

The BiomolBiomed publishes an “Advanced Online” manuscript format as a free service for authors in order to expedite the dissemination of scientific findings to the research community as soon as possible after acceptance following peer review and corresponding modifications (where appropriate). An “Advanced Online” manuscript is published online prior to copyediting, for publication and author proofreading, but is nonetheless fully citable through its Digital Object Identifier (doi®). Nevertheless, this “Advanced Online” version is NOT the final version of the manuscript. When the final version of this paper is published within a definitive issue of the journal, with copyediting, full pagination, etc., the new final version will be accessible through the doi and this “Advanced Online” version of the paper will disappear.

**Zhang et al: Regulatory cell signature in CRC**

# **Machine learning integration of single-cell and bulk transcriptomics identifies fibroblast-driven prognostic markers in colorectal cancer**

**Ning Zhang<sup>1,2</sup>, Ruiyan Liu<sup>1,2</sup>, Siya Wu<sup>1,2</sup>, Chenxi Feng<sup>1</sup>, Boxiang Wang<sup>1,2</sup>, Qiaoqiao Zheng<sup>1,2</sup>, Linru Jie<sup>1,2</sup>, Ruihua Kang<sup>1</sup>, Xiaoli Guo<sup>1</sup>, Xiaoyang Wang<sup>1\*</sup>, Shaokai Zhang<sup>1\*</sup>, and Jiangong Zhang<sup>1\*</sup>**

<sup>1</sup>Department of Cancer Epidemiology, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China;

<sup>2</sup>College of Public Health, Zhengzhou University, Zhengzhou, China.

\*Correspondence to **Xiaoyang Wang:** [xywang233@126.com](mailto:xywang233@126.com); **Shaokai Zhang:** [shaokaizhang@126.com](mailto:shaokaizhang@126.com) and **Jiangong Zhang:** [zhangjg@zzu.edu.cn](mailto:zhangjg@zzu.edu.cn)

**DOI:** <https://doi.org/10.17305/bb.2025.12038>

**Guest Editor:** Ziheng Wang

---

## ABSTRACT

Single-cell RNA sequencing (scRNA-seq) has significantly advanced our understanding of cellular heterogeneity and the complex interplay within the tumor microenvironment (TME) of colorectal cancer (CRC). However, translating these molecular insights into clinically actionable prognostic biomarkers and therapeutic strategies remains a considerable challenge. In this study, we conducted a comprehensive scRNA-seq analysis of 306 CRC samples comprising 448,255 cells to characterize the TME in depth. By constructing intercellular communication networks based on connection counts and communication probabilities, we identified fibroblasts as central regulatory hubs within the TME. Using Wilcoxon rank-sum tests and univariate survival analyses, we initially identified 23 prognostic fibroblast markers. These were refined to a seven-gene fibroblast-related prognostic signature via an integrated machine learning approach. The signature exhibited robust predictive performance in the The Cancer Genome Atlas - Colon Adenocarcinoma (TCGA-COAD) training cohort ( $n=351$ ; C-index=0.65) and was successfully validated in the GSE17536 dataset ( $n=177$ ; C-index=0.63). Functional enrichment analyses revealed that this signature is involved in immune regulation and multiple tumor-associated cellular pathways. Notably, high-risk patients displayed increased macrophage and NK cell infiltration, impaired immune function, and elevated immune rejection scores, while low-risk patients demonstrated heightened sensitivity to camptothecin and irinotecan. Together, our findings underscore the prognostic value of fibroblast-derived signatures in CRC and support their potential utility in risk stratification and the development of personalized therapeutic strategies, contributing to the advancement of precision oncology.

**Keywords:** Colorectal cancer; CRC; fibroblasts; prognosis signature; machine learning; therapy.

---

## INTRODUCTION

According to GLOBOCAN 2022, colorectal cancer (CRC) ranks fourth in global cancer incidence and third in cancer-related mortality.[1]. Despite advances in multimodal treatments[2, 3], patient outcomes remain poor, particularly for advanced cases with only 14% five-year survival rates [4, 5]. The poor prognosis of CRC can be attributed to several key factors. In particular, tumor heterogeneity-induced differential treatment responses and acquired drug resistance have led researchers to increasingly focus on the determinant influence of the tumor microenvironment (TME) in tumor prognosis and therapeutic outcomes[6, 7].

Recent studies[8, 9] have demonstrated that TME elements significantly influence CRC progression and treatment outcomes. While researchers have developed multiple prediction models derived from survival- correlated genes, these approaches have largely focused on tumor cell characteristics alone[10], overlooking the crucial role of the TME. Cancer-associated fibroblasts (CAFs) within the TME secrete various growth factors and cytokines that promote tumor growth and metastasis[11]. The composition of tumor-infiltrating lymphocytes has also been found to be closely associated with patient survival rates, with higher CD8+ T cell density typically indicating a better prognosis[12]. In addition, Regulatory T cells and myeloid-derived suppressor cells create an environment that inhibits anti-tumor immune responses[13]. This immunosuppression particularly affects the efficacy of immunotherapy, with only 15% of microsatellite instability-high CRC patients responding to immune checkpoint inhibitors (ICIs). Based on these complex interactions within the TME, a deep understanding and specific targeting of the key regulatory factors of the TME are expected to provide an important theoretical foundation for the prognosis assessment and optimization of personalized treatment plans for CRC.

Recent advances in single-cell RNA sequencing (scRNA-seq) [14] have enhanced our analysis of TME complexity. scRNA-seq provides gene expression profiles at single-cell resolution[15], enabling detailed cell subpopulation analysis, as demonstrated in CRC research where it identified distinct T cell exhaustion states[16]. While bulk RNA-seq[17] lacks this resolution, it offers larger-scale data essential for clinical pattern identification. Integrating these complementary approaches - combining the high-resolution cellular analysis of scRNA-seq with the large-scale validation capabilities of bulk RNA sequencing - represents a powerful prognostic assessment strategy in contemporary cancer research[18].

---

Therefore, this study aims to integrate high-resolution scRNA-seq data with large-scale bulk RNA sequencing data to gain a deeper understanding of the complexity of the TME and to identify key components within the CRC TME. We will focus on key regulatory factors discovered in our research and utilize advanced machine learning methods to provide deeper insights into the prognostic risks for CRC patients, thereby offering new perspectives for the identification of novel therapeutic targets.

## **MATERIALS AND METHODS**

### **scRNA-sequencing data collection and analysis**

The scRNA-seq data used in this study were obtained from a comprehensive dataset integrating 15 independent colorectal cancer cohorts compiled by Zhang et al [19]. This integrated dataset, which comprised 671,192 cells and 51,971 genes, is publicly available on Figshare (<https://figshare.com/>). We extracted all 204 tumor core tissues and 102 adjacent tissues from the integrated dataset. The sample types were classified based on the metadata information from the original dataset. Batch effects had already been addressed and corrected in the original integrated dataset using the Harmony algorithm, ensuring the homogeneity of samples across different datasets. Quality control excluded cells with fewer than 1,000 detected genes, mitochondrial gene content exceeding 20%, or red blood cell gene content above 3%.

scRNA-seq analysis was performed using Seurat v4.0[20] on tumor core tissue samples. The RunUMAP function facilitated nonlinear dimensionality reduction and visualization of the cell gene matrix. Cell clustering employed the Louvain algorithm with a 0.2 resolution parameter. Cell annotation was based on original dataset metadata and differential expression gene analysis using the FindAllMarkers function.

Subsequently, we employed CellChat analysis on the tumor core tissue samples to delineate the intercellular communication networks within the CRC TME. Based on curated ligand-receptor interaction databases, we quantified communication probabilities across signaling pathways. Comparative analysis between the 204 tumor core tissues and 102 adjacent non-cancerous tissues were performed to identify key cellular components governing TME interactions.

### **Bulk RNA-seq data collection**

Bulk RNA-seq datasets TCGA-COAD[21] and GSE17536[22]) were obtained from the TCGA and GEO, respectively. The training dataset comprised 351 CRC patient samples from TCGA-

---

COAD, with sample selection predicated on survival duration exceeding one month and comprehensive genomic expression data availability. An external validation cohort of 177 samples from GSE17536 was identified using identical inclusion criteria.

### **Machine learning-driven integrative signature development**

Following CellChat analysis highlighting fibroblasts' critical role in TME, we investigated their prognostic value in CRC patients. Using Wilcoxon rank-sum test, we identified fibroblast-specific marker genes (FSM genes) with strict criteria (LogFC threshold = 2, min.pct = 0.25, FDR < 0.05). Then, these FSM genes were screened through univariate Cox regression analysis in TCGA-COAD to identify potential prognostic markers.

To develop a robust risk score, we implemented an integrated machine learning approach combining 10 algorithms: Random Survival Forest, Elastic Net, Lasso regression, Ridge regression, Stepwise Cox regression, CoxBoost, Partial Least Squares Regression Cox model, Supervised Principal Component Analysis, Gradient Boosting Machine, and Survival-SVM. The integration process evaluated 101 algorithm combinations using Leave-One-Out Cross-Validation (LOOCV). Risk scores were calculated as linear combinations of gene expression levels, with model performance assessed via Concordance index (C-index). Top-performing algorithm combinations were selected based on validation set C-index and clinical translational potential, leading to the establishment of a fibroblast-related signature (FRS) for predicting CRC patient overall survival risk.

To validate FRS as an independent prognostic factor, we compared its ROC curves with other clinical characteristics and developed an integrated nomogram merging FRS with clinical characteristics to estimate survival outcomes in patients with CRC.

### **Enrichment analysis**

To explore the potential functions of FRS and the associated biological pathways, we utilized the STRING[23] to predict genes that may interact with the FRS gene, defining these genes as FRS-related genes (FRSR genes). Subsequently, functional enrichment analysis using KEGG pathways and GO terms were conducted on the FRSR genes to investigate the biological functions these genes play in tumor development. In this study, the KEGG pathway and GO enrichment analyses were performed using the OmicShare platform, an integrated online tool that provides comprehensive bioinformatics analysis functions with user-friendly visualization capabilities.

