



## OPEN ACCESS

## EDITED BY

Zhouxiao Li,  
Ludwig Maximilian University of Munich,  
Germany

## REVIEWED BY

Changgang Sun,  
Affiliated Hospital of Weifang Medical  
University, China  
Hui Yee Kwan,  
Hong Kong Baptist University, Hong  
Kong SAR, China

## \*CORRESPONDENCE

Qing-Hua Yao,  
yaoqh@zjcc.org.cn

<sup>†</sup>These authors have contributed equally  
to this work

## SPECIALTY SECTION

This article was submitted to Cancer  
Genetics and Oncogenomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 26 September 2022

ACCEPTED 14 November 2022

PUBLISHED 25 November 2022

## CITATION

Zhang L, Xu C, Wang S-H, Ge Q-W,  
Wang X-W, Xiao P and Yao Q-H (2022),  
Cancer-associated fibroblast-related  
gene signatures predict survival and  
drug response in patients with  
colorectal cancer.  
*Front. Genet.* 13:1054152.  
doi: 10.3389/fgene.2022.1054152

## COPYRIGHT

© 2022 Zhang, Xu, Wang, Ge, Wang,  
Xiao and Yao. 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.

# Cancer-associated fibroblast-related gene signatures predict survival and drug response in patients with colorectal cancer

Lei Zhang<sup>1†</sup>, Chao Xu<sup>1†</sup>, Si-Han Wang<sup>2</sup>, Qin-Wen Ge<sup>3</sup>,  
Xiao-Wei Wang<sup>4</sup>, Pan Xiao<sup>1</sup> and Qing-Hua Yao<sup>1,5,6\*</sup>

<sup>1</sup>Department of Integrated Chinese and Western Medicine, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China, <sup>2</sup>The Second College of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China, <sup>3</sup>The First College of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China, <sup>4</sup>Department of Plastic Surgery, Zhejiang Hospital, Hangzhou, China, <sup>5</sup>Integrated Traditional Chinese and Western Medicine Oncology Laboratory, Key Laboratory of Traditional Chinese Medicine of Zhejiang Province, Hangzhou, China, <sup>6</sup>Key Laboratory of Head and Neck Cancer Translational Research of Zhejiang Province, Hangzhou, China

**Background:** Cancer-associated fibroblasts (CAFs) play an important role in the tumorigenesis, immunosuppression and metastasis of colorectal cancer (CRC), and can predict poor prognosis in patients with CRC. The present study aimed to construct a CAFs-related prognostic signature for CRC.

**Methods:** The clinical information and corresponding RNA data of CRC patients were downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. The Estimation of STromal and Immune cells in Malignant Tumor tissues (ESTIMATES) and xCell methods were applied to evaluate the tumor microenvironment infiltration from bulk gene expression data. Weighted gene co-expression network analysis (WGCNA) was used to construct co-expression modules. The key module was identified by calculating the module-trait correlations. The univariate Cox regression and least absolute shrinkage operator (LASSO) analyses were combined to develop a CAFs-related signature for the prognostic model. Moreover, pRRophetic and Tumor Immune Dysfunction and Exclusion (TIDE) algorithms were utilized to predict chemosensitivity and immunotherapy response. Human Protein Atlas (HPA) databases were employed to evaluate the protein expressions.

**Results:** ESTIMATES and xCell analysis showed that high CAFs infiltration was associated with adverse prognoses. A twenty-gene CAFs-related prognostic signature (CAFPS) was established in the training cohort. Kaplan-Meier survival analyses revealed that CRC patients with higher CAFs risk scores were associated with poor prognosis in each cohort. Univariate and multivariate Cox regression analyses verified that CAFPS was as an independent prognostic factor in predicting overall survival, and a nomogram was built for clinical utility in predicting CRC prognosis. Patients with higher CAFs risk scores tended to

not respond to immunotherapy, but were more sensitive to five conventional chemotherapeutic drugs.

**Conclusion:** In summary, the CAFPS could serve as a robust prognostic indicator in CRC patients, which might help to optimize risk stratification and provide a new insight into individual treatments for CRC.

#### KEYWORDS

colorectal cancer, cancer-associated fibroblasts, tumor microenvironment, prognosis, WGCNA

## Introduction

Colorectal cancer (CRC) is globally the third most commonly diagnosed cancer and the second leading cause of cancer-related deaths (Bray et al., 2018). Although 39% of patients diagnosed with localized CRC present 90% 5-year survival, this decreases to 71% for patients with tumors that have spread regionally and is less than 14% in those with advanced distant metastases (Thompson et al., 2022). According to the Global Cancer Observatory (<https://gco.iarc.fr/today>), there were 555,628 new cases and 283,751 deaths from CRC in China in 2020. The occurrence and development of CRC is a multi-step and complex process with multiple genes involved. Colorectal cancer cells have an extraordinary biological ability to adapt themselves to adverse environments, leading to their strong invasive and metastatic characteristics (Kleppe et al., 2018). Conventional assessment, including methods based on tumor-node-metastasis (TNM) staging and pathology, is intrinsically subjective and not sufficient to predict treatment response and prognosis. The development of a novel prognostic model is therefore imperative for CRC. Prognostic prediction models are widely utilized both in the clinic and research to predict the probability or the risk of a specific events or future outcomes (Toll et al., 2008).

Cancer arises from the accumulation of gene mutations within cancer cells, while both tumorigenesis and patients' response to therapies are strongly regulated by non-mutant cells and the extracellular matrix (ECM) within the tumor microenvironment (TME). Cancer-associated fibroblasts (CAFs) are a special type of fibroblasts that surround tumors and form a key part of the TME. In recent years, CAFs have received increasing attention due to their crucial roles in tumor invasion, angiogenesis, and ECM remodeling by promoting cell-cell interaction and the secretion of pro-invasive factors (Villaronga et al., 2018; Bertero et al., 2019). Targeting CAFs by altering their numbers, subtype or biological functionality is emerging as an attractive avenue to improve therapeutic strategies for cancer.

In this study, we identified the infiltration score of CAFs in CRC as a risk factor. The bulk transcriptome RNA-seq and relevant clinical data of CRC patients were obtained from The

Cancer Genome Atlas (TCGA) datasets through the UCSC Xena browser (<https://xenabrowser.net/datapages/>). (Goldman et al., 2020) In addition, through a variety of bioinformatics methods, we aimed to discover promising CAFs-targeting therapeutic hallmarks and constructed a robust CAFs-related gene signature to predict the prognosis and drug response of CRC patients. Figure 1 illustrates the workflow of the study.

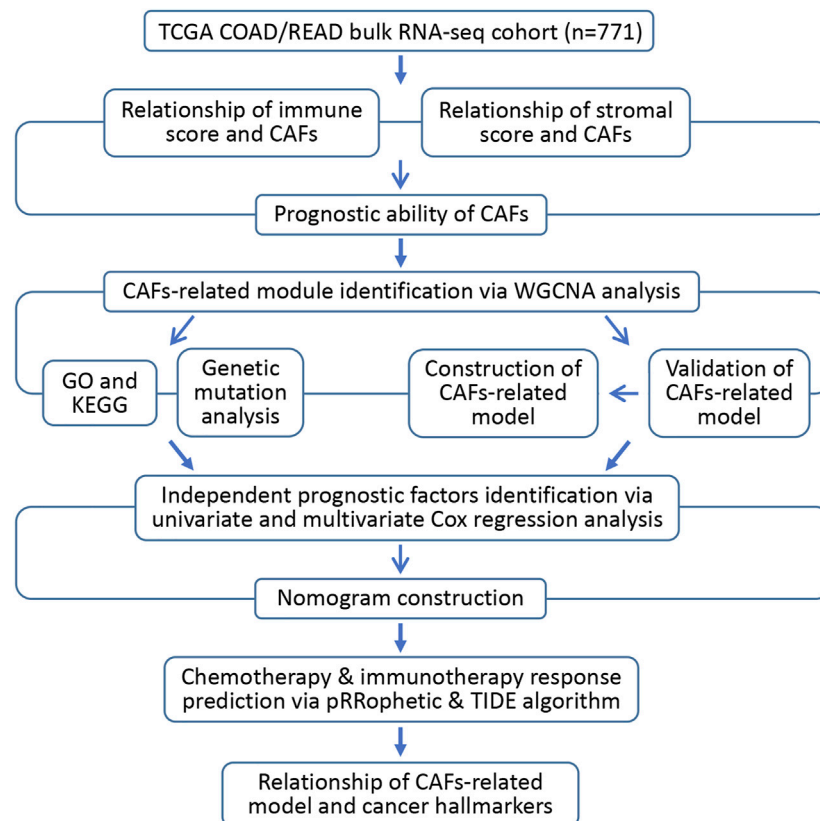
## Materials and methods

### Data source and preprocessing

Data that containing the RNA expression profiles and relevant clinical information of colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) patients were accessed through TCGA datasets. Following the removal of the batch effects, the two parts of data were merged using the "limma" R package (version 3.52.2). The GSE159216 and GSE72968 datasets were obtained from the Gene Expression Omnibus (GEO) data portal (<https://www.ncbi.nlm.nih.gov/geo/>) for further validation. In total, the data of 771 patients from the TCGA cohort, 283 patients from the GSE159216 cohort, and 585 patients from the GSE72968 cohort were recorded for utilization in the current study.

