Identification and validation of a prognostic signature based on six immune-related genes for colorectal cancer

Lifeng Zheng
the Nanjing Jiangbei Hospital

Ziyu Xu
the Nanjing Jiangbei Hospital

Wulou Zhang
the Nanjing Jiangbei Hospital

Shu Zhou
the Nanjing Jiangbei Hospital

Hao lin
the Nanjing Jiangbei Hospital

Yepeng Zhang
the Nanjing Jiangbei Hospital

Zonghang Liu
the Nanjing Jiangbei Hospital

Xi Gu
xigu19930327@163.com
the Nanjing Jiangbei Hospital

Research Article

Keywords: Colorectal cancer, Immune genes, Risk signature, Prognosis

Posted Date: September 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3277656/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Abstract

Background

Colorectal cancer (CRC) is a prevalent malignancy with high death and morbidity rates. Even though the significant efficacy of immunotherapy is well-established, it is only beneficial for a limited number of individuals with CRC.

Methods

Differentially expressed immune-related genes (DE-IRGs) were retrieved from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and Immport databases. The prognostic signature comprising DE-IRGs was developed by means of Univariate, Lasso, and Multivariate Cox-regression analyses. Following this, a nomogram integrating the independent prognostic factors was also developed. CIBERSORT was applied to ascertain the immune cell infiltration (ICI). Furthermore, wound healing, colony formation, migration, and invasion assays were executed to study the involvement of ACTG1 in CRC.

Results

A signature including six DE-IRGs was developed. It could estimate the rate of overall survival (OS) accurately for the TCGA and GSE38832 cohorts. The risk score (RS) of the signature was an independent factor for OS. Moreover, a nomogram encompassing age, RS, and pathological T could accurately predict the long-term OS probability of individuals with CRC. The high-risk group had an elevated proportion of ICI relative to the low-risk group, including native B cells. Additionally, ACTG1 expression was upregulated, which supported the proliferation, migration, and invasion abilities of CRC cells.

Conclusions

An immune-related prognostic signature was developed for predicting OS and to ascertain the immune status in individuals with CRC. It could provide new insights into accurate immunotherapy for individuals with CRC. Moreover, ACTG1 can possibly serve as a new immune biomarker.

Introduction

The incidence and death rates of colorectal cancer (CRC) rank second among all malignant tumors globally\(^1\). CRC incidence and death rates are increasing annually in China\(^2\). It is a multifactorial disease with a genetic component and unhealthy habits such as alcohol abuse, smoking, physical inactivity, and obesity acting as risk factors\(^3\). Regarding treatment, surgery is commonly recommended for early-stage
patients with CRC. However, despite surgery, approximately 20% of patients with advanced CRC report recurrence and metastasis. For such patients, neoadjuvant chemotherapy, radiotherapy, and immunotherapy have been used with significant therapeutic benefits. Nonetheless, these secondary treatments do have side effects. Therefore, patients with CRC should undergo personalized systemic treatment regimens, which could both improve quality of life and prolong survival.

Over the previous decade, immunotherapy has been extensively explored and subsequently established as an effective therapy for CRC with microsatellite instability. Checkpoint inhibitors are one of the immunotherapeutic approaches, and checkpoint inhibitors targeting programmed cell death 1 (PD1) or programmed cell death-ligand 1 (PD-L1) are typically recommended for mismatch repair-deficient metastatic CRC cases. Recently, Cercek et al. found that dostarlimab, a PD-1 inhibitor, was effective in mismatch repair-deficient, locally advanced rectal cancer. In the study, all 12 patients who received single-agent dostarlimab had a clinical complete response for at least six months. This indicates that immunotherapy can lead to long-term remission for certain patients with CRC. Hence, precision immunotherapy is essential for investigating the molecular attributes of immunotherapy.

Moreover, recent studies have found that immune-related genes (IRGs) can regulate the immune system to promote or suppress tumor development. Prediction tools based on IRGs have been reported to reflect tumor progression and predict patient survival with better accuracy than existing tools. Such prediction tools can also aid in identifying effective treatments that can better balance side effects and survival benefits for patients with CRC. For instance, Wang et al. established an immune-linked prognostic signature comprising eight genes that could depict the dysregulated tumor immune microenvironment in CRC. Similarly, Song et al. constructed a seven circRNA-related signature that could assess survival and reflect the immune status of individuals with CRC. However, a reliable signature for CRC is still rare, and most of those already developed lack validation. Thus, this study aims to develop a better and more accurate signature than those previously reported.