---

### **Exploration of immune characteristics**

To investigate the correlation between FRS and immune cell infiltration in the CRC TME, we quantified the infiltration levels of 22 immune cell types using the CIBERSORT method[24]. To validate the reliability and accuracy of the CIBERSORT analysis results, we also performed cross-validation using 5 supplementary algorithms: EPIC[25], Estimate[26], MCP-counter[27], QuantiSeq[27], TIMER[28].

To determine the immunogenicity based on immunomodulators, immunosuppressive cells, and effector cells, the immune response profile was characterized through TIDE score computation, which assesses patient responses to immunotherapy based on integrated gene expression data. The TIDE scores for TCGA-COAD patient samples were obtained from the TIDE (<http://tide.dfci.harvard.edu/>).

### **Drug discovery and sensitivity analysis**

To identify candidate therapeutic agents, we leveraged the Drug Signature Database (DSigDB) [29] to screen for compounds targeting FRS-associated genes. We then employed the oncoPredict package to evaluate chemotherapeutic sensitivity in CRC patients stratified by FRS risk scores. This approach enabled estimation of drug-specific IC50 values based on gene expression profiles, facilitating individualized drug response prediction.

### **Statistical analysis**

R software (version 4.4.0, R Foundation for Statistical Computing, Vienna, Austria) was used for the main analytical procedures, including data manipulation, statistical computations, and visualization. The analysis workflow incorporated a comprehensive suite of specialized R packages: Seurat for processing single-cell RNA sequencing data and correcting batch effects; dplyr, stringr, tidyverse, and reshape2 for data manipulation; scRNAtoolVis, ggplot2, ggpubr, and ComplexHeatmap for data visualization; DoubletFinder for doublet detection; CellChat for intercellular communication analysis; limma and Mime1 for integrating and constructing machine learning models; oncoPredict for drug sensitivity prediction; IOBR for immune infiltration analysis; survival, survminer, and ggDCA for survival analysis; and org.Hs.eg.db and msigdb for annotation and pathway analysis. Enrichment analyses were performed using the OmicShare platform.

For statistical analyses, the normality of the data was first assessed using the Shapiro-Wilk test, which revealed a non-normal distribution ( $p < 0.05$ ). Consequently, the Wilcoxon rank-sum test was employed for paired group comparisons, with data presented as median  $\pm$  interquartile

---

range (IQR). Nomograms were constructed using multivariate Cox regression analysis, and enrichment analysis applied hypergeometric tests to identify pathways or terms significantly enriched in differentially expressed genes compared to the whole genome background. All statistical tests were two-sided, with  $\alpha < 0.05$  considered statistically significant.

## **RESULTS**

### **T cells and epithelial cells constitute the primary cellular components of CRC tissues**

A graphic abstract of this study was presented in Figure 1. To comprehensively characterize the TME of CRC and delineate its cellular heterogeneity, we performed an integrated analysis of scRNA-seq data derived from 204 tumor core samples encompassing 15 independent datasets. The UMAP algorithm identified 21 distinct cell clusters (Figures S1A-C). Our analysis identified 15 distinct cell types (Figure 2A), with T cells and malignant/epithelial cells dominating the tumor tissue, alongside significant immune cell populations including monocytes/macrophages (11.6%) and NK cells (11.0%) (Figure 2B). To examine the precision of cell annotations, we constructed a heatmap of key marker gene expression across the various cell types (Figure 2C). Moreover, Gene Set Variation Analysis (GSVA) revealed that proliferating myeloid cells and proliferating T cells demonstrated significant enrichment in cell cycle-associated gene sets. Notably, fibroblasts showed marked upregulation of epithelial-mesenchymal transition signatures, implicating their potential involvement in facilitating tumor progression and metastasis (Figure 2D).

### **Cellchat analysis highlights fibroblasts as key regulators in CRC TME**

To identify the essential regulatory components in the CRC TME, we performed CellChat analysis on 204 tumor core tissues to systematically investigate intercellular communication patterns. Our analysis revealed that in the CRC tumor tissues, fibroblasts and endothelial cells exhibited higher net counts and interaction weights in intercellular interactions compared to other cell types (Figures 3A, B). These findings suggest that these two cell types serve as central regulators in modulating the functions and behaviors of other cells in the TME. Further investigation demonstrated that compared to other cells, fibroblasts contribute the most to the outgoing signals in the cellular communication network (Figures S2A, B). Significant variations in incoming and outgoing signal contributions were observed across different cellular groups (Figures 3C, D).

---

In addition, to elucidate key differences in cellular communication networks between tumor and normal tissues, we conducted comparative analysis between 204 tumor core tissues and 102 adjacent normal tissues and found that, compared to adjacent normal tissues, the net counts and interaction weights between fibroblasts, endothelial cells, and malignant/epithelial cells in tumor tissues were significantly increased (Figures 3E, F). The changes in fibroblasts are particularly notable (Figures 3G, H), indicating that in the CRC TME, fibroblasts may play an important regulatory role in tumor development by enhancing their interactions with other key cell types.

### **A robust 7-gene fibroblast-related signature predicts CRC prognosis using the LOOCV framework**

Acknowledging the central role of fibroblasts in cellular interactions, we explored the potential value of FRS in predicting the prognosis of CRC patients. Through the Findmarkers function, 435 FSM genes that are highly expressed in fibroblasts within tumor tissue have been identified. Then, we performed univariate Cox regression analysis on the FSM genes in TCGA-COAD, which has yielded 23 potential prognostic biomarkers. Subsequently, these markers were fitted into 101 combination models using LOOCV framework. The predictive performance of each model was evaluated by calculating C-index in both the training and validation set (Figure 4A).

All models were ranked based on their C-indices in the validation set. While four combination models incorporating all 23 biomarkers demonstrated optimal predictive performance (C-index = 0.64) (Figure 4A), we sought to develop a more clinically applicable signature. To enhance translational potential and minimize overfitting effects caused by multiple correlated genes, we conducted a comprehensive model selection process. This analysis revealed that a more parsimonious model, combining Lasso regression with forward stepwise Cox regression and including only 7 genes, achieved comparable predictive performance (C-index = 0.63). Based on these findings, we selected the Lasso+StepCox[forward] model as the optimal approach and developed a 7-gene FRS for predicting prognosis in CRC patients.

The Lasso+StepCox model-derived FRS stratified patients into high- and low-risk groups using median scores. Survival analyses demonstrated significantly poorer outcomes in high-risk patients across both training and validation sets (HR = 2.39 and 2.41 respectively, both  $P < 0.001$ ; Figures 4B, C). The robustness of the FRS as a prognostic tool was further supported by time-dependent ROC analysis, which demonstrated consistent predictive accuracy for

---

progression-free survival at 1-year (AUC in TCGA = 0.655, AUC in GSE17536 = 0.644), 3-year (AUC in TCGA = 0.612, AUC in GSE17536 = 0.639), and 5-year (AUC in TCGA = 0.678, AUC in GSE17536 = 0.68) intervals (Figures 4D-F).

The expression validation of 7 FRS genes in fibroblasts across independent scRNA-seq datasets was shown in Figures S4A-I.

### **Integration of FRS with clinical characteristics improves prognostic accuracy in CRC**

To assess the prognostic value of FRS in comparison with conventional clinical characteristics, we compared its prognostic value with characteristics including age, gender, tumor grade, and stage in GSE17536. Univariate Cox regression analysis revealed that FRS exhibited superior predictive accuracy (AUC=0.68) compared to these clinical characteristics, as evidenced by higher AUC values (Figure 5A). DCA demonstrated that FRS exhibited superior net clinical benefit across low-risk threshold (threshold < 0.5) probabilities (Figure 5B). Moreover, stratification analysis revealed differences in the distribution of tumor stage and grade among FRS-defined risk subgroups (Figures 5C, D).

To enhance the clinical utility of FRS, we developed a nomogram incorporating FRS with clinical characteristics using Cox regression analysis (Figure 5E). The model achieved a C-index of 0.81 (95% CI: 0.76-0.85), with FRS remaining an independent prognostic factor in multivariate analysis ( $P < 0.001$ ). The calibration curves demonstrated that the nomogram exhibited better predictive performance for 1-year survival compared to 3-year and 5-year survival predictions (Figure 5F). Additionally, these findings were independently validated in the TCGA-COAD (Figures S3A-F).

To further investigate the individual prognostic contributions of FRS genes, we performed univariate Cox analysis of the 7 signature genes. The results revealed that elevated expression of CSRP2, DBN1, FSTL3, GPX3, PAM, and RGS16 correlated with poor prognosis, while CXCL14 showed protective effects (Figure S5A). CSRP2 emerged as the strongest predictor (HR=1.94, 95% CI: 1.34-2.82). Risk score-based clustering demonstrated concordant patterns between gene expression, risk scores, and survival outcomes (Figures S5B, C).

### **FRS-related genes are enriched in immune regulation and cellular signaling pathways**

To elucidate the basic mechanisms by which FRS affects the prognosis of CRC patients, we analyzed genes interacting with FRS (designated as FRSR genes) using the STRING database (Figure 6A). According to the KEGG enrichment analysis of each module, the main classes

---

were signal transduction, immune system and signaling molecules and interaction (Figure 6B). Meanwhile, further analysis indicated that Cytokine-cytokine receptor interaction (ko04060) and Chemokine signaling pathway (ko04062) were the most significantly enriched pathways for FRSR genes (Figure 6C), suggesting that FRSR genes may regulate cancer-related processes through modulating immune system functions.

