scientific reports

Check for updates

OPEN

Development and validation of a cancer-associated fibroblast gene signature-based model for predicting immunotherapy response in colon cancer

Daoyang Zou[1], Xi Xin[2], Huangzhen Xu[1], Yunxian Xu[1] & Tianwen Xu[1]✉

The efficacy of immune checkpoint inhibitors in colon cancer has been established, and there is an urgent need to identify new molecular markers for colon cancer immunotherapy to guide clinical decisions. Using the "EPIC" and "MCPcounter" R packages to conduct cancer-associated fibroblast (CAF) infiltration scoring on colon cancer samples from the TCGA database and the GEO database, the WGCNA analysis was performed on the two databases' samples based on the CAF infiltration scores to screen for CAF-related genes. LASSO regression analysis was used to construct a risk model with these genes. Comprehensive bioinformatics analysis was conducted on the constructed model to evaluate the stability of its prediction of CAF infiltration abundance and the stability of its prediction of immunotherapy efficacy. The newly constructed risk model could well reflect the abundance of CAF infiltration in colon cancer, with a correlation coefficient of 0.91 in the training cohort TCGA-COAD and 0.88 in the validation cohort GSE39582. GSEA analysis revealed that CAF is closely related to functions associated with extracellular matrix remodeling. The constructed risk model can predict the efficacy of immunotherapy in colon cancer well, with the high-risk group showing significantly poorer immunotherapy response than the low-risk group, with an expected effective rate of immunotherapy of 68 vs. 24% in the training group ($P < 0.001$) and 64 vs. 26% in the validation group ($P < 0.001$). The AUC value for predicting immunotherapy response by the risk model in the training group was 0.780 (95% CI 0.736–0.820), and in the validation group, the AUC value was 0.774 (95% CI 0.735–0.810). Drug sensitivity analysis showed that the expected chemotherapeutic effect in the low-risk group was superior to that in the high-risk group. CAF is associated with immunosuppression and drug resistance. Predicting the efficacy of immunotherapy in colon cancer based on the abundance of CAF infiltration is a feasible approach. For the high-risk population identified by our model, clinical consideration should be given to prioritizing non-immunotherapy approaches to avoid potential risks associated with immunotherapy.

Keywords Cancer-associated fibroblasts (CAF), Risk stratification model, Immunotherapy response prediction, Tumor microenvironment, Colon cancer

Colorectal cancer (CRC) maintains its position as the third most prevalent and second most lethal malignancy globally[1]. The latest epidemiological data shows that in the United States, colorectal cancer is still the third most common cancer and the second leading cause of cancer-related deaths[2]. Paradoxically, while overall CRC incidence demonstrates a declining trajectory[3], surveillance data reveal concerning escalations in both early-onset presentations (age < 50 years) and advanced-stage diagnoses[3]. Colon cancer is the most common subtype of colorectal cancer, accounting for about 70% of all pathological cases and posing a serious threat to human health and life[3]. Like most cancers, chemotherapy and surgery have been the mainstay of treatment for colon cancer. The choice of treatment regimen varies depending on the stage of the tumor. Surgery is the preferred treatment for localized colon cancer (stage I-III)[4]. Adjuvant chemotherapy is also administered to patients with locally advanced colon cancer or lymph node-positive (stage III) to eliminate any residual cancer cells and

[1]The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. [2]Ganzhou People's Hospital, Ganzhou, China. ✉email: xutianwen53@163.com

reduce the risk of recurrence[5]. However, for advanced or metastatic colon cancer, a multimodal approach is often employed. In 2013, cancer immunotherapy was recognized as the top scientific breakthrough by SCIENCE magazine[6]. In the past decade, immune checkpoint inhibitors (ICIs) have revolutionized the treatment approach for many malignant tumors. However, only a small subset of patients with tumors respond well to ICIs[7]. Based on the results of the Keynote-177 study, pembrolizumab has been approved by the FDA for the treatment of MSI-H/dMMR colon cancer patients[8]. The CheckMate-142 study also demonstrated the feasibility of nivolumab in combination with ipilimumab for treating MSI-H/dMMR colon cancer[9]. Immunotherapy has now entered the clinical practice guidelines for colon cancer. MSI-H/dMMR is often used as a predictive biomarker for treatment efficacy in colon cancer immunotherapy[8–10]. However, MSI status alone is insufficient to identify potential immunotherapy beneficiaries in colon cancer. In the Keynote-177 study, even among the selected MSI-H/dMMR patients, the overall response rate to immunotherapy was only 43%[11]. Therefore, there is an urgent need for more precise biomarkers to predict the clinical efficacy of immunotherapy.

Current research indicates that different tumor microenvironments (TME) lead to varying outcomes in ICIs therapy[12]. TME refers to the cellular environment in which tumors or tumor stem cells exist. TME consists of various immune-infiltrating cells (such as macrophages, dendritic cells, and lymphocytes), blood vessels, extracellular matrix (ECM), cancer-associated fibroblasts (CAFs), myeloid-derived suppressor cells (MDSCs), cytokines, and signaling molecules[13–15]. CAFs, one of the most important stromal cell components in TME, have been shown in the majority of studies to promote tumor progression. CAFs remodel the ECM, weaken the tumor immune microenvironment, and the reshaped ECM protein network acts as a physical barrier to hinder immune cells from entering cancer cells[16,17]. This, in turn, promotes immune escape[18–21]. CAFs express programmed death ligand-1 and −2 (PD-L1 and PD-L2), which bind to programmed death receptor 1 (PD-1) on the surface of T cells, inhibiting the anti-tumor immune response of T cells[22]. CAFs can also inhibit innate and adaptive anti-tumor immune responses through the secretion of various chemokines and cytokines, such as transforming growth factor-beta (TGF-β), interleukin-6 (IL-6), and chemokine ligand 12 (CXCL12)[23,24]. Additionally, CAFs can convert the Th1 phenotype to Th2 and recruit regulatory T cells (Tregs) to suppress the cytotoxicity of CD8 + T cells[25]. Moreover, in colorectal cancer, CAF-derived prostaglandin E2 (PGE2) and indoleamine 2,3-dioxygenase (IDO) can down-regulate the expression of NK activating receptors, perforin, and granzyme B, thereby inhibiting the cytotoxicity and cytokine production of NK cells[26]. Studies have also shown a close relationship between CAFs and CRC prognosis, with genes that are positively correlated with CRC recurrence and poor prognosis being mainly upregulated in CAFs rather than tumor cells[19]. The above research indicates the central role of CAFs in the tumor immune-suppressive microenvironment. Therefore, we speculate that a high infiltration of CAFs in colon cancer tissue implies a poor response to immune therapy, and assessing CAF abundance for predicting the efficacy of immune therapy in colon cancer is a feasible approach.

This study aims to use the "EPIC" and "MCPcounter" R packages to score the CAF infiltration abundance in colon cancer samples from the GEO database and TCGA-COAD samples. Based on the scoring results, the Weighted Gene Co-Expression Network Analysis (WGCNA) will be applied to perform correlation analysis on the two groups of samples to identify CAF-related genes. A CAF risk model will be constructed using the obtained CAF-related genes, where the expression levels of the CAF-related genes in colon cancer tumor tissue will reflect the abundance of CAF infiltration. The correlation between the risk scores obtained from the model and CAF infiltration abundance will be assessed. Based on the constructed risk model, the samples will be divided into high and low-risk groups, and the survival and expected immune therapy responses between the two groups will be evaluated, along with the assessment of the immune microenvironment and immune cell infiltration between the two groups. GSEA analysis will be conducted on the high and low-risk groups to explore the feasibility of predicting immune therapy based on CAF infiltration scoring. Finally, the obtained risk model will be validated using an external dataset. Through comprehensive bioinformatics analysis, the potential relationship between CAF infiltration abundance and the expected immunotherapy efficacy in colon cancer was explored, seeking a new immunotherapy auxiliary decision-making tool for colon cancer patients, and providing preliminary research foundations for CAF infiltration abundance as a predictive marker for the efficacy of immunotherapy in colon cancer.

Method

The workflow of the whole study is presented in (Fig. 1).

Data processing and preprocessing

Transcriptome data (TSV format) and clinical information data (XML format) of colon adenocarcinoma (COAD) samples were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and processed using a Perl script to obtain the required gene expression matrix and clinical information file. The dataset contains a total of 476 tumor samples after excluding some normal tissue samples before analysis. Raw microarray data of colon cancer datasets GSE39582, GSE44076, and GSE143985 were downloaded from the Gene Expression Omnibus (GEO) website (http://www.ncbi.nlm.nih.gov/geo/). The original data were processed and transformed into gene expression matrix using Perl script. Among GSE44076 and GSE143985 transcriptome data, some normal tissue samples were excluded before analysis, resulting in a total of 768 tumor tissue samples. The three datasets were normalized and adjusted using the "limma" and "sva" packages in R language and integrated into a comprehensive gene matrix. The GSE39582 dataset is a large cohort of colon cancer patients collected by the French National Cancer Institute, containing detailed patient clinical information and survival data files. We selected this cohort as an independent validation set for subsequent prognostic model analysis.

