Identification and Validation of a Novel Inflammatory Response-Related Gene Signature for the Prognosis of Colon Cancer

Yichao Liang
Xin Wu
Qi Su
Yujie Liu
Hong Xiao
Department of General Surgery, Shengjing Hospital of China Medical University, Shenyang, People’s Republic of China

Purpose: The inflammatory response plays a crucial role in the occurrence and development of colon cancer. In this study, we aimed to explore a novel prognostic model for patients with colon cancer (COAD) based on inflammatory response-related genes.

Methods: Inflammatory response-related genes were obtained from Molecular Signatures database. Univariate and multivariate Cox regression analyses were used for model construction based on TCGA dataset. GSE39582 dataset and qRT-PCR dataset were used for validation. Gene set variation analysis and gene set enrichment analysis were performed to explore the potential regulatory pathways. The immune cell infiltration level was analyzed via CIBERSORT. Immunohistochemistry analysis and experiments were used to explore the function of genes in model.

Results: In this study, a novel prognostic signature was identified using stepwise Cox proportional hazards regression analysis based on TCGA dataset. The results were subsequently validated in 562 patients from GSE39582 and a qRT-PCR data set from 70 tumor samples. Functional analysis indicated that the tumor microenvironment and immune cell infiltrate were different between high- and low-risk groups. Additionally, IHC results showed that the protein levels of prognostic genes were significantly different between COAD tissues and adjacent non-tumorous tissues, and prognostic genes could regulate the malignant phenotype of COAD cells.

Conclusion: Overall, the inflammation-related gene signature can be used for prognostic prediction in patients with COAD.

Keywords: colon cancer, prognosis, inflammation

Introduction
Colon cancer is a common malignant tumor and one of the main causes of cancer-related death. Presently, surgery is the primary approach to treat colon cancer, and the metastasis or postoperative recurrence of colon cancer is a key factor affecting the long-term survival of patients. Therefore, a more effective prognostic evaluation system is required to provide individualized treatment for colon cancer patients and improve the prognosis of patients. The rapid development of bioinformatics analysis and high-throughput sequencing technology has helped identify a series of markers and risk prediction models related to the prognosis of colon cancer patients, assisting clinicians in evaluating the prognosis of colon cancer patients and formulating personalized treatment plans. However, most of the analysis parameters in these studies are derived from the entire genome or transcriptome and...
do not consider changes in specific genomes or symptoms. Therefore, these genetic models only represent statistical models and do not reflect the inherent characteristics of colon cancer.

Inflammation is one of the important characteristics of tumors. Tumor-related inflammation includes local inflammation and systemic inflammation. The local inflammatory response refers to the inflammatory microenvironment, which can promote tumor growth and progression by promoting angiogenesis and metastasis, subverting the anti-tumor immune response, and changing the sensitivity of tumor cells to chemotherapeutics. The persistent non-controllable inflammatory microenvironment can also trigger gene mutations to cause tumors. The systemic inflammatory response, including changes in neutrophils and lymphocytes numbers and albumin levels, is closely related to the response to tumor treatment. In colon cancer, patients with an elevated neutrophil-lymphocyte ratio (NLR) have a high risk of recurrence, and this ratio can predict the survival of patients with colon cancer after surgery. Additionally, increased neutrophils and decreased albumin can be used as independent predictors of the prognosis of colon cancer liver metastasis. Recent studies have highlighted that local chronic inflammatory cell infiltration combined with systemic inflammatory response analysis can predict the prognosis of colon cancer patients independently of clinicopathological staging and assist patients with stage II colon cancer in choosing postoperative adjuvant chemotherapy. However, the expression and clinical relevance of inflammatory response-related genes in colon cancer remain unclear.

In this study, we constructed an inflammation-related gene signature (IRGS) in the TCGA cohort and verified the stability and reliability of the model. Next, functional enrichment analysis was conducted to explore the potential mechanism. Finally, experiments verified the protein expression level and function of prognostic genes.