### Estimation of tumor microenvironment infiltration

The R package "xCell" (version 1.1.0) estimates the comprehensive levels of 64 immune and stromal cell types (Aran et al., 2017). The estimation of STromal and Immune cells in MAlignant Tumor tissues using the Expression data (ESTIMATES) algorithm (version 1.34.0) can accurately quantify the immune score and stromal score to identify the infiltration degree of immune cells and predict the immune status (Yoshihara et al., 2013). We applied xCell and ESTIMATES to separately calculate the abundance scores for stromal cells and immune cells for patient samples from different CRC stages.



**FIGURE 1**  
Flow diagram of the current investigation.

## Construction of prognostic tumor immune cells

Spearman's correlation analyses were conducted to determine the correlation between TME infiltration (immune score and stromal score) and CAFs levels in CRC samples. The “ggpubr” R package (version 0.4.0) was used to produce correlation plots. After selecting the *cut-off* values with the best sensitivity and specificity, CRC samples in the TCGA cohort were divided into two groups according to CAFs infiltration score using the “survival” R packages (version 3.6.1). The Kaplan-Meier survival curve was plotted using “survminer” R packages (version 0.4.6) to compare the survival rate.

## Identification of the hub CAFs-related module by weighted gene co-expression network analysis

In order to find genes that closely related to CAFs, we performed WGCNA by utilizing the “WGCNA” R package

(version 1.69) to find modules highly correlated with CAFs levels and stromal score (Langfelder and Horvath, 2008). After calculating the Pearson correlation between each gene pair, the weighted adjacency matrix was constructed using the WGCNA function adjacency function. Then, we used topological overlap matrix analysis to cluster the adjacency matrix of CRC patients' gene expression data. Next, the dynamic tree cut algorithm was applied to identify modules on the dendrogram. Finally, we calculated the correlation between the identified gene modules identified and CAFs levels to mine the hub module for subsequent analysis.

## Function and pathway enrichment analysis of genes in the hub module

In order to explore the biological function and pathway of genes in the hub module, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were analyzed and visualized through the “clusterProfiler” (version 3.18.0) and “org.Hs.eg.db” (version 3.1.0) R packages.

## CAFs-related genetic mutation analysis

The genetic landscape of CAF-related genes with copy number variations (CNV) and single nucleotide variations (SNV) from the TCGA datasets was generated with the “maftools” R package (version 2.6.05). Then, CNV and SNV correlation heatmaps were drawn using the “Complexheatmap” R package (version 2.6.2).

## Construction and validation of a CAFs-related prognostic model

We designed a prognostic signature for CRC patients by focusing on CAFs marker genes, which were identified from the hub module. Univariate Cox analysis of the overall survival (OS) was applied to screen the prognostic values of CAFs-related genes. Genes with  $p < 0.05$  in the univariate Cox analysis were regarded as candidate prognostic genes. Next, we displayed the prognostic genes in a forest plot using the “forestplot” R package (version 1.9). To minimize the risk of overfitting, we used LASSO-penalized Cox regression analysis to eliminate genes with an overfitting tendency and built a prognostic signature using the “glmnet” R package (version 2.2.1) (Simon et al., 2011). The signature of CAFs was established as follows: CAFs risk score =  $\sum(\beta_i \cdot \text{Exp}_i)$ , where  $\beta_i$  represented the corresponding regression coefficients of each candidate prognostic gene, and  $\text{Exp}_i$  was the candidate gene’s expression value. According to the median value of CAFs risk scores, we divided CRC patients into high-risk and low-risk groups. The OS curve was plotted *via* Kaplan-Meier analysis. Meanwhile, time-dependent receiver-operating characteristic (ROC) analysis was carried out by the “survivalROC” R package (version 1.34.0). Finally, heatmaps were generated to visualize the association between the risk scores of CAFs and candidate genes. Similarly, we validated our CAF prognostic model on the GSE159216 and GSE72968 external validation cohorts.

## Construction of predictive nomogram

Univariate and multivariate Cox regression analyses were performed to identify the independent prognostic factors. A nomogram was then constructed based on CAFs signature, clinical stage, TNM stage and lymphatic invasion using the “rms” R package (version 6.0.1). Afterwards, the ROC curve and calibration curve were employed to evaluate the nomogram’s predictive performance and accuracy.

## Prediction of patients’ drug response based on CAFs signature

We predicted the chemosensitivity/resistance for the high and low risk groups *via* the “pRRophetic” R package (version

0.5). According to ridge regression, the half-maximal inhibitory concentrations (IC50) were estimated for TCGA samples (Geeleher et al., 2014a; Geeleher et al., 2014b). Furthermore, the Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu/>) algorithm was employed to predict the potential response to immune checkpoint blockade (ICB) therapy between the two groups (Jiang et al., 2018).

## Association of prognostic CAFs signature with cancer hallmarks

A total of 50 hallmark gene sets were downloaded from the molecular signature database (MSigDB, <http://software.broadinstitute.org/gsea/msigdb>). These 50 hallmark gene sets were subjected to the gene set variation analysis (GSVA) R package (version 1.32.0) to further obtain the GSVA scores of each gene set for each sample from the TCGA cohort (Hänzelmann et al., 2013). The Pearson correlation between CAFs signature and 50 hallmark gene sets were calculated by the “Hmisc” R package (version 4.4.1).

## Human protein atlas database and immunohistochemistry verification

In order to further validate the protein expressions of CAFs signature genes, the immunohistochemistry staining images of prognosis-related genes in CRC tissues were retrieved from the HPA online database (<http://www.proteinatlas.org/>) (Uhlén et al., 2015).

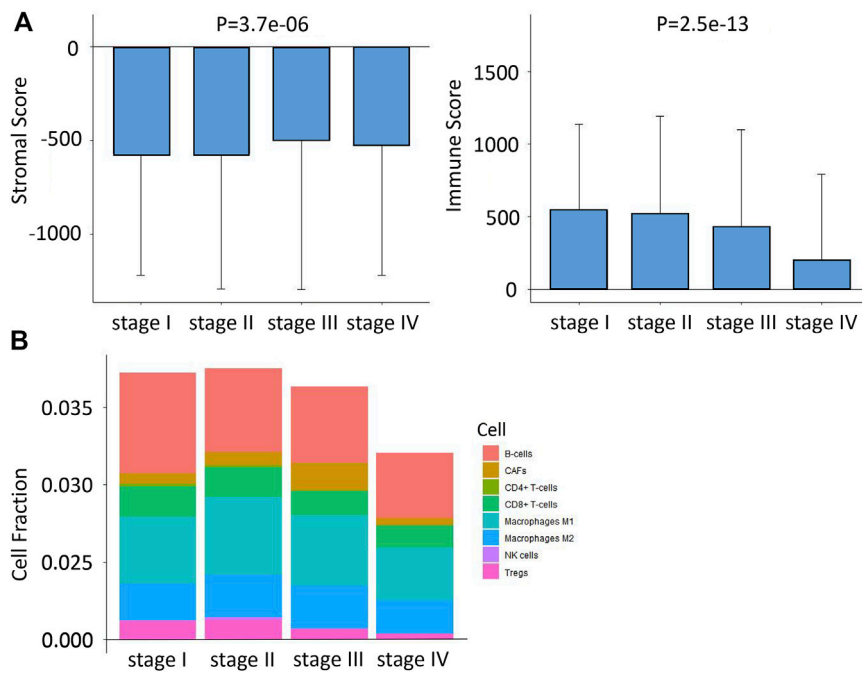
## Statistical analysis

All statistical analyses were carried out using R software (version 4.0.3). The Wilcoxon signed rank test was applied for comparisons between two groups, and the Kruskal-Wallis test for comparisons between three or more groups. Statistical significance was determined as two-sided with  $p < 0.05$ .