**Materials and methods**

**Data source**

Gene expression profiles of six GEO datasets (GSE90524, GSE84983, GSE134525, GSE104364, GSE109454, and GSE115856) were obtained from the GPL16956 platform. Additionally, the GSE38832 chip data with survival time were accessed at GEO (https://www.ncbi.nlm.gov/geo/). IRGs were derived utilizing ImmPort (http://www.immport.org). Additionally, the clinical information and transcriptomic data of 546 individuals with CRC were retrieved from TCGA (http://portal.gdc.cancer.gov). Furthermore, the protein expression data of six DE-IRGs were acquired from the human protein atlas (HPA) (http://www.proteinatlas.org/) database.
The major sources of bias and variability in integrating microarray data are heterogeneity and potential variables. Thus, to eliminate these biases, the data were integrated and batch-normalized with the help of the 'sva' and 'limma' packages in R software, respectively. Following this, the 'limma' package assisted in the identification of DEGs. The log2 |fold change (FC)| >1 and adjusted $P< 0.05$ were deemed statistically significant. Finally, DE-IRGs were acquired by overlapping IRGs and DEGs and then visualized using the 'gplots' package.

**Bioinformatic analysis of DE-IRGs**

Gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were executed with the aid of the packages 'org.Hs.eg.db' and 'clusterProfiler' in R software.

**Establishment and validation of the prognostic risk model**

Following the exclusion of 42 individuals with an unsatisfactory follow-up period (less than 30 days), a total of 506 individuals from TCGA were screened for further analysis. The individuals were randomly classified into training ($n = 253$) and testing ($n = 253$) cohorts. The clinicopathological attributes of the included individuals are presented in Table 1. The training cohort aided in the identification of DE-IRGs related to OS. The univariate Cox regression analysis distinguished prognostic DE-IRGs. Next, the lasso-penalized regression analysis was applied to decrease the number of prognostic DE-IRGs with the help of the 'glmnet' package. Finally, multivariate Cox regression was applied to screen optimal prognostic DE-IRGs. The risk score (RS) of each individual was computed utilizing the coefficient and mRNA expression of prognostic DE-IRGs as stated: $RS = \sum_{i=1}^{N} (\text{Expi} \times \text{Coei})$ (N: gene number; Expi: gene expression level; Coei: coefficient value).
Table 1
Clinical characteristics of 506 colorectal cancer patients.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Group</th>
<th>Total cohort</th>
<th>Training cohort</th>
<th>Testing cohort</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>229</td>
<td>121</td>
<td>108</td>
<td>0.246</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>277</td>
<td>132</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 65</td>
<td>207</td>
<td>99</td>
<td>108</td>
<td>0.416</td>
</tr>
<tr>
<td></td>
<td>≥ 65</td>
<td>299</td>
<td>154</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td>I</td>
<td>90</td>
<td>43</td>
<td>47</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>187</td>
<td>97</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>140</td>
<td>67</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>73</td>
<td>38</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknow</td>
<td>16</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Vital status</td>
<td>Dead</td>
<td>92</td>
<td>45</td>
<td>47</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>Alive</td>
<td>414</td>
<td>208</td>
<td>206</td>
<td></td>
</tr>
</tbody>
</table>

The median of the RSs was selected as the cut-off threshold to segregate individuals into high and low-risk groups. Survival analysis between the high and low-risk groups was conducted with the aid of the 'survival' and 'survminer' packages. Furthermore, to assess risk model accuracy, receiver operating characteristic (ROC) analysis was executed to obtain the 1-, 3-, and 5-year area under the curves (AUC) with the help of the 'survivalROC' package. Additionally, the entire TCGA cohort, the testing cohort, and GSE38832 were utilized to validate the model. Univariate and multivariate Cox regression analyses aided the evaluation of the independent prognostic value of RS and clinical attributes in the entire TCGA cohort. Prognostic nomograms were generated using Cox regression coefficients with the 'rms' package. Moreover, calibration plots assessing the performance of these nomograms were illustrated with the help of the 'regplot' package.