**Materials and Methods**

**Data Collection**

The Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) application performing interface (API) was used to download the latest expression data and clinical follow-up information of patients with colon adenocarcinoma (COAD). This cohort 277 samples. The GSE39582 chip expression data in MINiML format were downloaded from the National Center for Biotechnology Information (NCBI). GSE39582 contained 536 samples with clinical characteristics. These two cohorts were included because they were the largest sample sets in the same platform with detailed follow-up information of colon cancer. Inflammatory response-related genes were selected from hallmark gene sets in the Molecular Signatures database, and the link is [http://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_INFLAMMATORY_RESPONSE.html](http://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_INFLAMMATORY_RESPONSE.html).

**Construction and Validation of the IRGS**

The independent prognostic predictors among 200 inflammatory response-related genes were selected via univariate and multivariate Cox regression analyses. The risk score of the IRGS for each sample was calculated using the expression values of the selected genes weighted by their corresponding coefficients following multivariate Cox regression analysis. The high- and low-risk groups were divided following the optimized risk value. Kaplan-Meier survival analyses were used to analyze the difference in overall survival between the high- and low-risk groups. Time-dependent ROC curve analysis was used to evaluate the predictive value of the IRGS.

**Analysis of Potential Regulatory Pathways**

GSVA was used to observe the relationships between the risk score and pathway as described previously. Gene set enrichment analysis (GSEA) was performed to compare the expression of genes in the high-risk and low-risk score groups, and the MSigDB H: hallmark gene set was used as a reference.

**Immune Infiltration Analysis**

The immune cell infiltration level was estimated using the CIBERSORT website. CIBERSORT is a novel tool widely used for characterizing the cell composition of tissues through the gene expression profile.

**RNA Isolation and RT-PCR Analysis**

The surgically resected tumor samples of 70 patients with colon cancer at Shengjing Hospital of China Medical University were included in this study. All the patients were followed up by telephone. The study was approved by the Human Ethics Review Committee of the Shengjing Hospital of China Medical University. All patients provided informed consent, in accordance with the Declaration of Helsinki. RNA extraction from the tissues was performed using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). RNA was reverse-
transcribed into cDNA using the QuantiTect Reverse Transcription Kit (Qiagen, Valencia, CA, USA). Real-time PCR analyses were quantified by SYBR Green (Takara, Otsu, Shiga, Japan), and the levels were normalized to GAPDH levels. The sequences of the upstream and downstream primers were as follows: SLC4A4-forward: 5′-TTCACGGAACTGATGAGCT-3′; SLC4A4-reverse 5′-ACGTGGGAGAAGAAGACC-3′; TIMP1-forward: 5′-CGCAGCGGGTGGTTCTCAT-3′; TIMP1-reverse: 5′-GGCACGTGATGTCAAATTTCC-3′; NMUR1-Forward: 5′-CCTCAAGGATTACAGCCTG-3′; NMUR1-reverse: 5′-GTTCCTGAGGCTTTGGTAG-3′; TACR3-forward: 5′-CCTCCAACTACTGCCGCTTC-3′; TACR3-reverse: 5′-ACA GTCTGGGTTCTCAAGGATCA-3′.

Immunohistochemistry Analysis
Surgically resected colon cancer samples and matched nontumorous tissue were fixed in 10% formalin, embedded in paraffin, and processed as 4-µm continuous sections. IHC staining was performed according to the manufacturer’s instructions (UltraSensitive™ SP; MXB, China). The antibodies used were as follows: SLC4A4 (1:100; ab187511; Abcam), TIMP1 (1:1000; ab211926; Abcam), NMUR1 (1:50; ab121959; Abcam), and TACR3 (1:100; ab124025; Abcam). Each sample was independently assessed by two pathologists and scored using a semiquantitative scoring system. The histoscores ranged from zero (minimum) to 300 (maximum).

Cell Culture
Colon cancer cells (HCT-116) were purchased from the Chinese Academy of Medical Sciences and CAMS & PUMC Medical College (Beijing, China). The cell lines were cultured in 1640 medium supplemented with 10% fetal bovine serum and 100 units/mL of penicillin at 37°C in a humidified 5% CO₂ incubator.

