included dendritic cells (DCs), immature DCs, mast cells, natural killer cells, plasmacytoid DCs, T helper cells, T follicular helper cells, T helper 1 cells, and tumor-infiltrating lymphocytes. In the TCGA cohort, significant differences in immune function were observed between the high- and low-risk groups. These differences were most evident in chemokine receptor signaling, immune checkpoint pathways, human leukocyte antigen expression, para-inflammation, T cell co-stimulation, T cell inhibition, and type II interferon response ($p<0.05$, Figure 5B). In the GEO external validation cohort, there were notable differences in immunological functioning and total immune cell infiltration between the low- and high-risk groups ($p<0.05$; Figure 5C and D). These findings imply that the eight-mRNA signature is highly correlated

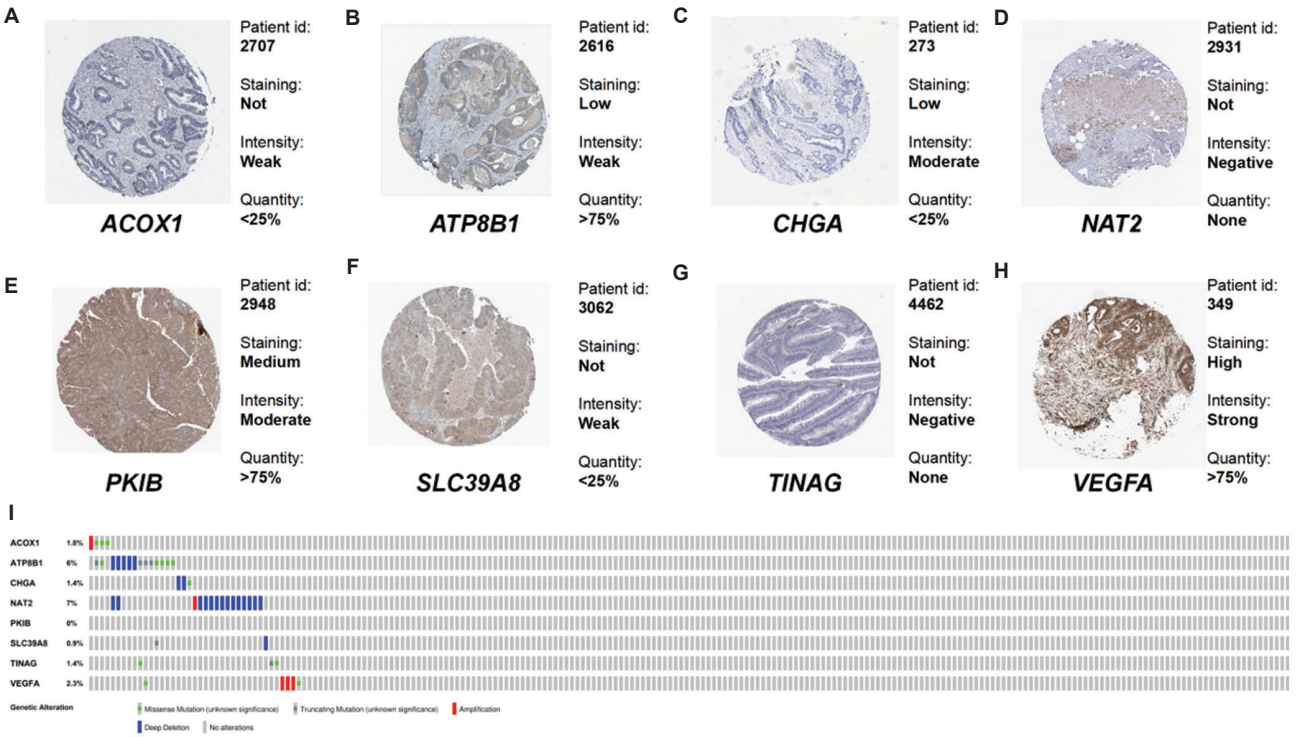


Figure 4. External validation of protein expression levels and genetic alterations. (A-H) The representative protein expression of the six mRNAs in colon adenocarcinoma. Data derived from the human protein atlas (http://www.proteinatlas.org) online database. (I) Genetic alterations of the eight mRNAs in colon adenocarcinoma.