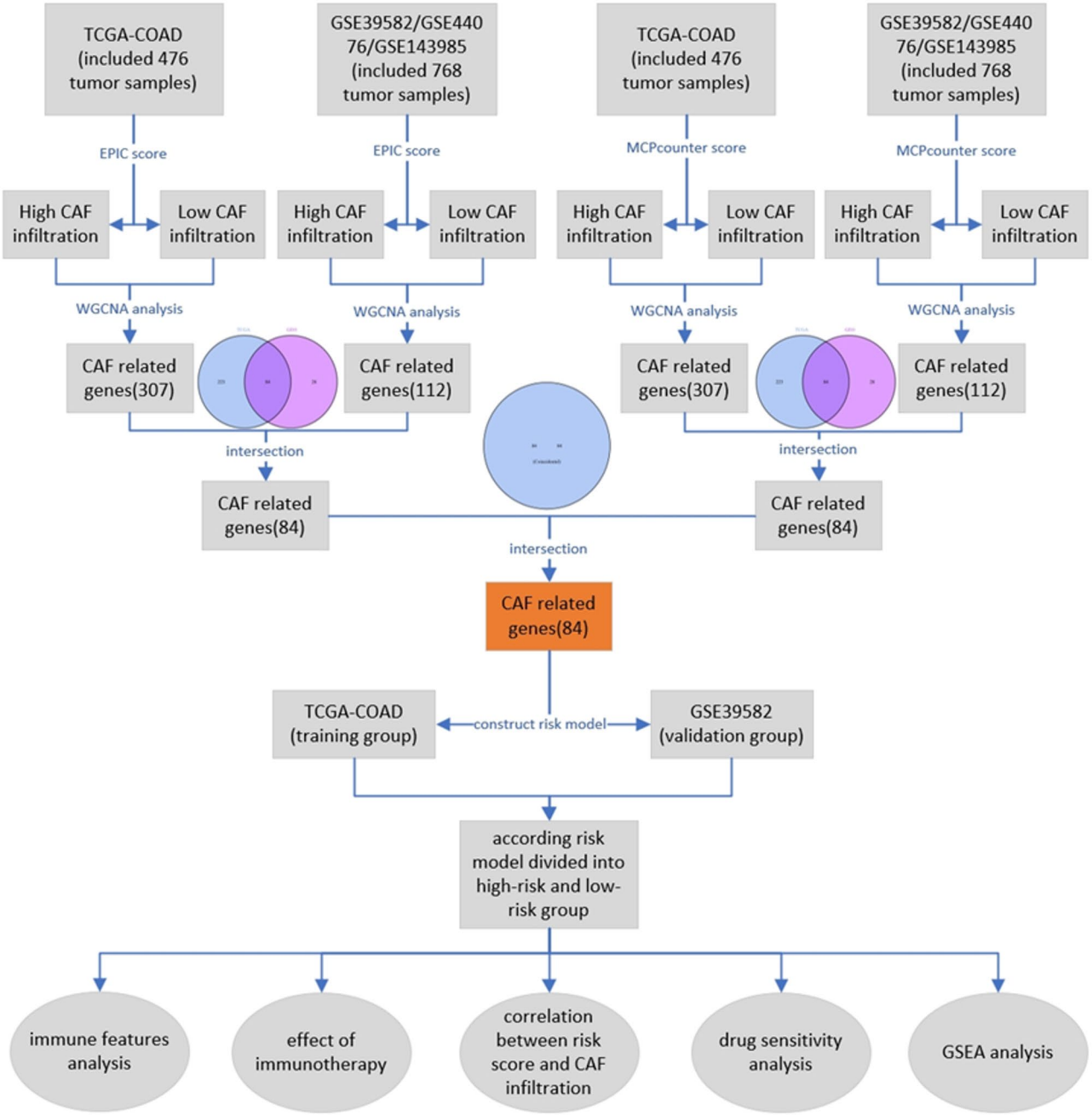

Fig. 1. The workflow of the study.

Use WGCNA analysis to screen for CAF-related genes

The "EPIC"[27]and "MCPcounter"[28] R language packages are commonly used in bioinformatics analysis to assess the abundance of CAF infiltration in tumor tissues based on gene expression data. These two R language packages were respectively used to score the CAF infiltration in colon cancer samples obtained from the GEO database and the TCGA database. The Weighted Gene Co-expression Network Analysis (WGCNA) method aims to identify modules of genes that are co-expressed and explore the relationships between gene networks and the phenotypes of interest, as well as determine core genes within the network[29]. The "WGCNA", "limma" and "gplots" software package in R language was used to conduct WGCNA analysis on the GEO samples and TCGA samples separately to identify gene modules most relevant to CAF and obtain the corresponding core genes of the modules. The intersection of core genes obtained from the GEO and TCGA samples were considered as CAF-related genes. The CAF-related genes obtained were subjected to Gene Ontology (GO)[30] and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis[31–34].

Construct risk model using CAF-related genes

Firstly, perform univariate Cox regression and survival analysis on CAF-related genes to identify prognostic-related genes, with a significance threshold set at $p < 0.05$. Subsequently, a risk model was constructed using LASSO regression analysis, and LASSO regression graphs were plotted. To evaluate the model's predictive performance and prevent overfitting, the cv.glmnet() function from the "glmnet" R package was used to plot cross-validation graphs and identify the point with the smallest cross-validation error. TCGA dataset will be used as the training set, and the GSE39582 dataset will be used as the validation set.

The risk score is calculated using the following formula:

$$Risk\ score = \sum [\mathrm{Exp}\ (\mathrm{Gene}) * \mathrm{coef}\ (\mathrm{Gene})]$$

where Exp(Gene) is the expression level of the CAF-related gene, and coef(Gene) is the corresponding regression coefficient. Finally, based on the median risk score, all samples in the TCGA-COAD cohort and the GSE39582 cohort were divided into high-risk and low-risk groups. Survival curves were plotted using the "survival" and "survminer" R language packages.

Analyze the correlation between the risk score and CAF infiltration abundance

Use the "GGally" R package to analyze the correlation between the risk scores obtained from the model and CAF infiltration abundance. Use the "limma" and "pheatmap" R packages to analyze the correlation between the risk scores and the expression levels of CAF-related genes reported in previous literature. Perform ssGSEA analysis to screen for pathways that are most correlated with the risk scores. Validate the expression differences of the genes involved in the model construction between fibroblasts and colon cancer using the CCLE database.

Analyze the drug sensitivity and immunotherapy sensitivity between the high-risk and low-risk groups

Upload the gene expression data from the TCGA-COAD and GSE39582 cohorts to the TIDE website (http://tide.dfci.harvard.edu/) to obtain TIDE scores and evaluate the response to immunotherapy in the high-risk and low-risk groups. Obtain the database files (**GDSC2_Expr.rds** and **GDSC2_Res.rds**)[35] from the GDSC website (http://www.cancerrxgene.org/) for further analysis. Use the "oncoPredict" R package[36] to evaluate the sensitivity of the high-risk and low-risk groups to drugs based on the GDSC database files.

Perform GSEA enrichment analysis for the high-risk and low-risk groups

GSEA enrichment analysis is a commonly used bioinformatics method for assessing whether there is a significant enrichment of predefined gene sets between two biological states. Based on the constructed risk model, the samples from the TCGA cohort and GSE39582 cohort were divided into high and low-risk groups. Gene sets "**c5.go.symbols.gmt"** and "**c2.cp.kegg.Hs.symbols.gmt"** were downloaded from the MSigDB database (https://www.gsea-msigdb.org/gsea/index.jsp). The R language packages "clusterProfiler", "enrichplot", "limma", and "org.Hs.eg.db" were used to conduct GSEA enrichment analysis on the high and low-risk groups. Enrichment analysis was conducted to observe which functions and pathways were enriched in the high-risk and low-risk groups. Subsequently, single-sample Gene Set Enrichment Analysis (ssGSEA) was performed on different signaling pathways to analyze the correlation between different pathways and the risk scores obtained from the model. For the ssGSEA analysis, consider pathways with $p < 0.001$ and correlation coefficient > 0.7 to be significant.

Analyze the immune characteristics of the high-risk and low-risk groups

Divide the TCGA-COAD and GSE39582 cohorts into high-risk and low-risk groups based on the risk model. Evaluate the tumor microenvironment characteristics of the high-risk and low-risk groups using the ESTIMATE algorithm[37] Analyze the infiltration profiles of 22 immune cell types in the high-risk and low-risk groups using the CIBERSORT algorithm[38].

