



## OPEN ACCESS

## EDITED BY

Abhirup Banerjee,  
University of Oxford, United States

## REVIEWED BY

Satya Ranjan Dash,  
KIIT University, India  
Valeria Pusceddu,  
University of Cagliari, Italy

## \*CORRESPONDENCE

Xin Wang  
✉ wangxin213@sina.com

†These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 10 November 2023

ACCEPTED 24 January 2024

PUBLISHED 09 February 2024

## CITATION

Shu P, Liu N, Luo X, Tang Y, Chen Z, Li D,  
Miao D, Duan J, Yan O, Sheng L, Ouyang G,  
Wang S, Jiang D, Deng X, Wang Z, Li Q and  
Wang X (2024) An immune-related gene  
prognostic prediction risk model for  
neoadjuvant chemoradiotherapy in  
rectal cancer using artificial intelligence.  
*Front. Oncol.* 14:1294440.  
doi: 10.3389/fonc.2024.1294440

## COPYRIGHT

© 2024 Shu, Liu, Luo, Tang, Chen, Li, Miao,  
Duan, Yan, Sheng, Ouyang, Wang, Jiang, Deng,  
Wang, Li and Wang. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# An immune-related gene prognostic prediction risk model for neoadjuvant chemoradiotherapy in rectal cancer using artificial intelligence

Pei Shu<sup>1,2,3†</sup>, Ning Liu<sup>4†</sup>, Xu Luo<sup>5,6†</sup>, Yuanling Tang<sup>1,2</sup>,  
Zhebin Chen<sup>5,6</sup>, Dandan Li<sup>1,2</sup>, Dong Miao<sup>5,6</sup>, Jiayu Duan<sup>1,2</sup>,  
Ouying Yan<sup>1,2</sup>, Leiming Sheng<sup>1,2</sup>, Ganlu Ouyang<sup>1,2</sup>, Sen Wang<sup>5,6</sup>,  
Dan Jiang<sup>7,8</sup>, Xiangbing Deng<sup>9</sup>, Ziqiang Wang<sup>8,9</sup>, Qingyun Li<sup>10</sup>  
and Xin Wang<sup>1,2,8\*</sup>

<sup>1</sup>Department of Radiation Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China, <sup>2</sup>Department of Abdominal Tumor Multimodality Treatment, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, <sup>3</sup>Clinical Trial Center, National Medical Products Administration Key Laboratory for Clinical Research and Evaluation of Innovative Drugs, West China Hospital, Sichuan University, Chengdu, Sichuan, China, <sup>4</sup>Division of Thoracic Tumor Multimodality Treatment, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, <sup>5</sup>Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu, China, <sup>6</sup>School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China, <sup>7</sup>Department of Pathology, West China Hospital, Sichuan University, Chengdu, China, <sup>8</sup>Sichuan University–University of Oxford Huaxi Joint Center for Gastrointestinal Cancer, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, Sichuan, China, <sup>9</sup>Division of Gastrointestinal Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu, China, <sup>10</sup>Genecast Biotechnology Co., Ltd., Xishan District, Wuxi, Jiangsu, China

**Background:** This study aimed to establish and validate a prognostic model based on immune-related genes (IRGPM) for predicting disease-free survival (DFS) in patients with locally advanced rectal cancer (LARC) undergoing neoadjuvant chemoradiotherapy, and to elucidate the immune profiles associated with different prognostic outcomes.

**Methods:** Transcriptomic and clinical data were sourced from the Gene Expression Omnibus (GEO) database and the West China Hospital database. We focused on genes from the RNA immune-oncology panel. The elastic net approach was employed to pinpoint immune-related genes significantly impacting DFS. We developed the IRGPM for rectal cancer using the random forest technique. Based on the IRGPM, we calculated prognostic risk scores to categorize patients into high-risk and low-risk groups. Comparative analysis of immune characteristics between these groups was conducted.

**Results:** In this study, 407 LARC samples were analyzed. The elastic net identified a signature of 20 immune-related genes, forming the basis of the IRGPM. Kaplan–Meier survival analysis revealed a lower 5-year DFS in the high-risk group compared to the low-risk group. The receiver operating characteristic (ROC) curve affirmed the model's robust predictive capability. Validation of the model

was performed in the GSE190826 cohort and our institution's cohort. Gene expression differences between high-risk and low-risk groups predominantly related to cytokine–cytokine receptor interactions. Notably, the low-risk group exhibited higher immune scores. Further analysis indicated a greater presence of activated B cells, activated CD8 T cells, central memory CD8 T cells, macrophages, T follicular helper cells, and type 2 helper cells in the low-risk group. Additionally, immune checkpoint analysis revealed elevated PDCD1 expression in the low-risk group.

**Conclusions:** The IRGPM, developed through random forest and elastic net methodologies, demonstrates potential in distinguishing DFS among LARC patients receiving standard treatment. Notably, the low-risk group, as defined by the IRGPM, showed enhanced activation of adaptive immune responses within the tumor microenvironment.

#### KEYWORDS

artificial intelligence, prognostic model, immune related gene, rectal carcinoma, neoadjuvant chemoradiotherapy

## Introduction

Colorectal cancer is the third most common cancer and the second leading cause of cancer-related mortality worldwide (1). Locally advanced rectal cancer (LARC) constitutes up to 15% of all colorectal cancer cases (2). For LARC, neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision (TME) and adjuvant chemotherapy is the standard treatment recommended by the National Comprehensive Cancer Network (NCCN) guidelines (3). However, rectal cancers are a widely heterogeneous group. The prognosis of patients is different even after the same standard treatment. Therefore, there is a strong need to predict the prognosis of patients treated with the standard therapy and to guide treatment alterations when necessary (4, 5).

Tumor behavior is controlled not only by the epithelial component but also by the tumor immune microenvironment (TIME) (6). Several studies have indicated that tumor-infiltrating lymphocytes are predictive of the response to treatment and prognosis (7, 8). Thus, the relationship between the TIME and treatment response/prognosis

was explored to find biomarkers or to construct models that could predict the response to treatment and prognosis (9). Along with this, the TIME can be modified by radiotherapy (10). Some studies have explored the correlation between the TIME and the clinical outcome of rectal cancer patients treated with neoadjuvant radiotherapy by immunohistochemical (IHC) methods (11). The results showed that different tumor infiltrating cell subsets (CD3+ lymphocytes, CD4+ lymphocytes, CD8+ lymphocytes, etc.) or the expression level of PD-L1 were predictive factors of the treatment response or prognosis (11–14). Due to the inconsistent results of IHC, it is hoped that further findings could be derived from gene analysis. Qian et al. explored the relationship between the gene expression profile and response to nCRT in LARC, and they developed an immune gene predictive model that could predict the response to nCRT by support vector machine (SVM) (15). Therefore, the value of immune-related gene signatures closely related to the TIME in predicting prognosis is worth studying.

Machine learning (ML), a branch of artificial intelligence, has had outstanding performance in disease diagnosis, prognosis prediction, and treatment response assessment. In previous studies, only the LASSO algorithm has been used to establish a prognostic model in LARC (16, 17). Timo M. Deist et al. suggested that random forest and elastic net logistic regression yield higher discriminative performance for chemoradiotherapy outcome than other classifier families (18). Thus, we planned to generate an optimized predictor of outcome in rectal cancer patients with the random forest and elastic net logistic regression methods in this study.