Cell Transfection
SiRNA (small interfering RNA) transfection was performed using Lipofectamine 2000 (Invitrogen, Shanghai, China). The sequences were as follows: 5′- GCAAUUGCUCAAG AGCUA-3′ for SLC4A4-specific siRNA; 5′- ATCAACC AGACCACCTTATA-3′ for TIMP1-specific siRNA, 5′- ATCAACCAGACCACCTTATA-3′ for NMUR1-specific siRNA and 5′- AUAGCCAUAUACCUGUCCAC-3′ for TACR3-specific siRNA. The sequence for the negative control (NC) was 5′- UUCUCCGAACGUGUCAGUTT-3′.

Cell Viability Assay
Cells (1500/well) were added to a 96-well culture plate and transfected with NC-siRNA, SLC4A4-siRNA, NMUR1-siRNA, TIMP1-siRNA, or TACR3-siRNA. After 0, 24, 48, or 72 h, the cells were cultured with 20 µL of CCK8 solution for another 2 h. Cell viability was expressed as an optical density (OD) value at 450 nm.

Colony Formation Assay
To explore the effects of SLC4A4, NMUR1, TIMP1 and TACR3 expression on cell proliferation, cells (1000/well) transfected with NC-siRNA or siRNA were added to each well of 6-well culture plates and incubated for two weeks. Cells were stained by Wright–Giemsa. Finally, the number of colonies was counted by light microscopy.

Statistical Analysis
Cox analyses were performed to screen the predictors for OS. Kaplan-Meier analysis was used to compare the differences in OS between the high- and low-risk score groups. Statistical comparisons between two groups were calculated using Student’s two-tailed t-test, and p-values < 0.05 were considered statistically significant.

Results
Construction of the IRGS in the TCGA Cohort
To comprehensively analyze the prognostic value of immune response-related genes in COAD, we first performed single-factor Cox analysis. The results showed that 16 genes (TACR3, SEMA4D, TIMP1, IL15RA, IRF7, DCBLD2, NMUR1, GABBR1, CX3CL1, PTGIR, SPHK1, SERPINE1, EREG, SLC4A4 and SRI) were significantly related to the prognosis of COAD patients (Figure S1). A Cox proportional hazards regression model was used to optimize and obtain IRGs. Subsequently, the following risk score was constructed: risk score = (0.498 × TIMP1 expression) - (0.188 × SLC4A4 expression) + (0.350 × NMUR1 expression) + (1.339 × TACR3 expression). The risk score was calculated, and the distribution of the TCGA training cohort is shown in Figure 1A. Following the optimized risk value, patients were assigned to the high-risk or low-risk group. Kaplan-Meier survival analyses showed that the overall survival of patients in the high-risk group was significantly lower than that of patients in the low-risk group (Figure 1B; P < 0.0001). Furthermore, the results based
on ROC curve analysis showed that the AUC values of the 1-, 3-, 5- and 7-year survival rates were 0.772, 0.731, 0.731 and 0.725, respectively (Figure 1C). Additionally, patients with low-risk scores had significantly longer DSS (disease-free survival) \( (p = 0.02; \text{Figure 1D}) \) and PFI (progression-free interval) \( (p = 0.0047; \text{Figure 1E}) \).
Verification of IRGS Based on GSE39582
To verify the accuracy of the 4-gene prognostic model, we downloaded the GSE39582 database as an external verification cohort. The risk score distribution of the validation cohort is shown in Figure 2A. Survival analysis showed that the OS of low-risk patients was significantly longer than that of high-risk patients (Figure 2B; p =0.0047). ROC curve analysis showed that the AUC values of the 1-, 3-, 5- and 7-year survival rates were 0.696, 0.624, 0.617 and 0.627, respectively (Figure 2C).