Statistical method

To perform statistical analysis using R software version 4.3.1, use a significance level of $p < 0.05$ to determine statistical significance. Additionally, we define the significance levels as follows: $p < 0.05$ as "*", $p < 0.01$ as "**", and $p < 0.001$ as "***".

Result

WGCNA analysis screened 84 CAF-related genes

Using the "EPIC" and "MCPcounter" R language packages, CAF infiltration scores were successfully assessed for colon cancer samples from the GEO database and TCGA-COAD samples. Based on the CAF infiltration scores obtained from these two software tools, WGCNA analysis was conducted to identify the most CAF-related gene modules in the samples. Topological Overlap Matrix (TOM) and module correlation heatmaps were generated for the WGCNA analysis based on the CAF scores obtained from the "EPIC" software(Fig. 2A–D) and the "MCPcounter" software (Fig. 2E–H). From the TOM, the turquoise module with the strongest correlation is shown in deeper color, and the module correlation heatmap results indicate that the turquoise module has the smallest correlation test p-value, suggesting that the turquoise module is the most CAF-related gene module, and the core genes within the blue-green module are considered CAF-related genes. Take the intersection of the genes obtained from the TCGA and GEO samples as the CAF-related genes for each scoring software. Finally, obtain the intersection of CAF-related genes obtained from both scoring software as the final

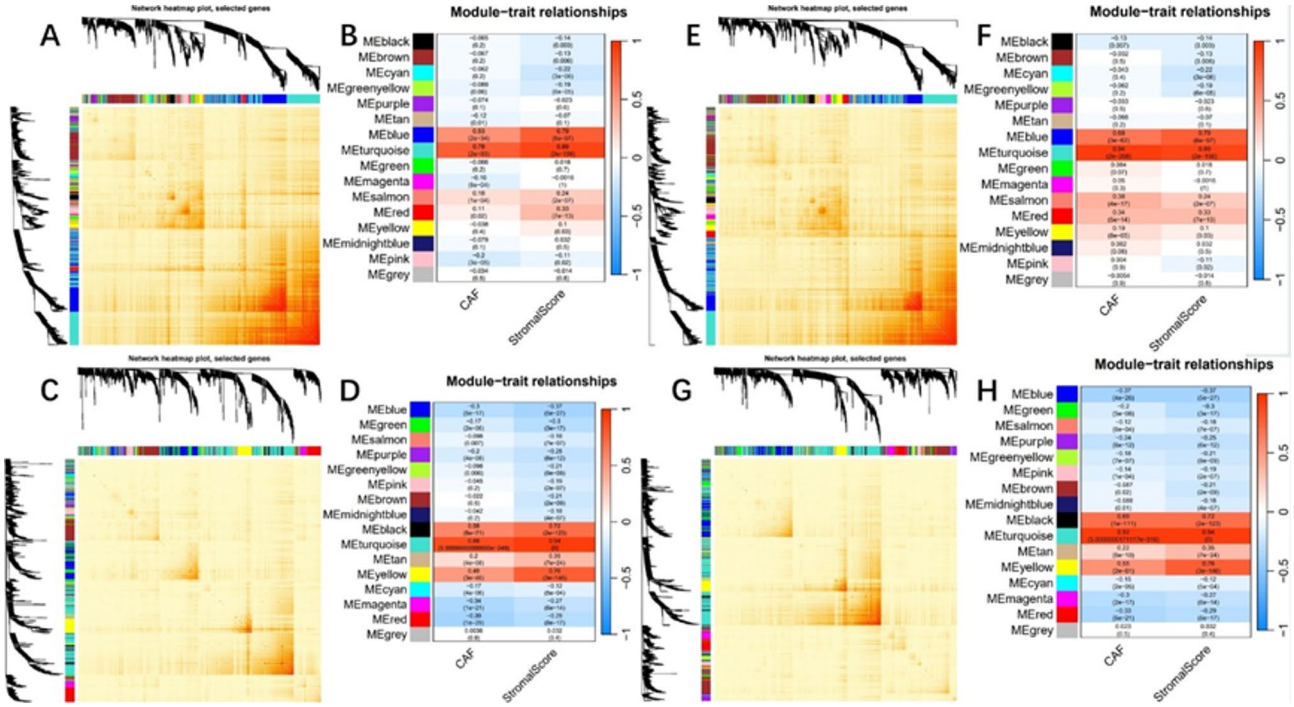

Fig. 2. (**A**) The topological overlap matrix (TOM) of TCGA samples using weighted gene co-expression network analysis (WGCNA) based on the EPIC score. (**B**) The heatmap of module-trait correlations. (**C**) The TOM of GEO samples using WGCNA based on the EPIC score. (**D**) The heatmap of module-trait correlations. (**E**) The TOM of TCGA samples using WGCNA based on the MCPcounter score. (**F**) The heatmap of module-trait correlations. (**G**) The TOM of GEO samples using WGCNA based on the MCPcounter score. (**H**) The heatmap of module-trait correlations.

set of CAF-related genes. Interestingly, the CAF-related genes obtained using the "EPIC" and "MCPcounter" R packages are consistent. In total, 84 CAF-related genes are identified (Table 1).

Perform GO enrichment analysis on the obtained CAF-related genes, and the results show enrichment in functions closely related to fibroblasts, such as extracellular matrix organization, extracellular structure organization, extracellular matrix structural constituent, collagen-containing extracellular matrix, and collagen fibril organization (Fig. 3A). KEGG enrichment analysis demonstrates enrichment in pathways such as protein digestion and absorption, PI3K-Akt signaling pathway, focal adhesion, and ECM-receptor interaction (Fig. 3B).

Constructed a risk model using the obtained CAF-related genes

Perform univariate and Cox regression to screen for prognostically relevant genes. Use the TCGA-COAD cohort as the training set to build the risk model, and validate the model's accuracy using the GSE39582 cohort. Apply lasso regression analysis using the selected CAF-related genes to construct the risk model. Use cross-validation to obtain the optimal value of lambda (λ) (Fig. 3C, D). Finally, a risk model with 4 CAF-related genes is determined, including *VEGFC*, *QKI*, *VIM*, and *ITGA5*. The risk score is calculated using the following formula: RiskScore = *VEGFC*(Exp)*0.0113 + *QKI*(Exp)*0.0016 + *VIM*(Exp)*0.087 + *ITGA5*(Exp)*0.033.

Based on the risk model, divide the samples into high-risk and low-risk groups. Survival curves show that both in the TCGA-COAD cohort (Fig. 3E) and the GSE39582 cohort (Fig. 3F), the low-risk group has a better prognosis than the high-risk group.

The risk score obtained from the model is significantly correlated with CAF infiltration abundance

We analyzed the correlation between the risk score and CAF infiltration abundance obtained from different scoring software to explore the correlation between the risk score and CAF infiltration abundance. The results show that in both the TCGA-COAD cohort (Fig. 4A) and the GSE39582 cohort (Fig. 4B), the risk score is positively correlated with CAF infiltration abundance obtained from all scoring software, indicating that higher risk scores are associated with higher CAF infiltration abundance. Furthermore, we analyzed the correlation between the risk score and the expression levels of CAF-related genes reported in previous literature. The results show that CAF-related genes reported in the literature are significantly upregulated in the high-risk group and downregulated in the low-risk group (Fig. 4C). Subsequently, we used the CCLE database to validate the expression differences of the genes involved in the model construction between fibroblasts and colon cancer. The results show that all the genes involved in the model construction are significantly upregulated in CAF and downregulated in colon cancer tissues (Fig. 4D, E).

The CAF-related genes:
CNRIP1
PTPRM
TCF4
ZNF521
FBXL7
TSPYL5
TGFB1I1
VEGFC
PECAM1
NUAK1
PLEKHO1
ATP8B2
SNAI2
CMTM3
QKI
AXL
CD93
GNB4
TUBB6
OSMR
MITF
VIM
KCNE4
FRMD6
FSTL1
FERMT2
EFEMP2
HEG1
SDC2
COL18A1
GLIS2
GLT8D2
THY1
GPC6
HTRA1
PDGFRB
SERPINF1
PXDN
NXN
C1R
OLFML2B
ITGA5
EMILIN1
CDH11
SPARC
COL15A1
CTSK
COL5A1
C1S
DDR2
MXRA8
TIMP2
CALD1
VCAN
COL6A2
COL5A2
Continued

The CAF-related genes:
FIBIN
FBN1
DCN
ANGPTL2
RAB31
GFPT2
SERPING1
COL3A1
FAP
ADAMTS2
MMP2
DPYSL3
COL1A2
COL6A3
CCDC80
MXRA5
PRRX1
BGN
LUM
ADAM12
ANTXR1
COL8A1
AEBP1
SULF1
GAS1
FNDC1
SPOCK1
THBS2

Table 1. CAF-related genes.