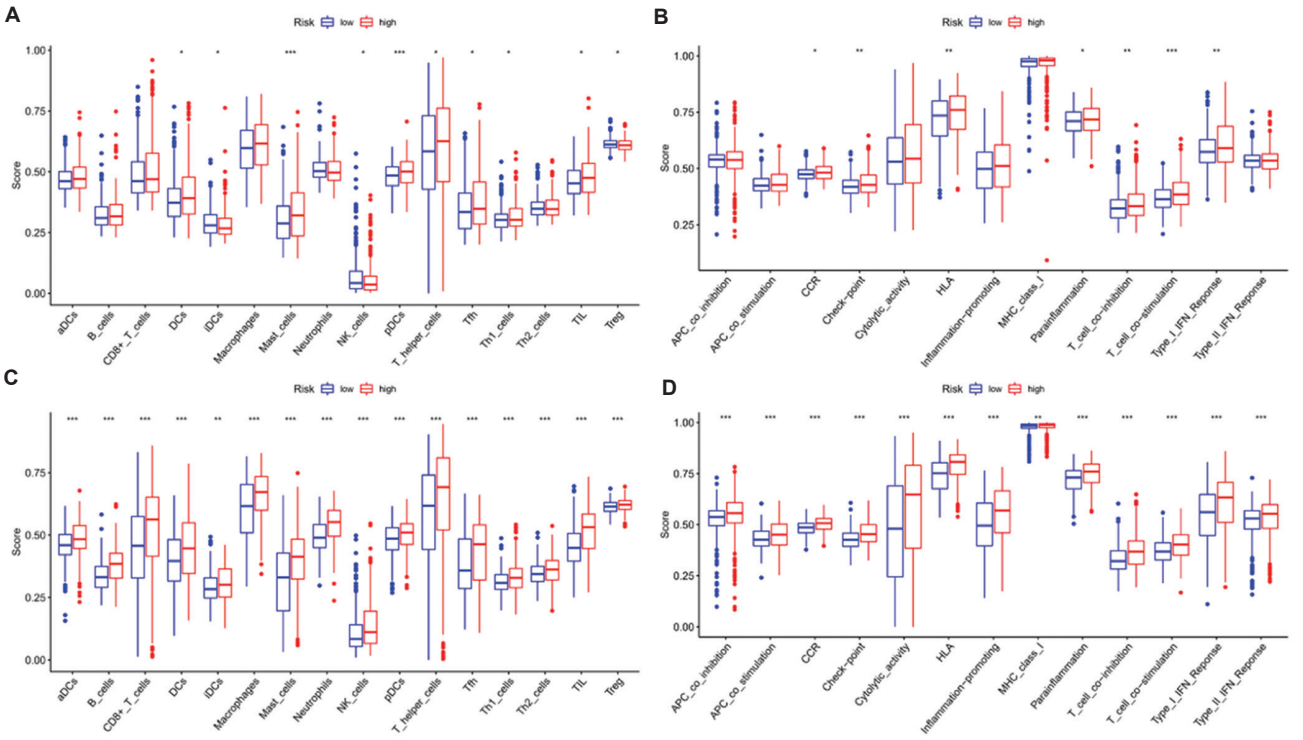


Figure 5. Distribution of immune cells in high- and low-risk groups. (A) Immune cell infiltration in the high- and low-risk groups of the total TCGA cohort. (B) Differences in immune function between high- and low-risk groups in the total TCGA cohort. (C) Immune cell infiltration between high- and low-risk groups in the GEO external validation cohort. (D) Differences in immune function between high- and low-risk groups in the GEO external validation cohort.
Abbreviations: APC: Antigen-presenting cell; CCR: Chemokine receptor; HLA: Human leukocyte antigen; MHC: Major histocompatibility class; TCGA: The cancer genome atlas; GEO: Gene Expression Omnibus.

with various immunological disorders and can be used to categorize individuals as high or low risk.

3.7. GSEA

The five most significant pathways enriched in the low-risk group are displayed in Figure 6. These pathways include calcium signaling, dilated cardiomyopathy, hypertrophic cardiomyopathy, neuroactive ligand-receptor interaction, and vascular smooth muscle contraction pathways.

3.8. Independent prognostic value of the risk signature

In addition to clinical factors, univariate and multivariate Cox regression analyses were conducted on the TCGA and GEO cohorts, respectively. Following multivariate correction for the variables above, the mRNA risk signature was found to be a robust and independent prognosis predictor for patients with colorectal cancer (hazard ratio=1.012, 95% confidence interval=1.008 – 1.017, p<0001; Figure 7A and B). Similar outcomes were observed in the GEO external validation group (Figure 7C and D; hazard ratio=1.028, 95% confidence interval=1.011 – 1.063, p<0.05).

3.9. Association between mRNA risk score and clinical characteristics and the role of risk stratification in response to chemotherapy

We analyzed risk differences for several pathological clinical characteristics, drawing on clinical data from the TCGA and GEO databases. The TCGA cohort showed that T stage (Figure 8C), N stage (Figure 8D), M stage (Figure 8E), and clinical stage (Figure 8F) significantly differed (p<0.05) between high- and low-risk groups, but age (Figure 8A) and sex (Figure 8B) differences were not significant (p>0.05). Results were similar in the TCGA and GEO queues (Figure 8G-L). We further investigated whether the mRNA risk signature could predict responses to commonly used chemotherapeutic agents in colorectal cancer patients. Notably, a significantly higher proportion of high-risk patients exhibited resistance to xeloda, camptothecin-11, and L-OHP (Figure A2B-D), with the exception of 5-fluorouracil (p<0.05; Figure A2A).

3.10. Nomogram construction and assessment of predictive value

To anticipate the prognosis of patients with colorectal cancer at 1, 3, and 5 years out, we constructed a Cox-factorial

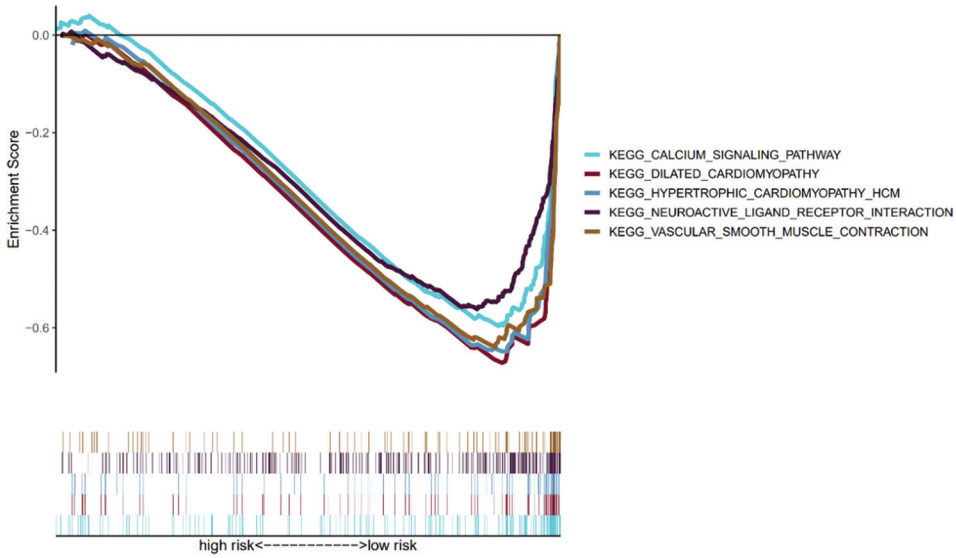


Figure 6. Gene set enrichment analysis of enriched pathways for high- and low-risk groups

