Research Article

Development of an Immune Infiltration-Related Eight-Gene Prognostic Signature in Colorectal Cancer Microenvironment

Beilei Wu, Lijun Tao, Daqing Yang, Wei Li, Hongbo Xu, and Qianggui He

1Medical Examination Center, Wenzhou Central Hospital, The Dingli Clinical College of Wenzhou Medical University, Wenzhou, 325000 Zhejiang, China
2Department of Traumatic Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000 Zhejiang, China
3Department of Colorectal Surgery, Wenzhou Central Hospital, The Dingli Clinical College of Wenzhou Medical University, Wenzhou, 325000 Zhejiang, China

Correspondence should be addressed to Qianggui He; heqianggui163126@163.com

Received 22 February 2020; Revised 23 May 2020; Accepted 16 July 2020; Published 27 August 2020

Copyright © 2020 Beilei Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Stromal cells and immune cells have important clinical significance in the microenvironment of colorectal cancer (CRC). This study is aimed at developing a CRC gene signature on the basis of stromal and immune scores.

Methods. A cohort of CRC patients (n = 433) were adopted from The Cancer Genome Atlas (TCGA) database. Stromal/immune scores were calculated by the ESTIMATE algorithm. Correlation between prognosis/clinical characteristics and stromal/immune scores was assessed. Differentially expressed stromal and immune genes were identified. Their potential functions were annotated by functional enrichment analysis. Cox regression analysis was used to develop an eight-gene risk score model. Its predictive efficacies for 3 years, 5 years, overall survival (OS), and progression-free survival interval (PFI) were evaluated using time-dependent receiver operating characteristic (ROC) curves. The correlation between the risk score and the infiltrating levels of six immune cells was analyzed using TIMER. The risk score was validated using an independent dataset.

Results. Immune score was in a significant association with prognosis and clinical characteristics of CRC. 736 upregulated and two downregulated stromal and immune genes were identified, which were mainly enriched into immune-related biological processes and pathways. An-eight gene prognostic risk score model was conducted, consisting of CCL22, CD36, CPA3, CPT1C, KCNE4, NFATC1, RASGRP2, and SLC2A3. High risk score indicated a poor prognosis of patients. The area under the ROC curves (AUC) s of the model for 3 years, 5 years, OS, and PFI were 0.71, 0.70, 0.73, and 0.66, respectively. Thus, the model possessed well performance for prediction of patients’ prognosis, which was confirmed by an external dataset. Moreover, the risk score was significantly correlated with immune cell infiltration. Conclusion. Our study conducted an immune-related prognostic risk score model, which could provide novel targets for immunotherapy of CRC.

1. Introduction

CRC, as a heterogeneous disease, is a common cause of cancer-related deaths worldwide [1]. TNM staging is usually considered to be one of the main tools for CRC prognosis [2]. However, the prognosis varies greatly among CRC patients with the same TNM stage, suggesting that the current TNM stage does not well provide complete prognostic information for CRC patients. Therefore, it is necessary to adopt a new strategy to increase the predictive efficiency of prognosis and survival outcomes of CRC patients.

Due to the considerable heterogeneity between CRCs, determination of the optimal treatment strategy at the individual level faces the large challenges. Thus, it is an urgent need to conduct robust models to identify high-risk CRC patients and to find novel molecular targets. In the tumor...
microenvironment (TME), stromal and immune cells are involved in the development of CRC [3, 4]. Increasing evidence suggests that stromal and immune cells possess critical clinical significance for CRC. It has been reported that stromal cells can contribute to transcriptome and clinical features of CRC subtype [5]. Furthermore, stromal gene expression can more robustly predict the prognosis of CRC subtypes compared to epithelial tumor cells [6]. In a large cohort of CRC patients, infiltrating immune cell data could better predict patients’ survival than histopathological methods [7]. Growing studies have found that infiltrating immune cells are involved in chemoresistance [8] and metastasis [9]. Thus, it is essential to further analyze the biological characteristics of stromal and immune genes and to determine their prognostic value for CRC patients. However, there is a lack of stromal and immune scores that can predict CRC patients’ prognosis based on multiple clinical factors. Moreover, robust prognostic models on the basis of stromal and immune scores are also lacking.

In this study, we established a reliable prognostic immune-related risk score for CRC. Our results could offer novel insights for prediction of CRC patients’ prognosis and development of individualized immunity therapy strategies.