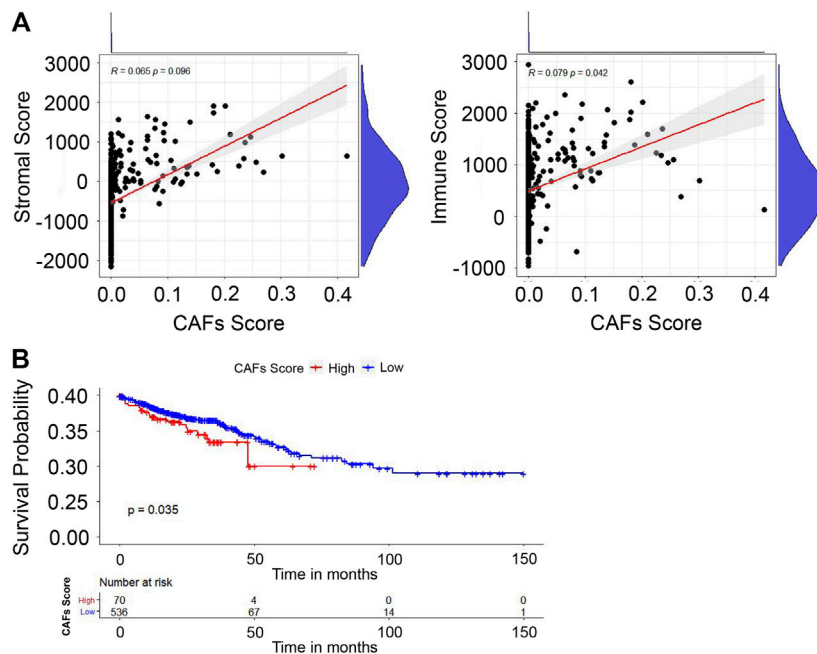
## Results

### TME infiltration patterns with different CRC stages

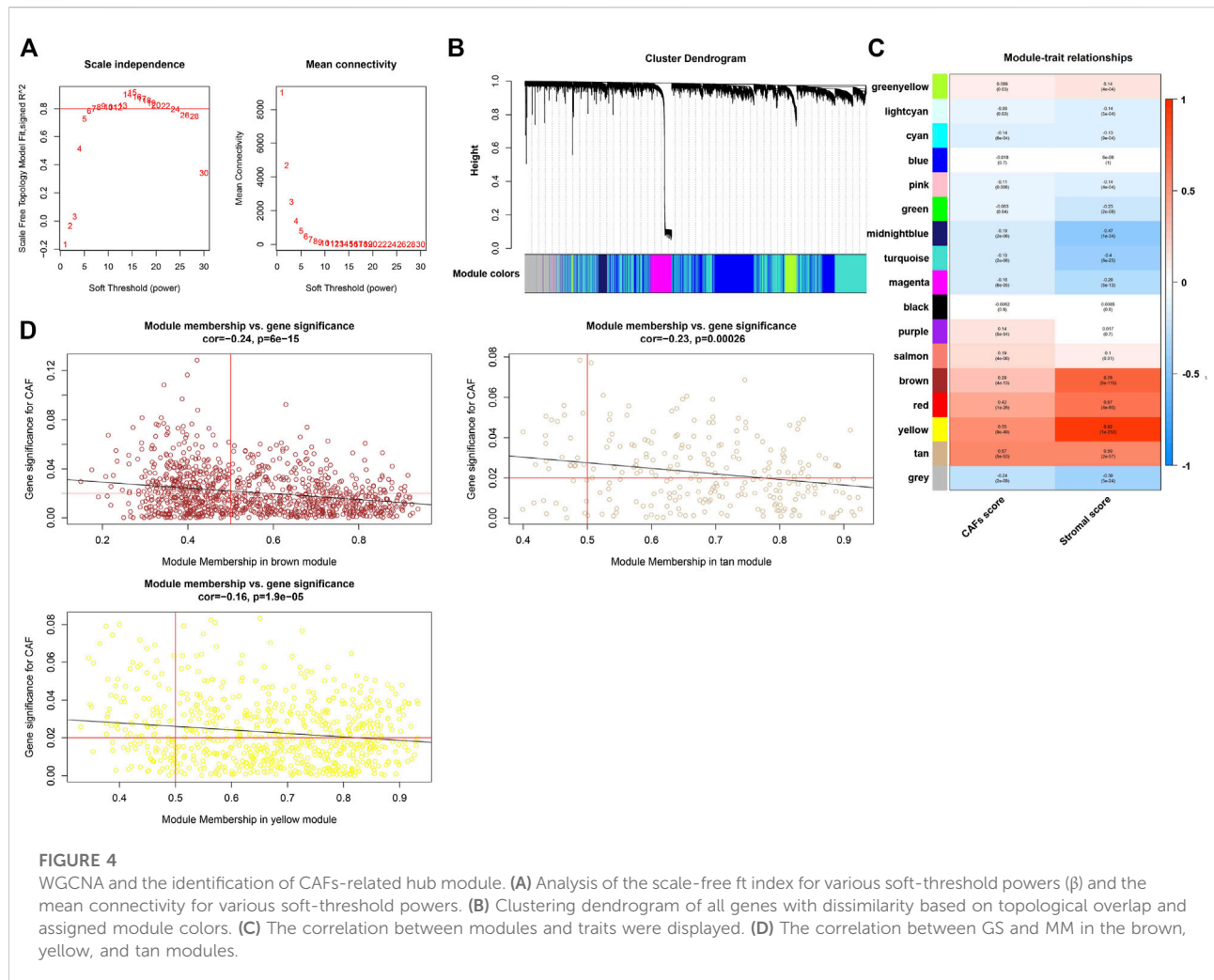
By running xCell and ESTIMATES algorithms, we measured the TME constituents in patients with different stages of CRC from the TCGA cohort. As shown in Figure 2, the pooled results of the stacked bar graph and Wilcoxon analyses on TCGA COAD/READ datasets revealed the stromal and immune scores; the infiltrations of several TME contents such as



**FIGURE 2** Differences in immune infiltration characteristics among the four stages. **(A)** Differences in stromal score and immune score among the four stages based on xCell algorithms. **(B)** Differences in eight immune cell scores among the four stages based on ESTIMATE algorithms.



**FIGURE 3** Evaluation of stromal/immune scores in CRC tissues. **(A)** The correlation between the stromal/immune scores and CAFs infiltration score. **(B)** The correlation between CAFs infiltration score and OS.



B cells, CD8<sup>+</sup> T cells, M1 Macrophages, M2 Macrophages, NK cells and Tregs were lower at later stages of CRC. However, the infiltration of CAFs increased first and then decreased in stage IV.

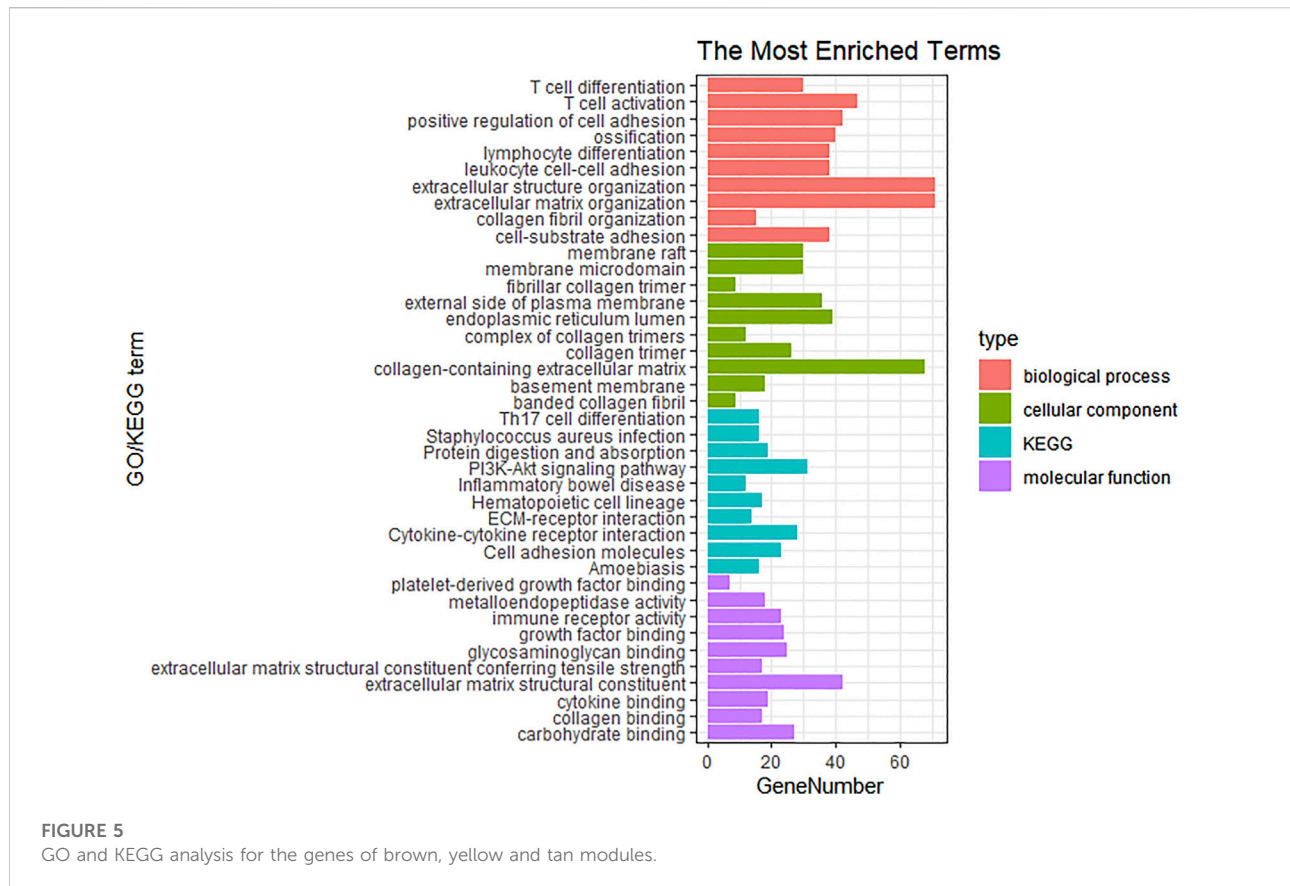
### Clinical correlation of tumor-infiltrating CAFs

We first investigated the relationship between the stromal/immune scores and CAFs infiltration score. The results in Figure 3A reveal that stromal score was not correlated with CAFs infiltration score; however, the immune score showed a positive correlation with the CAFs infiltration score ( $p = 0.042$ ). To investigate the potential relationship between OS and CAFs infiltration score, we further divided CRC patients into high- and low-score groups based on infiltration scores and the constructed Kaplan-Meier survival curves. We found that

CAF's infiltration score was significantly negatively correlated with OS ( $p = 0.035$ ) (Figure 3B). Overall, all these results suggested that CAFs infiltration score is associated with CRC patients' prognoses.

### WGCNA for key gene module associated with CAFs

We constructed the WGCNA analysis for all genes. With the power value selected as 7, the scale independence approached 0.8 (Figure 4A), suggesting a gene coexpression network with an inherent modularity and a scale-free topology. A total of 17 modules were identified through hierarchical clustering (Figure 4B). Next, we examined the correlation between the 17 modules and CAFs/stromal scores (Figure 4C). All of the brown, yellow, and tan modules had higher correlation with the CAFs/stromal scores. Thus, they



were considered as key modules because of the high correlation with traits. Under the condition of module membership (MM) > 0.5 and gene significance (GS) > 0.03, 559 genes in the brown, yellow and tan modules were taken out (Figure 4D).