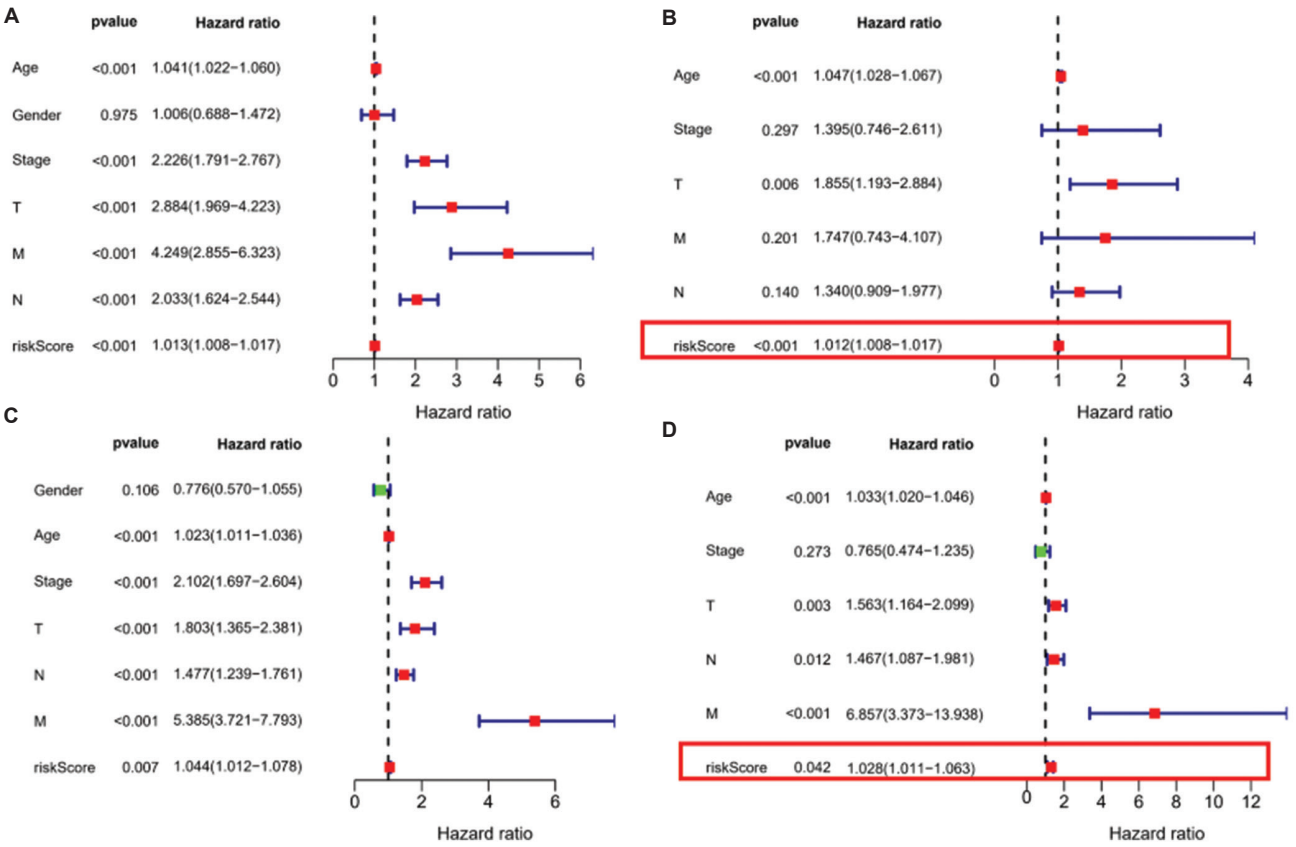


Figure 7. Univariate and multivariate Cox regression analysis of the mRNA signature of colon adenocarcinoma patients. (A and B) The Cancer Genome Atlas cohorts and (C and D) the Gene Expression Omnibus cohorts.

regression graph. The patient outcomes from the TCGA and GEO databases are displayed in Figure 9A and B, respectively. For the 1-, 3-, and 5-year projected OS probability in each cohort, the calibration curve similarly shows a high degree of agreement (Figure 9C-H) between the anticipated and observed survival periods.