Similarly, GO enrichment analysis was performed on FRSR genes, categorizing terms with  $P < 0.05$  into three main aspects: biological processes, cellular components, and molecular functions (Figure 6D). The analysis revealed that FRSR genes were predominantly enriched in cellular process (GO:0009987), cellular anatomical entity (GO:0110165), and binding (GO:0005488). Additionally, bubble plot visualization demonstrated that FRSR genes were mainly involved in cellular signaling transduction and immune response, highlighting their broad participation in fundamental cellular activities (Figure 6E). These findings, together with the KEGG pathway analysis, further support the possibility that FRSR genes may influence tumor progression through immune regulation.

### **FRS correlates with immune microenvironment characteristics and predicts immunotherapy efficacy**

To evaluate the impact of the FRS on immune cell infiltration in CRC patients, we employed the CIBERSORT algorithm to quantify immune cell abundance in TCGA-COAD samples. Cross-validation using 5 additional algorithms confirmed the reliability of the results (Figure 7A). The results showed consistency across different algorithms. Notably, under the Estimate algorithm, the high-risk group had significantly higher stromalscore, immunescore, and estimatescore, but lower tumorpurity (Figure 7A). Wilcoxon rank-sum test analysis revealed statistically substantial differences in the infiltration of 10 immune cell types between risk subgroups (Figure 7B), notably increased infiltration of activated NK cells and macrophage subsets in high-risk group, suggesting that the infiltration of different immune cells may influence the prognosis of CRC patients. Therefore, we analyzed the relationship between immune cell infiltration and overall survival of CRC patients. The results showed that 6 cell types were significantly associated with the prognosis of CRC patients (Figures S6B-G). Integrating differential (Figure 7B) and survival analyses (Figures S6B-G) identified three key immune microenvironment cell types: M1 macrophages, activated NK cells, and resting memory CD4<sup>+</sup> T cells (Figure 7C).

---

To investigate the potential association between the prognostic biomarkers we developed and the response to immunotherapy, we employed the TIDE algorithm to predict immune escape scores in the training set. The results indicated that patients in the high-risk group exhibited significantly higher ( $P < 0.001$ ) immune escape scores (Figure 7D), suggesting that tumors in the high-risk group are more capable of evading immune system surveillance and attack. Further analysis of the two key TIDE indicators demonstrated that immunological disorders manifested as T cell dysfunction and T cell rejection, with both scores being significantly higher in the high-risk group compared to the low-risk group ( $P < 0.001$ ) (Figure 7E). These findings suggest that patients in the high-risk group might have a poorer response to immunotherapy, which is consistent with their unfavorable survival outcomes.

Moreover, to better understand the interplay between FRS genes and the immune microenvironment, we constructed a network diagram to display the correlations between these genes and different immune cell subsets (Figure 7F). We found that genes such as FSTL3 and CSRP2 showed a high positive correlation with multiple immune cell subsets, especially macrophage subsets and neutrophils. These results suggest that specific genes may play a key role in immune regulation.

### **FRS-targeted therapeutic agents reveal distinct drug response patterns in different risk groups**

To identify potential therapeutic agents targeting FRS, we analyzed the DSigDB database and ranked candidate drugs by p-value. The top 10 most significant candidates are listed in Table 1. Among these candidate drugs, four anticancer agents widely used in clinical practice were identified: camptothecin, irinotecan, sanguinarine, and daunorubicin. It is worth noting that camptothecin and Irinotecan, as topoisomerase I inhibitors, are commonly used chemotherapeutic agents in clinical treatment of CRC.

Considering that the dynamic and heterogeneous nature of the TME may lead to drug resistance, we assessed the sensitivity of FRS risk subgroups to five commonly used chemotherapy drugs for CRC, including camptothecin, Irinotecan, and three other standard treatments (5-Fluorouracil, paclitaxel, and oxaliplatin) (Figure 8A). The results indicated that enhanced sensitivity was observed in the low-risk group ( $P < 0.01$ ) for camptothecin and irinotecan. Conversely, the high-risk group demonstrated increased sensitivity to paclitaxel ( $P < 0.05$ ). The scatter plot (Figure 8B) between FRS risk scores and drug sensitivity reveals similar results: as the risk score increases, the sensitivity of tumor cells to camptothecin ( $P <$

---

0.001) and irinotecan ( $P < 0.001$ ) significantly decreases, while their sensitivity to paclitaxel markedly increases ( $P < 0.001$ ). Detailed correlation studies demonstrated a significant positive association between oxaliplatin sensitivity and multiple FRS gene expression patterns (Figure 8C), suggesting these FRS genes may serve distinct functions in modulating tumor cell responses to various chemotherapeutic agents.

## **DISCUSSION**

In this study, we extracted integrated scRNA-seq data and employed methods such as ssGSEA and CellChat to explore the heterogeneity of TME of CRC. The systematic analysis based on single-cell sequencing not only revealed the functional diversity of fibroblasts in the TME, but also facilitated the identification of specific marker genes. Building on these findings, we innovatively developed a prognostic signature composed of 7 fibroblast-related genes by integrating multiple machine learning algorithms. Importantly, the FRS demonstrated strong prognostic predictive capabilities in both the training set and independent validation cohorts, outperforming traditional clinical variables and exhibiting robustness across various algorithms (Lasso + CoxBoost and Lasso + plsRcox). Additionally, our study not only highlights the pivotal role of FRS in CRC prognosis but also elucidates its potential mechanism in affecting patient outcomes by modulating the tumor immune microenvironment. This provides a theoretical foundation for developing personalized treatment strategies based on FRS expression patterns.

Fibroblasts, abundant components in the tumor microenvironment, play crucial roles in modulating CRC initiation and progression[30]. Consistent with our enrichment analysis, multiple studies[31, 32] have demonstrated that fibroblasts create a favorable microenvironment for tumor growth and metastasis through promoting angiogenesis, immune modulation, and matrix remodeling, highlighting their prognostic significance in CRC patients. Furthermore, fibroblast-derived ECM proteins and matrix-remodeling MMPs not only form physical barriers but also increase matrix stiffness and interstitial pressure[33], impeding the penetration of chemotherapeutic and targeted agents. The accumulation of aberrant ECM components further exacerbates immunosuppression and interferes with immune checkpoint inhibitor efficacy[34], emphasizing the value of patient stratification based on fibroblast signatures and the development of targeted therapies for specific fibroblast-defined subgroups.

In contrast to Zhang et al[35], who developed a 20-marker CRC fibroblast-related prognostic signature using bulk RNA-seq and WGCNA with validation cohort AUCs of 0.638 and 0.55,

---

our study employed a more sophisticated approach. We utilized scRNA-seq to derive an FRS through a LOOCV framework integrating 10 machine learning algorithms and their combinations. Our methodology successfully reduced variable complexity, ultimately identifying a consensus 7-gene FRS with enhanced predictive performance. The signature demonstrated superior prognostic accuracy, achieving a 3-year AUC of 0.639, and exhibited strong clinical translational potential. Critically, integrating the FRS with clinical features in a nomogram revealed significant net benefit improvements, positioning it as a promising precision medicine tool for colorectal cancer survival prediction.

Our research reveals significant differences in different FRS subgroups regarding immune responses and cancer progression. The high-risk group exhibited increased infiltration of activated NK cells and specific macrophage subsets, suggesting compensatory activation of innate immunity in response to compromised adaptive immune function. However, this compensatory mechanism appears insufficient to inhibit tumor progression and may promote TME formation through chronic inflammation[36]. Decreased infiltration of plasma cells and resting memory CD4<sup>+</sup> T cells indicated impaired adaptive immunity, affecting tumor-specific antibody production, immune memory, and ICI efficacy[37]. High-risk patients demonstrated enhanced immune evasion during checkpoint blockade therapy, characterized by compromised T cell function and reduced tumor infiltration. This suggests a dual immunosuppressive mechanism in the high-risk TME: impaired T cell infiltration coupled with functional deficits in successfully infiltrating T cells.

Notably, all genes included in our constructed FRS have been reported to be closely associated with immune responses or tumor development. DBN1 is significantly overexpressed in the CRCE1 cell line, and immunohistochemical experiments have validated its association with CRC metastasis[38]. FSTL3, containing a follistatin-like domain[39], promotes tumor invasion and metastasis by modulating epithelial-mesenchymal transition (EMT) key molecules through the TGF- $\beta$ 1 signaling pathway. Studies have demonstrated that FSTL3 is significantly elevated in CRC tissues, particularly in high-grade tumors[40, 41]. Although GPX3 negatively correlates with cholesterol levels, it is significantly elevated in poorly differentiated and advanced CRC patients, influencing CRC development through regulation of the cholesterol-T cell immune axis[42]. PAM, a bifunctional enzyme commonly dysregulated in cancer[43], has been studied by Zhang et al. [44], who identified three PAM patterns with distinct prognoses and tumor microenvironment characteristics from 1,224 CRC samples. The high PAM level subgroup correlates with advanced stages, immune-suppressive

---

cell infiltration, and poor prognosis. RGS16, characterized by a conserved RH domain and  $\alpha$ -helix[45], exhibits high expression associated with poor overall survival in CRC patients [46]. Research has revealed that RGS16 inhibits JNK/P38-mediated apoptosis in CRC cells by disrupting TAB2/TAK1 recruitment to TRAF6[47]. CSRP2, a member of the CSRP protein family, shows lower expression in CRC tissues compared to adjacent non-tumor tissues, with functional experiments confirming its inhibitory effects on CRC cell proliferation, migration, and invasion[48]. Interestingly, our findings in this study regarding CSRP2's tumor-promoting role are inconsistent with previous reports, which may be attributed to tumor heterogeneity and multifactor analysis. CXCL14, a crucial member of the chemokine family, plays a vital role in immune regulation and tumor microenvironment[49]. In colorectal cancer, CXCL14 is primarily downregulated due to epigenetic silencing and exerts tumor-suppressive effects by inhibiting EMT and regulating cell cycle progression. Its low expression has been associated with poor prognosis in CRC patients[50].