## GO and KEGG functional downstream analyses of CAFs-related genes

In order to investigate the biological functions and pathways of the above 559 genes in key modules, GO and KEGG pathway enrichment analyses were carried out. As shown in Figure 5, extracellular matrix organization, collagen fibril organization, collagen-containing extracellular matrix, and extracellular matrix structural constituent were the main significantly enriched GO terms. Moreover, the top 10 enriched KEGG pathways were also exhibited, which included Th17 cell differentiation, *Staphylococcus aureus* infection, protein digestion and absorption, PI3K-Akt signaling pathway, inflammatory bowel disease, hematopoietic cell lineage, ECM-receptor interaction, cytokine-cytokine receptor interaction, cell adhesion molecules, and amoebiasis.

## Alterations of CAFs-related genes in CRC samples

In order to investigate alterations in CAFs-related genes in the CRC samples, especially SNV and CNV, we analyzed CRC patients with SNV and CNV data after extraction from the TCGA. The top 30 genes with the highest mutation counts were shown in Supplementary Figure S1.

## Construction and verification of twenty-gene prognostic CAFs signature

In the TCGA cohort, by performing univariate Cox regression analysis on the 559 CAFs marker genes identified above, a total of 51 genes were found with  $p < 0.05$ . LASSO Cox regression algorithm was then performed to select hub genes. The minimal log(lambda) was determined as the optimal value by tenfold cross-validations (Figures 6A,B). Finally, a twenty-gene CAFs prognostic signature was constructed based on the expression levels of each gene and the coefficient, with the following formula: risk score = (0.0179849395327801 \* expression of CYTH3) + (0.133678301953266 \* expression of NAV3) + (0.97605105976877 \* expression of EPHA6) -

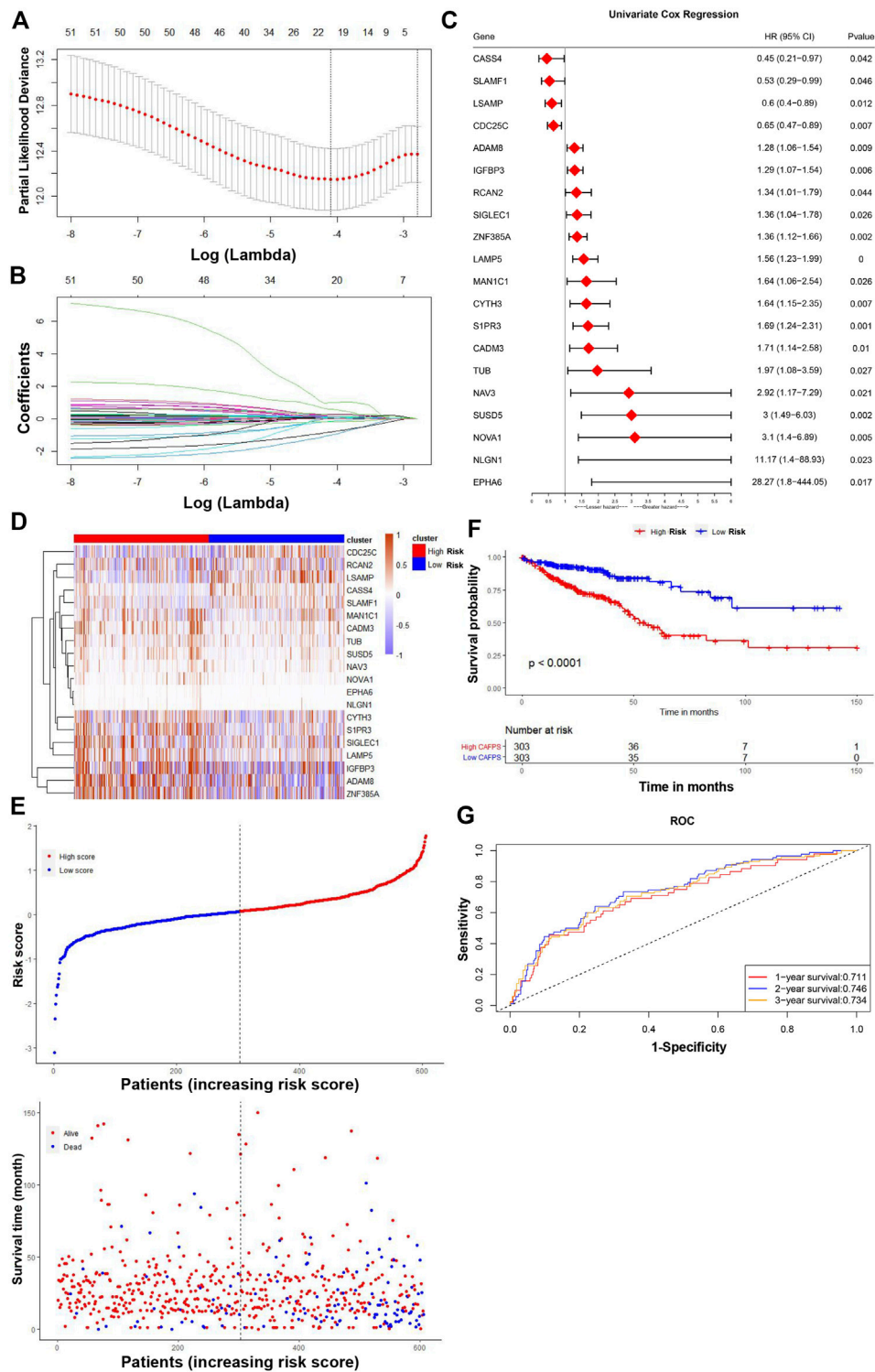


FIGURE 6

Screening of signature genes and the prognostic value of the CAFs-related signature in TCGA cohorts. (A) Ten-fold cross-validations for the screening of optimal parameter (lambda). (B) LASSO coefficient profiles determined by the optimal lambda. (C) The forest plot presented the HRs and p-values from the univariate Cox regression as well as the LASSO coefficient of the twenty prognostic signature genes. (D) Heatmap visualizing the expression levels of twenty prognostic CAFs genes with the CAFs risk scores in the TCGA cohort. (E) The distribution of patient survival status ranked by corresponding risk scores. (F) The Kaplan-Meier survival curves of OS between high and low risk score groups. (G) The time-dependent ROC curves of the prognostic signature for 1-, 2-, and 3-year overall survival.



(1.01688005937021 \* expression of CASS4) + (0.15388612377183 \* expression of SIGLEC1)—(0.162312764949028 \* expression of SLAMF1) + (0.0151723782273217 \* expression of MAN1C1) + (0.107689308053911 \* expression of LAMP5) + (0.0193478751759346 \* expression of NOVA1) + (0.127100030231241 \* expression of IGFBP3) + (0.0420090450866688 \* expression of ADAM8)—(0.194826817771919 \* expression of CDC25C) + (0.0834898746044337 \* expression of ZNF385A) + (0.0492899593436562 \* expression of CADM3) + (0.131501239268304 \* expression of TUB) + (0.584400505367543 \* expression of NLGN1) + (0.05334380595326 \* expression of RCAN2) + (0.231247598432235 \* expression of SUSD5) - (0.82291587246458 \* expression of LSAMP) + (0.286394869616561 \* expression of S1PR3). Among the 20 prognostic genes, sixteen (CYTH3, NAV3, EPHA6, SIGLEC1, MAN1C1, LAMP5, NOVA1, IGFBP3, ADAM8, ZNF385A, CADM3, TUB, NLGN1, RCAN2, SUSD5 and S1PR3) were regarded as risk-related genes, while CASS4, SLAMF1, CDC25C and LSAMP were considered as protective genes (Figure 6C). Kaplan-Meier survival curves revealed the relationship between prognosis and the expression levels of 20 genes (Supplementary Figure S2). Based on this risk formula, we calculated the CAFs risk score for each patient. The heatmap exhibited the risk scores and expression differences between the 20 genes in the TCGA cohort (Figure 6D).

The patients in the TCGA cohort were divided into high- and low-CAF risk groups according to the median risk scores. The distribution of the risk score and patients' survival status were ranked by the risk score value (Figure 6E). According to the Kaplan-Meier survival curves, patients in the high CAFs risk group had significantly unfavorable OS compared with the low CAFs risk group (Figure 6F). A ROC curve was constructed in Figure 6G showing the prognostic accuracy of the signature, and the AUCs for 1-, 2-, and 3-year overall survival were 0.711, 0.746, and 0.734, respectively. Moreover, we also verified the predictive ability of the signature in another two independent cohorts, GSE159216 and GSE72968. Patients in the high CAFs risk group had a worse prognosis than those in the low CAFs risk group (Figures 7A–D).