The comprehensive analysis suggests that the risk score obtained from the model can effectively differentiate the CAF infiltration abundance in different samples, with higher risk scores associated with higher CAF infiltration abundance. Finally, we further analyzed the correlation between the risk score and the expression levels of immune therapy-related genes reported in previous literature. The results show that the risk score is significantly correlated with the expression levels of most immune therapy-related genes (Fig. 4F), indicating that the newly constructed risk model may predict the effectiveness of immune therapy in colon cancer patients.

The risk model can effectively predict the efficacy of immunotherapy and chemotherapy in colon cancer

To evaluate the ability of the constructed risk model to predict the efficacy of immunotherapy in colon cancer, TIDE scores were obtained from the TIDE website (http://tide.dfci.harvard.edu/) to assess the response to immunotherapy in the high and low-risk groups. The results showed that in both the TCGA cohort (Fig. 5A) and the GSE39582 cohort (Fig. 5B), the TIDE scores in the low-risk group were significantly lower than those in the high-risk group. Subsequent bar charts of the expected response rates to immunotherapy also showed that the response rate in the low-risk group was significantly higher than that in the high-risk group ($P < 0.001$), with the TCGA cohort showing 68 vs. 24% ($P < 0.001$) (Fig. 5C) and the GSE39582 cohort showing 64 vs. 26% ($P < 0.001$) (Fig. 5D). This suggests that the low-risk group has a significantly better response to immunotherapy than the high-risk group. ROC curves demonstrate that the AUC values for predicting the effectiveness of immunotherapy based on the risk model are 0.780(95% CI 0.736–0.820) in the TCGA cohort (Fig. 5E) and 0.774(95% CI 0.735–0.810) in the GSE39582 cohort (Fig. 5F), suggesting that the newly constructed risk model can effectively predict the response to immunotherapy in colon cancer.

Furthermore, using the "oncoPredict" R package, we predicted the expected treatment response to commonly used chemotherapy drugs in the high-risk and low-risk groups. The results show that the sensitivity of the most commonly used drug for colon cancer, oxaliplatin, is significantly higher in the low-risk group compared to the high-risk group (Fig. 5G). However, there is no significant difference in the sensitivity to the other two commonly used chemotherapy drugs, 5-fluorouracil (Fig. 5H) and irinotecan (Fig. 5I). This suggests that the expected chemotherapy response is better in the low-risk group compared to the high-risk group.

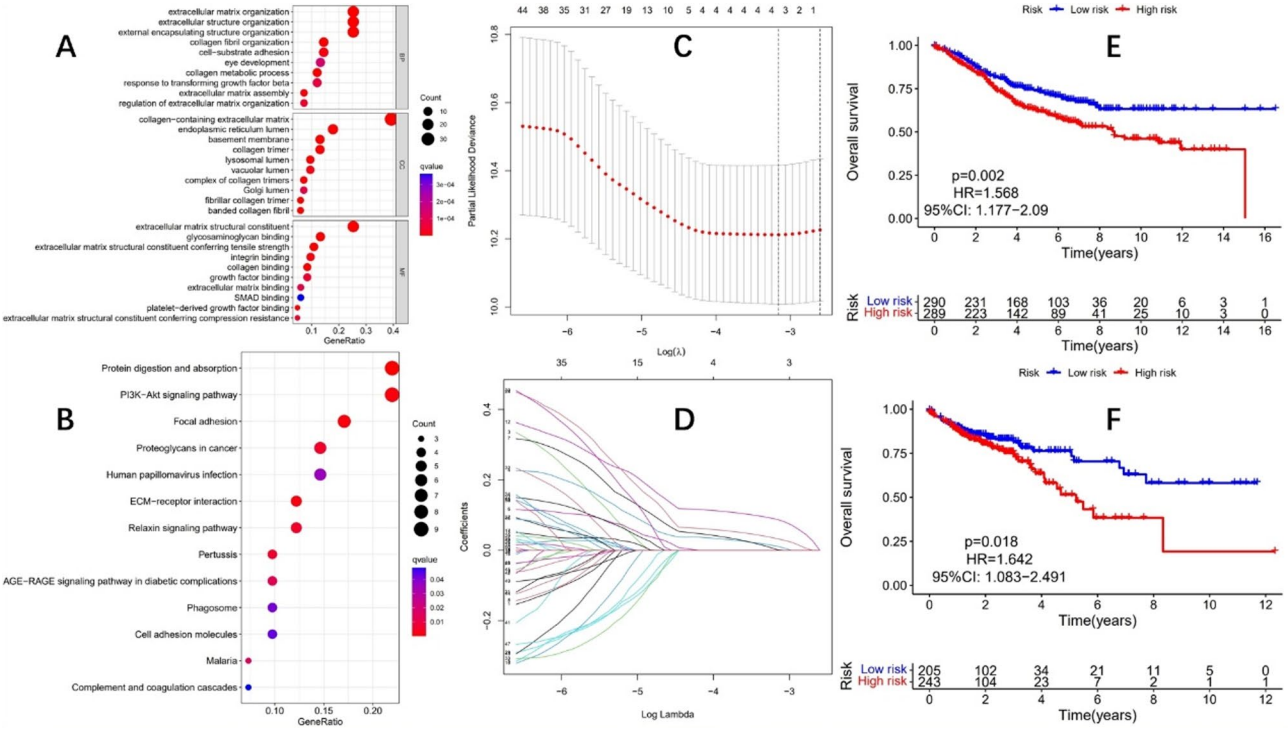

Fig. 3. Perform GO analysis and KEGG analysis on the screened CAF-related genes. (**A**) GO analysis bubble chart. (**B**) KEGG analysis bubble chart. (**C**) The tuning parameter (λ) in the LASSO model (**D**) LASSO coefficient distribution of CAF-related genes. The samples were divided into high-risk and low-risk groups based on the constructed risk model. (**E**) There was a significant difference in the survival curves between the high-risk and low-risk groups in the TCGA-COAD cohort. ($P=0.002$). (**F**) There was a significant difference in the survival curves between the high-risk and low-risk groups in the GSE39582 cohort. ($P=0.018$)

The high-risk group in the TCGA cohort and the GSE39582 cohort show similar enrichment in terms of functional and pathway analysis

According to the model, we divided the TCGA cohort and GSE39582 cohort into high-risk and low-risk groups. GSEA analysis showed similar functional enrichments in the high-risk groups of both cohorts (Fig. 6A, B). In both cohorts, the high-risk groups exhibited enrichment in functions closely related to CAF, such as collagen-containing extracellular matrix, collagen trimer, and extracellular matrix structural constituent. Similarly, the KEGG pathway analysis showed similar pathway enrichments in the high-risk groups of both cohorts (Fig. 6C, D). These pathways included cell adhesion molecules, cytokine-cytokine receptor interaction, ECM-receptor interaction, and focal adhesion, which are all closely related to CAF. However, in the low-risk groups, the TCGA cohort and GSE39582 cohort showed different functional (Fig. 7A, B) and pathway enrichments (Fig. 7C, D). The GSEA enrichment analysis results indicated significant enrichment in CAF-related pathways in the high-risk group, while no enrichment was observed in the low-risk group. This further suggests that the CAF infiltration abundance is significantly higher in the high-risk group compared to the low-risk group.

Conduct SsGSEA analysis to screen for pathways that are most correlated with the risk score

We performed ssGSEA enrichment analysis on both the TCGA-COAD cohort and GSE39582 cohort, with a correlation coefficient threshold set at 0.7. The results show that in both cohorts, the following six pathways: cell adhesion molecules, ECM-receptor interaction, focal adhesion, hypertrophic cardiomyopathy, leukocyte transendothelial migration, and regulation of actin cytoskeleton, are significantly associated with the risk score (Fig. 8A, B). These six pathways are identified as the most closely related to the risk score. Moreover, we observed that the correlation coefficients for ECM-receptor interaction and focal adhesion are both greater than 0.8 in both cohorts. Previous studies have shown that CAF is closely associated with the remodeling of the extracellular matrix. The ECM-receptor interaction and focal adhesion pathways are both ECM-related pathways. This further suggests that the risk model constructed in this study is closely associated with CAF infiltration.