**Immune cell infiltration (ICI)**

Tumour ICI was determined with the help of the CIBERSORT algorithm, based on the normalized gene expression data as well as the annotated gene signature matrix outlining 22 immune cell subtypes. The data were retrieved from the CIBERSORT web portal (http://cibersort.stanford.edu/), and the perm was set to 1000. Samples having *P* < 0.05 were deemed significant and utilized for subsequent investigation.

**Sample collection**

A total of eight individuals with CRC were selected from the Nanjing Jiangbei Hospital. The patients had a definite diagnosis of CRC by a professional pathologist. The tumor and adjacent non-tumor tissues were extracted via surgical resection and subsequently preserved in liquid nitrogen during
transportation. This was followed by refrigeration at -80°C until further analysis. This research was under approval by the Ethics Committee of Nanjing Jiangbei Hospital. Informed consent was acquired from all the individuals involved.

Real-time polymerase chain reaction (RT-PCR) assay

Trizol Reagent (Invitrogen, CA, USA) aided the extraction of total RNA from CRC tissues and adjacent non-tumor tissues. High-Capacity cDNA Reverse Transcription Kit (ThermoFisher, USA) was utilized in the reverse transcription of total RNA to complementary DNA (cDNA) strands. RT-PCR was conducted in an LC480II (Roche, Switzerland) thermocycler using an SYBR Green kit (ThermoFisher, USA). The primer pairs of the six DE-IRGs are listed in Table S1. Additionally, the $2^{-\Delta\Delta Ct}$ value was computed to determine gene expression levels, and glyceraldehyde-3-phosphate dehydrogenase served as a negative control.

Culturing of cells and transient transfection

Cell lines SW480, SW620, HT29, and LOVO, were subjected to incubation under 5% CO$_2$ and at 37°C in Roswell Park Memorial Institute medium or Dulbecco’s modified Eagle’s medium (DMEM) (Gibco BRL, Rockville, Maryland, USA) with 10% fetal bovine serum. For transient transfection, 1×10$^5$ cells/ml of SW480 and SW620 cells were seeded in 24-well plates prior to the day of transfection. This was followed by cell incubation for 24 h to allow for 60% confluence. These cells were then transfected with short hairpin RNA (shRNA) to ACTG1 (Sangon Biotech, Shanghai, China) with the help of Lipofectamine 2000 (Invitrogen, Carlsbad, USA) and subsequently subjected to incubation for 48 h. The target sequences for shRNAs were listed below:

ACTG1-shRNA1: GCTGGCAAGAACCAGTTGTTT;
ACTG1-shRNA2: CCGAGCCGTGTTTCCTTCCAT;
ACTG1-shRNA3: CGCATCCTCCTCTTCTCTGGA.

Colony formation and cell proliferation assay

SW480 and SW620 (5×10$^4$ cells/well) were inoculated into 96-well plates. Following this, cells were transfected with shRNA NC or ACTG1-shRNA1. Six replicate wells were set in each group, and 10 µl 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2-H-tetrazolium bromide (MTT) solution was added to each well at 24, 48, and 72 h post-transfection. Each well was measured spectrophotometrically at 492 nm after incubation for 2 h. For the colony formation assay, 500 SW480 and SW620 cells transfected with shRNA NC and ACTG1-shRNA1 were inoculated into each well of a 6-well plate. After 14 days, the colonies were fixed with methanol and stained with 0.1% crystal violet to facilitate counting the number of colonies. Moreover, all assays were repeated thrice.

Cell invasion and migration assay

Matrigel pre-coated Transwell inserts were placed into 24-well plates. The plates were then filled with 600 µl DMEM. A total of 1×10$^6$ of SW480 and SW620 cells transfected with shRNA NC or ACTG1-
shRNA1 were added to the upper chamber of the Transwell units. Cells that traversed the membrane between the upper and lower chambers of the units were fixed with methanol for 30 mins, followed by staining with 0.1% crystal violet. Wet cotton swabs were used to remove the cells that did not traverse the membrane. Finally, the plates were photographed in three different fields.

**Wound healing assay**

Approximately $1 \times 10^6$ of SW480 and SW620 cells transfected with shRNA NC or ACTG1-shRNA1 were added into 6-well plates. Following this, a 1-mm-wide scratch was made across the cell layer using a sterile pipette tip. The suspension cells were discarded, and images at a bright field were used as a control. The plates were photographed at 0 and 24 h post-scratching at an identical location, with the width of the scratch also measured.