## Independent prognostic value of CAFs risk score

There were no differences in CAFs risk score between COAD and READ patients from the TCGA cohort (Supplementary Figure S3A). However, the CAFs risk score was significantly related to clinical stage, TNM stage, and lymphatic invasion (Supplementary Figures S3B–F). We next performed univariate and multivariate Cox regression

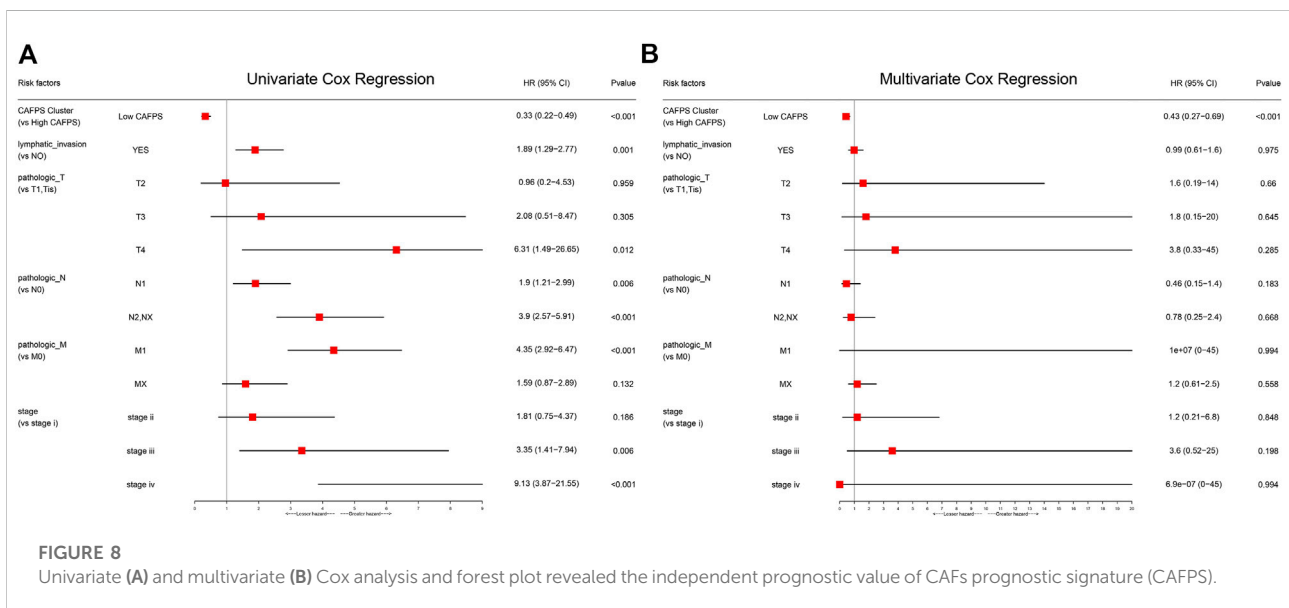
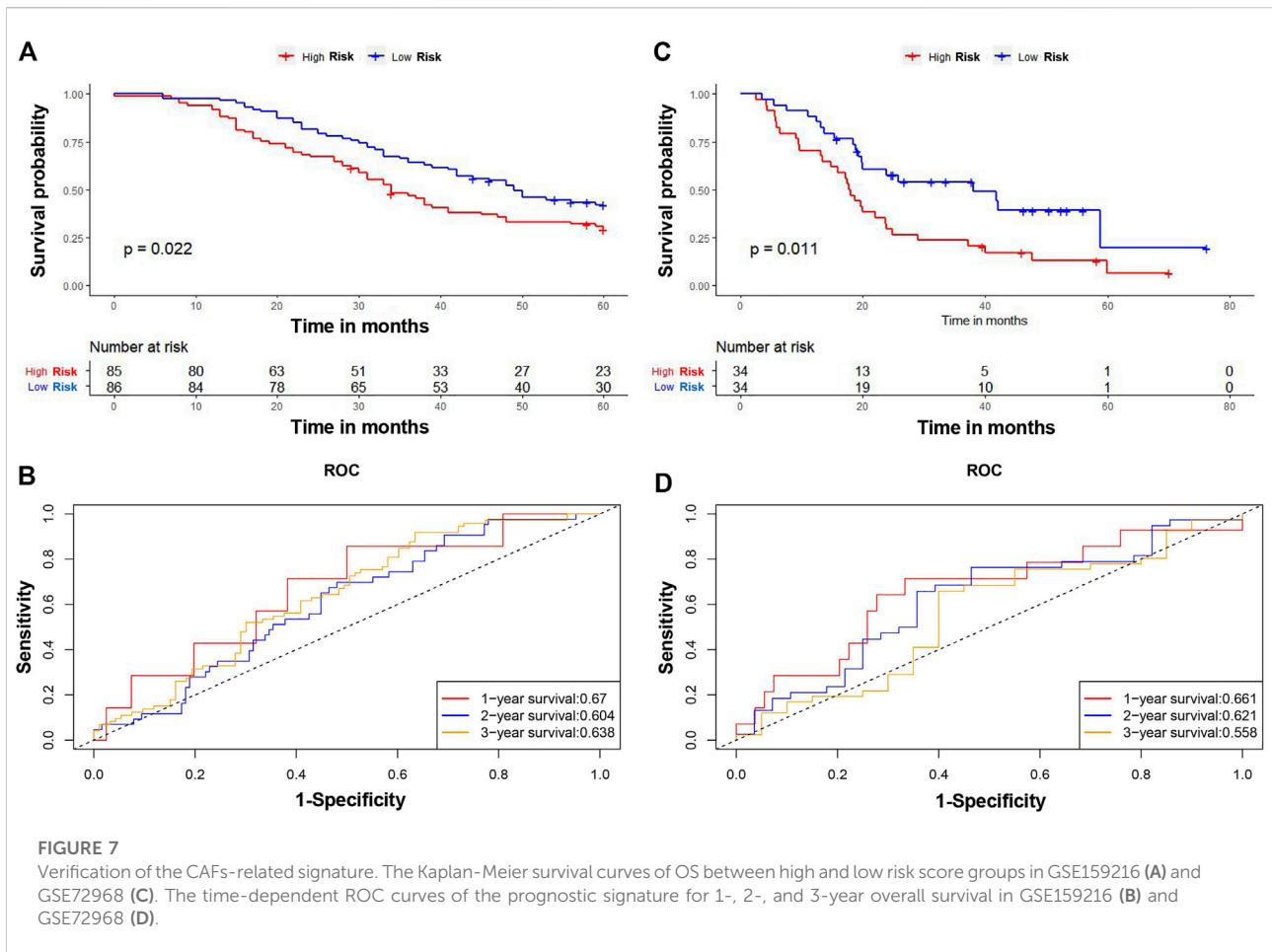
analyses on the clinical variables to identify whether CAFs prognostic signature (CAFPS) was an independent prognostic predictor of OS. We found that CAFPS was significantly associated with OS in the univariate Cox regression analysis (HR = 0.33; 95% CI = 0.22–0.49;  $p < 0.001$ ; Figure 8A). Furthermore, multivariate Cox regression analysis was carried out to correct the confounding factors. The CAFPS was nevertheless proved to be an independent predictor for OS (HR = 0.43; 95% CI = 0.27–0.69;  $p < 0.001$ ; Figure 8B).

## Establishment of predictive nomogram for CRC patients

According to the regression analysis results, we developed a nomogram including our CAFPS and multiple clinical factors. In the TCGA cohort, clinical stage, TNM stage, lymphatic invasion, and CAFPS were eventually selected to establish an accurate predictive nomogram (Figure 9A). Next, we evaluated the discriminative ability of the nomogram using the ROC-related AUC. The AUC of CAFPS was 0.711, and the calibration plots of 1-, 2-, and 3-year OS showed no deviations from the Platt calibration curves, indicating the high predictive accuracy of the nomogram (Figures 9B,C).

## Prognostic value of CAFPS for drug response

We next examined the correlation between CAFPS and cancer hallmark-related pathways. As shown in Supplementary Figure S4, CAFPS was significantly associated with 32 cancer relevant pathways among the total of 50 pathways. To explore the difference between low-risk and high-risk groups regarding drug resistance potential, we estimated the IC50 levels of 138 chemotherapy drugs or inhibitors in the two groups. We found that AZD.0530, JNK.9L, PD.0332991, shikonin, and Z.LLNle.CHO could be candidate drugs for treating patients in the high-risk group (Figures 10A–E). The bubble chart shows the top 30 most relevant drugs for 20 prognostic genes (Supplementary Figure S5). We also predicted the response of CRC patients in the TCGA cohort to immunotherapy by the TIDE algorithm. The CAFPSs were significantly different between the non-responder group and the responder group (Figure 10F;  $p < 0.01$ ). The proportion of responders in the low CAFPS group was significantly higher than that in the high CAFPS group (Figure 10G). The AUC of CAFPS for 1-year overall survival was 0.739 (95% CI = 0.689–0.79; Figure 10H). These evidences indicated that the CAFPS based on the signatures of 20 genes was helpful to assess patients' response to chemotherapy and immunotherapy.