In this study, we focused on immune genes in an RNA immunology panel and screened immune-related genes related to prognosis. The aim of this study was to construct an immune-

**Abbreviations:** IRGPM, Immune-related gene prognostic model; DFS, Disease-free survival; LARC, Locally advanced rectal cancer; GEO, Gene expression omnibus; ROC, Receiver operating characteristic; AUCs, Areas under the curve; nCRT, Neoadjuvant chemoradiotherapy; TME, Total mesorectal excision; NCCN, The National Comprehensive Cancer Network; TIME, The tumor immune microenvironment; IHC, Immunohistochemical; SVM, Support vector machine; ML, Machine learning; WCHSC, West China Hospital, Sichuan University; FFPE, Formalin-fixed, paraffin-embedded; DEGs, Differentially expressed genes; GSEA, Gene set enrichment analysis; ssGSEA, Single-sample gene set enrichment analysis; mDFS, Median disease-free survival; HR, Hazard ratio.

related gene prognostic model (IRGPM) with the random forest and elastic net logistic regression methods for LARC patients treated with nCRT. We further characterized the immune microenvironment of patients with different prognostic risks. The results showed that IRGPM, which was established by using the random forest and elastic net methods, was a promising prognostic biomarker for patients receiving neoadjuvant treatment. Lack of effective immune infiltration in tumor microenvironment was observed in the high risk group with poorer survival.

## Materials and methods

### Data acquisition and processing

A public database was queried for patients with LARC. Only patients who received a standard strategy (neoadjuvant treatment followed by TME and adjuvant chemotherapy) and had prognostic data were selected. The gene expression data with matched clinical information were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), with accession numbers GSE87211, GSE190826 and GSE119409. Patients with missing disease-free survival (DFS) were excluded to reduce statistical bias in this analysis. The GSE87211 dataset was used as the training cohort, and the GSE190826 dataset was used as the validation cohort. These three cohorts were all included in the immune analysis. Another validation cohort was created from LARC patients at West China Hospital, Sichuan University (WCHSC). Patients with LARC who underwent standard nCRT treatment between January 2013 and January 2018 were recruited. The study was approved by the ethical committee of WCHSC. An RNA immune-oncology panel was used to detect the expression of 395 genes at the RNA level from formalin-fixed, paraffin-embedded (FFPE) tumor specimens (details are provided in [Supplementary Table S1](#)). The raw gene expression data in GSE87211, GSE190826, GSE119409 and WCHSC were normalized by using the rank-in algorithm (19).

### Identification of prognosis-related genes

Univariable Cox regression analyses were applied to identify the prognosis-related genes. Univariable Cox regression analysis was performed to explore the potential confounders. We use three hypothesis testing methods (Wald.test, Logrank.test and likelystest). Gene expression levels with p values less than 0.05 in the three hypothesis testing methods were considered significant. Subsequently, we shall include prognostic variables that exhibit statistical significance with  $p < 0.05$  from the univariate analysis into the ensuing analysis.

### Construction and validation of the prognostic signature

After screening with univariable analyses, elastic net was adopted to ascertain the optimal panel of prognostic genes (20).

Compared with the LASSO algorithm used in previous studies, the elastic net has two advantages. It can select more features than the number of samples in the dataset, which is problematic when dealing with high-dimensional data. Additionally, if data contain a group of features that are highly correlated, the LASSO penalty is going to randomly choose one feature from this group, whereas the elastic net penalized model would tend to select all. The elastic net penalty overcomes these problems by using a weighted combination of the  $L_1$  and  $L_2$  penalties as represented by the following mathematical formula in [Equation 1](#):

$$\hat{\beta} = \operatorname{argmin} \left( \sum_{i=1}^n \left( y_i - \sum_j^p \beta_j x_{ij} \right)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|^2 \right) \quad (1)$$

The IRGPM was established with random forest using the Python package scikit-survival 0.17.2 on the GSE87211 cohort (21, 22). The expression status of 20 selected prognostic-related genes was used as the model input. The goal is to predict DFS time. We employed grid search to optimize the model's hyperparameters for best performance, adjusting the number of trees, maximum depth, and minimum number of branch nodes with specific step sizes. We evaluated the model's general performance using a stratified 10-fold cross-validation. For each patient, the model assigns a prognostic risk score. Using the X-tile program, we divided the patients into low-risk and high-risk groups. We then compared the DFS of rectal cancer patients in these groups using the Kaplan–Meier method and log-rank test with the survival R package (23). Data from the GSE190826 cohort and WCHSC cohort, were used to validate the prognostic risk model. Due to the absence of DFS data, GSE119404 cohort is excluded from the prognostic risk model validation.

### Exploration of the molecular mechanisms underlying the prognostic signature

The limma R software package was used to analyze differentially expressed genes (DEGs) between the low- and high-risk groups (24). We used the cluster Profiler R package to carry out gene set enrichment analysis (GSEA) and to identify significantly enriched pathways between the low- and high-risk groups (25). A  $P$  value  $< 0.05$  and  $|\log_2 \text{Fold Change}| > 0.58$  were regarded as the cutoff values for statistical significance.

### Evaluation of the TIME

To determine the relationship between the IRGPM and the TIME, we explored the abundance of immune cells and stromal cells between different groups. The stromal score, immune score and ESTIMATE score of each patient were calculated using ESTIMATE method (26). Their differences were compared using the Wilcoxon signed-rank test, and  $p < 0.05$  was considered to indicate significance. Then, we calculated the infiltration values for rectal cancer dataset samples based on 28 immune gene sets. Using single-sample gene set enrichment analysis (ssGSEA), the proportions of immune cells were quantified (27). The scores of

immune cells in different groups are shown on multi-boxplots. We also made comparisons of immune checkpoint activation between the low- and high-risk groups by the ggpubr R package.

## Statistical analysis

Most statistical analyses were performed using R software (Version 4.1.3; R Foundation for Statistical Computing, Vienna, Austria). The cutoff points for age ranges were determined using the X-tile program (<http://www.tissuearray.org/rimmlab/>). The survival rate was plotted using the Kaplan–Meier method and analyzed using the log-rank test.  $P < 0.05$  was considered statistically significant. Graphs related to R statistical analyses were drawn using the ggplot2 package in R.

## Results

### Patient information

The detailed flow diagram of our study is shown in Figure 1. A total of 407 LARC samples were used in this study. There were 180 LARC cases in the GSE87211 cohort that were used for prognostic risk model training. The GSE190826 dataset with 86 LARC cases and the WCHSC cohort with 75 LARC patients were used for model validation. For immune infiltration analysis, a total of 332 LARC samples in three GEO cohorts (GSE87211, GSE190826 and GSE119409) were included in the study. The detailed information (age, gender, tumor stage and treatment method) of patients from the GEO database and our hospital is shown in Table 1.

### Prognostic risk model construction

To determine the independent prognostic genes, univariate Cox regression analyses for DFS were performed. As shown in Supplementary Table S3, 23 genes were identified and further included in the next analysis. To obtain the optimal panel, elastic net logistic regression was carried out on the candidate genes. We screened a 20-gene signature with the best concordance index (Figure 2A). Then, we constructed the optimized prognostic

model by using the random forest method, which assigned each sample a prognostic risk score based on the identified coefficient of each gene (Figure 2B; details in Supplementary Table S4). To obtain the optimized cutoff value for the prognostic risk score, we applied the X-tile program to assess the statistical significance and avoid arbitrary cut point selection. The patients were then divided into high- and low-risk groups based on a cutoff value of 18.00. The Kaplan–Meier curves showed that the low-risk group was associated with a significantly longer median disease-free survival (mDFS) than the high-risk group (unreached vs. 20 months;  $p < 0.0001$ ; hazard ratio (HR), 0.05; 95% confidence interval (95% CI), 0.03–0.11; Figure 2C). As shown in Figure 2D, receiver operating characteristic (ROC) analysis indicated that the areas under the curve (AUCs) of the IRGPM for the training cohort reached 0.87, 0.94, and 0.95 at 1, 3, and 5 years, respectively.