The high-risk and low-risk groups exhibit different immune characteristics

We used the ESTIMATE algorithm[37] to score the tumor microenvironment in both the TCGA cohort and the GSE39582 cohort. We analyzed the differences in the immune microenvironment between the high-risk and low-risk groups. The results show that in both the TCGA cohort (Fig. 9A) and GSE39582 cohort (Fig. 9B), significant differences are observed in the stromal score, immune score, and ESTIMATE score between the high-risk and low-risk groups. We further analyzed the infiltration of 22 immune cell types in the high-risk and low-risk groups of both the TCGA cohort and GSE39582 cohort using the CIBERSORT algorithm[38]. The

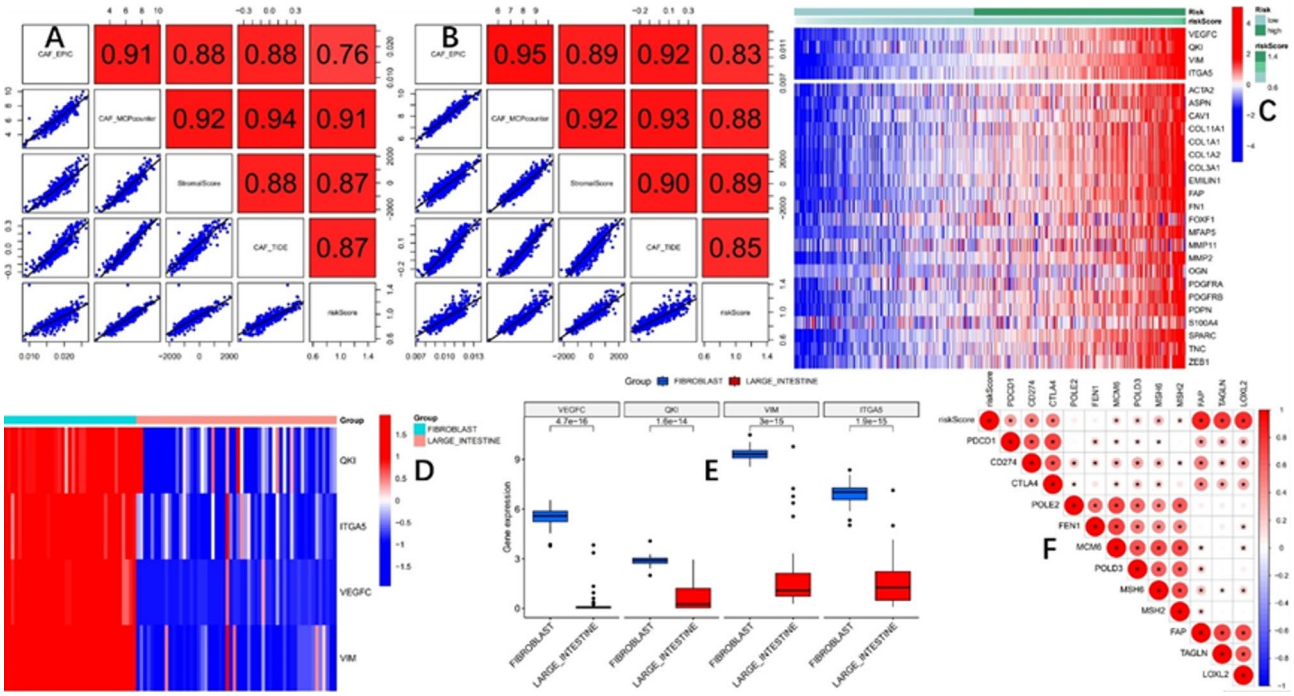

Fig. 4. (**A**) Correlation between risk scores and CAF infiltration abundance from different scoring software in the TCGA cohort. (**B**) Correlation between risk scores and CAF infiltration abundance from different scoring software in the GSE39582 cohort. (**C**) The expression pattern of CAF-related genes reported in previous studies in patients of both high and low-risk groups. (**D**,**E**) Validation in the CCLE database shows that the genes involved in model construction are highly expressed in fibroblasts and lowly expressed in large intestine tissues. (**F**) Risk scores are significantly correlated with most immune-related genes associated with immunotherapy.

results indicate similar immune cell infiltration patterns between the two cohorts (Fig. 9C, D). Plasma cells and CD4 memory T cells show significantly higher infiltration in the low-risk groups of both cohorts, while M0 macrophages and neutrophils exhibit significantly higher infiltration in the high-risk groups of both cohorts.

Discussion

Currently, the effectiveness of immune checkpoint inhibitors (ICIs) in cancer treatment has been established. However, only a small proportion of cancer patients have shown favorable responses to ICIs. In colon cancer, it is widely recognized that patients with dMMR/MSI-H tumors may benefit from immune therapy, but this subtype accounts for only about 15% of all colon cancer cases[39]. Nevertheless, pooled analysis of clinical trials reveals heterogeneous outcomes even within this molecular cohort, with objective response rates oscillating between 32 and 63%, indicating that at least 30% of dMMR/MSI-H patients derive suboptimal therapeutic benefit[40]. Additionally, there have been clinical studies demonstrating that a subset of MSS patients also shows good responses to ICIs[41]. This suggests that MSI status is not a perfect predictive marker for immune therapy outcomes in colon cancer. Therefore, there is a clinical need to identify new predictive markers for immune therapy efficacy.

Previous studies have shown that solid tumors can exhibit two phenotypes in response to immune therapy: "hot tumors" with favorable responses characterized by T lymphocyte infiltration, and "cold tumors" with no or limited T lymphocyte infiltration and poor response[16]. Cancer-associated fibroblasts have been proposed to play a critical role in mediating T lymphocyte exclusion[42]. Studies have also shown that high CAF infiltration in colon cancer is associated with poor response to immunotherapy[43]. Therefore, we hypothesize that the level of CAF infiltration in colon cancer tissue may be predictive of the effectiveness of immune therapy.

Firstly, this study identified CAF-related genes in colon cancer through comprehensive bioinformatics analysis and successfully constructed a risk model using these identified genes. The constructed risk model can effectively reflect the infiltration abundance of CAF in colon cancer tissues. In both the training group TCGA-COAD cohort and the validation group GSE39582 cohort, the risk scores were positively correlated with the CAF infiltration abundance obtained from all scoring software, indicating that the higher the risk score, the greater the abundance of CAF infiltration. Further analysis of the expression levels of CAF-related genes reported in previous literature in the high-risk and low-risk groups showed that these genes were significantly upregulated in the high-risk group and downregulated in the low-risk group. Subsequently, validation using the CCLE database showed that the genes involved in the model construction were significantly upregulated in fibroblasts and downregulated in colon cancer tissues. The comprehensive bioinformatics analysis suggests that