2. Materials and Methods

2.1. CRC Datasets. TCGA RNA-seq data (including Counts and FPKM data) of GDC CRC (including 469 CRC tissue samples and 41 adjacent normal tissue samples) were downloaded from the xenabrowser website (https://xenabrowser.net/). Among all CRC samples, 433 samples contained complete clinical information, including gender, age, TNM stage, tumor grade, microsatellite instability (MSI), and mismatch repair (MMR). The clinical information of 433 CRC patients is listed in Table 1. Survival information including OS status, OS time, progression-free survival (PFS) status, and PFS time was derived from the pan-cancer on the GDC website (https://gdc.cancer.gov/about-data/publications/PanCan-Clinical-2018). Furthermore, mutation data (including BRAF, KRAS, and TP53) were from CRC MuTect. An overview of the workflow is shown in Figure 1.

2.2. Estimation of Stromal/Immune Scores. ESTIMATE algorithm was used to calculate the stromal/immune scores on the basis of unique expression profiles of stromal/immune cells by the ESTIMATE package in R (https://bioinformatics.mdanderson.org/estimate/) [10].

2.3. Kaplan-Meier Survival Analysis. According to the optimal cutoff of stromal/immune scores, CRC samples were classified into high and low stromal/immune score groups. Kaplan-Meier plot of overall survival between the two groups was analyzed, and the results were evaluated by log-rank test.

2.4. Correlation between Clinical Characteristics and Stromal/Immune Scores. To probe into the clinical significance of stromal/immune scores, we analyzed the correlation between clinical characteristics (including pathologic T, pathologic N, pathologic M, and tumor stage) and stromal/immune scores.

2.5. Differential Expression Analysis. Differential expression analysis between high and low stromal/immune score groups was carried out using the edgeR package in R, following the screening criteria of \( \log_{2} \) fold change (FC) \( >1 \) and FDR (adjusted \( p \) value) \( <0.05 \). Then, up- or downregulated stromal/immune genes were intersected by the VennDiagram package in R, respectively.

2.6. Functional Enrichment Analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of differentially expressed stromal and immune genes were carried out through the clusterProfiler package in R [11]. GO analysis contains three terms, cellular component (CC), molecular function (MF), and biological process (BP), \( p \) value after adjustment \( <0.05 \) was significantly enriched.

2.7. Protein-Protein Interaction (PPI) Analysis. PPI analyses of differentially expressed stromal and immune genes were carried out via The Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/; version 11) [12]. Then, the PPI network was visualized through Cytoscape (version 3.7.2) [13].

2.8. Univariate and Multivariate Cox Regression Analyses. Univariate cox regression analysis of differentially expressed
stromal and immune genes was carried out via the survival package in R. Genes with p value < 0.05 were screened for multivariate cox regression analysis. To validate the sensitivity and accuracy of the risk score for prediction of CRC, an ROC curve was drawn to evaluate the predictive performance of the risk core for 3 years, 5 years, OS, and PFI using the “tdROC” package in R. The results were visualized with the “ggplot2” package in R. The AUC was then calculated. The GSE39582 dataset from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov) was used to validate the prognostic value of the risk score. The dataset was composed of 566 CRC samples.

2.9. Immune Infiltration Analysis. The tumor-immune infiltration cells including B cells, CD4+ T cells, CD8+ T cells, macrophages, neutrophils, and dendritic cells were estimated via the TIMER (https://cistrome.shinyapps.io/timer/) [14]. Spearman’s correlation between the risk score and the infiltrating levels of immune cells was evaluated through the psych package in R. Furthermore, we also assessed the correlation between the genes in the risk score and marker genes of immune cells. The strength of correlation followed the criteria: 0.7 ≤ |r| ≥ 1 suggested a high correlation, 0.3 ≤ |r| < 0.7 suggested a moderate correlation, and 0 < |r| < 0.3 suggested a weak correlation [15].