Another key finding, computational analysis identified camptothecin and its semisynthetic derivative irinotecan as potential targeted therapeutics for FRS in CRC, consistent with clinical practice where irinotecan-based FOLFIRI[51] and oxaliplatin-based FOLFOX[52] regimens serve as standard treatments. Our analysis suggests enhanced efficacy of camptothecin and its derivative irinotecan in CRC patients with low FRS scores, offering a molecular classification-based approach to treatment selection. The therapeutic efficacy of drugs in colorectal cancer is strongly supported by Douillard et al[53], whose multicenter randomized controlled trial (n=387) demonstrated that in metastatic colorectal cancer patients who failed 5-FU treatment, irinotecan monotherapy significantly prolonged median overall survival compared to best supportive care (9.2 months vs. 6.5 months), and improved objective response rate (13% vs. 0%) and disease control rate (49% vs. 21%). However, TME heterogeneity frequently leads to treatment resistance, compromising therapeutic efficacy; therefore, these findings warrant comprehensive clinical validation, particularly through prospective clinical trials stratifying patients by FRS risk score to directly compare sensitivity to both camptothecin and its derivative irinotecan across molecular subtypes, which would provide definitive evidence for implementing this molecular classification-based approach in personalized treatment decisions.

Our investigation underscores the pivotal role of FRS in informing targeted prevention and personalized medicine for CRC. The findings suggest that FRS can furnish vital insights to support clinicians' individualized treatment decisions, thereby improving patient outcomes

---

while potentially curbing unnecessary costs. Nonetheless, this study has certain limitations. First, although the FRS were evaluated and validated in both training and external cohorts, further confirmation through large-scale, multicenter prospective investigations is needed. Second, additional *in vitro* and *in vivo* research will be crucial for elucidating the biological mechanisms of FRS-related genes in CRC. Third, although we assessed the sensitivity of different FRS risk subgroups to various small-molecule drugs, these predictions still require validation through *in vitro* drug experimentation and clinical trials. Fourth, while our current study identified fibroblasts as an important regulatory component in the tumor microenvironment, our single-cell analysis pipeline did not delve into the heterogeneity of fibroblast populations. A more refined characterization of fibroblast subtypes, including their distinct molecular signatures and functional states, might provide additional insights into their diverse roles in tumor progression and potentially enhance the prognostic value of our signature. Future research should focus on more nuanced classification of stromal cell populations to further optimize risk stratification strategies. Collectively, these limitations point to key directions for future work.

## CONCLUSION

This study provides a comprehensive single-cell RNA sequencing analysis of colorectal cancer, revealing the critical role of fibroblasts in the tumor microenvironment. By developing a novel 7-gene fibroblast-related signature, we have demonstrated a robust prognostic tool that not only predicts patient survival with high accuracy but also offers insights into potential personalized treatment strategies. The FRS signature highlights the complex interactions between fibroblasts, immune cells, and cancer progression, opening new avenues for precision medicine in colorectal cancer management.

## ACKNOWLEDGMENTS

We extend our sincere gratitude to our colleagues at the Department of Cancer Epidemiology, Henan Cancer Hospital, especially Dr. Cheng and Hongwei Liu, for their expertise in data analysis and visualization that greatly enhanced the quality of this study.

**Conflicts of interest:** Authors declare no conflicts of interest.

**Funding:** This work was supported by the National Natural Science Foundation of China (82204128), Henan Provincial Medical Science and Technology Research Program—Provincial-Ministerial Co-constructed Youth Project (SBGJ202303015), and Henan Provincial

---

Science and Technology Research Program (242102310040) for Xiaoyang Wang; the China Postdoctoral Science Foundation (2024M750824) for Xiaoli Guo.

**Data availability:** The data acquisition methods for this study have been mentioned in the text, and researchers can obtain additional data and analysis scripts by contacting the corresponding author.

**Submitted:** 10 January 2025

**Accepted:** 12 April 2025

**Published online:** 22 April 2025

## REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-63.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30.
3. Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. NCCN Guidelines Insights: Colon Cancer, Version 2.2018. *J Natl Compr Canc Netw.* 2018;16(4):359-69.
4. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet.* 2019;394(10207):1467-80.
5. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(3):145-64.
6. Källberg J, Harrison A, March V, Bērziņa S, Nemazanyy I, Kepp O, et al. Intratumor heterogeneity and cell secretome promote chemotherapy resistance and progression of colorectal cancer. *Cell Death Dis.* 2023;14(5):306.
7. Zhang A, Miao K, Sun H, Deng CX. Tumor heterogeneity reshapes the tumor microenvironment to influence drug resistance. *Int J Biol Sci.* 2022;18(7):3019-33.
8. Fadlallah H, El Masri J, Fakhreddine H, Youssef J, Chemaly C, Doughan S, et al. Colorectal cancer: Recent advances in management and treatment. *World J Clin Oncol.* 2024;15(9):1136-56.
9. Novoa Díaz MB, Martín MJ, Gentili C. Tumor microenvironment involvement in colorectal cancer progression via Wnt/ $\beta$ -catenin pathway: Providing understanding of the complex mechanisms of chemoresistance. *World J Gastroenterol.* 2022;28(26):3027-46.
10. Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J Gastroenterol.* 2016;22(5):1745-55.
11. Wright K, Ly T, Kriet M, Czirok A, Thomas SM. Cancer-Associated Fibroblasts: Master Tumor Microenvironment Modifiers. *Cancers (Basel).* 2023;15(6).
12. Chatzopoulos K, Kotoula V, Manoussou K, Markou K, Vlachtsis K, Angouridakis N, et al. Tumor Infiltrating Lymphocytes and CD8+ T Cell Subsets as Prognostic Markers in

- 
- Patients with Surgically Treated Laryngeal Squamous Cell Carcinoma. *Head Neck Pathol.* 2020;14(3):689-700.
13. Heregger R, Huemer F, Steiner M, Gonzalez-Martinez A, Greil R, Weiss L. Unraveling Resistance to Immunotherapy in MSI-High Colorectal Cancer. *Cancers (Basel).* 2023;15(20).
  14. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5):e1001453.
  15. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol.* 2018;18(1):35-45.
  16. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, et al. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell.* 2019;176(4):775-89.e18.
  17. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13.
  18. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell.* 2017;171(7):1611-24.e24.
  19. Chu X, Li X, Zhang Y, Dang G, Miao Y, Xu W, et al. Integrative single-cell analysis of human colorectal cancer reveals patient stratification with distinct immune evasion mechanisms. *Nat Cancer.* 2024;5(9):1409-26.
  20. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573-87.e29.
  21. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell.* 2018;173(2):283-5.
  22. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* 2010;138(3):958-68.
  23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-d13.
  24. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453-7.
  25. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife.* 2017;6.
  26. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
  27. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
  28. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48(W1):W509-w14.
  29. Musa A, Ghorraie LS, Zhang SD, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform.* 2018;19(3):506-23.

- 
30. Luo H, Xia X, Huang LB, An H, Cao M, Kim GD, et al. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat Commun.* 2022;13(1):6619.
  31. Abudukelimu S, de Miranda N, Hawinkels L. Fibroblasts in Orchestrating Colorectal Tumorigenesis and Progression. *Cell Mol Gastroenterol Hepatol.* 2024;17(5):821-6.
  32. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer.* 2016;16(9):582-98.
  33. Provenzano PP, Cuevas C, Chang AE, Goel VK, Von Hoff DD, Hingorani SR. Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. *Cancer Cell.* 2012;21(3):418-29.
  34. Ireland L, Santos A, Ahmed MS, Rainer C, Nielsen SR, Quaranta V, et al. Chemoresistance in Pancreatic Cancer Is Driven by Stroma-Derived Insulin-Like Growth Factors. *Cancer Res.* 2016;76(23):6851-63.
  35. Zhang L, Xu C, Wang SH, Ge QW, Wang XW, Xiao P, et al. Cancer-associated fibroblast-related gene signatures predict survival and drug response in patients with colorectal cancer. *Front Genet.* 2022;13:1054152.
  36. T GS. Innate and adaptive immune cells in Tumor microenvironment. *Gulf J Oncolog.* 2021;1(35):77-81.
  37. Yap TA, Parkes EE, Peng W, Moyers JT, Curran MA, Tawbi HA. Development of Immunotherapy Combination Strategies in Cancer. *Cancer Discov.* 2021;11(6):1368-97.
  38. Lin Q, Tan HT, Lim TK, Khoo A, Lim KH, Chung MC. iTRAQ analysis of colorectal cancer cell lines suggests Drebrin (DBN1) is overexpressed during liver metastasis. *Proteomics.* 2014;14(11):1434-43.
  39. Tian S, Xu X, Yang X, Fan L, Jiao Y, Zheng M, et al. Roles of follistatin-like protein 3 in human non-tumor pathophysiologicals and cancers. *Front Cell Dev Biol.* 2022;10:953551.
  40. Li CH, Fang CY, Chan MH, Chen CL, Chang YC, Hsiao M. The cytoplasmic expression of FSTL3 correlates with colorectal cancer progression, metastasis status and prognosis. *J Cell Mol Med.* 2023;27(5):672-86.
  41. Li H, Zheng N, Guo A, Tang W, Li M, Cao Y, et al. FSTL3 promotes tumor immune evasion and attenuates response to anti-PD1 therapy by stabilizing c-Myc in colorectal cancer. *Cell Death Dis.* 2024;15(2):107.
  42. Chen J, Wu Y, Zhou Q, Song Y, Zhuang J, Lu K, et al. GPX3 is a key cholesterol-related gene associated with prognosis and tumor-infiltrating T cells in colorectal cancer. *Neoplasma.* 2023;70(5).
  43. Park SW, Kang J, Kim HS, Yoon S, Kim BS, Lim C, et al. Predicting prognosis through the discovery of specific biomarkers according to colorectal cancer lymph node metastasis. *Am J Cancer Res.* 2023;13(7):3221-33.
  44. Zhang E, Ding C, Li S, Aikemu B, Zhou X, Fan X, et al. Polyamine metabolism patterns characterized tumor microenvironment, prognosis, and response to immunotherapy in colorectal cancer. *Cancer Cell Int.* 2023;23(1):96.
  45. Bansal G, Druey KM, Xie Z. R4 RGS proteins: regulation of G-protein signaling and beyond. *Pharmacol Ther.* 2007;116(3):473-95.
  46. Miyoshi N, Ishii H, Sekimoto M, Doki Y, Mori M. RGS16 is a marker for prognosis in colorectal cancer. *Ann Surg Oncol.* 2009;16(12):3507-14.
  47. Shen H, Yuan J, Tong D, Chen B, Yu E, Chen G, et al. Regulator of G protein signaling 16 restrains apoptosis in colorectal cancer through disrupting TRAF6-TAB2-TAK1-JNK/p38 MAPK signaling. *Cell Death Dis.* 2024;15(6):438.