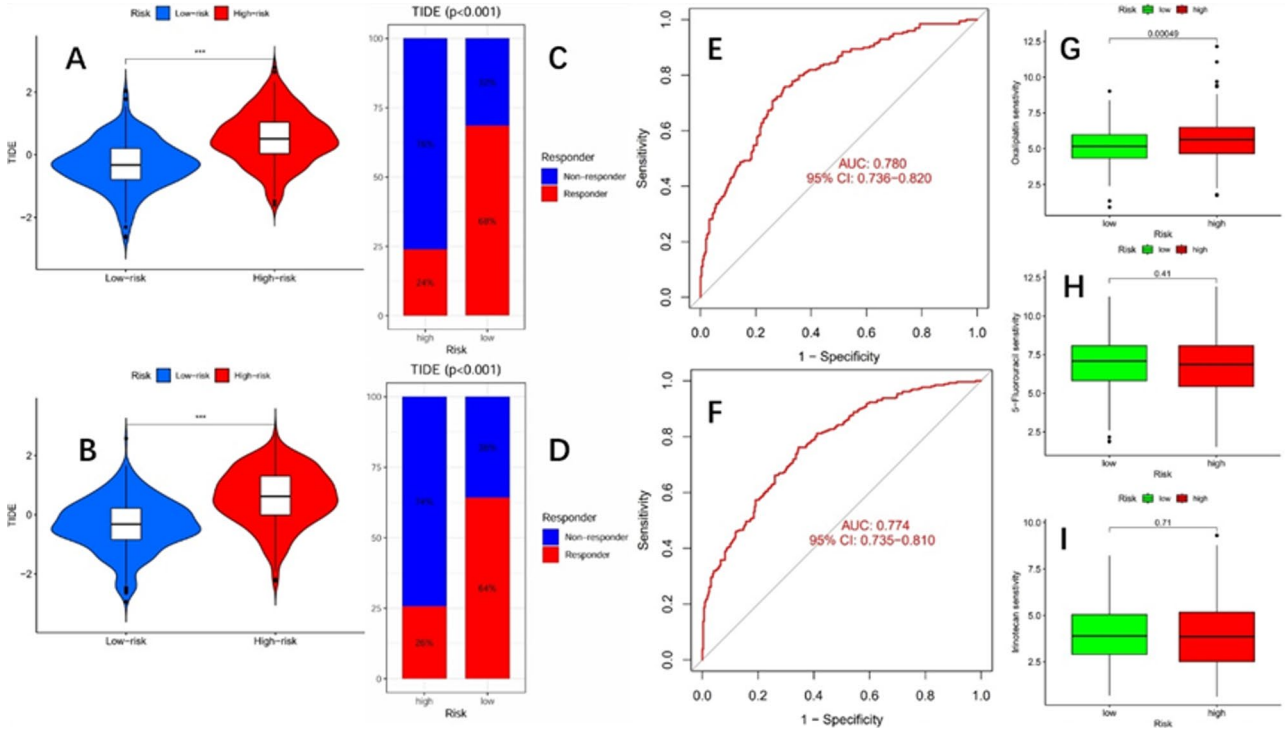

Fig. 5. (**A**) In the TCGA cohort, the TIDE scores in the low-risk group are significantly lower than those in the high-risk group ($P<0.001$ is * * *). (**B**) In the GSE39582 cohort, the TIDE scores in the low-risk group are significantly lower than those in the high-risk group ($P<0.001$ is * * *). Suggesting that the low-risk group exhibits a significantly better response to immunotherapy compared to the high-risk group. (**C**) Bar graph of expected response rate in TCGA cohort (68 vs. 24%, $P<0.001$). (**D**) Bar graph of expected response rate in GSE39582 cohort (64 vs. 26%, $P<0.001$). (**E**) In the TCGA cohort, the area under the curve (AUC) for predicting the efficacy of immunotherapy based on the risk model is 0.780. (**F**) In the GSE39582 cohort, the AUC for predicting the efficacy of immunotherapy based on the risk model is 0.774. Different drug sensitivity of the high and low-risk groups. (**G**) Oxaliplatin. (**H**) 5-fluorouracil. (**I**) irinotecan.

the constructed risk model can effectively reflect the abundance of CAF infiltration in colon cancer tissue, with higher risk scores associated with higher CAF infiltration abundance.

Subsequently, to evaluate the immunotherapy prediction ability of the risk model, patients in the training group TCGA-COAD cohort and the validation group GSE39582 cohort were divided into high and low-risk groups based on the risk model. The TIDE scores were used to predict the immunotherapy efficacy in both high and low-risk groups. The results showed that the expected immunotherapy efficacy in the low-risk group was significantly better than that in the high-risk group, with an expected effective rate of 68 vs. 24% in the training group ($P<0.001$) and 64 vs. 26% in the validation group ($P<0.001$). The Keynote-177 study indicates that even in the selected MSI-H/dMMR patient population, the overall effective rate of immunotherapy is only 43%[11], which suggests that the newly constructed model can predict the efficacy of immunotherapy in colon cancer relatively well. The AUC value for predicting the effective rate of immunotherapy based on the risk model in the training group was 0.780, and in the validation group, the AUC value was 0.774. Many studies use CAF-related genes to construct models for predicting the survival rate of cancer patients, with AUC values of 0.65[44], 0.74[45], and 0.73[46], respectively. The AUC of the model constructed in this study is slightly higher than that of the aforementioned studies, which further highlights the reliability of the model constructed in this research. Further analysis of immune cell infiltration between the high-risk and low-risk groups in the TCGA-COAD and GSE39582 cohorts revealed that in the low-risk group, there was significantly higher infiltration of plasma cells and T lymphocytes compared to the high-risk group. Previous studies have shown that CD8 T cells have difficulty infiltrating into tumors rich in CAF, but accumulate at the tumor edge, leading to immune therapy resistance[42,47]. Single-cell studies have also shown that CAF is associated with immune therapy resistance in tumors[48,49]. The risk model constructed in this study suggests that high CAF infiltration levels in colorectal cancer tumor tissue are associated with decreased T lymphocyte infiltration and poor expected immune therapy response, which is consistent with previous research. Therefore, this study also suggests that predicting the efficacy of immune therapy in colon cancer based on CAF infiltration levels is a feasible approach.

There is growing evidence that CAF is associated with resistance to various cancer therapies and plays a role in enhancing tumor resistance to drugs[50–52]. In colon cancer, CAF has been shown to promote drug resistance as well[53,54]. A recent study demonstrated that CAF-derived exosomal lncRNA FAL1 promotes resistance to the chemotherapy drug oxaliplatin in colon cancer by regulating autophagy[53]. In this study, we used the "oncoPredict" R package to predict the expected treatment response to commonly used chemotherapy drugs in

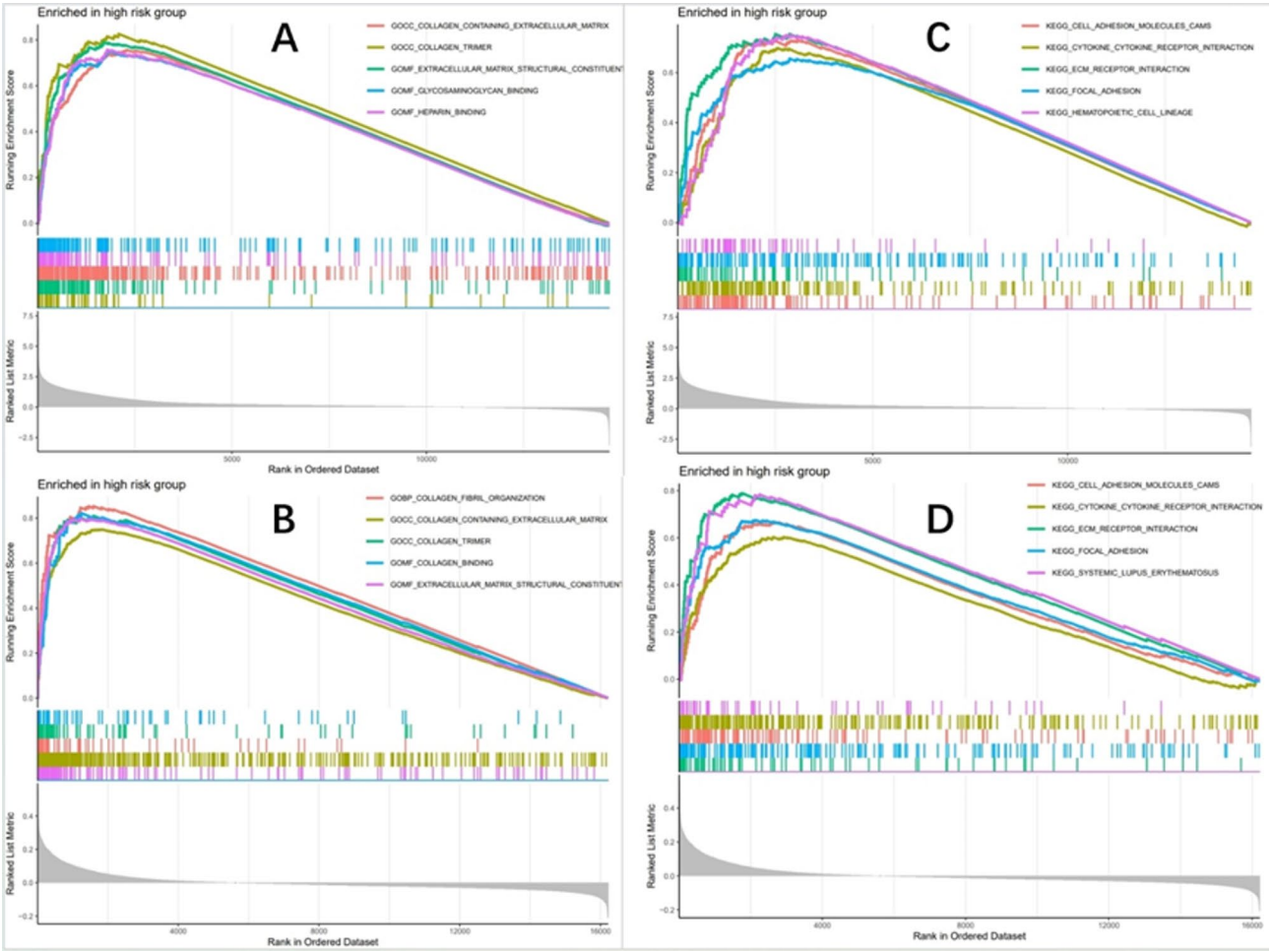

Fig. 6. Similar functional enrichment between the TCGA cohort and the GSE39582 cohort in the high-risk group. (**A**) TCGA cohort. (**B**) GSE39582 cohort. Similar pathway enrichment between the TCGA cohort and the GSE39582 cohort in the high-risk group. (**C**) TCGA cohort. (**D**) GSE39582 cohort.

the high-risk and low-risk groups. The results showed that, in the low-risk group, sensitivity to oxaliplatin was significantly higher compared to the high-risk group. This is consistent with previous studies, as the previous bioinformatics analysis in this study suggested that high risk is associated with higher CAF infiltration levels, indicating a relationship between CAF and oxaliplatin resistance in colon cancer. Overall, the findings of this study support the notion that CAF is indeed associated with oxaliplatin resistance in colon cancer, in line with previous research.