3. Results

3.1. Immune Score Is in Significant Association with Prognosis and Clinical Features of CRC Patients. According to the optimal cutoff of stromal/immune scores, the CRC patients were divided into two groups. Kaplan-Meier OS analysis results showed that patients with high stromal score had shorter OS time than those with low stromal score; however, it was not statistically significant (Figure 2(a); p value = 0.057). As depicted in Figure 2(b), we found that patients with low immune score implied a poor prognosis (p value = 0.0094). Furthermore, we analyzed the correlation between stromal/immune scores and clinical features. As depicted in the results, stromal score was not significantly associated with pathologic T (Figure 3(a); p value = 0.61), pathologic N (Figure 3(b); p value = 0.28), pathologic M (Figure 3(c); p value = 0.63), tumor stage (Figure 3(d); p value = 0.68), and age (Figure 3(e); p value = 0.76). Similarly, we also found that there was no statistical significance between immune score and pathologic T (Figure 3(f); p value = 0.88) and pathologic N (Figure 3(g); p value = 0.17). As expected, immune score was in significant association with pathologic M (Figure 3(h); p value = 0.0045) and tumor stage (Figure 3(i); p value = 0.0093). However, no significant correlation between immune score and age was found in Figure 3(j) (p value = 0.29). Furthermore, ESTIMATE scores were not correlated with pathologic T (Figure 3(k); p value = 0.98), pathologic N (Figure 3(l); p value = 0.73), pathologic M (Figure 3(m); p value = 0.095), tumor stage (Figure 3(n); p value = 0.28), and age (Figure 3(o); p value = 0.74). These findings indicated that immune score was in significant association with CRC patients’ prognosis and clinical features.

3.2. Identification of Differentially Expressed Stromal and Immune Genes for CRC. We analyzed differentially expressed genes (DEGs) with |log2FC| > 1 and FDR < 0.05 between the high and low stromal/immune score groups. As volcano plots, there were 1197 up- and 28 downregulated stromal genes in the high stromal score group (Figure 4(a)). Furthermore, 899 immune genes were upregulated and eight immune genes were downregulated in the high immune score group (Figure 4(b)). Hierarchical clustering analysis...
Figure 2: The correlation between stromal/immune scores and CRC patients’ survival outcomes. (a) Stromal score. (b) Immune score. The x-axis suggests overall survival time and y-axis represents survival probability.
<table>
<thead>
<tr>
<th>Pathologic_T</th>
<th>Pathologic_N</th>
<th>Pathologic_M</th>
<th>Tumor_stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>N0</td>
<td>M0</td>
<td>I</td>
</tr>
<tr>
<td>T2</td>
<td>N1</td>
<td>M1</td>
<td>II</td>
</tr>
<tr>
<td>T3</td>
<td>N2</td>
<td>MX</td>
<td>III</td>
</tr>
<tr>
<td>T4</td>
<td></td>
<td></td>
<td>IV</td>
</tr>
</tbody>
</table>

Kruskal–Wallis, $p = 0.61$

Kruskal–Wallis, $p = 0.28$

Kruskal–Wallis, $p = 0.63$

Kruskal–Wallis, $p = 0.68$

Figure 3: Continued.
Figure 3: Continued.
Figure 3: Continued.
results showed that both stromal DEGs and immune DEGs could distinguish high stromal/immune score from low stromal/immune score (Figures 4(c) and 4(d)). 736 genes were upregulated both in high stromal and immune scores (Figure 4(e)). Moreover, among eight downregulated immune genes, two genes were also downregulated in the

**Figure 3:** The correlation between stromal/immune/ESTIMATE scores and CRC patients’ clinical features. Stromal score is not associated with (a) pathologic T ($p = 0.61$), (b) pathologic N ($p = 0.28$), (c) pathologic M ($p = 0.63$), (d) tumor stage ($p = 0.68$), and (e) age ($p = 0.76$). Immune score is not correlated with (f) pathologic T ($p = 0.88$) and (g) pathologic N ($p = 0.17$). Immune score significantly associated with (h) pathologic M ($p = 0.0045$) and (i) tumor stage ($p = 0.0093$). Immune score is not correlated with (j) age ($p = 0.29$). ESTIMATE scores are not correlated with (k) pathologic T ($p = 0.98$), (l) pathologic N ($p = 0.73$), (m) pathologic M ($p = 0.095$), (n) tumor stage ($p = 0.28$), and (o) age ($p = 0.74$).
Figure 4: Continued.
Identification of differentially expressed stromal and immune genes for CRC. Volcano plot showing up- (red) and downregulated (blue) stromal/immune genes in the high stromal score group (a) and immune score group (b). Heat maps depicting all DEGs in the high/low stromal score group (c) and high/low immune score group (d). Venn diagram showing common upregulated (e) and downregulated (f) stromal and immune genes.
GeneRatio

p adjust
3.5e − 025
7.0e − 025
1.1e − 024

Extracellular structure organization
Leukocyte migration
Extracellular matrix organization
Regulation of inflammatory response
Leukocyte cell-cell adhesion
Positive regulation of response to external stimulus
Response to molecule of bacterial origin
Cell chemotaxis
Leukocyte chemotaxis
Myeloid leukocyte migration