48. Chen L, Long X, Duan S, Liu X, Chen J, Lan J, et al. CSRP2 suppresses colorectal cancer progression via p130Cas/Rac1 axis-mediated ERK, PAK, and HIPPO signaling pathways. *Theranostics*. 2020;10(24):11063-79.
49. Yang XY, Ozawa S, Kato Y, Maehata Y, Izukuri K, Ikoma T, et al. C-X-C Motif Chemokine Ligand 14 is a Unique Multifunctional Regulator of Tumor Progression. *Int J Mol Sci*. 2019;20(8).
50. Gowhari Shabgah A, Haleem Al-Qaim Z, Markov A, Valerievich Yumashev A, Ezzatifar F, Ahmadi M, et al. Chemokine CXCL14; a double-edged sword in cancer development. *Int Immunopharmacol*. 2021;97:107681.
51. Tournigand C, André T, Achille E, Lledo G, Flesh M, Mery-Mignard D, et al. FOLFIRI Followed by FOLFOX6 or the Reverse Sequence in Advanced Colorectal Cancer: A Randomized GERCOR Study. *J Clin Oncol*. 2023;41(19):3469-77.
52. Neugut AI, Lin A, Raab GT, Hillyer GC, Keller D, O'Neil DS, et al. FOLFOX and FOLFIRI Use in Stage IV Colon Cancer: Analysis of SEER-Medicare Data. *Clin Colorectal Cancer*. 2019;18(2):133-40.
53. Chai Y, Liu JL, Zhang S, Li N, Xu DQ, Liu WJ, et al. The effective combination therapies with irinotecan for colorectal cancer. *Front Pharmacol*. 2024;15:1356708.

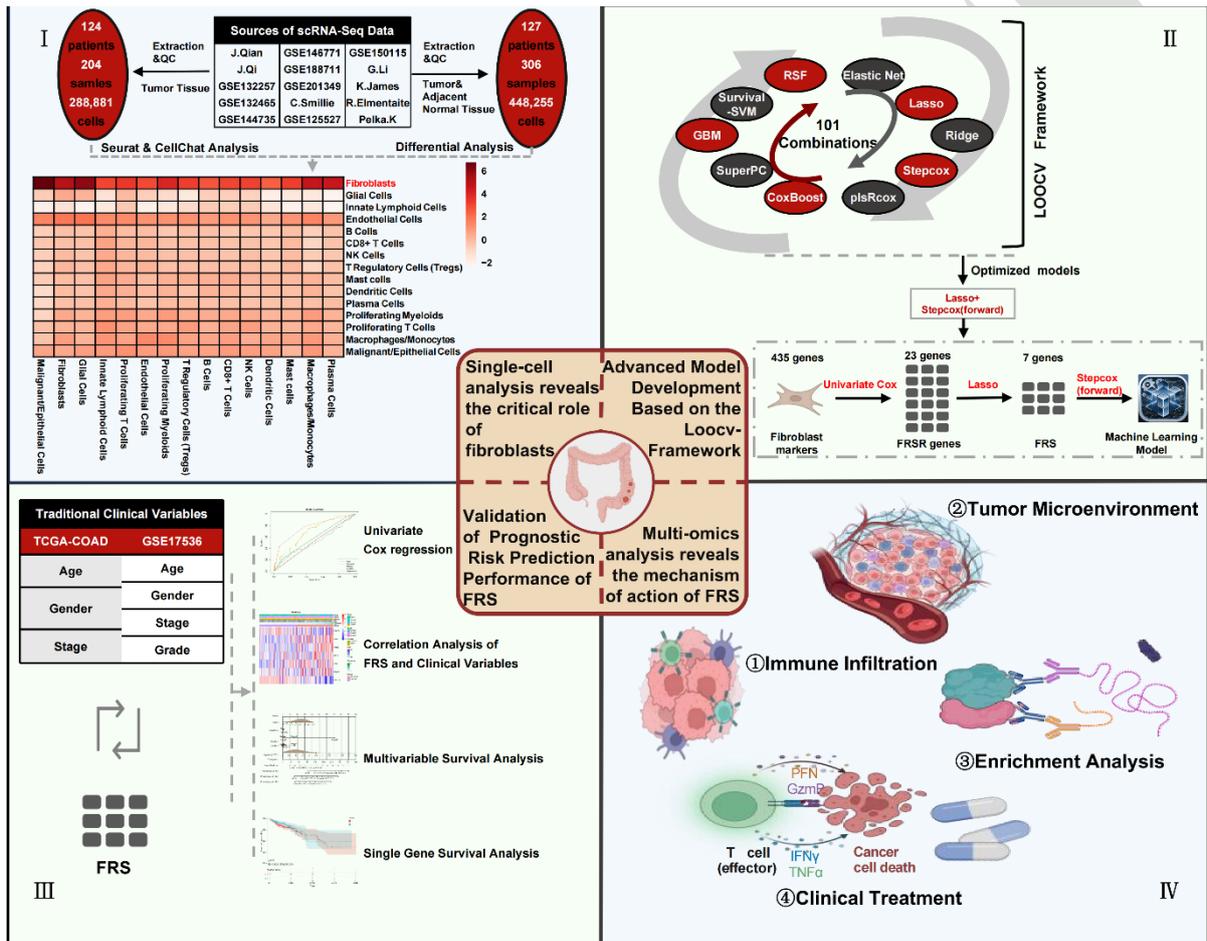
## TABLES AND FIGURES WITH LEGENDS

**Table 1. Drug prediction based on intrinsic target genes of FRS from DSigDB**

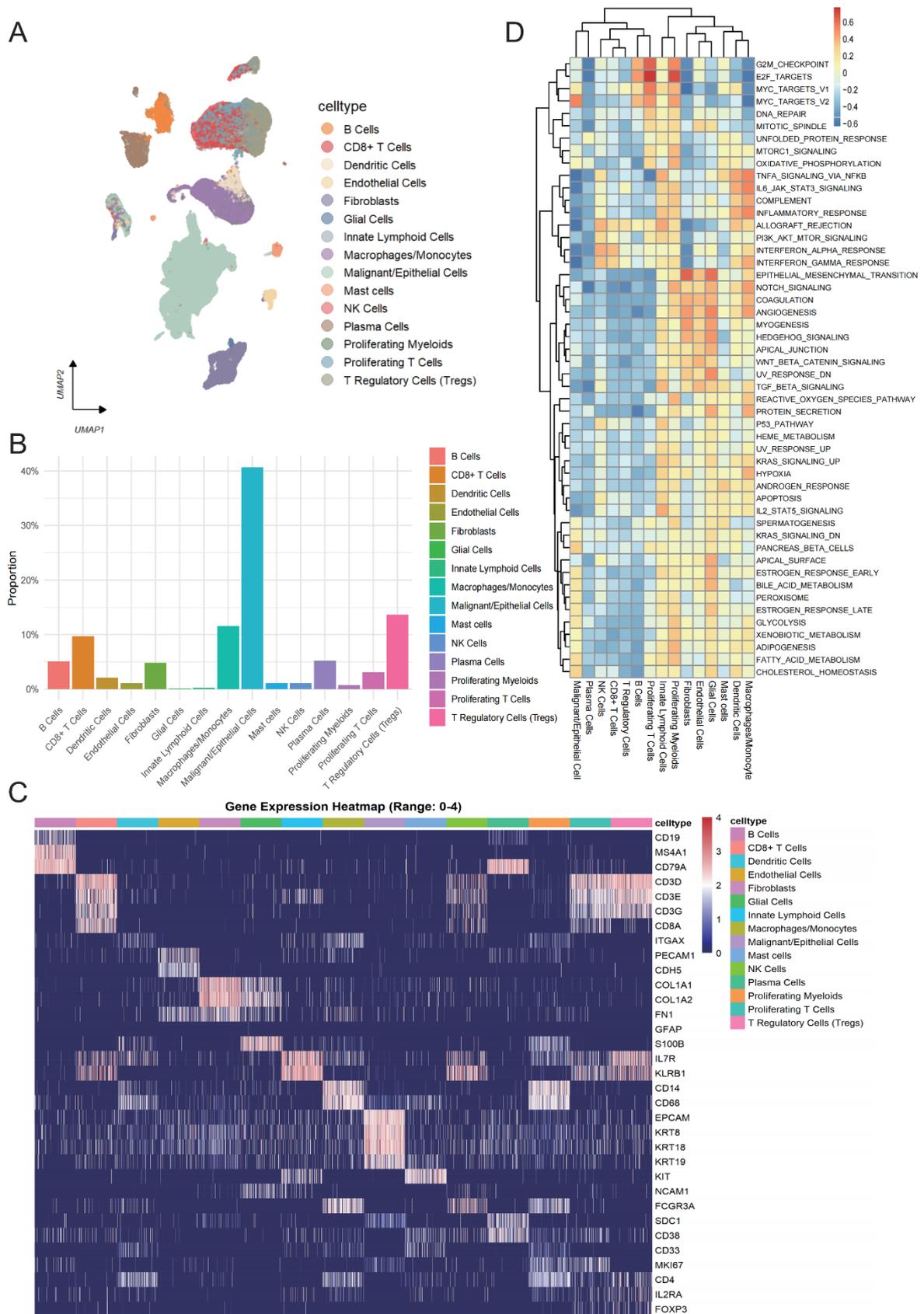
Term	P-value	Odds Ratio <sup>a</sup>	Combined Score <sup>b</sup>	Genes <sup>c</sup>
estradiol	<0.001	109648	1173816	CSRP2; GPX3; RGS16; CXCL14; PAM; DBN1; FSTL3
camptothecin	<0.001	67.7	676.8	GPX3; RGS16; FSTL3
ellipticine	<0.001	100.8	806.6	CSRP2; RGS16
15-delta prostaglandin	<0.001	66.2	475.8	CSRP2; GPX3
irinotecan	0.0012	21.7	146.1	GPX3; RGS16; FSTL3
sanguinarine	0.0019	40.8	254.9	CSRP2; RGS16
progesterone	0.0023	12.6	76.6	CSRP2; GPX3; CXCL14; DBN1
daunorubicin	0.0027	34.4	203.2	CSRP2; RGS16
3-Nitrobenzanthrone	0.0038	333.1	1852.2	CXCL14

Arecoline hydrobromide	0.0042	302.8	1657.4	RGS16
------------------------	--------	-------	--------	-------

\*<sup>a</sup>Odds Ratio: a measure of the association strength between the term and target genes. <sup>b</sup> Combined Score: a comprehensive indicator, higher scores indicate more reliable associations between the term and target genes. <sup>c</sup> Genes: genes that interact with the drug.