Many studies suggest that CAFs can remodel the extracellular matrix (ECM), improving the tumor microenvironment, inducing drug resistance, and promoting tumor immune evasion[21,55,56]. GO analysis of the identified CAF-related genes in our study revealed significant functional enrichment in ECM-related processes, including extracellular matrix organization, extracellular structure organization, extracellular matrix structural constituent, collagen-containing extracellular matrix, and collagen fibril organization. This indicates that CAFs indeed play a crucial role in ECM remodeling. Subsequently, we performed ssGSEA analysis on the TCGA-COAD and GSE39582 cohorts to identify the pathways most correlated with the risk score. The results showed a significant correlation between the risk score and ECM-related pathways in both cohorts, and the correlation coefficients between ECM receptor interaction and focal adhesion pathways were greater than 0.8. Functional and pathway enrichment analysis has shown that CAFs are closely related to functions and pathways associated with ECM remodeling, which is consistent with previous research findings. In the end, GSEA analysis was performed on the high and low-risk groups in both cohorts. The results showed that in the high-risk group, there was a similar enrichment of functions and pathways, which were closely related to CAFs. In contrast, the functions and pathways in the low-risk group were diverse. This suggests that the high-risk patients selected by the model constructed in this study may have a significantly high level of CAF infiltration. The correlation analysis between the risk score and CAF infiltration abundance indicated a significant positive correlation, which once again highlights the reliability of the model constructed in this research.

In summary, this study utilized public databases to identify CAF-related genes through comprehensive bioinformatics analysis and successfully constructed a risk model using these identified genes. The newly constructed model effectively reflects the abundance of CAF infiltration in colon cancer tissues and predicts the

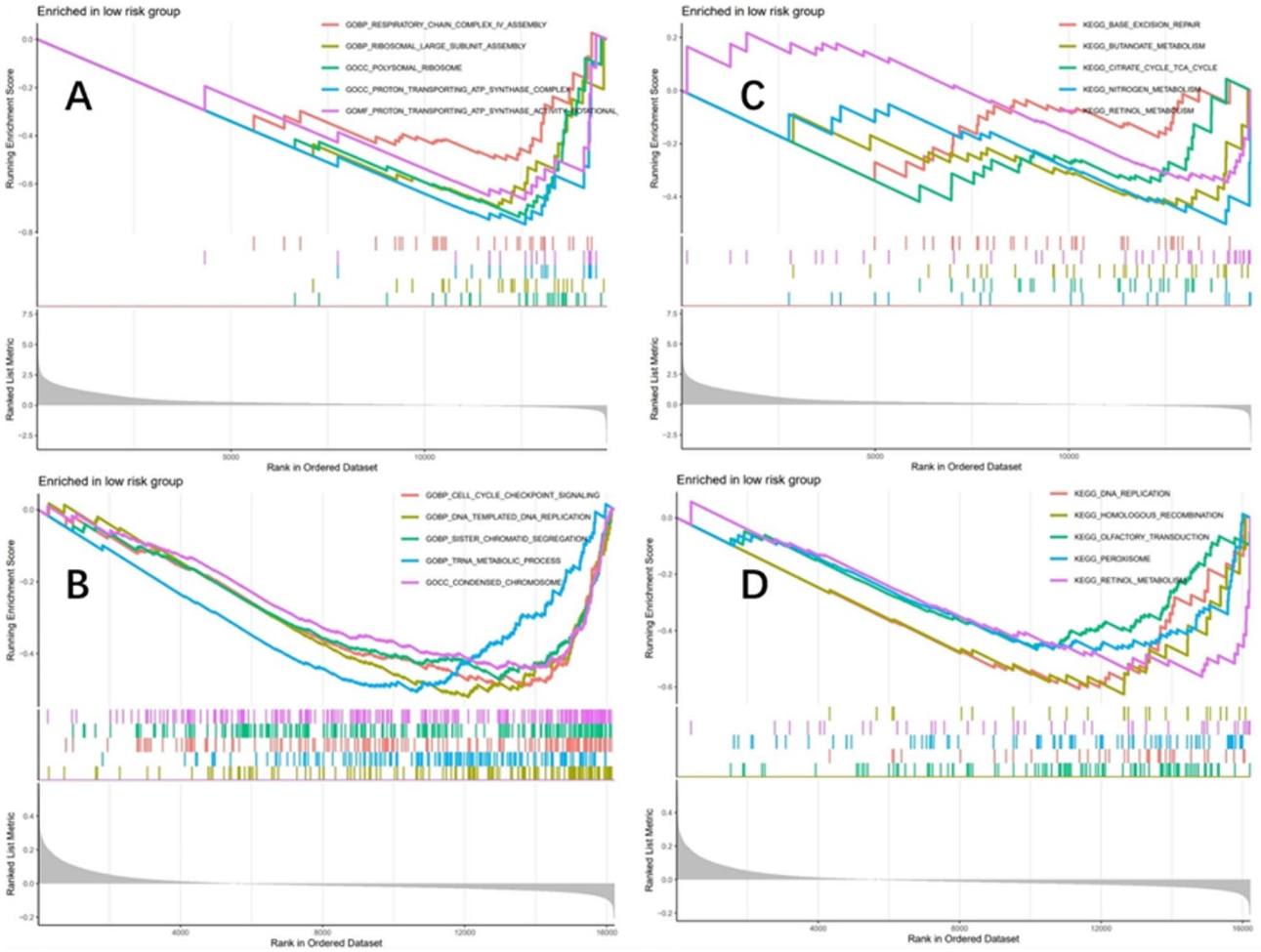

Fig. 7. The TCGA cohort and the GSE39582 cohort have different functional enrichments in the low-risk group. (**A**) TCGA cohort. (**B**) GSE39582 cohort. The TCGA cohort and the GSE39582 cohort have different pathway enrichments in the low-risk group. (**C**) TCGA cohort. (**D**) GSE39582 cohort.

efficacy of immunotherapy in colon cancer, both in the training group TCGA-COAD cohort and the validation group GSE39582 cohort. The results of this study indicate that predicting the efficacy of immunotherapy in colon cancer based on the abundance of CAF infiltration is a feasible approach, which paves the way for the exploration of new immunotherapy molecular markers in colon cancer. The findings also highlight again the role of CAFs as immunosuppressive factors in the immunological microenvironment of colon cancer. Additionally, the prognostic model constructed in this study has good predictive power for immunotherapy, which can help clinicians further screen potential beneficiaries of colon cancer immunotherapy and assist in clinical treatment decision-making.

However, our study also has limitations. Firstly, it is a bioinformatics study based on public databases, and although the results were consistently validated in an independent external dataset, the findings still need to be translated into clinical practice for further validation. Secondly, the model constructed in our study requires genetic sequencing for clinical application, which is not convenient for widespread clinical use. Exploring a method to use immunohistochemistry to detect the relative protein expression levels of the four genes involved in the model construction to predict the efficacy of immunotherapy in colon cancer would be a simpler approach. This is the direction of our research team for the next research phase.

Conclusion

In conclusion, this study constructs a risk model that can well reflect the infiltration abundance of CAF in colon cancer tissue and has good predictive power for immunotherapeutic effects, which is worthy of being promoted to clinical assistance for therapeutic decision-making. The results of this study also suggest that CAF plays the role of an immunosuppressor in the immunological microenvironment of colon cancer. Using the infiltration abundance of CAF to predict the immunotherapeutic effects of colon cancer is a potentially feasible method, which is worth further exploration of its potential molecular mechanisms.