Count
50
59
68
77
86

GeneRatio

p adjust
1.3e − 012
2.6e − 012
3.9e − 012

Extracellular matrix
Collagen−containing extracellular matrix
External side of plasma membrane
Secretory granule membrane
Receptor complex
Endoplasmic reticulum lumen
Tertiary granule
Collagen trimer
Extracellular matrix component

Count
19
37
56

Figure 5: Continued.
Figure 5: Continued.
We further performed functional enrichment analysis of these common stromal and immune genes. These genes were significantly correlated with inflammatory biological processes like regulation of inflammatory response and pathways such as cytokine-cytokine receptor interaction and chemokine signaling pathway (Figures 5(a)–5(d)). As shown in the PPI network, COL6A2, COL6A1, COL5A2, C1S, and C1R were the top five genes, which were considered hub genes (Figure 5(e)).

3.3. Construction of an Eight-Gene Prognostic Signature for CRC. Among 738 differentially expressed stromal and immune genes, 23 genes were significantly associated with CRC patients’ prognosis according to univariate Cox regression analysis results. Of them, 20 genes were risk factors, and the remaining three (CCL22, CPA3, and MMP1) were protective factors (Table 2). These genes were used for multivariate Cox regression analysis. Finally, an eight-gene signature was constructed for CRC. The risk score was calculated on the basis of the coefficients and expression values of the eight genes. All CRC patients were divided into two groups in accordance with the median value of risk score (Figure 6(a)). Heat maps depicted the difference in expression patterns of the eight genes (including CD36, KCNE4, CPT1C, SLC2A3, RASGRP2, NFATC1, CCL22, and CPA3).

Figure 5: Functional enrichment and PPI analyses of common stromal and immune DEGs. (a) The top the BP terms of GO enrichment analysis. (b) The top ten CC terms of GO enrichment analysis. (c) The top ten MF terms of GO enrichment analysis. (d) The top ten KEGG pathway analysis. (e) PPI network construction.

between the high and low risk scores (Figure 6(b)). As shown in Figure 6(c), the risk score was capable of predicting CRC patients’ prognosis. High risk score implied a poor prognosis (p value < 0.0001). Among the eight genes, KCNE4 and CCL22 were protective factors of CRC, while CD36, CPT1C, SLC2A3, RASGRP2, NFATC1, and CPA3 were risk factors of CRC, as shown in the forest diagram (Figure 6(d)). We further validated the sensitivity and accuracy of the model. AUCs of the model for 3 years, 5 years, OS, and PFI were 0.71, 0.70, 0.73, and 0.66, respectively (Figures 6(e) and 6(f)). Thus, the risk score model could well predict CRC patients’ prognosis, with high sensitivity and accuracy. As shown in the multivariate Cox regression analysis of these genes between CRC samples and normal samples were visualized (Figures 8(a)–8(g)). CD36 (p value < 2.22e-16), CPA3 (p value < 2.22e-16), NFATC1 (p value = 9.1e-08), and RASGRP2 (p value < 2.22e-16) were highly expressed and SLC2A3 (p value = 0.0015) was lowly expressed in tumor samples. As shown in Figures 9(a)–9(h), low expression of CCL22 (p value = 0.0047) and CPA3 (p value = 0.035) indicated shorter OS time than high expression. Moreover, we found that highly expressed CPT1C (p value = 0.0017), KCNE4 (p value = 0.002), and SLC2A3 (p value = 0.0048) was significantly correlated with poor PFS (Figures 9(i)–9(p)).