Verification of the IRGS in an Independent Group of Colon Cancer Tissue Samples
To evaluate the robustness of the IRGS for patients with COAD in clinical practice, we used qRT-PCR to further validate the specific signature in an independent cohort comprising tissue samples from 70 patients with colon cancer. Using the same formula, the risk score of each patient was calculated. The risk score distribution of the validation cohort is shown in Figure 3A. Survival analysis showed that the OS of low-risk patients was significantly longer than that of high-risk patients (Figure 3B; p =0.0017). ROC curve analysis showed that the AUC values of the 1-, 3- and 5-year survival rates were 0.551, 0.724 and 0.636, respectively (Figure 3C). These findings were consistent with the results of the TCGA cohort and GSE39582 cohort, and verified the reliability of our prognostic model.

Prognostic Model Risk Score and Clinical Features
Patients with different clinical features require different therapeutic strategies and have different prognoses; therefore, subgroup analysis was applied. In Figure 4A and B, a higher risk score was associated with a significantly worse
OS, regardless of whether the patient was female or male. In Figure 4C and D, a higher risk score was associated with a significantly worse OS in patients with N0 stage or N1-2 stage disease. In T1+T2 stage patients and T3+T4 patients, a higher risk score was associated with a significantly worse OS (Figure 4E and F). Additionally, a higher risk score was associated with a significantly worse OS of patients with stage 3–4 disease but was not significantly associated with that of patients with stage 1–2 disease (Figure 4G and H).

Next, we analyzed the relationship between the risk score and clinical features.

The risk score between T1+T2 and T3+T4 was not significantly different (Figure 4I). The risk score was significantly increased in patients with lymph node metastasis (Figure 4J), metastasis (Figure 4K), advanced TNM stage (Figure 4L), lymphatic invasion (Figure 4M), and vascular invasion (Figure 4N).

Association Between Risk Score and the Tumor Microenvironment

We attempted to analyze the relationship between the risk score and tumor immune microenvironment. The ESTIMATE score, stromal score and immune score were used to estimate the tumor immune microenvironment. In Figure 5A–C, the risk score was positively correlated with the ESTIMATE score (R=0.44; p<0.001), immune score (R=0.12; p=0.04) and stromal score (R=0.52; p<0.001).

Next, we checked the correlation between the risk score and immune infiltrate cells. Compared with the low-risk group, the fractions of B cell memory cells, M0 macrophages, M1 macrophages, M2 macrophages and resting mast cells were higher in the high-risk group (Figure 5D). Additionally, the fractions of activated mast cells and plasma cells were higher in the low-risk group. These results
indicate the differences in macrophage cell and mast cell regulation between the high- and low-risk groups.

Identifying Pathways Related to IRGS

Next, we analyzed the relationship between the risk scores of different samples and biological functions. First, we calculated the ssGSEA scores via GSEA to obtain the scores of all patients in terms of the different pathways (Figure 6A). The activity of the hedgehog signaling pathway and basal cell carcinoma increased as the risk score rose, while the activity of the citrate cycle TCA cycle, glyoxylate and dicarboxylate metabolism decreased. Furthermore, GSEA was used to perform pathway enrichment analyses between the high- and low-risk groups, and the MSigDB H: hallmark gene set was used as a reference in this step. Eleven pathways, such as “HALLMARK_ANGIOGENESIS” and “HALLMARK_APICAL_JUNCTION”, were significantly enriched in the high-risk group (Figure 6B).

Protein Level Validation and Functional Analysis of the Genes in the IRGS

To verify the protein expression of SLC4A4, NMUR1, TIMP1 and TACR3 in colon cancer tissues, 30 colon cancer tissues and paired normal tissues were tested. Immunohistochemistry (Figure 7A) showed that SLC4A4 was downregulated in colon cancer tissues, and NMUR1, TIMP1, and TACR3 were highly expressed in colon cancer tissues. Next, we analyzed the potential function of SLC4A4, NMUR1, TIMP1 and TACR3 in colon cancer. Silencing SLC4A4 promoted the proliferation of colon cancer cells in vitro, and silencing NMUR1, TIMP1 and TACR3 was highly expressed in colon cancer tissues. Next, we analyzed the potential function of SLC4A4, NMUR1, TIMP1 and TACR3 in colon cancer. Silencing SLC4A4 promoted the proliferation of colon cancer cells in vitro, and silencing NMUR1, TIMP1 and TACR3 was highly expressed in colon cancer tissues. Next, we analyzed the potential function of SLC4A4, NMUR1, TIMP1 and TACR3 in colon cancer. Silencing SLC4A4 promoted the proliferation of colon cancer cells in vitro, and silencing NMUR1, TIMP1 and TACR3 inhibited the proliferation of colon cancer cells in vitro (Figure 7B and C). These results suggest that SLC4A4 may function as a tumor suppressor gene and NMUR1, TIMP1 and TACR3 may function as oncogenes in colon cancer.
Discussion