**Figure 1. Abstract figure of this study.** Integrated analysis of colorectal cancer tumor microenvironment and development of key cellular signatures for overall survival prediction and personalized therapy. QC refers to quality control, FRS refers to fibroblast-related signature, FRSR genes refers to FRS-related genes.

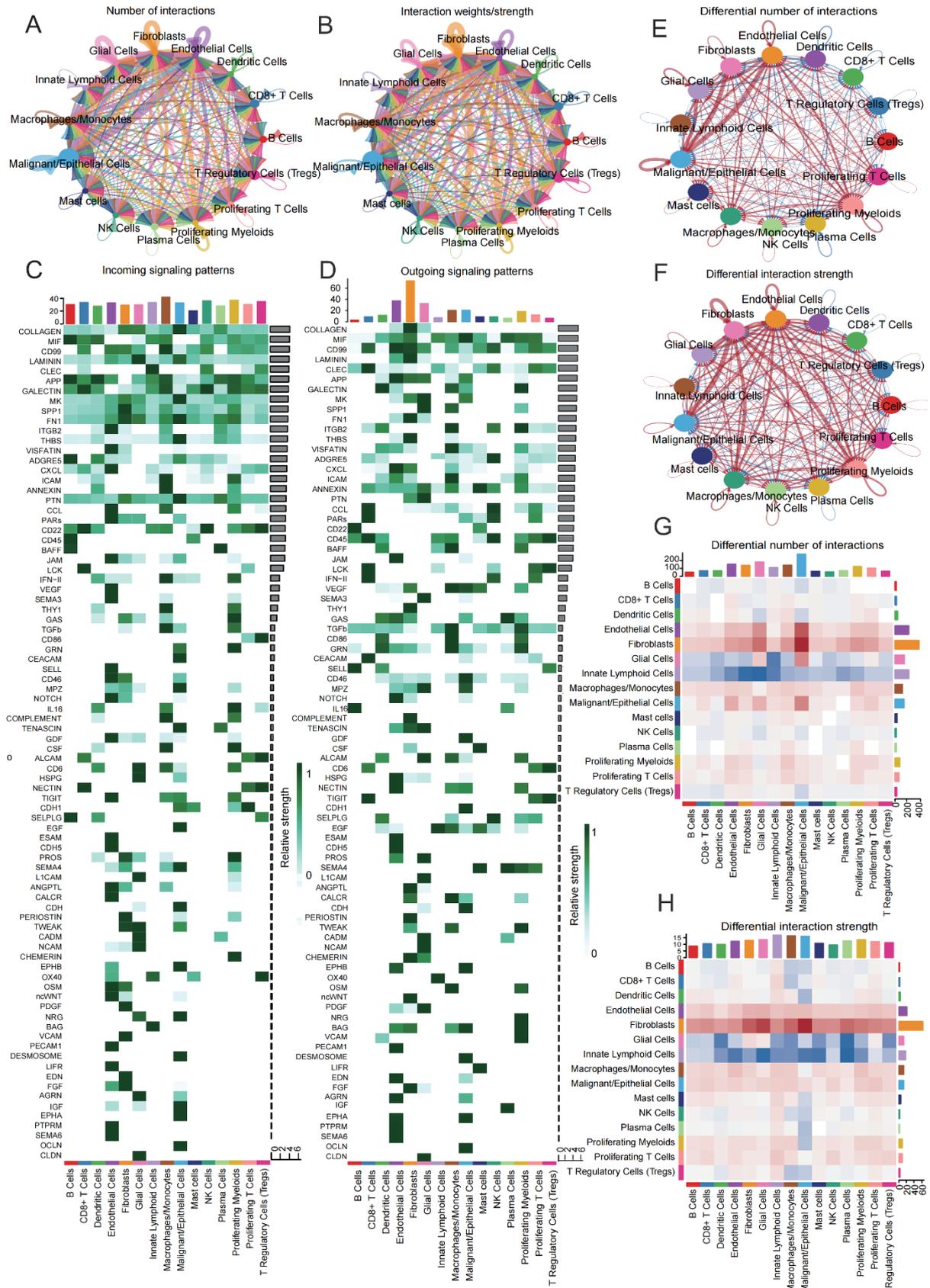


**Figure 2. Overview of various cell types in the TME of CRC at the scRNA transcriptome level. (A)** Identification of 15 cell types in 204 CRC tumor center tissues. **(B)** Bar plot showing

---

the distribution proportion of each cell type. **(C)** Heatmap showing the expression distribution of classic marker genes for each cell type. **(D)** ssGSEA score from hallmark gene sets of Human MSigDB Collections, Changes in the heatmap cell colors indicate the level of activity of the corresponding pathways in specific cell types.

EARLY ACCESS

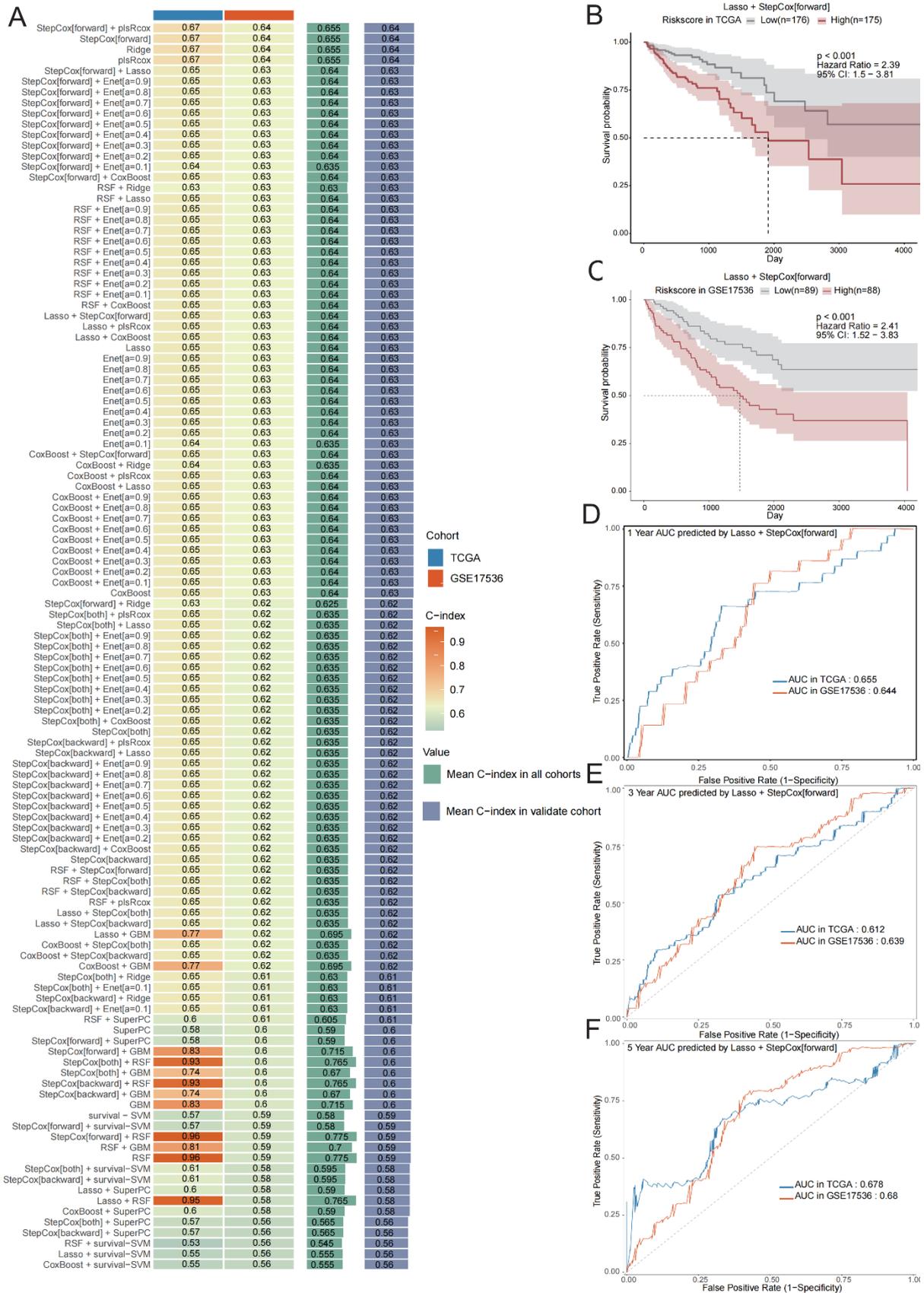


**Figure 3. CellChat reveals the crucial role of fibroblasts in the TME of CRC.**

---

The circular network diagram displays the net number (**A**) and strength (**B**) of intercellular communication among different cell types in tumor tissue. colors represent different cell types, while the thickness of the lines indicates the magnitude of change. Heatmaps detail the contribution of signaling molecules, separately showing incoming(**C**) and outgoing(**D**) signals for each cell type. Circular plots display the differential networks of net number (**E**) and interaction strengths (**F**) of cell-cell interactions between tumor core and adjacent normal tissues. Blue indicates decrease, red indicates increase, and the thickness of the lines represents the magnitude of change. The heatmap provides a more intuitive representation of the variations in the quantity(**G**) and intensity(**H**) of interactions between tumor core and adjacent normal tissues.