3.5. The Eight-Gene Model Is in Significant Correlation with Immune Cell Infiltration. The correlation between the model and the infiltrating levels of six immune cells was analyzed. We found that the model was in significant association with the infiltrating levels of six immune cells, including B cell (Figure 10(a); R = 0.13, p value = 0.0064) and CD4+ T cell (Figure 10(b); R = 0.21, p value = 4.3e-06). However, no significant correlation between the model and CD8+ T cell was found in Figure 10(c) (R = 0.045, p value = 0.34). Furthermore, there were distinct correlations between the model and dendritic cell (Figure 10(d); R = 0.12, p value = 0.0072), macrophage (Figure 10(e); R = 0.19, p value = 3.3e-05), neutrophil (Figure 10(f); R = 0.18, p value = 9.4e-05). We also found that the expression levels of the eight genes in the model were significantly correlated with the infiltrating levels of six immune cells, including CCL22 (Figure 11(a)), CD36 (Figure 11(b)), CPA3 (Figure 11(c)), CPT1C (Figure 11(d)), KCNE4 (Figure 11(e)), NFATC1 (Figure 11(f)), RASGRP2 (Figure 11(g)), and SLC2A3 (Figure 11(h)). Moreover, the eight genes were in significant association with markers of immune cells (Supplementary Table 1). These results suggested that the model could be in association with immune cell infiltration.

3.6. Validation of the Risk Score Using an Independent Dataset. Based on 566 CRC samples from the GSE39582 dataset, the prognostic value of the risk score was validated. The risk score distribution and survival status of CRC patients are shown in Figure 12(a). Heat maps showed the expression differences of CD36, KCNE4, CPT1C, SLC2A3, RASGRP2, NFATC1, CCL22, and CPA3 between the high and low risk scores (Figure 12(b)). As expected, CRC patients with high risk score had a poorer prognosis than those with low risk score (Figure 12(c)). Among the eight genes, CD36, NFATC1, and CCL22 were significantly associated with prognosis of CRC patients (Figure 12(d)). AUCs of the model for 3 years and 5 years were 0.65 and 0.66, respectively (Figure 12(e)), indicating that the risk score could well predict CRC patients’ prognosis. The expression levels of CCL22 (Figure 13(a)), CD36 (Figure 13(b)), CPA3 (Figure 13(c)), CPT1C (Figure 13(d)), KCNE4 (Figure 13(e)), NFATC1 (Figure 13(f)), RASGRP2 (Figure 13(g)), and SLC2A3

Table 2: Univariate Cox regression analysis results of 23 differentially expressed stromal and immune genes.

<table>
<thead>
<tr>
<th>Variables</th>
<th>HR</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD36</td>
<td>1.381</td>
<td>1.046</td>
<td>1.823</td>
<td>0.02284</td>
</tr>
<tr>
<td>KCNE4</td>
<td>1.346</td>
<td>1.004</td>
<td>1.805</td>
<td>0.047303</td>
</tr>
<tr>
<td>VEGFC</td>
<td>1.39</td>
<td>1.011</td>
<td>1.912</td>
<td>0.042907</td>
</tr>
<tr>
<td>PDE1B</td>
<td>2.242</td>
<td>1.361</td>
<td>3.693</td>
<td>0.001518</td>
</tr>
<tr>
<td>BGN</td>
<td>1.166</td>
<td>1.003</td>
<td>1.355</td>
<td>0.046135</td>
</tr>
<tr>
<td>CPT1C</td>
<td>2.561</td>
<td>1.476</td>
<td>4.446</td>
<td>8.26E-04</td>
</tr>
<tr>
<td>GPX3</td>
<td>1.231</td>
<td>1.03</td>
<td>1.471</td>
<td>0.022285</td>
</tr>
<tr>
<td>NGFR</td>
<td>1.344</td>
<td>1.014</td>
<td>1.779</td>
<td>0.039388</td>
</tr>
<tr>
<td>SERPINE1</td>
<td>1.173</td>
<td>1.01</td>
<td>1.362</td>
<td>0.037066</td>
</tr>
<tr>
<td>CHST1</td>
<td>1.448</td>
<td>1.02</td>
<td>2.055</td>
<td>0.038579</td>
</tr>
<tr>
<td>FBLN7</td>
<td>2.648</td>
<td>1.003</td>
<td>6.994</td>
<td>0.049317</td>
</tr>
<tr>
<td>KCN8</td>
<td>1.372</td>
<td>1.003</td>
<td>1.876</td>
<td>0.047805</td>
</tr>
<tr>
<td>SLC2A3</td>
<td>1.225</td>
<td>1.007</td>
<td>1.49</td>
<td>0.041957</td>
</tr>
<tr>
<td>CD72</td>
<td>1.714</td>
<td>1.127</td>
<td>2.606</td>
<td>0.011724</td>
</tr>
<tr>
<td>APLP1</td>
<td>1.631</td>
<td>1.031</td>
<td>2.58</td>
<td>0.036535</td>
</tr>
<tr>
<td>SIGLEC1</td>
<td>1.348</td>
<td>1.019</td>
<td>1.783</td>
<td>0.036725</td>
</tr>
<tr>
<td>RASGRP2</td>
<td>1.641</td>
<td>1.044</td>
<td>2.579</td>
<td>0.031925</td>
</tr>
<tr>
<td>SPHK1</td>
<td>1.218</td>
<td>1.002</td>
<td>1.481</td>
<td>0.048069</td>
</tr>
<tr>
<td>NFATC1</td>
<td>1.674</td>
<td>1.157</td>
<td>2.421</td>
<td>0.006228</td>
</tr>
<tr>
<td>LRRN2</td>
<td>1.653</td>
<td>1.139</td>
<td>2.398</td>
<td>0.008126</td>
</tr>
<tr>
<td>CCL22</td>
<td>0.686</td>
<td>0.51</td>
<td>0.923</td>
<td>0.012756</td>
</tr>
<tr>
<td>CPA3</td>
<td>0.81</td>
<td>0.68</td>
<td>0.966</td>
<td>0.019315</td>
</tr>
<tr>
<td>MMP1</td>
<td>0.902</td>
<td>0.816</td>
<td>0.996</td>
<td>0.042396</td>
</tr>
</tbody>
</table>
Figure 6: Continued.
Figure 6: Continued.
Hazard ratio