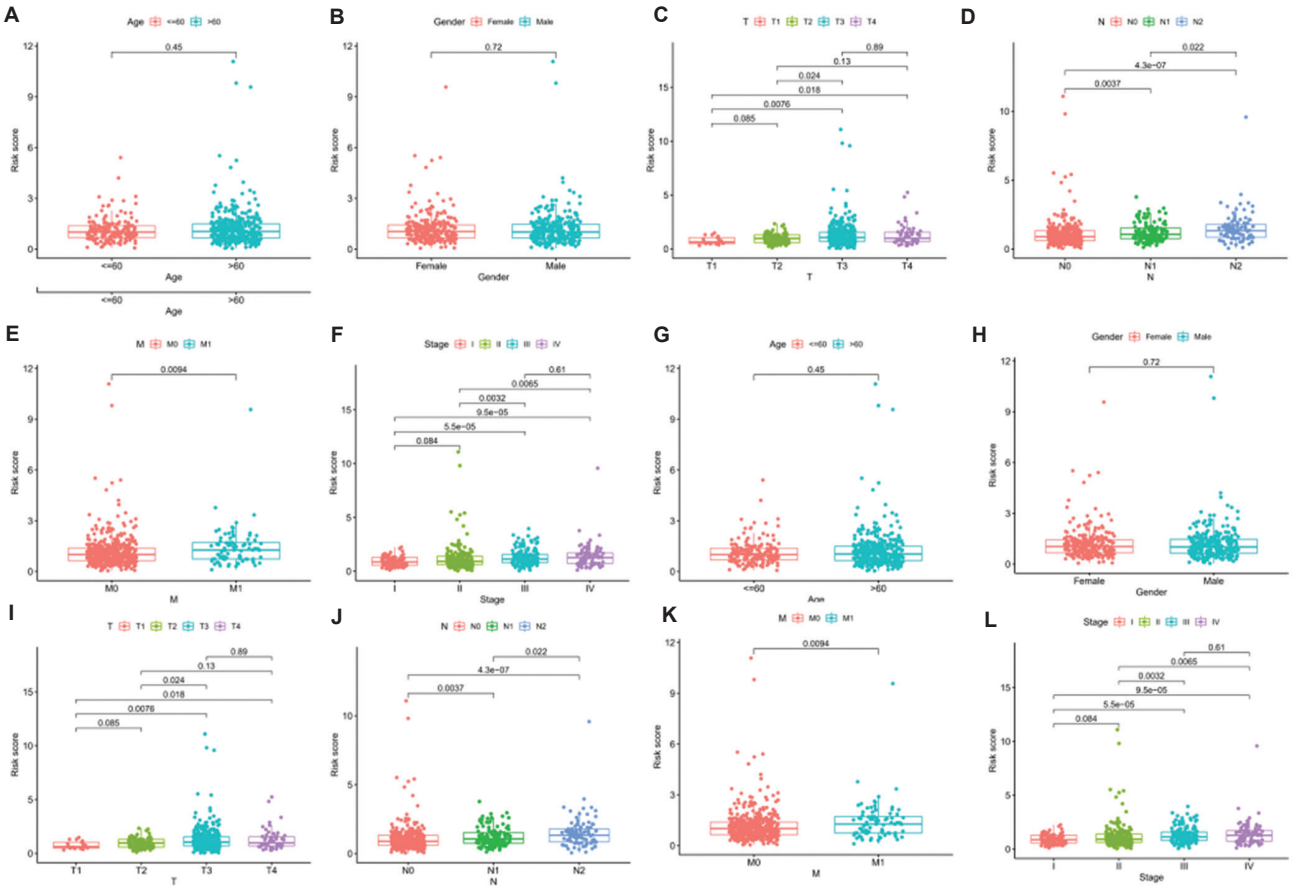


Figure 8. mRNA risk score and clinical characteristics. Differences in (A) age, (B) gender, (C) T-stage, (D) M-stage, (E) N-stage, and (F) clinical stage between high- and low-risk groups in the Cancer Genome Atlas cohort. (G-L) show the differences in these same variables between high- and low-risk groups in the GEO cohorts.

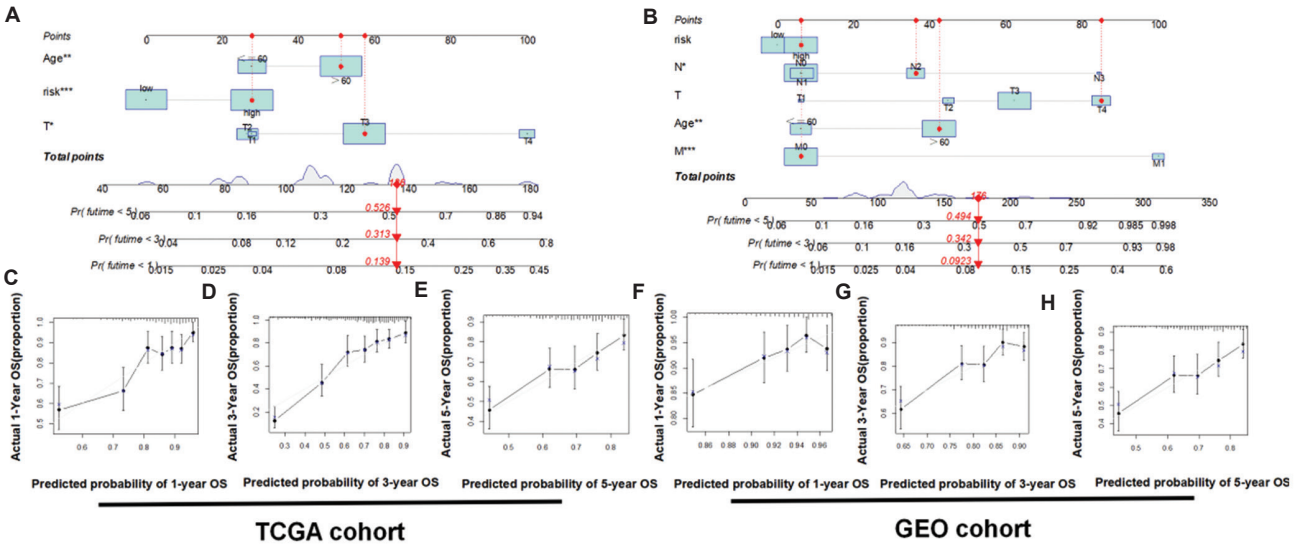


Figure 9. Nomograms and calibration plots for the TCGA and GEO cohorts. Predictions of survival for colon adenocarcinoma patients at (A-B) 1, 3, and 5 years and (C-H) calibration curves for these predictions.
Abbreviations: GEO: Gene Expression Omnibus; OS: Overall survival; TCGA: The cancer genome atlas.