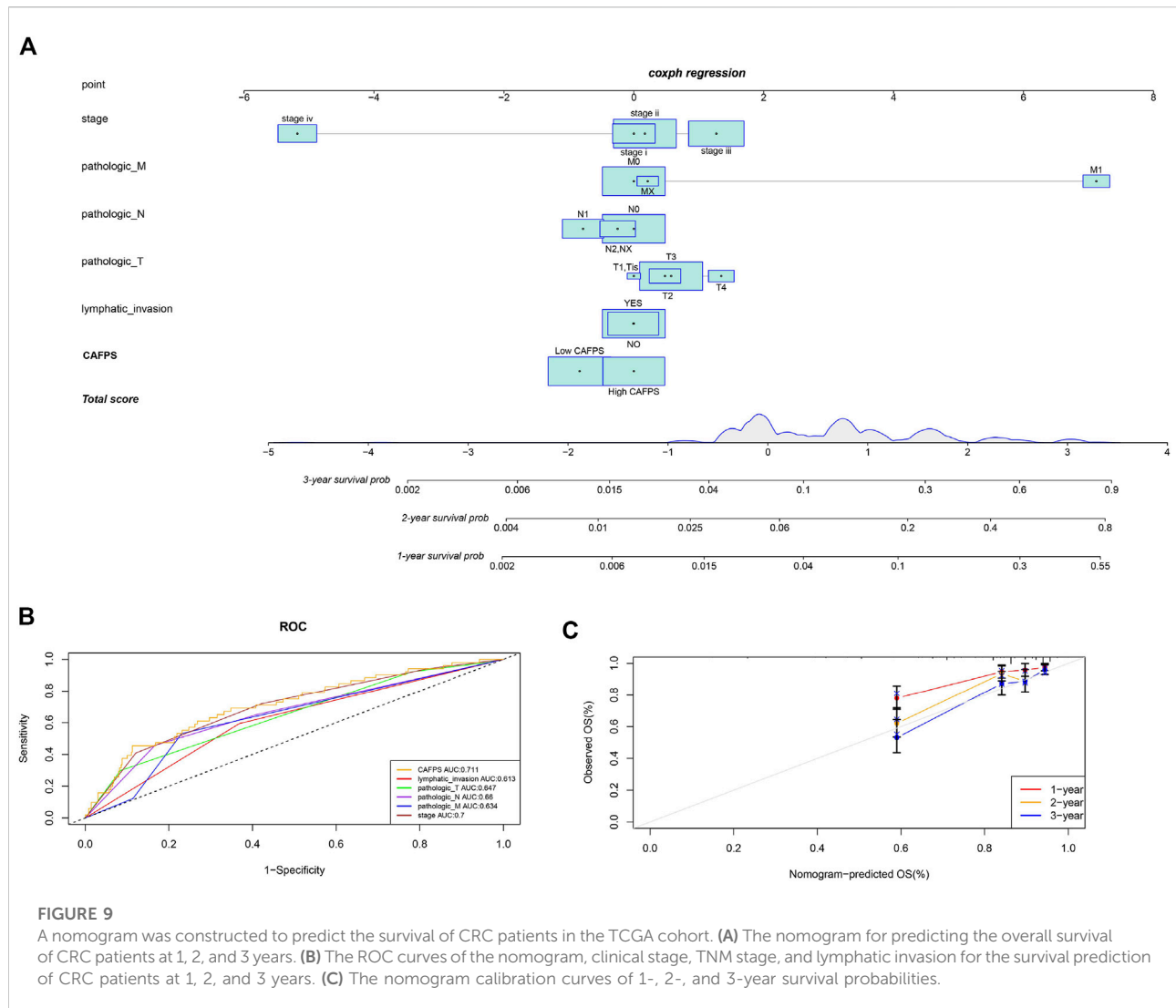


FIGURE 9

A nomogram was constructed to predict the survival of CRC patients in the TCGA cohort. (A) The nomogram for predicting the overall survival of CRC patients at 1, 2, and 3 years. (B) The ROC curves of the nomogram, clinical stage, TNM stage, and lymphatic invasion for the survival prediction of CRC patients at 1, 2, and 3 years. (C) The nomogram calibration curves of 1-, 2-, and 3-year survival probabilities.

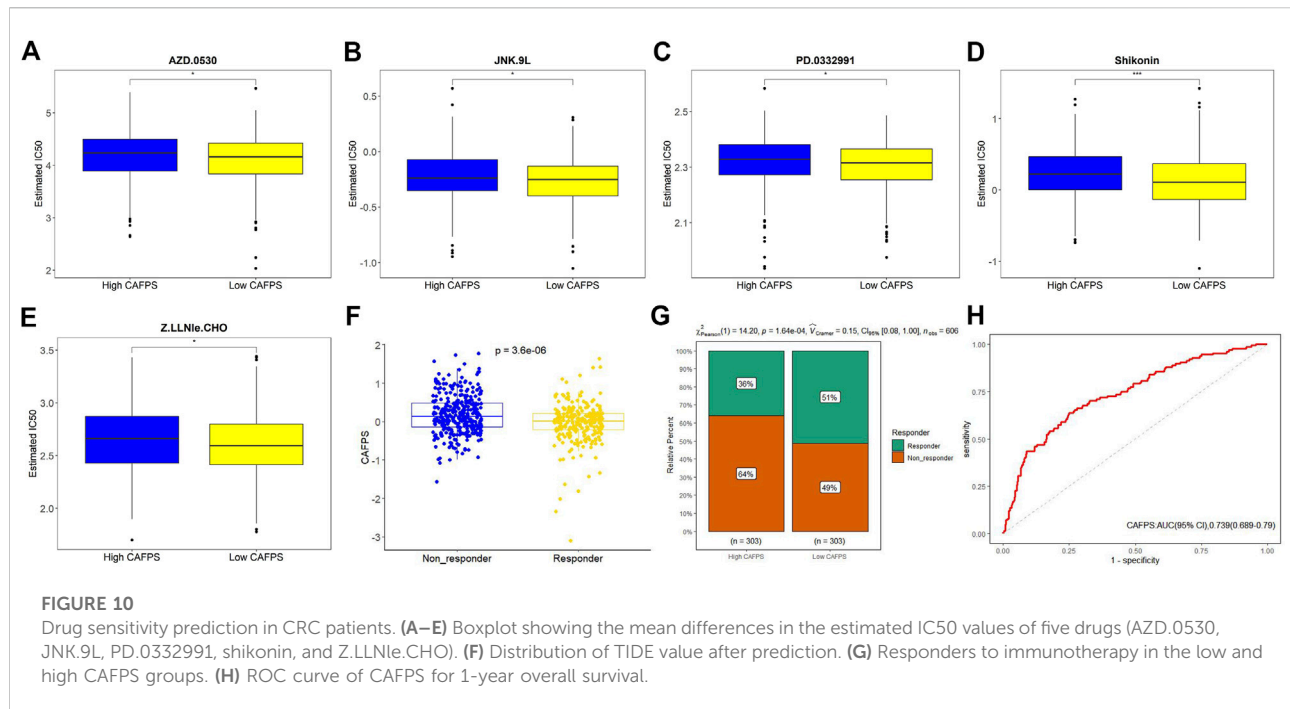
## Evaluation of the expression patterns of CAFs-related signature genes at the protein levels *via* HPA database

Finally, we used the IHC data from the HPA database to validate our previous findings to evaluate the expression of risk model genes associated with CAFs in tumor and normal tissues. Since IGFBP3, LSAMP, S1PR3 and ZNF385A have not been included in the HPA database, we provided the IHC results for ADAM8, CADM3, CASS4, CDC25C, CYTH3, EPHA6, LAMP5, MAN1C1, NAV3, NLGN1, NOVA1, RCAN2, SIGLEC1, SLAMF1, SUSD5, and TUB. The results showed that the protein expression levels confirmed the majority of our previous findings at the mRNA levels. Moreover, the IHC results from HPA database indicated that the protein expressions of ADAM8, CYTH3, and TUB were higher in CRC stroma (Figures 11A,E,P), while those of CADM3, EPHA6, MAN1C1, NAV3, NLGN1, NOVA1, and RCAN2 were higher in CRC interstitial

areas (Figures 11B,F,H-L). No expressions of LAMP5, SLAMF1, and CDC25C were observed either in stroma or interstitial areas (Figures 11G,N,D). Moreover, CASS4, SIGLEC1, and SUSD5 were weakly expressed, like CASS4 and SUSD5 in interstitial areas (Figures 11C,O) and SIGLEC1 in stroma (Figure 11M).

## Discussion

CAFs, as one of the most abundant cell type in the TME, facilitate the development, propagation and invasiveness of tumors (Chan et al., 2017). They have been previously reported to support the TME, which can lead to poor prognosis and drug resistance (Chen et al., 2014; Shiga et al., 2015). On the one hand, CAFs can produce proteases to remodel the tumor extracellular matrix (ECM) and increase the stiffness of tumor tissue, creating a pathway for tumor cells to invade more easily (Kechagia et al., 2019). Increased tumor