EARLY ACCESS

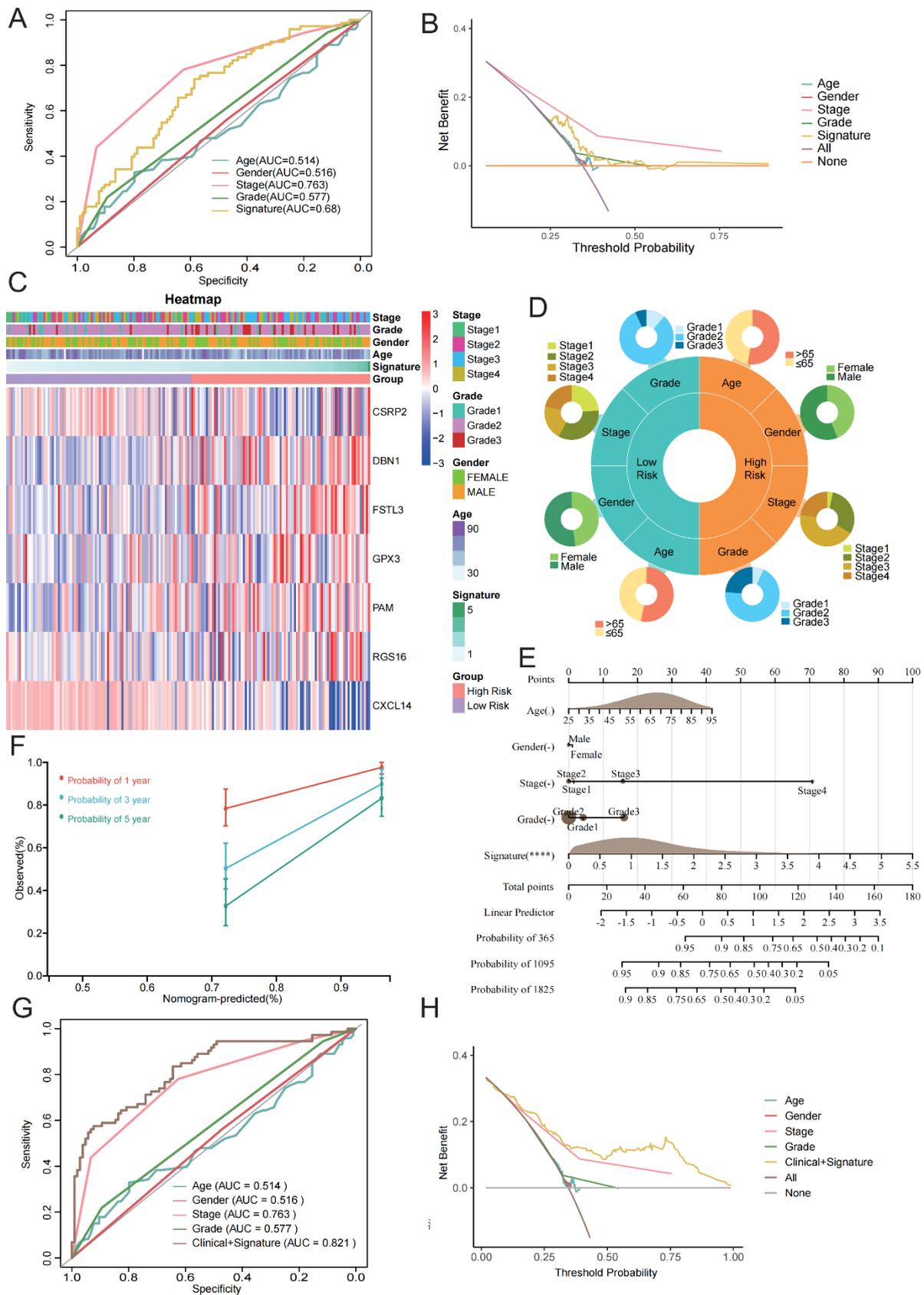


**Figure 4. Establishment and validation of a consensus FRS through a machine learning-based integrated pipeline. (A)** total of 101 types of prediction models using the LOOCV

---

framework and further calculated the C-index of each model in all validation datasets. Kaplan-Meier survival curves for different risk groups in TCGA-COAD (**B**) and GSE17536 (**C**). The ROC curves illustrate the OS performance at 1 year (**D**), 3 years (**E**), and 5 years (**F**) in the two datasets.

EARLY ACCESS

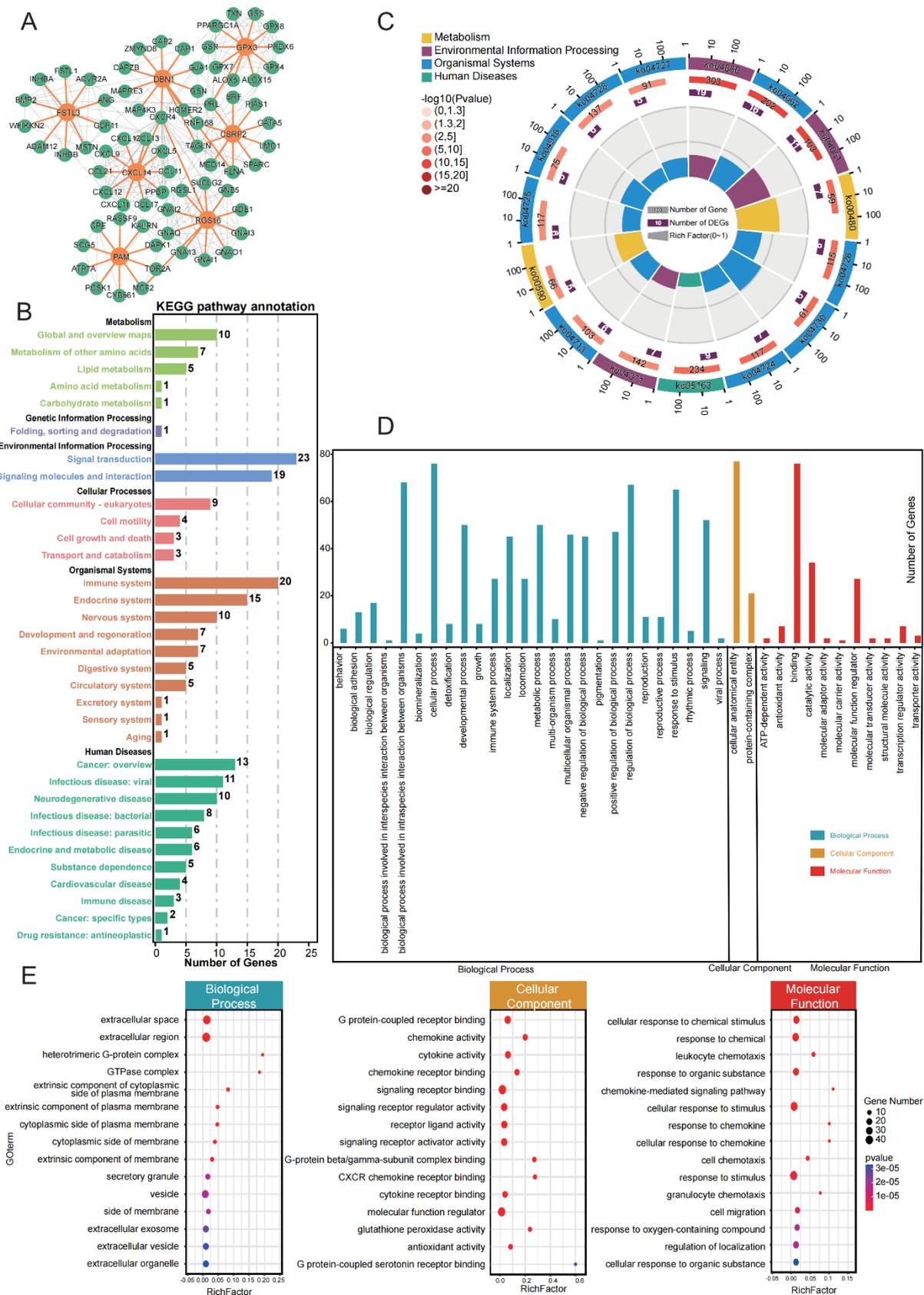


**Figure 5. Compared to traditional clinical variables, FRS demonstrates potential as an independent prognostic factor in GSE17536. Univariate ROC Curve (A) and DCA Curve**

---

(B) of Clinical characteristics and Signature. (C)The distribution of clinical characteristics and the expression of model genes according to the FRS risk score. (D)Correlation between the FRS low- and high-risk groups and clinical characteristics. (E)Construction of the nomogram based on the FRS and clinical characteristics, including age, gender, grade and stage. (F)Calibration curve of the nomogram for 1, 3, and 5-year OS. ROC curve(G) and DCA Curve(H) shows the prediction performance between the nomogram and Clinical characteristics.