**Statistical analyses**

All statistical analyses were executed with the help of the R software (version 4.0.5). The Kaplan–Meier survival curves were compared utilizing the log-rank test. The false discovery rate correction was applied to adjust the $P$-value for multiple tests. The significance of gene expression across the normal and tumor tissues was assessed with the help of the Student’s t-test. Univariate and multivariate Cox regression analyses were performed to identify independent prognostic factors related to OS. $P < 0.05$ was deemed statistically significant.

**Results**

**Detection of DE-IRGs**

The design of the current research is presented as a flowchart (Fig. 1). Six datasets (GSE90524, GSE84983, GSE134525, GSE104364, GSE109454, and GSE115856) were downloaded, and their platform information was retrieved from GEO (Table S2). All the datasets had 32,386 genes after reannotation except for GSE84983. Moreover, there were 2557 genes in GSE84983, which indicated that it was not consistent with the other datasets. Hence, it was excluded owing to the limited number of genes obtained. Following this, DEGs analysis was carried out in each dataset. Five datasets of DEGs (GSE90524, GSE134525, GSE104364, GSE109454, and GSE115856) comprising 3116, 2434, 2037, 3187, and 2029 DEGs, respectively, were identified between tumor and adjacent non-tumor tissues (Fig. 2a; Fig. S1a-e). Furthermore, integrated analysis with a batch normalization method was performed to merge five datasets into one. A total of 1610 DEGs, including 1146 upregulated and 464 downregulated genes, were identified in the merged dataset (Fig. 2b). DEGs overlapped with IRGs, which revealed 118 DE-IRGs, including 73 upregulated and 45 downregulated genes (Fig. 2c). The top 10 upregulated and downregulated DE-IRGs are listed in Table S3.

**Functional enrichment analysis of DE-IRGs**
GO and KEGG analyses aided in the identification of the likely role of DE-IRGs. The findings of the GO analysis indicated that the most enriched terms of cellular components, biological processes, and molecular functions were secretory granule lumen, leukocyte migration, and receptor-ligand activity, respectively. The most significant enrichment pathway of KEGG analysis was the cytokine-cytokine receptor interaction. The top ten most enriched terms of GO and KEGG analysis are visualized and presented in Figs. 2d-e.

Establishment of the risk model

The univariate Cox regression analysis aided the identification of DE-IRGs that were related to OS. Using the training cohort, 15 DE-IRGs were obtained (Fig. 3a) which were reduced to 12 through lasso-cox regression analysis (Fig. 3b-c). After multivariate cox regression, six optimal DE-IRGs such as C-X-C motif chemokine ligand 1 (CXCL1), catalase (CAT), gastrin-releasing peptide (GRP), actin gamma 1 (ACTG1), raf-1 proto-oncogene, serine/threonine kinase (RAF1), and peptidoglycan recognition protein 2 (PGLYRP2) were selected to develop the prognostic risk model (Fig. 3d). The RS of each individual was computed with the equation below: \[ RS = (-0.32029 \times \text{expression of CXCL1}) + \text{(0.84476 \times \text{expression of CAT}) + (0.67235 \times \text{expression of PGLYRP2}) + (0.33094 \times \text{expression of GRP}) + (-1.4025 \times \text{expression of ACTG1}) + \text{(-1.0255} \times \text{expression of RAF1).} \]

Individuals in the training cohort were classified into high (n = 126) and low (n = 127) risk groups per their median value of the RS. The Kaplan–Meier survival curve indicated that the individuals with higher risk had poor OS relative to the lower-risk individuals (\( P = 1.159e-05 \), Fig. 3e). RS distribution curve, survival status plot, and heatmap of the risk groups are illustrated in Figs. 3f-h. The respective AUCs of the training cohort at 1, 3, and 5 years were 0.799, 0.803, and 0.828 (Fig. 3i). These findings demonstrated the robust prediction ability of the six IRG-based model concerning the prognosis of individuals with CRC.