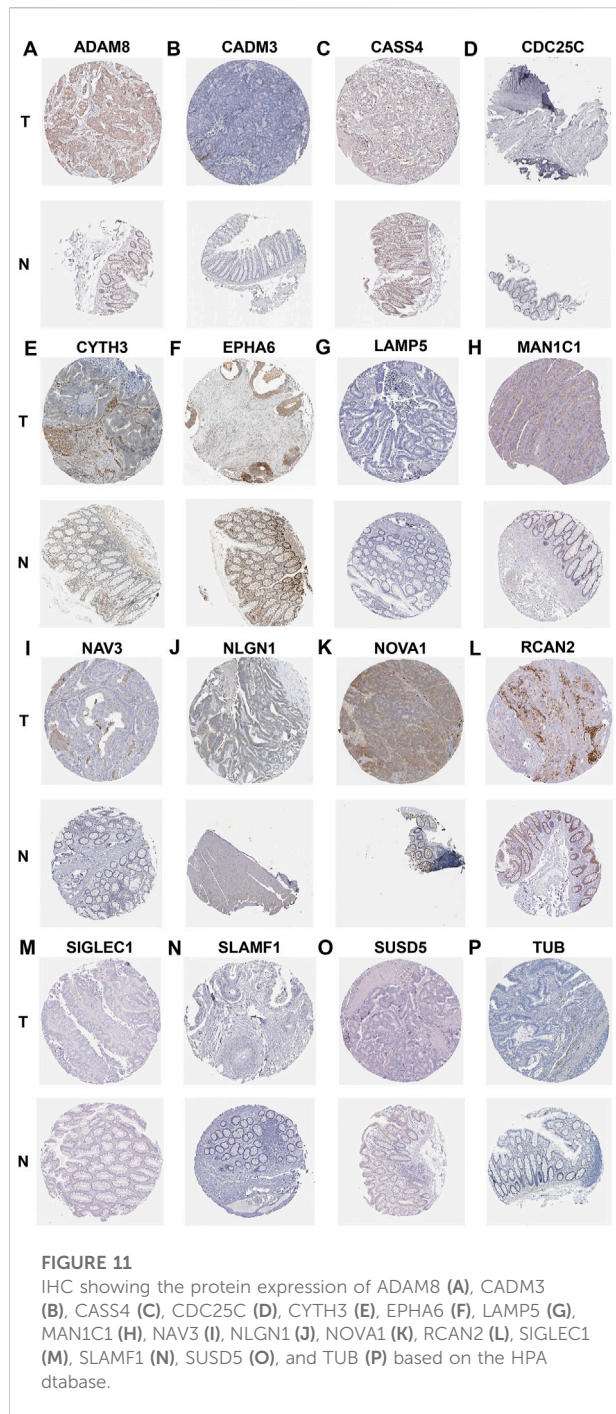


tissue stiffness in ECM also makes blood vessels collapse and lead to hypoxia, which promotes the survival and proliferation phenotype of tumor cells and reduces drug releases (Zhang et al., 2021). On the other hand, the roles of CAFs are not limited to ECM regulation but also include communicating with other cells to establish an immunosuppressive TME, enabling tumor cells to evade antitumor immunity (Barrett and Pure, 2020). CAFs are a rich source of secretomes and may influence the process of tumor-specific immune cell differentiation. Recently, several studies have indicated that CAFs skew infiltrating tumor-associated macrophage (TAM) populations towards the M2 immunosuppressive phenotype (Mao et al., 2021). However, emerging evidence suggests that WNT-2 secreted by CAFs may inhibit the differentiation and activation of dendritic cells, facilitating immune evasion by esophageal squamous cell carcinoma and colorectal cancer (Huang et al., 2022). Notably, the secretory function of CAFs could produce extensive crosstalk with surrounding cells in the TME, eventually leading to drug resistance (Monteran and Erez, 2019). In recent years, many studies have demonstrated the cancer-promoting effect of CAFs; therefore, the ability to target CAFs could be an attractive strategy for anti-cancer therapy.

In the present study, *via* the analysis of transcriptome data of TCGA cohorts, we estimated the proportion of CAFs in patients with CRC, and confirmed that it was closely associated with prognosis. Since CAFs exhibit a high degree of heterogeneity (Kalluri, 2016), integrating multiple biomarkers into an aggregated model would considerably improve their prognostic value. Thus, we established a CAFs-related gene signature including twenty genes (ADAM8, CADM3, CASS4, CDC25C,

CYTH3, EPHA6, IGFBP3, LAMP5, LSAMP, MAN1C1, NAV3, NLGN1, NOVA1, RCAN2, S1PR3, SIGLEC1, SLAMF1, SUSD5, TUB, and ZNF385A) through WGCNA, univariate, LASSO, and multivariate Cox regression analysis for predicting the prognosis and therapy response of CRC patients. In our study, the risk score derived from CAFs-related gene signature was abbreviated as CAFPS in our study. Moreover, CAFPS' predictive value has been validated in two additional independent cohorts, suggesting the reliability of the CAFPS-based model. We subsequently generated a nomogram based on clinical stage, TNM stage, lymphatic invasion, and CAFPS for clinical application.

Twenty CAFs-related genes were used to construct a new prognostic model through WGCNA. According to the risk value of each gene, ADAM8, IGFBP3, RCAN2, SIGLEC1, ZNF385A, LAMP5, MAN1C1, CYTH3, S1PR3, CADM3, TUB, NAV3, SUSD5, NOVA1, NLGN1, and EPHA6 were regarded as risk genes related to the poor prognosis of patients with CRC, whereas CASS4, SLAMF1, LSAMP, and CDC25C were associated with favorable prognosis. The biological functions of these genes involved in our signature have been elucidated more or less in previous studies. It was shown that the high levels of ZNF385A, LAMP, CADM3, NAV3, and NLGN1 indicate the poor prognosis of CRC patients (Martinez-Romero et al., 2018; Chang et al., 2021; Chen et al., 2021; Yu et al., 2021; Li et al., 2022), and these results were in accordance with our findings. Similarly, low expressions of CYTH3, NOVA1, and EPHA6 were highly correlated with longer OS in patients with other types of cancers (Zhang et al., 2014; Zhou et al., 2018; Xu et al., 2022). Furthermore, high expression of ADAM8 has been reported in



various tumor types and is related to invasiveness and poor prognosis (Conrad et al., 2019). ADAM8 has been found to cleave and remodel the ECM components of the tumor stroma (Zack et al., 2009; Schlomann et al., 2015), and thus could directly contribute to tumor invasiveness and metastasis. IGFBP3 and SUSD5 promote epithelial-mesenchymal transition (EMT) through the upregulation of a major cell surface receptor of

hyaluronic acid (CD44H) (Vicent et al., 2008; Du et al., 2022). SIGLEC1 is a sialic binding receptor mainly expressed by macrophages; the infiltration of SIGLEC1<sup>+</sup> macrophages in CRC was associated with tumor progression (Cassetta et al., 2019). In addition, the involvement of S1PR3 has been demonstrated in tumor growth. The S1P/S1PR3 axis is considered to promote tumor cell proliferation, migration and angiogenesis (Lee et al., 2017).

In the present study, RCAN2 was identified as a harmful predictor; however, our results are the contrary to the finding of Niitsu et al. (Niitsu et al., 2016) This may be related to the KRAS mutation in CRC leading to the decreased expression of RCAN2. Similar to our results, low expressions of CASS4 and LSAMP indicated poor prognosis in other types of cancers (Zhao et al., 2020; Gong et al., 2022). SLAMF1 and CDC25C were also identified as anti-tumor biomarkers in CRC (Qi et al., 2021; Song et al., 2022).

Some limitations of our results have to be recognized. First, although some genes were expressed at very low levels in CRC tissues, 20 candidate hub genes could not be filtered out due to the restrictions of the applied bioinformatics methods. Second, ours was a retrospective study for the establishment of gene signatures using public databases, thus multi-center and large-sample studies are needed to prospectively verify the prognostic and predictive efficacy of our CAFPS. Finally, the verification by detection at the protein level is insufficient; the molecular mechanisms of how the 20 candidate genes of this study influence the prognosis of CRC patients and responses to treatments need to be further explored through basic research.

## Conclusion

In summary, we used WGCNA analysis to create a gene co-expression network, and identified and validated a twenty-gene CAFs-related signature associated with CRC progression and prognosis. Based on this signature, the CAFPS could identify CRC patients who might not benefit from chemotherapy or immunotherapy.

## 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.

## Author contributions

LZ and CX put forward the concept of the study, designed the study, processed the data, prepared the manuscript, and contributed to the statistical analysis. S-HW and Q-WG contributed to the data

acquisition and performed the literature search for the manuscript. X-WW and PX contributed to the quality control of data. Q-HY put forward the concept of the study, contributed to the data analysis and interpretation, and reviewed the manuscript. All authors read and approved the final manuscript.

## Funding

This work was financially supported by the National Natural Science Foundation of China for Young Scholars (Grant No. 82104564), the Traditional Chinese Medicine Science and Technology Plan Projects of Zhejiang Province (Grant No. 2022ZQ019), and the Medicine and Health Science and Technology Plan Projects of Zhejiang Province (Grant No. 2020KY070).

## Acknowledgments

The authors thank the contributors of TCGA and GEO for sharing the bulk transcriptome RNA-seq and relevant clinical data.

## Conflict of interest

The 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.

## References

- Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1
- Barrett, R. L., and Puré, E. (2020). Cancer-associated fibroblasts and their influence on tumor immunity and immunotherapy. *Elife* 9, e57243. doi:10.7554/Elife.57243
- Bertero, T., Oldham, W. M., Grasset, E. M., Bourget, I., Boulter, E., Pisano, S., et al. (2019). Tumor-stroma mechanics coordinate amino acid availability to sustain tumor growth and malignancy. *Cell Metab.* 29 (1), 124–140. e10. doi:10.1016/j.cmet.2018.09.012
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492
- Cassetta, L., Fraggogianni, S., Sims, A. H., Swierczak, A., Forrester, L. M., Zhang, H., et al. (2019). Human tumor-associated macrophage and monocyte transcriptional landscapes reveal cancer-specific reprogramming, biomarkers, and therapeutic targets. *Cancer Cell* 35 (4), 588–602. e10. doi:10.1016/j.ccell.2019.02.009
- Chan, J. S., Tan, M. J., Sng, M. K., Teo, Z., Phua, T., Choo, C. C., et al. (2017). Cancer-associated fibroblasts enact field cancerization by promoting extratumoral oxidative stress. *Cell Death Dis.* 8 (1), e2562. doi:10.1038/cddis.2016.492
- Chang, K., Yuan, C., and Liu, X. (2021). A new RBPs-related signature predicts the prognosis of colon adenocarcinoma patients. *Front. Oncol.* 11, 627504. doi:10.3389/fonc.2021.627504
- Chen, Y., Zou, L., Zhang, Y., Chen, Y., Xing, P., Yang, W., et al. (2014). Transforming growth factor- $\beta$ 1 and  $\alpha$ -smooth muscle actin in stromal

## 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/fgene.2022.1054152/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

Alterations of CAFs-related genes in CRC patients from TCGA.

### SUPPLEMENTARY FIGURE S2

The Kaplan-Meier survival curves of OS between highly expressed and lowly expressed CAFs-related genes.

### SUPPLEMENTARY FIGURE S3

The differences in CAFs risk scores between different cancer types (A), clinical stages (B), lymphatic invasion (C), and TNM stage (D,F).