<table>
<thead>
<tr>
<th>Gene</th>
<th>Hazard Ratio</th>
<th>CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD36</td>
<td>1.44</td>
<td>(1.01 - 2.05)</td>
<td>0.043*</td>
</tr>
<tr>
<td>KCNE4</td>
<td>0.57</td>
<td>(0.32 - 0.99)</td>
<td>0.047*</td>
</tr>
<tr>
<td>CPT1C</td>
<td>2.19</td>
<td>(1.04 - 4.59)</td>
<td>0.039*</td>
</tr>
<tr>
<td>SLC2A3</td>
<td>1.43</td>
<td>(1.04 - 1.95)</td>
<td>0.027*</td>
</tr>
<tr>
<td>RASGRP2</td>
<td>1.97</td>
<td>(1.06 - 3.65)</td>
<td>0.032*</td>
</tr>
<tr>
<td>NFATC1</td>
<td>1.77</td>
<td>(1.09 - 2.86)</td>
<td>0.02*</td>
</tr>
<tr>
<td>CCL22</td>
<td>0.47</td>
<td>(0.31 - 0.72)</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>CPA3</td>
<td>0.74</td>
<td>(0.57 - 0.94)</td>
<td>0.016*</td>
</tr>
</tbody>
</table>

# events: 95; Global p value (log-rank): 1.3005e−07
AIC: 978.3; concordance index: 0.72

Figure 6: Continued.
between the high risk score and low risk score were validated based on the 566 CRC samples. Univariate Cox regression analysis results showed that age, KRAS mutation, pathologic T, pathologic N, pathologic M, tumor stage, and risk score were notably associated with CRC patients’ prognosis. After multivariate Cox regression analysis, we found that age, KRAS mutation, pathologic M, and risk score could be independent prognostic factors for CRC (Table 4).