Verification of the model

To investigate the model concerning its accuracy, the RS of each individual in the testing cohort, the entire TCGA cohort, and GSE38832 were computed. Individuals were classified into high and low-risk groups per the median value of the RS in each cohort. In the three validation cohorts, high-risk individuals had poor OS relative to the low-risk individuals (testing cohort: \( P = 4.463e-01 \), Fig. 4a; entire TCGA cohort: \( P = 2.859e-04 \), Fig. 4b; GSE38832: \( P = 1.755e-01 \), Fig. 4c). RS distribution curve, survival status plot and heatmap of the risk groups in the three validation cohorts are depicted in Fig. 4d-l. The outcomes demonstrated that mortality was elevated in the high-risk group relative to the other group. Moreover, the gene expression profiles of the validation cohorts were similar to that of the training cohort. In the testing cohort, the AUCs at 1, 3, and 5 years were 0.561, 0.553, and 0.679, respectively (Fig. 4m). Concerning the entire TCGA cohort, the respective AUCs at 1, 3, and 5 years were 0.681, 0.680, and 0.753 (Fig. 4n). Furthermore, in GSE38832, the respective AUCs at 1, 3, and 5 years were 0.656, 0.671, and 0.677 (Fig. 4o). Thus, the findings of the validation cohorts were consistent with that of the training cohort, thereby demonstrating the predictive accuracy of the risk model.
Prognostic evaluation and clinical utility of the model

Univariate and multivariate Cox regression analyses facilitated ascertaining the prognostic value of the model (Fig. 5a-b). The results revealed age, pathological T, and RS as independent prognostic factors for OS in the entire TCGA cohort \((P < 0.05)\). The respective AUCs of age, gender, stage, and RS at 5 years were 0.643, 0.459, 0.752, and 0.753, which implied that RS was the most accurate concerning the 5-year OS prediction (Fig. 5c). On comparing the expression of the six genes with clinical features, ACTG1 expression was observed to be decreased when the TNM stage increased \((P < 0.05, \text{Fig. } 5d)\). Thus, ACTG1 can possibly aid early tumor diagnosis. Additionally, the expression profiles of the genes are illustrated in Fig. 5e.

Establishment of the prognostic nomogram

The independent prognostic factors, such as age, pathological T, and RS, were used to develop the nomogram (Fig. 6a). The nomogram enabled the visualization of the prognostic factors and aided in assessing the survival probability of the individuals with CRC. The calibration plots for the 1-, 3-, and 5-year OS prediction are illustrated in Fig. 6b-d. The corrected curve was closer to the ideal curve, which indicated strong coherence of the predicted and observed values.

Relationship of ICI with the risk model

The proportions of 22 immune cells in every sample were computed with the help of the CIBERSORT algorithm. A total of 232 samples with \(P > 0.05\) were obtained for subsequent investigation. The composition of the immune cell population varied across samples and between different groups (Fig. 7a). Furthermore, a comparison of the ICI levels across the high and low-risk groups revealed that naive B cells were enriched in the high-risk group \((P = 0.013, \text{Fig. } 7b)\).

Validation of the expression of the six genes in the model

The HPA database aided the exploration of the protein expression of the six genes. CXCL1 and GRP were excluded as they were not detected in HPA. Among them, ACTG1, CAT, and RAF1 were upregulated in tumor tissues relative to normal tissues. However, PGLYRP2 was downregulated in tumor tissues (Fig. S2a). Moreover, to further examine the expression of the six genes, RT-PCR using eight pairs of samples was performed. Among them, all genes except PGLYRP2 \((P < 0.05)\) were upregulated in tumor tissues, and CAT \((P < 0.01)\), CXCL1 \((P < 0.05)\), and RAF1 \((P < 0.05)\) were statistically significant (Fig. S2b-g). Moreover, these findings were congruent with the aforementioned results.

Knockout of ACTG1 suppressed the proliferation, migration, and invasion of CRC cells

To investigate the possible involvement of ACTG1 in the advancement of CRC, MTT, wound healing, colony formation, migration, and invasion assays were executed. Owing to upregulated ACTG1 expression in the SW480 and SW620 cell lines (Fig. 8a), they were selected for further investigation.
After knocking down ACTG1 (Fig. 8b), the proliferation ability of the cells was inhibited in the MTT assay (Fig. 8c). Furthermore, CRC cell migration and invasion abilities were also significantly suppressed on downregulating ACTG1 expression (Fig. 9a-c). Additionally, the colony formation assay depicted that ACTG1 promoted CRC cell cloning (Fig. 9d). Thus, high ACTG1 expression is speculated to indicate poor outcomes for individuals with CRC.