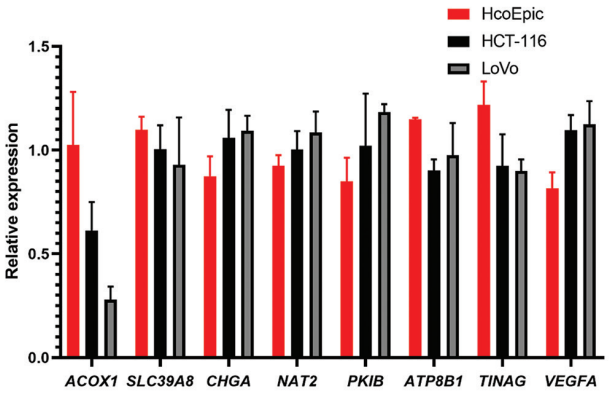


Figure 10. Quantitative real-time polymerase chain reaction. The expression levels of the eight genes were validated in human colonic epithelial cells (HCoEpiC), LoVo, and HCT-116 cell lines.

3.11. Quantitative real-time polymerase chain reaction validation

To verify the expression levels of the aforementioned genes in the tumor population to the healthy population, we compared two strains of human colon cancer cells to normal colon epithelial cells. Our results showed that *ACOX1*, *ATP8B1*, *TINAG*, and *SLC39A8* were expressed at higher levels in normal colon epithelial cells. In contrast, *CHGA*, *TINAG*, *VEGFA*, and *PKIB* were expressed at lower levels in normal colon epithelial cells than in colon cancer cells (Figure 10).

4. Discussion

Colon adenocarcinoma is a common malignancy, yet it remains challenging to classify in terms of risk stratification. The high cost of screening for high-risk individuals and the lack of clinically useful biomarkers to identify such patients further complicates the issue. Several studies have focused on developing multiple gene profiles using large public datasets to construct prognostic risk models that can predict tumor dynamics, prognosis, and responses to common clinical treatment strategies.[12-14] In this study, we aimed to risk-stratify COAD patients. We selected the TCGA-COAD dataset for modeling and internal validation and the GSE39582 dataset for external validation. Subsequently, we constructed an eight-gene prognostic prediction model for COAD. The established prognostic prediction model performed better in risk-stratifying COAD patients and predicting OS outcomes when integrated with clinical features. Multiple genes have previously been shown to perform better at predicting prognosis compared to single genes.[15] The modeling set, internal validation set, and GEO dataset demonstrated statistically significant differences in survival outcomes between high- and low-risk groups when the eight-mRNA signature for COAD patients was used for categorization.

Several of the genes in our model have been linked to colorectal cancer progression and prognosis. For example, acyl-CoA oxidase 1 (ACOX1), the initial rate-limiting enzyme in fatty acid β-oxidation, has been implicated in hepatocellular carcinoma development.[16] ACOX1 is a direct target of miR-15b-5p in colorectal cancer, and upregulation of miR-15b-5p by ACOX1 overexpression prevented colorectal cancer cell motility and invasion.[17] In this study, ACOX1 was found to be a protective prognostic variable. The balance of membrane phospholipids significantly influences cell carcinogenesis and the advancement of cancer. ATPase class I type 8b member 1 (ATP8B1), a P4-ATPase involved in phospholipid flipping, plays a role in cancer progression. In line with our findings, decreased ATP8B1 levels have been positively linked to epithelial-mesenchymal transition and favorable prognoses for patients with colorectal cancer.[18]

The solute carrier (SLC) group of membrane transport proteins regulates zinc inflow and is implicated in a number of physiological processes.[19] A previous study found that SLC family 39 (SLC39) members may be predictive indicators for breast cancer, with higher expression of *SLC39A8* expression in normal tissues than in breast cancer tissues.[19] In addition, in colorectal cancer, *SLC39A8* has been identified as a protective gene, particularly related to ferroptosis.[20] Our findings support this role, with increased *SLC39A8* levels being associated with lower risk scores. The four above-mentioned genes exhibited negative coefficients in the risk score equation, and their increased expression was associated with low-risk scores. The other four genes exhibited positive coefficients in the risk score equation, and their increased expression is hypothesized to be associated with higher risk scores.

Tumor angiogenesis is a hallmark of cancer, and vascular endothelial growth factor A is an angiogenesis-associated cytokine with pro-tumor angiogenic and pro-tumor invasive metastatic effects.[21] In this study, using the TCGA dataset, we found that *VEGFA* expression is a relevant high-risk marker for COAD. Interestingly, *VEGFA* was overexpressed in the low-risk group in the external validation set. However, our study evaluated the value of these eight mRNAs, as a whole, on risk stratification and prognosis rather than just the value of a single gene. Furthermore, in the GEO external validation set, *CHGA* overexpression was linked to high risk, but in the full TCGA cohort, it was a protective prognostic factor. According to bioinformatics analyses, patients with primary colorectal cancer had lower levels of

CHGA expression than the general population.[22] cAMP-dependent protein kinase inhibitor-beta is a type of protein kinase inhibitory peptide.[23] It has been shown to be highly expressed in colorectal cancer tissues. In addition, a study that compared the expression levels of *PKIB* in these tissues with those of matching normal tissues found that it positively correlated with T-stage and poor prognosis in colorectal cancer patients.[24]