3.7. The Eight Genes in the Risk Score Are Distinctly Correlated with Molecular Markers of CRC Prognosis. In...
Figure 7: Continued.
Figure 14(a), CCL22 was significantly correlated with BRAF mutation ($p$ value = 0.014) and KRAS mutation ($p$ value = 0.041). For CD36, there was a distinct correlation between its expression and KRAS mutation ($p$ value = 0.00034) and MMR ($p$ value = 0.025) in Figure 14(b). CPA3 ($p$ value = 0.0066; Figure 14(c)) and CPT1C ($p$ value = 0.005; Figure 14(d)) had higher expression levels in KRAS mutation. As shown in Figure 14(e), KCNE4 expression was in significant correlation with BRAF mutation ($p$ value = 0.0014), KRAS mutation ($p$ value = 0.049), and MSI ($p$ value = 0.05). NFATC1 expression was prominently correlated with BRAF mutation ($p$ value = $2.2 \times 10^{-11}$), KRAS mutation ($p$ value = 0.00051) and MSI ($p$ value = $1.1 \times 10^{-13}$) in Figure 14(f). In Figure 14(g), RASGRP2 expression was significantly decreased in KRAS mutation. For SLC2A3, we found that there was a distinct correlation between its expression and BRAF mutation ($p$ value = 0.0011) and MSI ($p$ value = 0.00013) in Figure 14(h).
Figure 8: Continued.
4. Discussion

In TME, stromal and immune cells are involved in the development of CRC. In this study; using the ESTIMATE algorithm, stromal and immune scores were calculated. A significant correlation between the immune score and CRC patients’ prognosis was observed. Both the stromal score and immune score were in significant correlation with clinical characteristics of CRC patients. Furthermore, we identified differentially expressed stromal and immune genes for CRC. Functional enrichment analysis results suggested that these genes were positively related with immune-related pathways like cytokine-cytokine receptor interaction [16, 17] and chemokine signaling pathway [18, 19].

Figure 8: Box plots showing the expression patterns of the eight genes in the risk score model between CRC samples and normal samples. (a) CCL22. (b) CD36. (c) CPA3. (d) CPT1C. (e) KCNE4. (f) NFATC1. (g) RASGRP2. (h) SLC2A3.
Figure 9: Continued.
\textbf{Figure 9:} Continued.
Figure 9: Continued.
Figure 9: Continued.
Survival probability

Number at risk

Strata

(i)

Number at risk

Strata

(j)

Figure 9: Continued.
Figure 9: Continued.
Figure 9: Continued.
Figure 9: Kaplan-Meier plot of OS and PFS for the eight genes in the risk score model between high and low risk score. Kaplan-Meier OS results for (a) CCL22, (b) CD36, (c) CPA3, (d) CPT1C, (e) KCNE4, (f) NFATC1, (g) RASGRP2, and (h) SLC2A3. Kaplan-Meier PFS results for (i) CCL22, (j) CD36, (k) CPA3, (l) CPT1C, (m) KCNE4, (n) NFATC1, (o) RASGRP2, and (p) SLC2A3.
Figure 10: The correlation between the eight-gene risk score and immune cell levels. (a) B cell. (b) CD4+ T cell. (c) CD8+ T cell. (d) Dendritic cell. (e) Macrophage. (f) Neutrophil.
Figure 11: Continued.
Individual prognosis for CRC patients varies widely. Individual genes often cannot accurately predict the prognosis of patients with CRC. Genes in most prognostic risk scores are screened via differential expression analyses [20–22]. Yet, there are few prognostic models associated with CRC immune infiltration. Therefore, in this study, we selected eight differentially expressed stromal and immune genes related to prognosis for constructing a risk score model. However, focusing only on CRC-related immune-related genes may limit its clinical value. For this reason, through multivariate regression analysis, after adjustment of the clinical characteristics of CRC, we assessed the association between the risk score and CRC prognosis. The results showed that the model may become an independent prognostic factor for CRC. Our risk score exhibited well efficiency in predicting CRC patients’ prognosis. Therefore, the risk score model possessed potential prognostic value, which was confirmed using an independent dataset. Among the eight genes in the model, both in the discovery and independent datasets, CCL22 was a protective factor of CRC, while CD36 and NFATC1 were two risk factors of CRC. However, other genes exhibited inconsistent results in the two datasets. This is partly due to the heterogeneity of the samples in the two datasets. Patients in the same pathological stage have different prognosis. Both in the discovery and independent datasets, CCL22 and CPA3 were lowly expressed and KCNE4, NFATC1, and SLC2A3 were highly expressed in the high-risk samples compared to the low-risk samples. However, there were inconsistent results about other genes between the high- and low-risk samples in the two datasets, partly due to the heterogeneity of the samples, different sequencing platforms, different background correction and normalization methods and so on. Thus, it is unreliable to predict CRC patients’ prognosis by an individual gene. However, our risk score composed of these genes may accurately suggest the patient’s prognosis.