**Discussion**

Approximately 590,000 new CRC cases and 310,000 deaths are speculated to be reported in China in 2022\(^2\). Moreover, to date, approximately half of the patients with CRC have died due to recurrence and metastasis. However, a precise signature and novel biomarkers should be found to improve the situation.

In the current study, five GEO datasets were integrated and batch-normalized into one, revealing 1610 DEGs in the merged dataset. After overlapping with the immune genes, 118 DE-IRGs were obtained. Furthermore, DE-IRGs were found to be strongly linked with leukocyte migration and cytokine-cytokine receptor interaction, which indicated that they were involved in tumor development and metastasis. Finally, an immune-related prognostic signature was developed, comprising six DE-IRGs (ACTG1, CAT, CXCL1, GRP, RAF1, and PGLYRP2).

As an actin protein, ACTG1 contributes to cell division, migration, chromatin remodeling, vesicle transport, and other cellular functions\(^17\). ACTG1 encodes for the cytoskeleton protein \(\gamma\)-actin, which serves an important function in non-muscle cells. It is also abundant in the auditory hair cells of the cochlea\(^18\). Furthermore, it acts as a tumor promoter in different cancers, such as skin cancer, hepatocellular cancer, and lung cancer\(^19\–21\). Liu et al. reported that ACTG1 acts as a target gene of miR-10a in CRC and is highly expressed in CRC cells\(^22\). However, the mechanism of ACTG1 in CRC carcinogenesis requires further research.

CAT is the key antioxidant enzyme that decomposes \(\text{H}_2\text{O}_2\) and counteracts the damaging effects of reactive oxygen species (ROS)\(^23\). The overproduction of ROS can lead to redox imbalance and cause tumor initiation and progression\(^24\). Justyna et al. reported that the activity of CAT in individuals with CRC was lower relative to the control group, and the expression of CAT could predict lymph node metastasis\(^25\). Furthermore, Adam et al. studied the immunoexpression of CAT in colorectal lesions and observed that elevated expression levels of CAT in adenoma and decreased expression in adenocarcinoma could lead to the development of CRC\(^26\). Notably, Reyhaneh et al. demonstrated that curcumin could inhibit tumor growth by enhancing the activity of CAT in colitis-related CRC\(^27\).

Chemokines are a group of soluble chemotactic cytokines and a key element of inflammation. As inflammation is a hallmark of cancer, chemokines have been accordingly observed to promote tumorigenesis via various pathways\(^28\). As per the position of key cysteine residues, chemokines are classified into four classes (CXC, CX3C, CC, and C)\(^29\). CXCL1 belongs to the C-X-C family of chemokines
and interacts with CXCR2\textsuperscript{30,31}. The function of CXCL1 in CRC has been widely researched. Ogata \textit{et al.} noted that CXCL1 expression was linked to a differentiated degree of CRC and could predict the prognosis of individuals with CRC\textsuperscript{32}. Likewise, Gong \textit{et al.} reported CXCL1 to be upregulated in tumor tissues and demonstrated it to be a diagnostic marker for colon cancer using a TCGA cohort and two Guangxi cohorts\textsuperscript{33}. Additionally, Zhuang \textit{et al.} reported miR-145-5p could inhibit colon cancer development by suppressing the expression of CXCL1 and ITGA2\textsuperscript{34}.

GRP is the mammalian counterpart of bombesin and can induce gastrin secretion from G cells in the gastric antrum\textsuperscript{35}. It is not only involved in the mitogenesis of cancer cells but also has paracrine and endocrine effects\textsuperscript{36}. Recently, many studies have focused on the function of GRP and its receptors in CRC. Li \textit{et al.} documented that GRP can predict the prognosis of individuals with colon cancer with high sensitivity and specificity\textsuperscript{37}. Additionally, GRP receptors have also been demonstrated to serve as biomarkers for early CRC diagnosis and colorectal tumor growth monitoring\textsuperscript{38}.

Raf-1, also called C-Raf, is part of the Raf family of protein kinases and contributes to the EGFR/Raf/MEK/ERK signaling pathway. The EGFR pathway is a classic antitumor therapy approach. It transmits extracellular signals into the nucleus through cell membrane receptors and participates in cell proliferation, differentiation, and other functions\textsuperscript{39,40}. The inhibitors of Raf-1 have been widely studied and promoted in the treatment of CRC. For instance, Regorafenib, a Raf-1 inhibitor, has been approved for the oral treatment of metastatic CRC in refractory to standard chemotherapy\textsuperscript{41–43}.