### Validation of the prognostic risk model

The robustness of IRGPM was validated in two independent cohorts of LARC patients with adequate information on genomic expression profiles and survival. The patients in both cohorts were classified into high- and low-risk groups in the same way as the training cohort, with the optimal cut point determined in the training cohort (details in Supplementary Table S4). As shown in Figure 3A, for the GSE190826 cohort ( $n = 86$ ), low-risk patients ( $n = 76$ ) had a better DFS than high-risk patients ( $n = 10$ ) (mDFS, unreached vs. 13 months;  $p < 0.0001$ ; HR, 0.52; 95% CI, 0.28–0.95). ROC analysis indicated that the AUCs of the IRGPM for the GSE190826 validation cohort were 0.79, 0.64, and 0.63 at 1, 3, and 5 years, respectively (Figure 3B). Consistently, in the SUWCH cohort ( $n = 75$ ), low-risk patients ( $n = 66$ ) also had a better DFS than high-risk patients ( $n = 9$ ) (mDFS, unreached vs. 22 months;  $p = 0.0031$ ; HR, 0.29; 95% CI, 0.12–0.69; Figure 3C), and the AUCs for the prognostic risk model were 0.64, 0.66, and 0.64 at 1, 3, and 5 years, respectively (Figure 3D).

### Differential expression analysis

The expression profiles of LARC patients with high prognostic risk scores were compared with those of patients with low risk

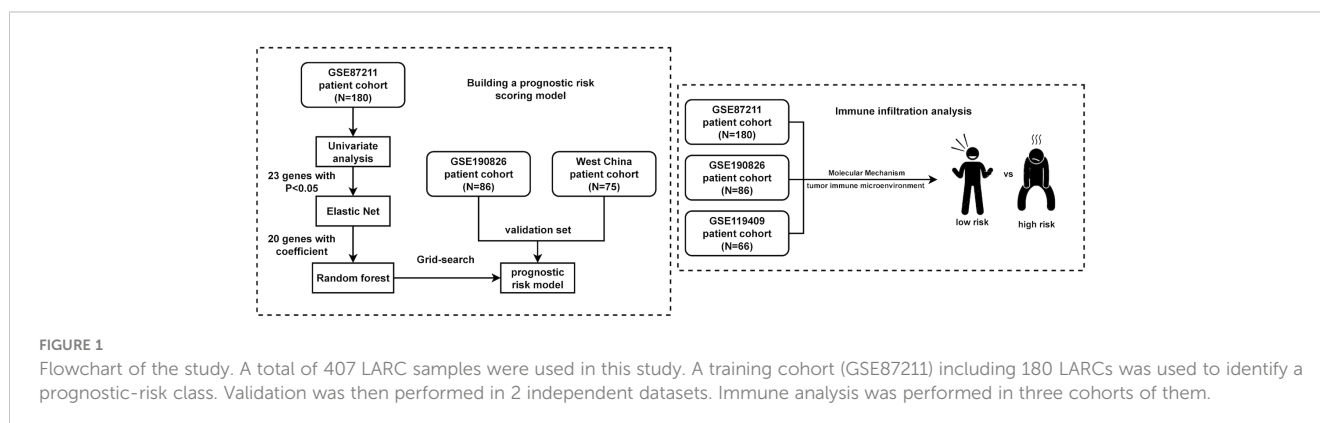
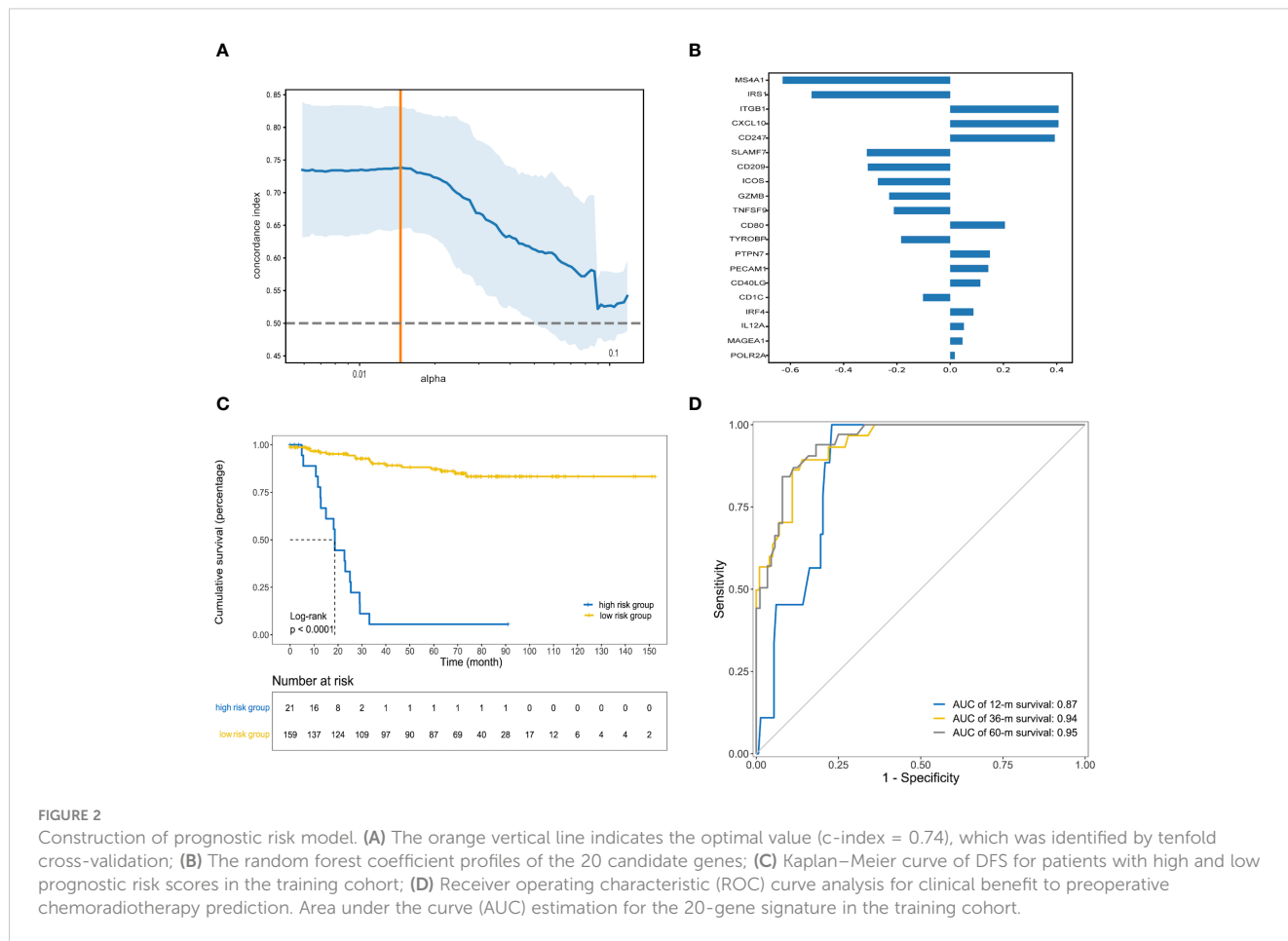