As described in a previous study, high CCL22 expression was found in CRC tissues [23]. Recent study has found that CCL22 secreted by M2 macrophages could mediate CRC 5-FU-mediated chemoresistance [24]. Furthermore, it has been reported that CCL22 was in significant correlation with the infiltrating levels of different T cell subsets for CRC [25]. Our results showed that CD36 was significantly downregulated in CRC tissues compared to normal tissues, which was validated in vitro and in vivo [26]. Genome-wide DNA methylation analysis revealed that hypermethylation of CD36 could contribute to its low expression [27]. Fang et al. found that CD36 expression gradually decreased from adenoma to cancer and CD36 loss implied a poor prognosis in patients with CRC [28]. NFATC1 was deregulated in CRC tissues, which was consistent with previous findings [29]. In vitro, its overexpression significantly promoted CRC cell invasion and metastasis [30]. Kumar et al. reported that NFATC1 indicated poor survival outcomes of CRC patients [31]. High SLC2A3 expression was observed in CRC tissues and its high expression indicated a poor prognosis, consistently with previous research [32, 33]. Furthermore, downregulated CPA3 and RASGRP2 and upregulated CPT1C and KCNE4 were found in CRC tissues, which implied poor prognosis.

As for immune cell infiltration, we found that the eight genes in the risk score model were moderately correlated with the infiltrating levels of CD4+ T cell, dendritic cell, macrophage, and neutrophil. It has been confirmed that TME affects the efficacy of immunotherapy, and immune cells in TME possess predictive value for immunotherapy treatment [34–36]. Increasing genes have been shown to participate in the regulation of immune cells [37–40]. Therefore, our risk score model could possess potential value to predict CRC.
Figure 12: Continued.
Figure 12: Continued.
Figure 12: Validation of the eight-gene prognostic signature for CRC using an independent dataset. (a) Risk score distribution and survival status. (b) Heat maps showing the expression patterns of the eight genes between high- and low risk score. (c) Kaplan-Meier survival analysis of the model. (d) Forest plot of the eight genes for CRC. (e) ROC curve of the model for 3 years and 5 years.
patients’ prognosis, and the eight genes could become promising immunotherapeutic targets, which deserve further study.

Our correlation analysis results confirmed that the eight genes in the risk score were distinctly correlated with molecular markers of CRC prognosis. However, our study has several limitations. First, our retrospective study limited the application of this risk score. Second, the heterogeneity of the immune microenvironment would inevitably contribute to result bias. Therefore, it is necessary to validate our findings in a prospective clinical study.

5. Conclusion

In this study, we conducted an immune-related prognostic model for CRC on the basis of stromal and immune scores.
Table 4: Univariate and multivariate Cox regression analyses in the GSE39582 dataset.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>p value</td>
</tr>
<tr>
<td>Age</td>
<td>1.455 (1.033-2.051)</td>
<td>0.031</td>
</tr>
<tr>
<td>Sex</td>
<td>1.310 (0.980-1.750)</td>
<td>0.068</td>
</tr>
<tr>
<td>BRAF mutation</td>
<td>1.111 (0.664-1.861)</td>
<td>0.688</td>
</tr>
<tr>
<td>KRAS mutation</td>
<td>1.361 (1.018-1.818)</td>
<td>0.037</td>
</tr>
<tr>
<td>TP53 mutation</td>
<td>1.197 (0.844-1.697)</td>
<td>0.312</td>
</tr>
<tr>
<td>MMR status</td>
<td>0.768 (0.471-1.252)</td>
<td>0.290</td>
</tr>
<tr>
<td>Pathologic T</td>
<td>2.005 (1.025-3.921)</td>
<td>0.042</td>
</tr>
<tr>
<td>Pathologic N</td>
<td>1.648 (1.228-2.213)</td>
<td>0.0008</td>
</tr>
<tr>
<td>Pathologic M</td>
<td>5.175 (3.621-7.395)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Tumor stage</td>
<td>1.767 (1.326-2.354)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Risk score</td>
<td>2.718 (2.056-3.595)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

NA: not available.
Figure 14: Continued.
The model had well predictive efficacy for CRC patients' prognosis. Our findings could provide novel biomarkers for predicting the prognosis of CRC patients and developing individualized immunity therapy strategies.

**Abbreviations**

CRC: Colorectal cancer  
TCGA: The Cancer Genome Atlas
References