Colon cancer is a common malignant tumor of the digestive tract and the third most common malignant tumor worldwide.\(^{17}\) Despite many advances in comprehensive treatment strategies, the prognosis of colon cancer patients remains poor, mainly because of the lack of effective diagnostic markers and molecular targeted therapies.\(^{18,19}\) The rapid development of omics sequencing technology enables researchers to analyze the mechanism of ovarian cancer progression through large-scale gene expression data and clinical information.\(^{15}\) An abnormal immune response is an important mechanism for the occurrence and development of colon cancer.\(^{20,21}\) Immune response-related indicators have shown the application value of evaluating the prognosis of colon cancer patients.

In this study, we systematically analyzed the expression of 200 inflammatory response-related genes in colon cancer tissues and their relationship with survival. A prognostic model integrating four inflammatory response-related genes was constructed through Cox regression analysis and verified in GSE39582 and PCR data from 70 samples. The prognostic value of inflammation-related gene signature in our collected clinical samples was consistent with the results of online analysis (TCGA and GEO). Inflammation-related gene signature can be used for prognostic prediction in patients with COAD in TCGA, GSE39582 dataset and our qRT-PCR dataset. Using different platforms and subgroups of colon cancer patients, the 4-gene signature showed good prognostic prediction performance.

Our IRGS was based on immune-related genes, including SLC4A4, NMUR1, TIMP1, and TACR3. SLC4A4 encodes a sodium bicarbonate cotransporter (NBC) involved in regulating bicarbonate secretion and absorption and intracellular pH.\(^{22,23}\) Compared with normal tissues, SLC4A4 mRNA and protein levels are reduced in colon cancer tissues. Low expression of SLC4A4 is closely related to a shorter overall survival of CRC patients.\(^{24}\) NMUR1 belongs to the G-protein coupled receptor 1
Figure 6 Gene set variation analysis and gene set enrichment analysis of pathways. (A) Heatmap showing the differentially expressed pathways between the low- and high-risk score groups based on GSVA analysis. (B) GSEA showed eleven pathways enriched in the high-risk group.
family and is the receptor for neuromedin-U and neuromedin-S neuropeptides. Additionally, NMUR1 has been identified as a prognostic biomarker of oropharyngeal cancer and breast cancer patients. The proteins encoded by TIMP1 are natural inhibitors of matrix metalloproteinases (MMPs), which are important regulators of the extracellular environment. TIMP1 in serum could identify colon cancer patients with a poor prognosis. In patients showing no systemic inflammatory response, TIMP1 is also associated with a poor prognosis. TACR3 encodes the receptor for tachykinin neurokinin 3, also referred to as neurokinin B, and is abnormally

Figure 7  Protein level validation and functional analysis of 4 genes in the IRGS. (A) Immunohistochemical results showed that SLC4A4 was downregulated in colon cancer tissues, and NMUR1, TIMP1, TACR3 were highly expressed in colon cancer tissues. Magnification: 10×20. Cell viability assays (B) and colony formation assays (C) showed that silencing SLC4A4 promoted the proliferation of HCT116 cells, and silencing NMUR1, TIMP1 and TACR3 inhibited the proliferation of HCT116 cells, data represent mean ± SD, *P < 0.05, **P < 0.01 and ***P < 0.001 (versus control group).
expressed in bladder cancer and oral squamous cell carcinoma. In this study, both NMUR1 and TACR3 reported their prognostic value in colon cancer for the first time, and the underlying mechanism of action depends on further exploration.