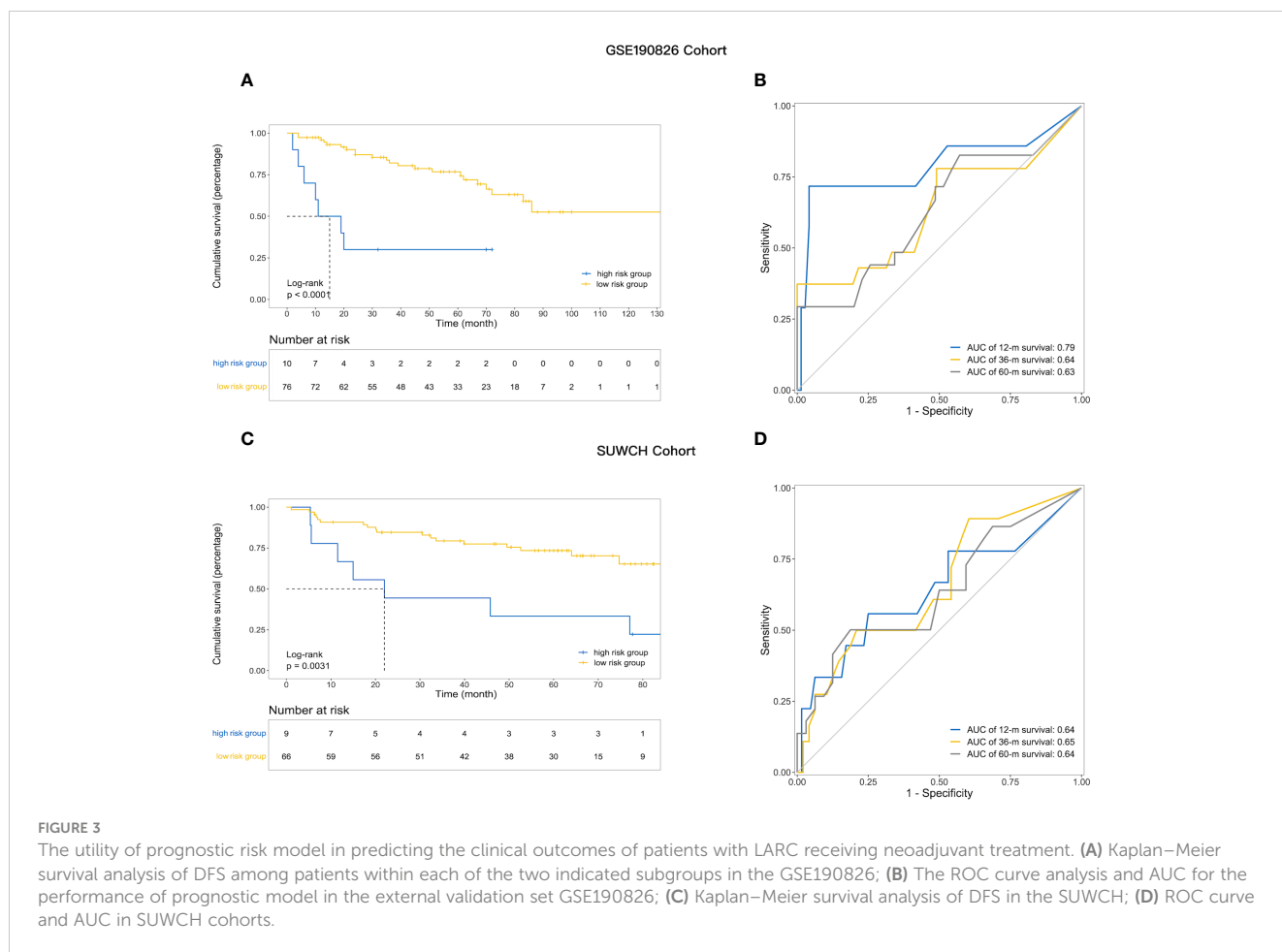
TABLE 1 Cohorts' information in the study.

Dataset	GSE87211	GSE190826	GSE119409	SUWCH
Institution	National Cancer Institute, USA	Georg Speyer Haus, Germany	Peking University Cancer Hospital & Institute, China	Sichuan university, West China Hospital, China
No. of patients	180	86	66	75
Sex	Female:54 Male:126	Unknown	Unknown	Female:32(42.7%) Male:43(57.3%)
Age	≤65:73(40.5%) >65:107(59.5%)	Unknown	≤65:20(33.3%) >65:45(75%) Unknown:1(1.7%)	≤65:24(32%) >65:51(68%)
Clinical stage before neoadjuvant treatment*	II:59 III:117 IV:2 Unknown:2	II:6 III:57 Unknown:3	Unknown	II:9 III:66
Pathological stage after surgery*	pCR:68 I:17 II:47 III:45 IV:3	Unknown	Unknown	pCR:7 I:12 II:23 III:33
Treatment	nCRT	nCRT	nCRT	nCRT or SCRT

nCRT, neoadjuvant chemoradiotherapy; SCRT, neoadjuvant short course radiotherapy. \*The staging is based on the 8th edition of the TNM staging system for colorectal cancer, as established by AJCC.

More detail are provided in Supplementary Table S2.