We integrated risk subtypes and clinical features to predict the prognosis of each COAD patient accurately. While age and gender did not significantly differ between high- and low-risk groups, significant differences were observed in T, N, and M staging, as well as clinical staging, suggesting that these criteria could be crucial clinical traits in identifying risk subtypes. We discovered that the eight-mRNA signature was an independent predictive factor after incorporating these clinical characteristics and risk signatures into univariate and multifactorial Cox regression analysis in the TCGA and GEO datasets. High-risk patients were more likely to develop resistance to xeloda, camptothecin-11, and L-OHP than low-risk patients. However, there were no significant differences in resistance between high- and low-risk groups for 5-fluorouracil, which may assist in optimizing drug dosage. Clinical characteristics and eight-mRNA profiles were created as column line plots to predict each COAD patient's prognosis precisely. The high accuracy of our eight-mRNA signature for COAD prognosis was further demonstrated by calibration curves, which displayed a high agreement between anticipated and actual observed survival times.

The prognosis of patients with tumors is significantly influenced by immune cell infiltration and immunological activities. With significant differences in immune cell infiltrations and immune function between high- and low-risk groups, a robust correlation between the eight-mRNA signature and various immune profiles was discovered in this study.

We selected the top five key pathways for analysis after performing pathway enrichment analysis using GSEA, which revealed significant enrichment in the low-risk group. Interestingly, calcium ion metabolism is implicated in four of the five key pathways. Calcium ions, the most prevalent intracellular second messenger, play a critical role in regulating various physiological processes through cell signaling.[25] Intracellular calcium ions homeostasis is tightly regulated, and its imbalance has been associated with tumor growth, invasion, and metastasis.[26] A study by Tsoi *et al.*[27] compared calcium ion concentrations in colon cancer cell lines (caco-2, DLD-1, HCT116, LoVo, HT29, LS180, SW480, SW620, and SW1116) and the normal colon cell line, NCM460.[27] They discovered that all colorectal cancer cells exhibited higher calcium ion concentrations than the normal cell line. Furthermore, Pan and Xie[28] reported that calcium ion-focal adhesion kinase signaling activation enhanced *FOXC2* mRNA stability, promoting colorectal cell invasion, migration, and proliferation.[28]

5. Conclusion

In this study, we developed a robust eight-gene signature and a corresponding column line graph to predict the prognosis of colorectal cancer patients. This tool could assist in identifying high-risk individuals who may benefit from more intensive treatment strategies.

Acknowledgments

The authors would like to thank Tianjin University of Traditional Chinese Medicine for supporting the experiment.

Funding

This study was supported by the Tianjin Municipal Health Commission Chinese Medicine and Western Medicine Research Project (2023193), the Tianjin Municipal Health Commission Chinese Medicine and Western Medicine Research Project (2023222), and the Tianjin Municipal Education Commission Subjects (2022KJ187).

Conflict of interest

The authors declare no competing interest.

Author contributions

Conceptualization: Yuying Yang, Cuiying Wang, Zhongqiu Chai

Formal analysis: Hongqian Wei, Bing Zhou, Yuying Yang, Xiaochen Pang

Methodology: Yuying Yang, Cuiying Wang, Zhongqiu Chai, Wenhai Dong

Writing – original draft: Yuying Yang, Cuiying Wang

Writing – review & editing: Zhongqiu Chai, Songtao Hou

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data analyzed in this study are available in the TCGA and GEO databases.

References

1. Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.

   doi: 10.3322/caac.21660

2. Wang H, Liu J, Li J, *et al*. Identification of gene modules and hub genes in colon adenocarcinoma associated with pathological stage based on WGCNA analysis. *Cancer Genet.* 2020;242:1-7.

   doi: 10.1016/j.cancergen.2020.01.05

3. Aaltonen LA. The multistep process of colon carcinogenesis. *Cytokines Mol Ther.* 1996;2(2):111-114.

4. Miller KD, Nogueira L, Mariotto AB, *et al*. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin.* 2019;69(5):363-385.

   doi: 10.3322/caac.21565

5. Arnold M, Abnet CC, Neale RE, *et al*. Global burden of 5 major types of gastrointestinal cancer. *Gastroenterology.* 2020;159(1):335-349.e15.

   doi: 10.1053/j.gastro.2020.02.068

6. Oberndorfer F, Moling S, Hagelkruys LA, *et al*. Risk reclassification of patients with endometrial cancer based on tumor molecular profiling: First real world data. *J Pers Med.* 2021;11(1):48.

   doi: 10.3390/jpm11010048

7. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* 2008;9:559.

   doi: 10.1186/1471-2105-9-559

8. Tu J, Kuang Z, Xie X, *et al*. Prognostic and predictive value of a mRNA signature in peripheral T-cell lymphomas: A mRNA expression analysis. *J Cell Mol Med.* 2021;25(1):84-95.

   doi: 10.1111/jcmm.15851

9. Marisa L, De Reyniès A, Duval A, *et al*. Gene expression classification of colon cancer into molecular subtypes: Characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5):e1001453.

   doi: 10.1371/journal.pmed.1001453

10. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol.* 2016;1418:93-110.

    doi: 10.1007/978-1-4939-3578-9-5

11. Flicek P, Amode MR, Barrell D, *et al*. Ensembl 2011. *Nucleic Acids Res.* 2011;39(S 1):D800-D806.

    doi: 10.1093/nar/gkq1064

12. Yuan Y, Chen J, Wang J, *et al*. Development and clinical validation of a novel 4-gene prognostic signature predicting survival in colorectal cancer. *Front Oncol.* 2020;10:595.