PGLYRP2 is a pattern recognition receptor and encodes a peptidoglycan recognition protein. It is primarily expressed in the liver and participates in the regulation of innate immunity and immunosurveillance\textsuperscript{44}. Moreover, PGLYRP2 gene polymorphisms are also associated with Parkinson's disease\textsuperscript{45,46}. Notably, PGLYRP2 also acts as a biomarker for the adequate immune response against hepatocellular carcinoma and improves patient outcomes\textsuperscript{47}.

In this study, PGLYRP2 was downregulated and recognized as a protective gene, whereas ACTG1, CAT, CXCL1, GRP, and RAF1 were upregulated and linked to poor survival. Moreover, ACTG1 expression was increased when the TNM stage increased. To subsequently study the involvement of ACTG1 in CRC, the ACTG1 expression in SW480 and SW620 cell lines was reduced, and the proliferation, migration, and invasion abilities in these cells were inhibited. The results indicate that ACTG1 plays an active role in CRC carcinogenesis. Nevertheless, the precise involvement of the six IRGs in CRC progression warrants further investigation.

In the training cohort, individuals in the low-risk groups had better OS relative to those in the high-risk groups. Then ROC analysis depicted that the prediction of a 5-year OS exerted a better performance than that of a 1-year or 3-year OS. The results were validated in the external cohorts and were observed to be consistent with the training cohort. Additionally, the RS could provide a better predictive value in terms of accuracy compared to the pathological stage, gender, and age. It was established as an independent risk
factor for predicting the OS of individuals with CRC. Moreover, RS and clinicopathology features aided in
developing a nomogram system that confirms the stability of the RS and the accuracy of the survival
prediction ability. In conclusion, the developed model and nomogram system may be useful in assessing
the OS of individuals with CRC.

Immune cells in the tumor immune microenvironment contribute significantly to the progression of
tumors, and the degree of ICI has been demonstrated to influence the survival of individuals with CRC. Additionally, the proportion of the 22 immune cells in every sample was recorded, and the resulting data
demonstrated that naive B cells were significantly elevated in the high-risk groups in comparison with the
low-risk groups. Naive B cells refer to white blood cells that have not been activated by an antigen. They
are also involved in manufacturing antibodies, assisting in eliminating harmful pathogens, and
preventing subsequent assaults from re-occurring. Yan et al. evaluated the B cell immunity in CRC cells
and observed attenuated antigen presentation and diminished antitumor immunity capacity of CD40 +
and CD27 + B cells in tumor tissues. Overall, the degree of ICI reflects the immune status of individuals
with CRC, which may lead to different survival outcomes in the risk groups.

Nevertheless, there are certain limitations in this study. First, the number of samples is limited because
they all come from public databases. Second, this signature was obtained from a retrospective study
and warrants verification by prospective clinical trials. Finally, more basic experiments must be
conducted to investigate the molecular mechanisms of the signature.

Conclusions

A six-IRG risk model was developed to predict OS and depict the immune status of individuals with CRC.
Furthermore, ACTG1 was highlighted as a possible new immune biomarker in CRC. Overall, these
findings aid in personalized treatment decisions and provide new immunotherapeutic measures for
individuals with CRC.

Abbreviations

CRC
Colorectal cancer
DE-IRGs
Differentially expressed immune-related genes
TCGA
The Cancer Genome Atlas
GEO
Gene Expression Omnibus
ICI
Immune cell infiltration
OS
Overall survival
RS
Risk score
PD1
Programmed cell death 1
PD-L1
Programmed cell death-ligand 1
IRGs
Immune-related genes
HPA
Human protein atlas
FC
Fold change
GO
Gene ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
ROC
Receiver operating characteristic
AUC
Area under the curves
RT-PCR
Real-time polymerase chain reaction
cDNA
Complementary DNA
DMEM
Dulbecco's modified Eagle's medium
shRNA
Short hairpin RNA
MTT
3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2-H-tetrazolium bromide
CXCL1
C-X-C motif chemokine ligand 1
CAT
Catalase
GRP
Gastrin-releasing peptide
ACTG1
Actin gamma 1
RAF1
Raf-1 proto-oncogene, serine/threonine kinase
PGLYRP2
Peptidoglycan recognition protein 2
ROS
Reactive oxygen species

Declarations

Acknowledgements

We appreciate the TCGA and GEO databases for providing the data. We thank Bullet Edits Limited for linguistic editing and proofreading of the manuscript.