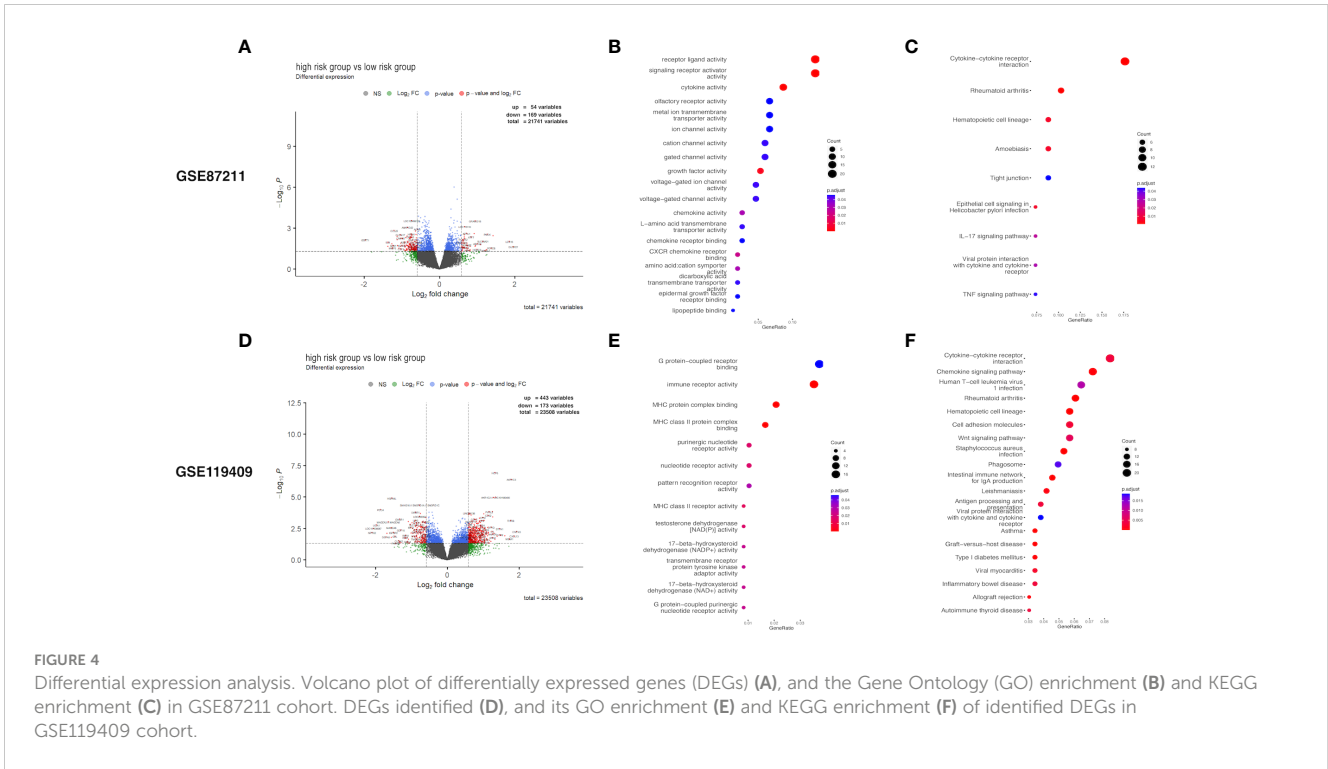


scores to identify DEGs. The DEGs related to prognostic risk status for each cohort are listed in detail in [Supplementary Table S5](#) and visualized in volcano plots ([Figure 4](#)). In the GSE87211 cohort, a total of 54 upregulated genes and 169 downregulated genes were identified in the high-risk group compared with the low-risk group ([Figure 4A](#)). Functional enrichment analysis showed that the DEGs were significantly associated with 19 GO terms and 9 KEGG pathways (details in [Supplementary Table S6](#)). GO enrichment analysis of biological processes was conducted for the DEGs, which demonstrated that receptor ligand activity, signaling receptor activity, cytokine activity growth factor activity and CXCR chemokine receptor binding were the most frequently enriched biological process terms in the high-risk group ([Figure 4B](#)). The top enriched KEGG pathways by the DEGs in GSE87211 were also related to cytokine–cytokine receptor interactions ([Figure 4C](#)). In the GSE119409 cohort, 616 DEGs were identified, with 443 unregulated and 173 downregulated genes ([Figure 4D](#)). The functional annotation results showed that the DEGs were significantly associated with 13 GO terms and 30 KEGG pathways ([Supplementary Table S6](#)). The DEGs were mostly enriched in the GO terms MHC protein complex binding, immune receptor activity, MHC class II protein complex binding, and MHC class II receptor activity ([Figure 4E](#)). In addition, the most enriched

immune-related KEGG pathways were cytokine–cytokine receptor interactions and chemokine signaling pathways ([Figure 4F](#)).

## GSEA of the DEGs

GSEA identified a number of KEGG pathways enriched in distinct prognostic risk subgroups ([Supplementary Table S7](#)). Typically, immune-associated pathways were highly active in the low-risk group, including antigen processing and presentation pathways, the hematopoietic cell lineage pathway, JAK–STAT signaling, T and B-cell receptor signaling, cytokine–cytokine receptor interactions, natural killer cell-mediated cytotoxicity and the intestinal immune network for IgA production ([Figures 5A, B](#)). In addition, we identified cancer-associated pathways that were hyperactivated in the low-risk group, including NOTCH signaling ([Figure 5B](#)). This suggests that the activities of these cancer-associated pathways are positively associated with LARC immunity. We also analyzed 13 immune-associated gene sets that represented diverse immune functions and pathways ([Supplementary Table S8](#)). The ssGSEA score was used to quantify the activity or enrichment levels of immune functions and pathways in the cancer samples. The gene sets of the low-risk samples were enriched in APC coinhibition, CCR,



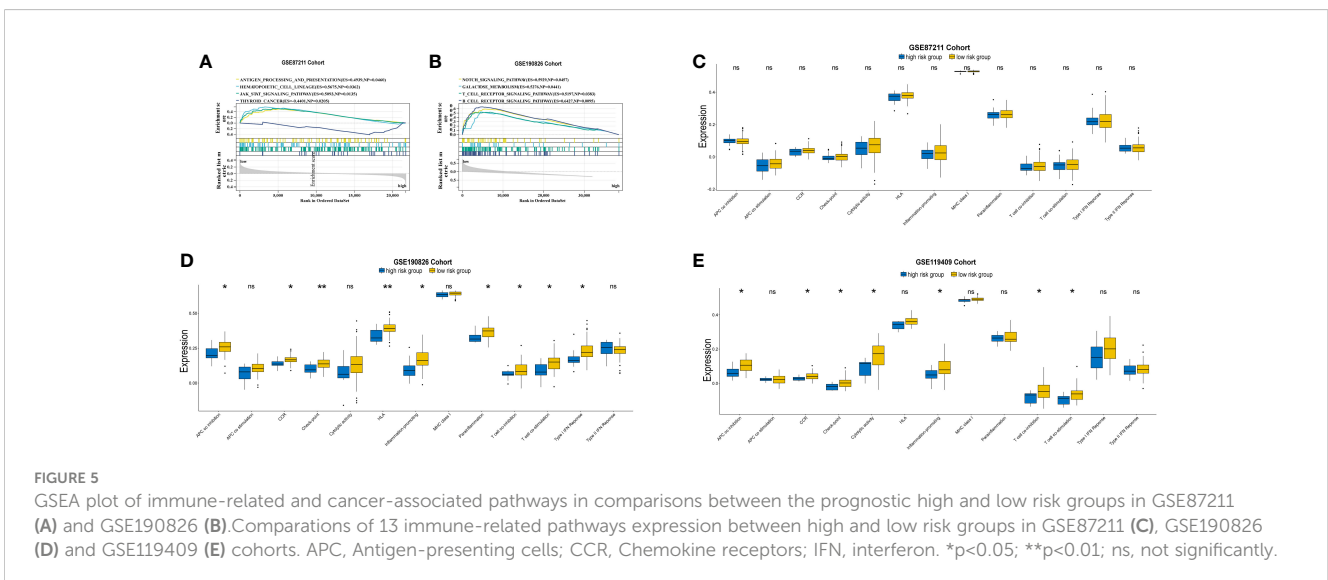
**FIGURE 4** Differential expression analysis. Volcano plot of differentially expressed genes (DEGs) (A), and the Gene Ontology (GO) enrichment (B) and KEGG enrichment (C) in GSE87211 cohort. DEGs identified (D), and its GO enrichment (E) and KEGG enrichment (F) of identified DEGs in GSE119409 cohort.

checkpoint, cytolytic activity, inflammation promotion, T-cell coinhibition, T-cell costimulation, type I IFN response and HLA (Figures 5C–E).

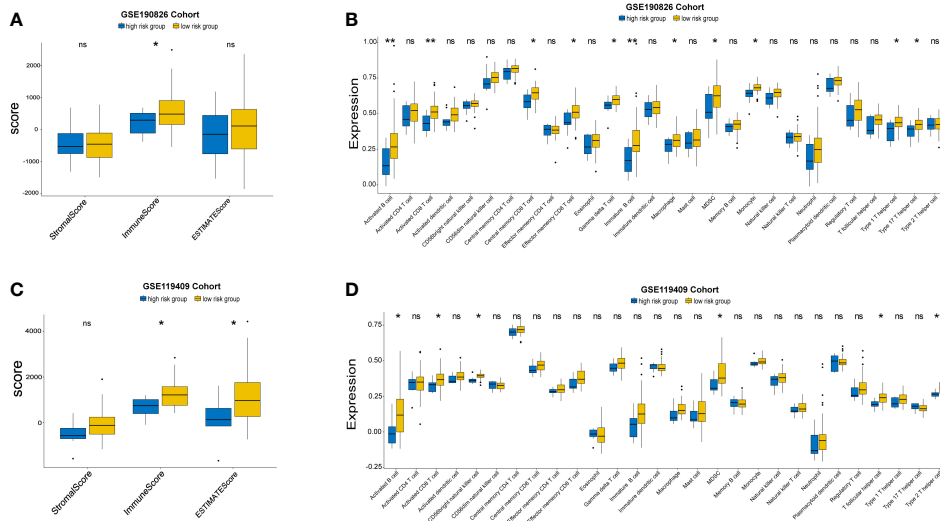
### Immune characteristics of different IRGPM subgroups

Given that the expression of immune genes is closely related to prognosis, it is necessary to investigate the role of the TIME in different groups. We compared the ESTIMATE scores, stromal scores, and immune scores between the high-risk group and the

low-risk group. Although no significant differences were found in the GSE87211 cohort (Supplementary Table S9), low-risk patients had higher immune scores than high-risk patients in the GSE190826 cohort (Figure 6A), and the GSE119409 results showed that low-risk patients had higher ESTIMATE and immune scores than high-risk patients (Figure 6C). We then investigated 28 immune-associated gene sets that represented diverse immune cell types, and the ssGSEA score was used to quantify the activity or enrichment levels of immune cells. The low-risk group had higher levels of activated B cells, activated CD8 T cells, central memory CD8 T cells, effector memory CD8 T cells, gamma delta T cells, immature B cells, MDSCs, monocytes, type 1



**FIGURE 5** GSEA plot of immune-related and cancer-associated pathways in comparisons between the prognostic high and low risk groups in GSE87211 (A) and GSE190826 (B). Comparisons of 13 immune-related pathways expression between high and low risk groups in GSE87211 (C), GSE190826 (D) and GSE119409 (E) cohorts. APC, Antigen-presenting cells; CCR, Chemokine receptors; IFN, interferon. \* $p < 0.05$ ; \*\* $p < 0.01$ ; ns, not significantly.