    doi: 10.3389/fonc.2020.00595

13. Mo X, Huang X, Feng Y, *et al*. Immune infiltration and immune gene signature predict the response to fluoropyrimidine-based chemotherapy in colorectal cancer patients. *Oncoimmunology.* 2020;9(1):1832347.

    doi: 10.1080/2162402X.2020

14. Miao YD, Wang JT, Yang Y, *et al*. Identification of prognosis-associated immune genes and exploration of immune. *Biomark Med*. 2020;14(14):1353-1369.

    doi: 10.2217/bmm-2020-0024

15. Cucchiara V, Cooperberg MR, Dall'Era M, *et al*. Genomic Markers in Prostate Cancer Decision Making. *Eur Urol.* 2018;73(4):572-582.

    doi: 10.1016/j.eururo.2017.10.036

16. Chen XF, Tian MX, Sun RQ, *et al*. SIRT5 inhibits peroxisomal ACOX1 to prevent oxidative damage and is downregulated in liver cancer. *EMBO Rep.* 2018;19(5):e45124.

    doi: 10.15252/embr.201745124

17. Sun LN, Zhi Z, Chen LY, *et al*. SIRT1 suppresses colorectal cancer metastasis by transcriptional repression of mir-15b-5p. *Cancer Lett.* 2017;409:104-115.

    doi: 10.1016/j.canlet.2017.09.001

18. Deng L, Niu GM, Ren J, Ke CW. Identification of ATP8B1 as a tumor suppressor gene for colorectal cancer and its involvement in phospholipid homeostasis. *Biomed Res Int.* 2020;2020:2015648.

    doi: 10.1155/2020/2015648

19. Liu L, Yang J, Wang C. Analysis of the prognostic significance of solute carrier (SLC) family 39 genes in breast cancer. *Biosci Rep.* 2020;40(8):BSR20200764.

    doi: 10.1042/BSR20200764

20. Shao Y, Jia H, Huang L, *et al*. An original ferroptosis-related gene signature effectively predicts the prognosis and clinical status for colorectal cancer patients. *Front Oncol.* 2021;11:711776.

    doi: 10.3389/fonc.2021.711776

21. Wang R, Ma Y, Zhan S, *et al*. B7-H3 promotes colorectal cancer angiogenesis through activating the NF-Kb pathway to induce VEGFA expression. *Cell Death Dis.* 2020;11(1):55.

    doi: 10.1038/s41419-020-2252-3

22. Han J, Zhang X, Liu Y, Jing L, Liu YB, Feng L. CLCA4 and MS4A12 as the significant gene biomarkers of primary colorectal cancer. *Biosci Rep.* 2020;40(8):BSR20200963.

    doi: 10.1042/BSR20200963

23. Wang HW, Duan ZJ, Hu SS, Wang S. Expression of cAMP-dependent protein kinase inhibitor beta in colorectal

carcinoma and its clinical significance. *Nan Fang Yi Ke Da Xue Xue Bao.* 2017;37(6):744-749.

doi: 10.3969/j.issn.1673-4254.2017.06.05

24. Fei L, Wang Y. Microrna-495 reduces visceral sensitivity in mice with diarrhea-predominant irritable bowel syndrome through suppression of the PI3K/AKT signaling pathway via PKIB. *IUBMB Life.* 2020;72(7):1468-1480.

    doi: 10.1002/iub.2270

25. Bootman MD, Bultynck G. Fundamentals of cellular calcium signaling: A primer. *Cold Spring Harb Perspect Biol.* 2020;12(1):a038802.

    doi: 10.1101/cshperspect.a038802

26. Marchi S, Giorgi C, Galluzzi L, Pinton P. $Ca2^{+}$ fluxes and cancer. *Mol Cell.* 2020;78(6):1055-1069.

    doi: 10.1016/j.molcel.2020.04.017

27. Tsoi H, Lam KC, Dong Y, *et al*. Pre-45s rRNA promotes colon cancer and is associated with poor survival of CRC patients. *Oncogene.* 2017;36(44):6109-6118.

    doi: 10.1038/onc.2017.86

28. Pan K, Xie Y. LncRNA FOXC2-AS1 enhances FOXC2 mRNA stability to promote colorectal cancer progression via activation of $Ca2^{+}$-FAK signal pathway. *Cell Death Dis.* 2020;11(6):434.