Authors’ contributions

Xi Gu wrote the article, analyzed the data and plotted the pictures; Lifeng Zheng came up with the ideal, acquired data and analyzed the data; Ziyu Xu added the idea, acquired data and analyzed the data. Shuzhou, Wulou Zhang and Zonghang Liu modified the article. Hao Lin and YePeng Zhang acquired data. All authors have read and approved the final manuscript.

Funding

The study was supported by Nanjing Medical Science and Technology Development Foundation (Grant No.YKK22258) and Wannan Medical College (JXYY202252).

Availability of data and materials


Ethics approval and consent to participate

This research was under approval by the Ethics Committee of Nanjing Jiangbei Hospital. Informed consent was acquired from all the individuals involved.

Consent for publication

Not applicable

Competing interests

The authors declare they have no competing interests
References

28. Cabrero-de Las Heras S, Martinez-Balibrea E. CXC family of chemokines as prognostic or predictive biomarkers and possible drug targets in colorectal cancer. World J Gastroenterol. 2018;24:4738–49.
34. Zhuang W, Niu T, Li Z. MicroRNA miR-145-5p regulates cell proliferation and cell migration in colon cancer by inhibiting chemokine (C-X-C motif) ligand 1 and integrin alpha2. Bioengineered 2021; 12:


Figures
Figure 1

Flowchart of the research.
Figure 3

Establishment of the prognostic risk model based on the training cohort. a. Fifteen differentially expressed immune-related genes (DE-IRGs) were identified by a univariate Cox regression analysis. b and c. Twelve DE-IRGs were identified by a lasso-cox regression analysis. d. Six optimal DE-IRGs were identified for the establishment of the prognostic risk model by a
Validation of the prognostic risk model in the external cohorts. a-c Kaplan-Meier survival curves of the prognostic risk model in the testing, TCGA and GSE38832 cohorts. d-l Risk score distribution, Survival status scatter plots, Heatmap of six DE-IRGs, and Time-dependent receiver operating characteristic (ROC) curve analysis of the prognostic risk model in the training cohort.
status scatter plots, Heatmap of six differentially expressed immune-related genes (DE-IRGs) of the prognostic risk model in the testing, TCGA and GSE38832 cohorts. m-o Time-dependent receiver operating characteristic (ROC) curve analysis of the prognostic risk model in the testing, TCGA and GSE38832 cohorts.

Figure 5
Independent prognostic value of the model in TCGA cohort. a-c Univariate, Multivariate Cox regression analyses and time-dependent receiver operating characteristic (ROC) curve analysis at five years of risk score and clinical features. d. Relationship between ACTG1 expression and pathological stage. e. Heatmap of six differentially expressed immune-related genes (DE-IRGs) expression and clinical features.

Figure 6
Nomogram predicts the survival of patients with colorectal cancer (CRC). a Nomogram to predict the probability of 1, 3, and 5 years overall survival (OS) in CRC patients. b-d Normative curve for predicting the probability of 1, 3, and 5 years OS in CRC patients.

Figure 7

Immune infiltration status. a The infiltration ratio of 22 immune cell types in high and low-risk groups samples. b Differences in immune cell types between the high and low-risk groups.
Figure 8

ACTG1 knockdown inhibits cell proliferation in vitro. a ACTG1 expression is upregulated in colorectal cancer (CRC) cells. b ACTG1 expression in ACTG1-shRNA SW480 and SW620 cells. c Growth curves for SW480 and SW620 cells after transfection with ACTG1 shRNA determined by MTT analysis. ** $P<0.01$
Figure 9

ACTG1 knockdown inhibits cell migration and invasion in vitro. a-c Decreased cell migration and invasion activity in ACTG1-shRNA knockdown SW480 and SW620 cells. d Clonogenic assay on ACTG1-shRNA knockdown SW480 and SW620 cell lines showing decreased colony formation. ** P<0.01

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.tif
- Additionalfile2.tif
- Additionalfile3.docx
- Additionalfile4.docx
- Additionalfile5.docx