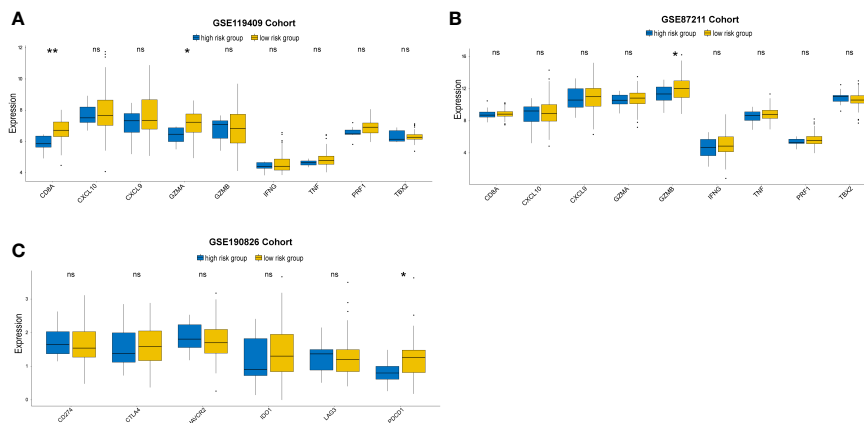
**FIGURE 6**  
The Landscape of Immuno-cell Infiltration in the TME of LARC. The comparison of estimate, immune and stromal scores of high and low risk groups in GSE190826 (A) and GSE119409 cohort (C). The expression of tumor-infiltrating immune cells in two subgroups in GSE190826 (B) and GSE119409 (D) cohort. The statistical difference of two groups was compared through the Kruskal-Wallis test and student-t test. \* $p < 0.05$ ; \*\* $p < 0.01$ . ns, not significantly.

helper cells and type 17 helper cells in GSE190826 (Figure 6B). Similarly, the low-risk group in GSE119409 had higher levels of activated B cells, activated CD8 T cells, CD56 bright natural killer cells, MDSCs, T follicular helper cells and type 2 helper cells (Figure 6D). The low-risk group had higher levels of activated CD4 T cells and type 2 helper cells in the GSE87211 cohort (Supplementary Table S9).

We further investigated the expression of immune-activity-related genes, including CD8A, CXCL10, CXCL9, GZMA, GZMB, IFNG, PRF1, TBX2, and TNF. We found that the expression levels of CD8A, GZMA and GZMB were higher in the low-risk group than in the high-risk group (Figures 7A, B).

In addition, we examined the expression of immune checkpoint genes related to the treatment response to immune checkpoint inhibitors in the two distinct subgroups. The expression status of seven genes previously identified as targets of immune checkpoint inhibitors were evaluated: CD274, CTLA4, LAG3, PDCD1, TIGIT, IDO1 and HAVCR2. The low-risk group had higher levels of PDCD1 than the high-risk group in GSE190826 (Figure 7C).

Based on the above analyses, we find that patients in low risk group had distinct TIME characterization compared with high risk group. Low risk group was characterized by adaptive immune cell infiltration and immune activation.



**FIGURE 7**  
Immune-activation-relevant genes (CD8A, CXCL10, CXCL9, GZMA, GZMB, PRF1, IFNG, TBX2, and TNF) and immune-checkpoint-relevant genes (IDO1, CD274, HAVCR2, PDCD1, CTLA4, TIGIT and LAG3) expressed in high and low prognostic risk subgroups. The statistical difference of two groups was compared through the Kruskal-Wallis test. \* $p < 0.05$ ; \*\* $p < 0.01$ . ns, not significantly.



## Discussion

Due to the heterogeneity of tumors, even after standard treatment, including neoadjuvant chemoradiotherapy, over 30% of patients experience recurrence within 5 years (28). Consequently, it is an urgent task to develop reliable biomarkers that could identify patients with a high risk of recurrence after the standard treatment method. Then, intensive or different treatment schemes, as well as different monitoring schemes, can be appropriately selected for these patients. A previous study reported that immune-related genes can be used to predict the response to neoadjuvant therapy (15). However, the relationship between immune-related genes and the long-term prognosis (DFS) of LARC patients receiving neoadjuvant treatment followed by TME and adjuvant chemotherapy is currently unclear. We used a machine learning strategy to identify a 20-immune-gene signature that can predict the prognosis of LARC patients after nCRT. The prognostic model showed excellent discrimination for the individualized prediction of DFS and had an AUC of 0.95 for the prediction of DFS at 5 years.

The predictive model constructed in this study demonstrates robust predictive performance (AUCs of 0.94 in this study) (29–31). The good performance could be partly explained by the use of different machine learning classification algorithms in our study. Compared with the LASSO Cox regression analysis used in previous studies, the elastic net used in this study can select more features. Additionally, when data contain a group of features that are highly correlated, the elastic net penalized model tends to select all features, and the LASSO penalty randomly chooses one feature from this group. Therefore, the elastic net is more likely to choose the optimal panel for constructing a prognostic model. A previous study also showed that random forest and elastic net logistic regression were better than other studied classifiers (decision tree, neural network, and SVM) in treatment response prediction (18). In this study, with immune-related genes selected by elastic net, we applied random forest to construct the prognostic classifier of LARC. As demonstrated in the [Supplementary Table S11](#), consistent with previous study outcomes, the model constructed via random forest exhibited superior performance.

To explore whether the immunological biomarkers at the RNA level could predict the prognosis of rectal cancer, an RNA immunology panel was selected because it covered the signaling pathways of immunological system activation, tumor immune response, immune cell differentiation, immune regulation, tumor antigens, antigen presentation, and so on (32, 33). Previous studies have shown that it is possible to efficiently and accurately investigate, analyze and identify prognostic markers related to immune factors in the tumor microenvironment with an RNA immune-oncology panel (33, 34). Thus, an RNA immune-oncology panel was used to identify biomarkers in this study. In addition to the gene expression data from the database, we also detected the expression of 395 genes in the RNA immune-oncology panel at the RNA level from FFPE tumor specimens at WCHSC. The data originating from our institution were included in the validation cohort. Kaplan–Meier survival analysis showed poorer 5-year DFS in the high-risk group than in the low-risk group, and the ROC curve suggested good model prediction (AUC=0.64 at 5 years). This further proved that the developed prognostic risk model had good predictive power.

The immune gene signature in the prognostic model consisted of 20 genes, of which 11 genes were associated with T-cell immunity. TNFSF9, CD1C and CD209 play an important role in presenting antigens for T cells (35–37). ICOS, CD247, CD80, CD40LG and GZMB are associated with T-cell activation (38). ICOS, CD247 and CD80 belong to the CD28 family, which is critical for controlling the adaptive arm of the immune response (39). It is suggested that immunity has an effect on prognosis. CXCL10 is an interferon-stimulated chemokine that attracts T cells (15). IL12A and IRF4 are associated with T-cell differentiation (40). Five of 20 genes were shown to be predictive in rectal cancer in previous studies. CXCL10, IL12A, CD247 and ITGB1 were identified as predictive markers for the response to neoadjuvant treatment in rectal cancer in previous studies (15, 41–45). Tumor IRS1 expression was related to the survival of rectal cancer patients (46).