The IRGS we constructed showed good prognostic ability in the training set (TCGA), validation set (GSE39582 and PCR) and colon cancer patients of different subtypes. Univariate and multivariate Cox analyses after including clinical features showed that the IRGS is an independent prognostic factor for colon cancer patients. To gain a deeper understanding of the relationship between risk scores and immune components, we studied the relationship between risk scores and immune infiltrating cells. The results showed that high-risk scores are closely related to macrophages (M0, M1, and M2). Macrophages are important immune cells that infiltrate tumor tissues. Macrophages play an important role in the local inflammatory response of tumors. Under the influence of the tumor microenvironment, macrophages exhibit a classic activation state (M1 type) and selective activation state (M2 type), play antitumor and tumor-promoting functions, respectively, and play an important role in the interaction between inflammation and tumors. This result further indicates the important role of macrophages in the immune response of colon cancer.

Therefore, we tried to further analyze the mechanism of IRGS in colon cancer. GSVA and GSEA showed that pathways such as WNT-β-catenin, focal adhesion, and TGF-β were enriched with an increasing risk score. The Wnt signaling pathway is evolutionarily conserved and can regulate embryonic development and maintain adult tissue homeostasis. Wnt signaling controls basic cell functions, including proliferation, differentiation, migration and stemness. In colon cancer, the WNT pathway is overactivated or overactivated due to genetic, epigenetic, or receptor/ligand changes. Focal adhesion is a type of cytoplasmic nonreceptor protein tyrosine kinase. FAK is the center of intracellular and extracellular signal entry and exit, mediating multiple signaling pathways. In colon cancer, the upregulation of FAK activity promotes the malignant biological behavior of tumor cells. The TGF_BETA pathway determines the specific fate of cells in the process of cell proliferation, embryogenesis, differentiation, and cell death. The results of this study suggest that methylation-related genes in colon cancer may function through pathways, such as the WNT-β-catenin, focal adhesion and TGF-β pathways.

Our inflammation-related gene signature was constructed based on TCGA-COAD dataset, GSE39582 dataset was used for validation. TCGA-COAD dataset was published in 2012, it suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression. GSE39582 dataset was published in 2013, researchers report a new classification of CC into six molecular subtypes that arise through distinct biological pathways. Different from previous research, we re-analysis these datasets for the initial screening of biomarkers and for validation.

Due to the advent of genomics, TCGA and GEO databases are available for exploring prognostic signatures in colorectal cancer. Liu et al report deoxyribonuclease 1-like 3 could be a potential prognostic biomarker in colon cancer based on the results from TCGA and GEO datasets. Cao et al combined three independent datasets and screened to determine that CXCL11 is a prognostic biomarker in colon cancer. However, the expression of a single gene can vary among populations whereas, a multi-gene signature can compensate for individual differences and thus, can be more predictive and higher accuracy in comparison to single gene-expression based markers. Zhang et al constructed a 17 immune-related gene prognostic signature for colon cancer. Lv et al constructed an 12 immune-related genes signature with prognostic value for colon cancer. Luo et al established a 17 immune-related genes model to predict colon adenocarcinoma prognosis. However, these models lack of sufficient advantages due to larger genes number and a lack of independent external verification. Compared with these models, our inflammation-related gene signature has a smaller number of genes, and shows better prediction performance in TCGA, GEO and our qRT-PCR data set from 70 tumor samples.

IRGS based on immune response-related genes can be used as an effective method to judge the prognosis of patients. However, there are still some limitations. First, the samples in our study were retrospective; prospective samples are required for verification. Second, we only analyzed the prognostic value and clinical significance of the IRGS. The mechanism of action of the four genes contained in the signature requires further experimental exploration.

**Conclusion IRGS**

Overall, we comprehensively analyzed the prognostic value of inflammatory response-related genes in colon
cancer. The IRGS we constructed for the first time has stable predictive power in predicting the prognosis of patients. These findings provide potential clinical application tools for better management of the prognosis of colon cancer patients.

Disclosure

The authors report no conflicts of interest in this work.

References