EARLY ACCESS

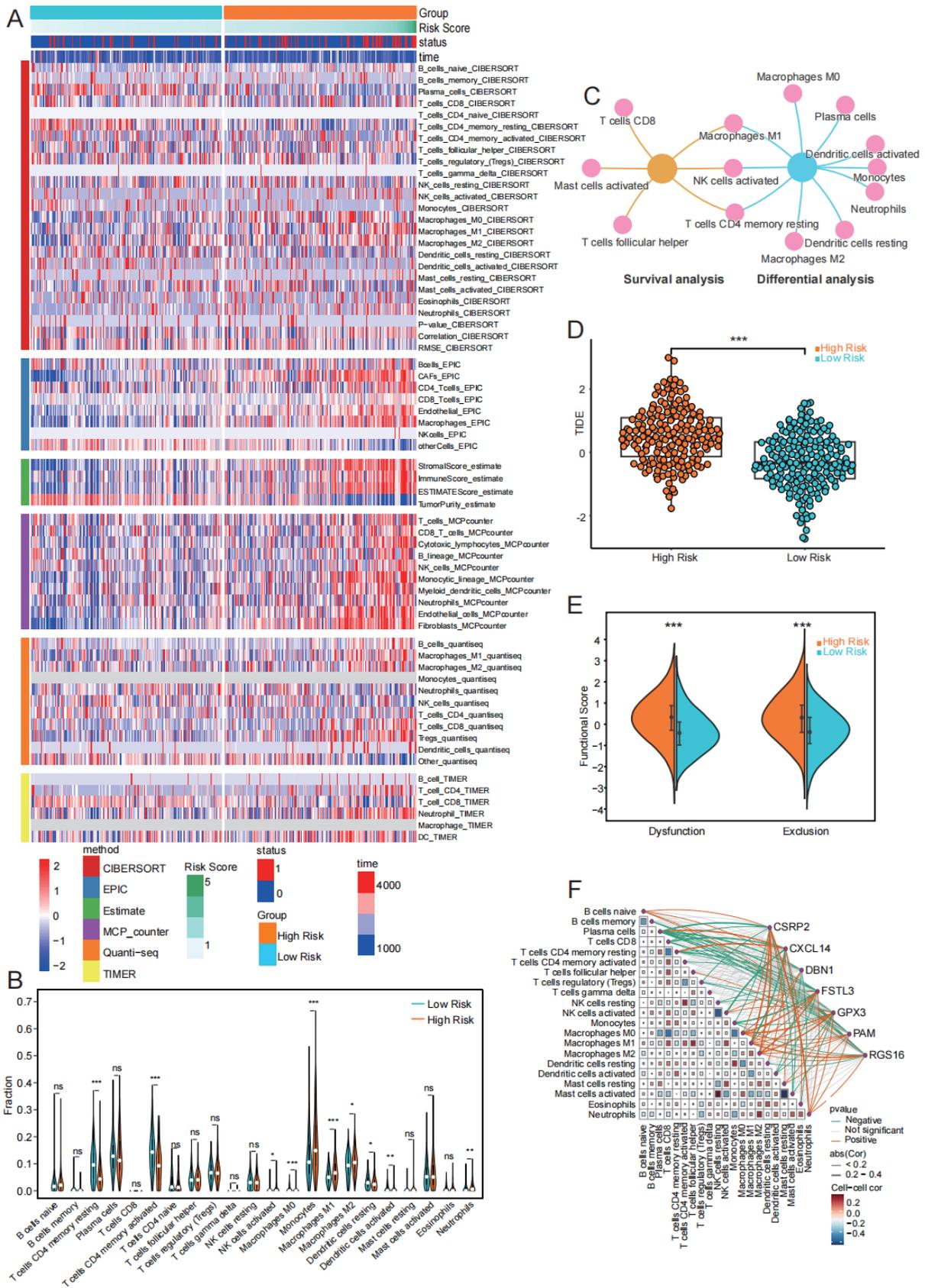


**Figure 6. Functional enrichment analysis reveals the potential molecular mechanisms by which FRS affects prognosis. (A)PPI network of FRSR genes predicted through String**

---

database. **(B)**The number of differentially expressed genes (DEGs) in various pathways in KEGG analysis. **(C)** KEGG enrichment circle diagram of cyan module (from the outside to the inside, the first circle represents the top 15 enrichment pathways, and the number outside the circle is the coordinate ruler of the number of genes; The second circle represents the number and Q value of background genes in this pathway, and the more genes, the longer the bar; The third circle represents the number of the DEGs in this pathway; The fourth circle represents the value of Rich Factor in each pathway). **(D)**The number of DEGs in various terms in GO enrichment analysis. **(E)**Enrichment of the top 15 gene sets (ranked by p-value) across the three primary categories of GO enrichment analysis.

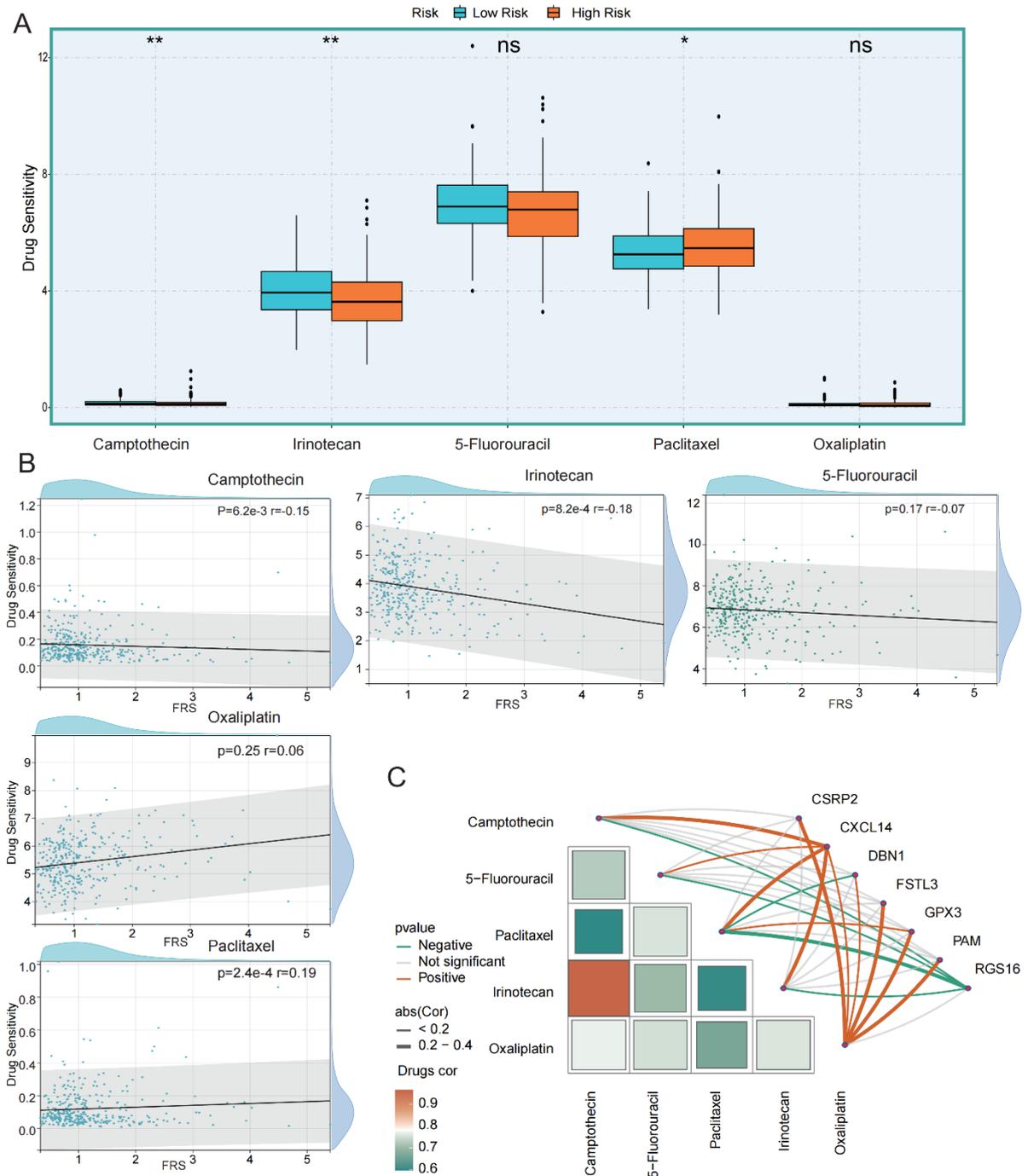
EARLY ACCESS



**Figure 7. The immune landscape associated with FRS in CRC. (A)**The heatmap shows the immune infiltration landscape of 6 methods under different FRS risk subgroups. **(B)**The

---

abundance of each type of tumor-infiltrating immune cell in high-risk and low-risk groups, ns (not significant), \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ). (C) Types of infiltrating cells in the TME intersecting with differential analysis and survival analysis. (D) TIDE scores for each sample within different risk subgroups, TIDE refers to the TIDE score derived from the TIDE database. ns (not significant), \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ). (E) T-cell Dysfunction and Exclusion Scores between High- and Low-Risk Groups, ns (not significant), \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ). (F) The network diagram of the correlation between immune cell subsets and gene expression (The correlation between specific cell and FRS genes is represented by different colored lines, with orange indicating positivity and green indicating negativity. The thickness of the lines reflects the strength of the correlation. Abs[52] represents the absolute value of the correlation coefficient, and Cell-cell cor indicates the correlation coefficient between different cell types.).



**Figure 8. The drug sensitivity associated with FRS in CRC. (A)** Drug sensitivity of five compounds calculated by the oncoPredict function across different FRS subgroups, ns (not

---

significant), \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ). **(B)** The scatter plot shows the sensitivity of the samples to five compounds as a function of FRS. **(C)** The network heatmap reflects the relationship between drug sensitivity and gene expression. (Abs[52] represents the absolute value of the correlation coefficient, and Drugs cor represents the correlation coefficient between different drugs.).

#### **SUPPLEMENTAL DATA**

Supplemental data are available at the following link:

<https://www.bjbms.org/ojs/index.php/bjbms/article/view/12038/3846>

EARLY ACCESS