To further understand the immunologic properties of subgroups classified according to the prognostic model, we studied the characteristics of the TIME of the different subgroups. Through the ESTIMATE algorithm, we found that the immune score in the low-risk group was significantly higher than that in the high-risk group, suggesting that the infiltration of immune cells was significantly different between the two groups. Our data revealed a significant difference in the distribution of some immune cells between the high-risk group and the low-risk group in the GSE119409 and GSE190826 cohorts. In the low-risk group, activated B cells, activated CD8 T cells, central memory CD8 T cells, macrophages, T follicular helper cells, and type 2 helper cells were significantly increased compared with those in the high-risk group. This is consistent with the current understanding that the presence of abundant immune cells in TIME was associated with good prognosis (47). A previous study indicated that the presence of B cells was associated with successful tumor regression following nCRT in LARC (48). Shinji e et al. reported that a high density of CD8+ T cells in tumors in baseline biopsy samples was associated with a good response to treatment (7). It was revealed that macrophages, T follicular helper cells, and type 2 helper cells were related to the prognosis of rectal cancer patients (49). However, these results could not be confirmed in GSE87211. The inconsistencies might be due to tumor heterogeneity, different detection methods and the complexity of the TIME.

There were several limitations in our study that should be acknowledged. Our study was based on publicly available datasets, and it was not possible to obtain complete clinical information and demographic data for each patient. In addition, the potential mechanisms, molecular interactions and clinical applications of the prognostic genes of rectal cancer need to be further explored. The association between better TRG and favorable clinical outcomes continues to be a matter of debate (50). In this study, uncertainty remains regarding the relationship between the low-risk group characterized by a high immunoscore and TRG 0-1. Due to the absence of TRG staging data in the included database, we did not investigate the potential correlation between the low-risk group and favorable TRG regression in this study. Nevertheless, the exploration of predictive models for TRG remains critically important, and we plan to conduct further research in this domain in the future.

## Conclusions

In conclusion, this study developed a prognostic risk prediction model by analyzing immune-related genes and using the random forest and elastic net methods. The IRGPM has good prediction accuracy and provides a good stratification strategy for further trials. Different groups separated by IRGPM had distinct TIME characterization. These findings suggest that the immune response might play a vital role in the prognosis of chemoradiotherapy.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

## Ethics statement

The studies involving humans were approved by Biomedical Ethics Review Committee of West China Hospital of Sichuan University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

PS: Data curation, Writing – original draft, Writing – review & editing, Conceptualization, Supervision. NL: Data curation, Writing – original draft, Writing – review & editing, Methodology, Validation. XL: Data curation, Methodology, Writing – original draft, Writing – review & editing, Visualization. YT: Conceptualization, Data curation, Writing – review & editing. ZC: Writing – review & editing. DL: Data curation, Writing – review & editing. DM: Methodology, Writing – review & editing. JD: Writing – review & editing. OY: Writing – review & editing. LS: Writing – review & editing. GO: Writing – review & editing. SW: Writing – review & editing. DJ: Writing – review & editing. XD: Writing – review & editing. ZW: Writing – review & editing. QL: Writing – review & editing. XW: Writing – original draft, Writing – review & editing.

## References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. Bailey CE, Hu CY, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975–2010. *JAMA Surg* (2015) 150(1):17–22. doi: 10.1001/jamasurg.2014.1756
3. Benson AB 3rd, Venook AP, Cederquist L, Chan E, Chen YJ, Cooper HS, et al. Colon cancer, version 1.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2017) 15(3):370–98. doi: 10.6004/jnccn.2017.0036
4. Gunderson LL, Sargent DJ, Tepper JE, Wolmark N, O'Connell MJ, Begovic M, et al. Impact of T and N stage and treatment on survival and relapse in adjuvant rectal

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by National Natural Science Foundation of China, No. 82073338; Sichuan Science and Technology Support Project, No. 2023YFS0319 and No. 2022YFS0217; The 1-3-5 Project for Disciplines of Excellence–Clinical Research Incubation Project West China Hospital, Sichuan University, No. 2020HXFH002; and The 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University, No. JCS2022034.

## Acknowledgments

Thank the staff of the Radiology Department of West China Hospital of Sichuan University for their help.

## Conflict of interest

QL was employed by Genecast Biotechnology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1294440/full#supplementary-material>

cancer: a pooled analysis. *J Clin Oncol* (2004) 22(10):1785–96. doi: 10.1200/jco.2004.08.173

5. Fokas E, Allgäuer M, Polat B, Klautke G, Grabenbauer GG, Fietkau R, et al. Randomized phase II trial of chemoradiotherapy plus induction or consolidation chemotherapy as total neoadjuvant therapy for locally advanced rectal cancer: CAO/ARO/AIO-12. *J Clin Oncol* (2019) 37(34):3212–22. doi: 10.1200/jco.19.00308

6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

7. Matsutani S, Shibutani M, Maeda K, Nagahara H, Fukuoka T, Nakao S, et al. Significance of tumor-infiltrating lymphocytes before and after neoadjuvant therapy for rectal cancer. *Cancer Sci* (2018) 109(4):966–79. doi: 10.1111/cas.13542