    doi: 10.1038/s41419-020-2633-7.2041-488

Appendices

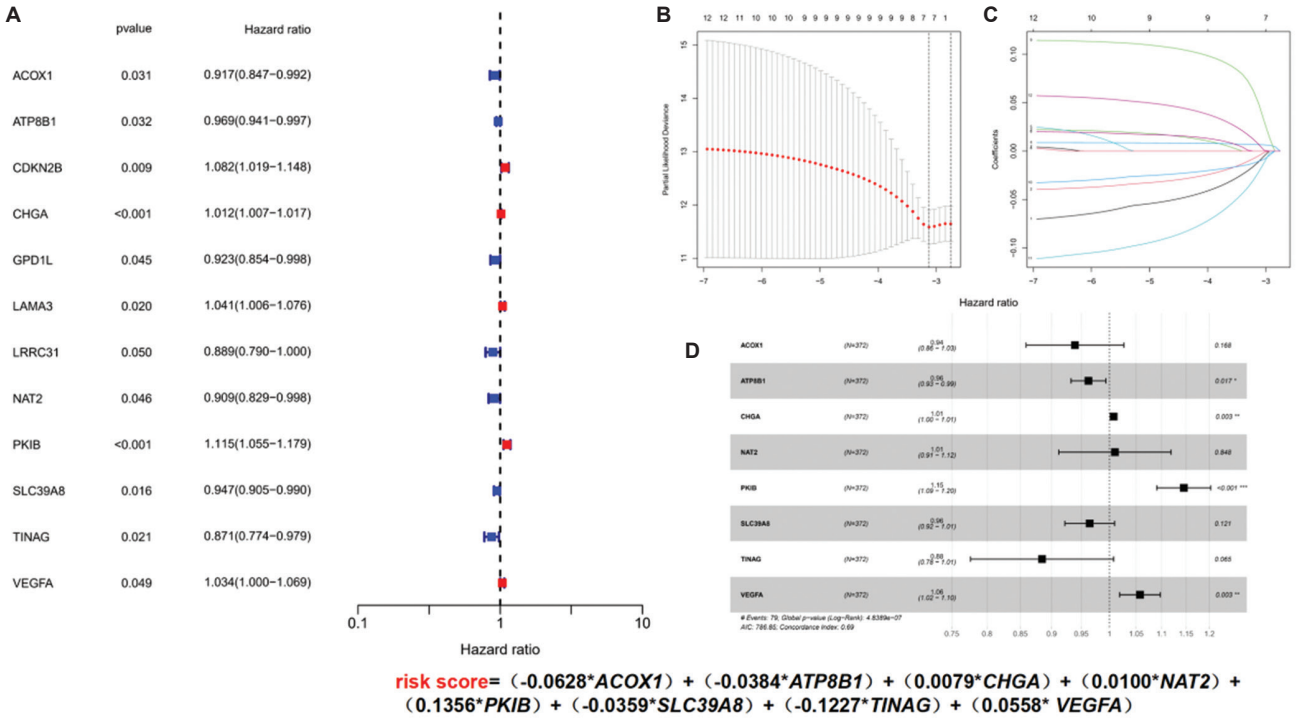


Figure A1. Identification of risk signatures for modeling sets of patients. (A) One-way Cox regression analysis identifies 12 genes associated with survival. (B) LASSO coefficient of the 20 selected genes. (C) Ten-fold cross-validation for tuning parameter selection in the LASSO module. (D) Overall information of eight genes identified using multivariate Cox proportional hazards regression in the TCGA-COAD dataset. Solid squares represent the hazard ratio of death.
Abbreviations: COAD: Colon adenocarcinoma; LASSO: Least absolute shrinkage and selection operator; TCGA: The cancer genome atlas.

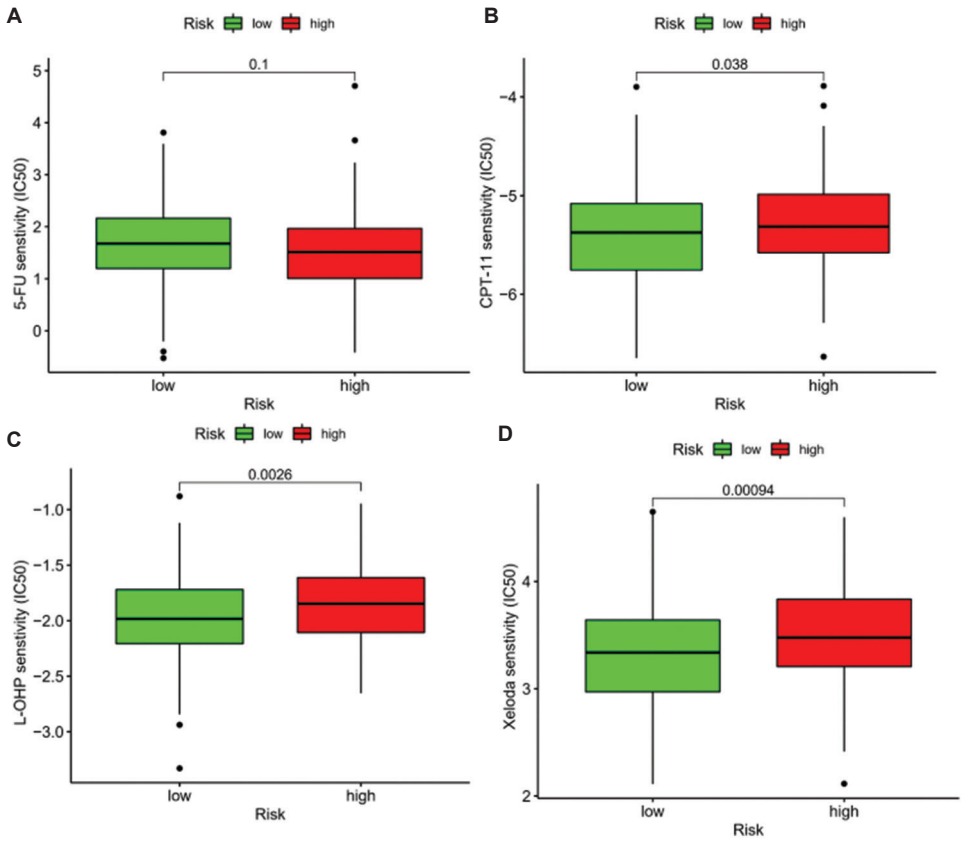


Figure A2. The role of mRNA risk stratification on chemotherapy response. (A) 5-fluorouracil (5-FU), (B) camptothecin-11 (CPT-11), (C) L-OHP, and (D) xeloda.
Abbreviation: IC_{50}: Inhibitory concentration at 50%.