### SUPPLEMENTARY FIGURE S4

Correlation between CAFPS and 50 cancer hallmark-related pathways.

### SUPPLEMENTARY FIGURE S5

The top 30 most relevant drugs for 20 prognostic genes were shown by a bubble chart.

fibroblasts are associated with a poor prognosis in patients with clinical stage I-IIIa non-small cell lung cancer after curative resection. *Tumour Biol.* 35 (7), 6707–6713. doi:10.1007/s13277-014-1908-y

Chen, W., Huang, J., Xiong, J., Fu, P., Chen, C., Liu, Y., et al. (2021). Identification of a tumor microenvironment-related gene signature indicative of disease prognosis and treatment response in colon cancer. *Oxid. Med. Cell. Longev.* 2021, 6290261. doi:10.1155/2021/6290261

Conrad, C., Benzel, J., Dorzweiler, K., Cook, L., Schlomann, U., Zarbock, A., et al. (2019). ADAM8 in invasive cancers: Links to tumor progression, metastasis, and chemoresistance. *Clin. Sci.* 133 (1), 83–99. doi:10.1042/CS20180906

Du, J., Tao, Q., Liu, Y., Huang, Z., Jin, H., Lin, W., et al. (2022). Assessment of the targeted effect of Sijunzi decoction on the colorectal cancer microenvironment via the ESTIMATE algorithm. *PLoS ONE* 17 (3), e0264720. doi:10.1371/journal.pone.0264720

Geeleher, P., Cox, N., and Huang, R. S. (2014a). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Geeleher, P., Cox, N. J., and Huang, R. S. (2014b). Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol.* 15 (3), R47. doi:10.1186/gb-2014-15-3-r47

Goldman, M. J., Craft, B., Hastie, M., Repčič, K., McDade, F., Kamath, A., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38 (6), 675–678. doi:10.1038/s41587-020-0546-8

Gong, W., Li, Y., Xian, J., Yang, L., Wang, Y., Zhang, X., et al. (2022). Long non-coding RNA LSAMP-1 is down-regulated in non-small cell lung cancer and predicts a poor prognosis. *Cancer Cell Int.* 22 (1), 181. doi:10.1186/s12935-022-02592-0

- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 147, 7. doi:10.1186/1471-2105-14-7
- Huang, T. X., Tan, X. Y., Huang, H. S., Li, Y. T., Liu, B. L., Liu, K. S., et al. (2022). Targeting cancer-associated fibroblast-secreted WNT2 restores dendritic cell-mediated antitumor immunity. *Gut* 71 (2), 333–344. doi:10.1136/gutjnl-2020-322924
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1
- Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* 16 (9), 582–598. doi:10.1038/nrc.2016.73
- Kechagia, J. Z., Ivaska, J., and Roca-Cusachs, P. (2019). Integrins as biomechanical sensors of the microenvironment. *Nat. Rev. Mol. Cell Biol.* 20 (8), 457–473. doi:10.1038/s41580-019-0134-2
- Kleppe, A., Albrechtsen, F., Vlatkovic, L., Pradhan, M., Nielsen, B., Hveem, T. S., et al. (2018). Chromatin organisation and cancer prognosis: A pan-cancer study. *Lancet. Oncol.* 19 (3), 356–369. doi:10.1016/S1470-2045(17)30899-9
- Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9559, 559. doi:10.1186/1471-2105-9-559
- Lee, H. M., Lo, K. W., Wei, W., Tsao, S. W., Chung, G., Ibrahim, M. H., et al. (2017). Oncogenic S1P signalling in EBV-associated nasopharyngeal carcinoma activates AKT and promotes cell migration through S1P receptor 3. *J. Pathol.* 242 (1), 62–72. doi:10.1002/path.4879
- Li, M., Wang, Z., Dong, S., and Xu, Y. (2022). NAV3 is a novel prognostic biomarker affecting the immune status of the tumor microenvironment in colorectal cancer. *J. Immunol. Res.* 2022, 8337048. doi:10.1155/2022/8337048
- Mao, X., Xu, J., Wang, W., Liang, C., Hua, J., Liu, J., et al. (2021). Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: New findings and future perspectives. *Mol. Cancer* 20 (1), 131. doi:10.1186/s12943-021-01428-1
- Martinez-Romero, J., Bueno-Fortes, S., Martín-Merino, M., Ramirez de Molina, A., and De Las Rivas, J. (2018). Survival marker genes of colorectal cancer derived from consistent transcriptomic profiling. *BMC Genomics* 19, 857. doi:10.1186/s12864-018-5193-9
- Monteran, L., and Erez, N. (2019). The dark side of fibroblasts: Cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front. Immunol.* 10, 1835. doi:10.3389/fimmu.2019.01835
- Niitsu, H., Hinoi, T., Kawaguchi, Y., Sentani, K., Yuge, R., Kitadai, Y., et al. (2016). KRAS mutation leads to decreased expression of regulator of calcineurin 2, resulting in tumor proliferation in colorectal cancer. *Oncogenesis* 5 (8), e253. doi:10.1038/oncsis.2016.47
- Qi, J., Crinier, A., Escalière, B., Ye, Y., Wang, Z., Zhang, T., et al. (2021). Single-cell transcriptomic landscape reveals tumor specific innate lymphoid cells associated with colorectal cancer progression. *Cell Rep. Med.* 2 (8), 100353. doi:10.1016/j.xcrm.2021.100353
- Schlomann, U., Koller, G., Conrad, C., Ferdous, T., Golfi, P., Garcia, A. M., et al. (2015). ADAM8 as a drug target in pancreatic cancer. *Nat. Commun.* 66, 6175. doi:10.1038/ncomms7175
- Shiga, K., Hara, M., Nagasaki, T., Sato, T., Takahashi, H., and Takeyama, H. (2015). Cancer-associated fibroblasts: Their characteristics and their roles in tumor growth. *Cancers (Basel)* 7 (4), 2443–2458. doi:10.3390/cancers7040902
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39 (5), 1–13. doi:10.18637/jss.v039.i05
- Song, D., Zhang, D., Chen, S., Wu, J., Hao, Q., Zhao, L., et al. (2022). Identification and validation of prognosis-associated DNA repair gene signatures in colorectal cancer. *Sci. Rep.* 12 (1), 6946. doi:10.1038/s41598-022-10561-w
- Thompson, C. A., Begi, T., and Parada, H., Jr (2022). Alarming recent rises in early-onset colorectal cancer. *Cancer* 128 (2), 230–233. doi:10.1002/cncr.33919
- Toll, D. B., Janssen, K. J., Vergouwe, Y., and Moons, K. G. (2008). Validation, updating and impact of clinical prediction rules: A review. *J. Clin. Epidemiol.* 61 (11), 1085–1094. doi:10.1016/j.jclinepi.2008.04.008
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419
- Vicent, S., Luis-Ravelo, D., Antón, I., García-Tuñón, I., Borrás-Cuesta, F., Dotor, J., et al. (2008). A novel lung cancer signature mediates metastatic bone colonization by a dual mechanism. *Cancer Res.* 68 (7), 2275–2285. doi:10.1158/0008-5472.CAN-07-6493
- Villaronga, M. Á., Teijeiro, S. Á., Hermida-Prado, F., Garzón-Arango, M., Sanz-Moreno, V., and García-Pedrero, J. M. (2018). Analysis of invasive activity of CAF spheroids into three dimensional (3D) collagen matrices. *Methods Mol. Biol.* 1731, 145–154. doi:10.1007/978-1-4939-7595-2\_14
- Xu, T., Xu, W., Zheng, Y., Li, X., Cai, H., Xu, Z., et al. (2022). Comprehensive FGFR3 alteration-related transcriptomic characterization is involved in immune infiltration and correlated with prognosis and immunotherapy response of bladder cancer. *Front. Immunol.* 13, 931906. doi:10.3389/fimmu.2022.931906
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 42, 2612. doi:10.1038/ncomms3612
- Yu, Q., Wang, X., Yang, Y., Chi, P., Huang, J., Qiu, S., et al. (2021). Upregulated NLGN1 predicts poor survival in colorectal cancer. *BMC Cancer* 21 (1), 884. doi:10.1186/s12885-021-08621-x
- Zack, M. D., Malfait, A. M., Skepner, A. P., Yates, M. P., Griggs, D. W., Hall, T., et al. (2009). ADAM-8 isolated from human osteoarthritic chondrocytes cleaves fibronectin at Ala(271). *Arthritis Rheum.* 60 (9), 2704–2713. doi:10.1002/art.24753
- Zhang, Y. A., Zhu, J. M., Yin, J., Tang, W. Q., Guo, Y. M., Shen, X. Z., et al. (2014). High expression of neuro-oncological ventral antigen 1 correlates with poor prognosis in hepatocellular carcinoma. *PLoS ONE* 9 (3), e90955. doi:10.1371/journal.pone.0090955
- Zhang, Y., Coleman, M., and Brekken, R. A. (2021). Perspectives on hypoxia signaling in tumor stroma. *Cancers (Basel)* 13 (12), 3070. doi:10.3390/cancers13123070
- Zhao, R., Ding, D., Yu, W., Zhu, C., and Ding, Y. (2020). The lung adenocarcinoma microenvironment mining and its prognostic merit. *Technol. Cancer Res. Treat.* 19, 1533033820977547. doi:10.1177/1533033820977547
- Zhou, D., Ren, K., Wang, J., Ren, H., Yang, W., Wang, W., et al. (2018). Erythropoietin-producing hepatocellular A6 overexpression is a novel biomarker of poor prognosis in patients with breast cancer. *Oncol. Lett.* 15 (4), 5257–5263. doi:10.3892/ol.2018.7919