8. Shibutani M, Maeda K, Nagahara H, Fukuoka T, Iseki Y, Matsutani S, et al. Tumor-infiltrating lymphocytes predict the chemotherapeutic outcomes in patients with stage IV colorectal cancer. *In Vivo* (2018) 32(1):151–8. doi: 10.21873/invivo.11218
9. Taube JM, Galon J, Sholl LM, Rodig SJ, Cottrell TR, Giraldo NA, et al. Implications of the tumor immune microenvironment for staging and therapeutics. *Mod Pathol* (2018) 31(2):214–34. doi: 10.1038/modpathol.2017.156
10. Frey B, Rubner Y, Kulzer L, Werthmüller N, Weiss EM, Fietkau R, et al. Antitumor immune responses induced by ionizing irradiation and further immune stimulation. *Cancer Immunol Immunother* (2014) 63(1):29–36. doi: 10.1007/s00262-013-1474-y
11. Noshio K, Baba Y, Tanaka N, Shima K, Hayashi M, Meyerhardt JA, et al. Tumor-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *J Pathol* (2010) 222(4):350–66. doi: 10.1002/path.2774
12. Teng F, Meng X, Kong L, Mu D, Zhu H, Liu S, et al. Tumor-infiltrating lymphocytes, forkhead box P3, programmed death ligand-1, and cytotoxic T lymphocyte-associated antigen-4 expressions before and after neoadjuvant chemoradiation in rectal cancer. *Transl Res* (2015) 166(6):721–32.e1. doi: 10.1016/j.trsl.2015.06.019
13. Anitei MG, Zeitoun G, Mlecnik B, Marliot F, Haicheur N, Todosi AM, et al. Prognostic and predictive values of the immunoscore in patients with rectal cancer. *Clin Cancer Res* (2014) 20(7):1891–9. doi: 10.1158/1078-0432.Ccr-13-2830
14. Farchoukh L, Hartman DJ, Ma C, Celebrezze J, Medich D, Bahary N, et al. Intratumoral budding and automated CD8-positive T-cell density in pretreatment biopsies can predict response to neoadjuvant therapy in rectal adenocarcinoma. *Mod Pathol* (2021) 34(1):171–83. doi: 10.1038/s41379-020-0619-8
15. Qian L, Lai X, Gu B, Sun X. An immune-related gene signature for predicting neoadjuvant chemoradiotherapy efficacy in rectal carcinoma. *Front Immunol* (2022) 13:784479. doi: 10.3389/fimmu.2022.784479
16. Zhang Z, Zhang X. Identification of m6A-related biomarkers associated with prognosis of colorectal cancer. *Med Sci Monit* (2021) 27:e932370. doi: 10.12659/msm.932370
17. Zhang ZY, Yao QZ, Liu HY, Guo QN, Qiu PJ, Chen JP, et al. Metabolic reprogramming-associated genes predict overall survival for rectal cancer. *J Cell Mol Med* (2020) 24(10):5842–9. doi: 10.1111/jcmm.15254
18. Deist TM, Dankers F, Valdes G, Wijsman R, Hsu IC, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. *Med Phys* (2018) 45(7):3449–59. doi: 10.1002/mp.12967
19. Tang K, Ji X, Zhou M, Deng Z, Huang Y, Zheng G, et al. Rank-in: enabling integrative analysis across microarray and RNA-seq for cancer. *Nucleic Acids Res* (2021) 49(17):e99. doi: 10.1093/nar/gkab554
20. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B: Stat Method* (2005) 67(2):301–20. doi: 10.1111/j.1467-9868.2005.00503.x
21. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat* (2008) 2(3):841–860. doi: 10.1214/08-AOAS169
22. Pölsterl S. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res* (2020) 21(212):1–6.
23. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer (2000), ISBN: . doi: 10.1111/j.1467-9868.2005.00503.x
24. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl Genet Mol Biol* (2004) 3. doi: 10.2202/1544-6115.1027
25. Mootha V, Lindgren C, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* (2003) 34:267–73. doi: 10.1038/ng1180
26. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
27. Yi M, Nissley DV, McCormick F, Stephens RM. ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Sci Rep* (2020) 10:10258. doi: 10.1038/s41598-020-66986-8
28. Roh MS, Colangelo LH, O'Connell MJ, Yothers G, Deutsch M, Allegra CJ, et al. Preoperative multimodality therapy improves disease-free survival in patients with carcinoma of the rectum: NSABP R-03. *J Clin Oncol* (2009) 27(31):5124–30. doi: 10.1200/jco.2009.22.0467
29. Zhu Q, Wu X, Ma L, Xue H. Apoptosis-associated gene expression profiling is one new prognosis risk predictor of human rectal cancer. *Dis Markers* (2022) 2022:4596810. doi: 10.1155/2022/4596810
30. Liu J, Lan Y, Tian G, Yang J. A systematic framework for identifying prognostic genes in the tumor microenvironment of colon cancer. *Front Oncol* (2022) 12:899156. doi: 10.3389/fonc.2022.899156
31. Wang D, Xie B. Prognostic and tumor immunity implication of inflammatory bowel disease-associated genes in colorectal cancer. *Eur J Med Res* (2022) 27(1):91. doi: 10.1186/s40001-022-00720-0
32. Cui Y, Yu S, Zhu M, Cheng X, Yu Y, Tang Z, et al. Identifying predictive factors of recurrence after radical resection in gastric cancer by RNA immune-oncology panel. *J Cancer* (2020) 11(3):638–47. doi: 10.7150/jca.38536
33. Paluch BE, Glenn ST, Conroy JM, Papanicolau-Sengos A, Bshara W, Omilian AR, et al. Robust detection of immune transcripts in FFPE samples using targeted RNA sequencing. *Oncotarget* (2017) 8(2):3197–205. doi: 10.18632/oncotarget.13691
34. Chen B, Yang M, Li K, Li J, Xu L, Xu F, et al. Immune-related genes and gene sets for predicting the response to anti-programmed death 1 therapy in patients with primary or metastatic non-small cell lung cancer. *Oncol Lett* (2021) 22(1):540. doi: 10.3892/ol.2021.12801
35. Cheuk AT, Mufti GJ, Guinn BA. Role of 4-1BB:4-1BB ligand in cancer immunotherapy. *Cancer Gene Ther* (2004) 11(3):215–26. doi: 10.1038/sj.cgt.7700670
36. Brigl M, Brenner MB. CD1: antigen presentation and T cell function. *Annu Rev Immunol* (2004) 22:817–90. doi: 10.1146/annurev.immunol.22.012703.104608
37. Zhou T, Chen Y, Hao L, Zhang Y. DC-SIGN and immunoregulation. *Cell Mol Immunol* (2006) 3(4):279–83.
38. Kondou R, Akiyama Y, Iizuka A, Miyata H, Maeda C, Kanematsu A, et al. Identification of tumor microenvironment-associated immunological genes as potent prognostic markers in the cancer genome analysis project HOPE. *Mol Clin Oncol* (2021) 15(5):232. doi: 10.3892/mco.2021.2395
39. Riley JL, June CH. The CD28 family: a T-cell rheostat for therapeutic control of T-cell activation. *Blood* (2005) 105(1):13–21. doi: 10.1182/blood-2004-04-1596
40. Wong RWJ, Ong JZL, Theardy MS, Sanda T. IRF4 as an oncogenic master transcription factor. *Cancers (Basel)* (2022) 14(17):4314. doi: 10.3390/cancers14174314
41. Li L, Du X, Fan G. Identifying potential biomarkers of prognostic value in colorectal cancer via tumor microenvironment data mining. *Front Genet* (2021) 12:787208. doi: 10.3389/fgenet.2021.787208
42. Park JJ, Yu YS, Mustafa B, Park JY, Seo YB, Kim GD, et al. A nine-gene signature for predicting the response to preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Cancers (Basel)* (2020) 12(4):800. doi: 10.3390/cancers12040800
43. Park JJ, An S, Kim SY, Lim HM, Hong SM, Kim MJ, et al. Prediction of radio-responsiveness with immune-profiling in patients with rectal cancer. *Oncotarget* (2017) 8(45):79793–802. doi: 10.18632/oncotarget.19558
44. Nishioka M, Shimada M, Kurita N, Iwata T, Morimoto S, Yoshikawa K, et al. Gene expression profile can predict pathological response to preoperative chemoradiotherapy in rectal cancer. *Cancer Genomics Proteomics* (2011) 8(2):87–92.
45. Sun Y, Zhang Y, Wu X, Chi P. A four gene-based risk score system associated with chemoradiotherapy response and tumor recurrence in rectal cancer by co-expression network analysis. *Onco Targets Ther* (2020) 13:6721–33. doi: 10.2147/ott.S256696
46. Hanyuda A, Kim SA, Martínez-Fernández A, Qian ZR, Yamauchi M, Nishihara R, et al. Survival benefit of exercise differs by tumor IRS1 expression status in colorectal cancer. *Ann Surg Oncol* (2016) 23(3):908–17. doi: 10.1245/s10434-015-4967-4
47. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* (2017) 541(7637):321–30. doi: 10.1038/nature21349
48. Sendoya JM, Iseas S, Coraglio M, Golubicki M, Robbio J, Salanova R, et al. Pre-existing tumoral B cell infiltration and impaired genome maintenance correlate with response to chemoradiotherapy in locally advanced rectal cancer. *Cancers (Basel)* (2020) 12(8):2227. doi: 10.3390/cancers12082227
49. Chandra R, Karalis JD, Liu C, Murimwa GZ, Voth Park J, Heid CA, et al. The colorectal cancer tumor microenvironment and its impact on liver and lung metastasis. *Cancers (Basel)* (2021) 13(24):6206. doi: 10.3390/cancers13246206
50. Li JY, Huang XZ, Gao P, Song YX, Chen XW, Lv XE, et al. Survival landscape of different tumor regression grades and pathologic complete response in rectal cancer after neoadjuvant therapy based on reconstructed individual patient data. *BMC Cancer* (2021) 21(1):1214. doi: 10.1186/s12885-021-08922-1