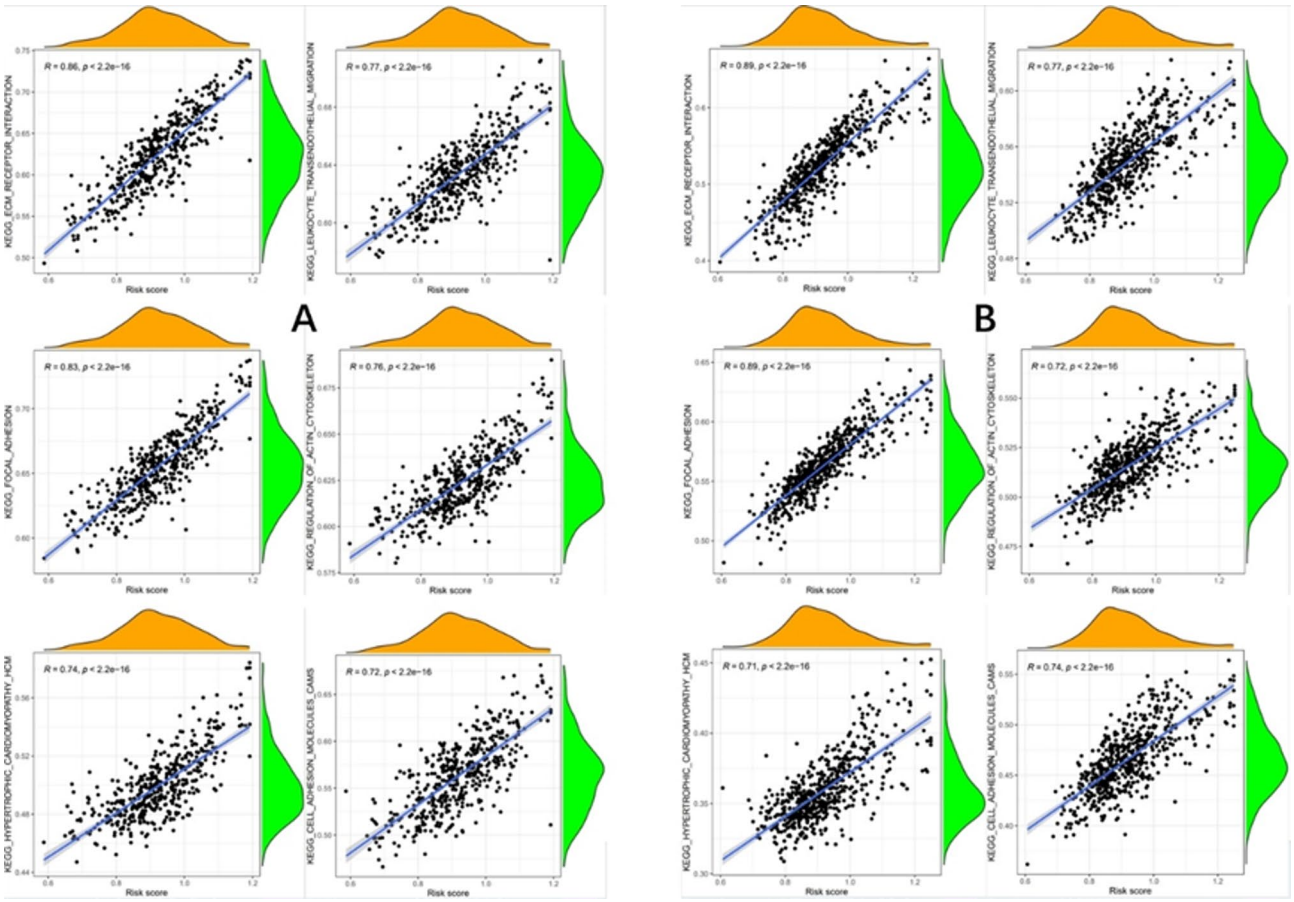

Fig. 8. ssGSEA analysis was performed to identify pathways most correlated with the risk score. (**A**) TCGA cohort. (**B**) GSE39582 cohort.

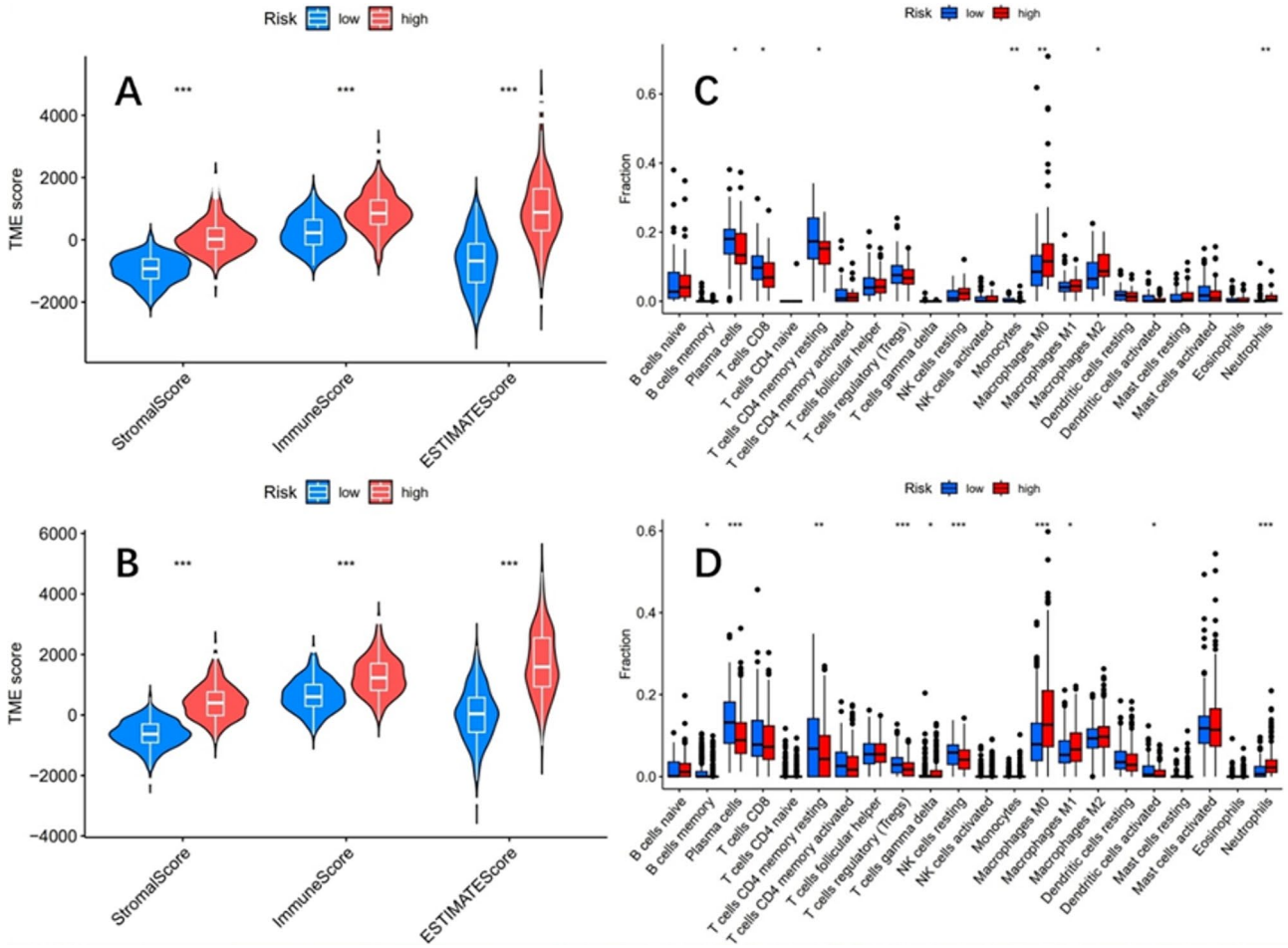

Fig. 9. The ESTIMATE scores of the high-risk and low-risk groups are significantly different. (**A**) TCGA cohort. (**B**) GSE39582 cohort. The CIBERSORT algorithm evaluated the infiltration of 22 types of immune cells in both the high-risk and low-risk groups. (**C**) TCGA cohort. (**D**) GSE39582 cohort.

Data availability

The data that support the findings of this study are openly available inTCGA database: https://portal.gdc.cancer.gov/repositoryGEO database: https://www.ncbi.nlm.nih.gov/geo/.

Received: 17 March 2024; Accepted: 5 May 2025
Published online: 13 May 2025

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71** (3), 209–249 (2021).
2. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* 12 (2023).
3. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* 1 (2023).
4. Vogel, J. D., Eskicioglu, C., Weiser, M. R., Feingold, D. L. & Steele, S. R. The American society of colon and rectal surgeons clinical practice guidelines for the treatment of colon cancer. *Dis. Colon Rectum.* **60** (10), 999–1017 (2017).
5. Taieb, J. & Gallois, C. Adjuvant chemotherapy for stage III colon cancer. *Cancers (Basel)* **12** (9) (2020).
6. Breakthrough of the year 2013. How we did in 2013 and. *Science* **20** (6165), 1442 (2013).
7. Dammeijer, F. et al. The PD-1/PD-L1-checkpoint restrains T cell immunity in tumor-draining lymph nodes. *Cancer Cell* **9** (5), 685–700e688 (2020).
8. André, T. et al. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N Engl. J. Med. Dec.* **3** (23), 2207–2218 (2020).
9. André, T. et al. Nivolumab plus low-dose ipilimumab in previously treated patients with microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: 4-year follow-up from checkmate 142. *Ann. Oncol.* **25** (2022).
10. Le, D. T. et al. Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: KEYNOTE-164. *J. Clin. Oncol. Jan.* **1** (1), 11–19 (2020).
11. Diaz, L. A. Jr. et al. Pembrolizumab versus chemotherapy for microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer (KEYNOTE-177): final analysis of a randomised, open-label, phase 3 study. *Lancet Oncol.* **23** (5), 659–670 (2022).
12. Petitprez, F., Meylan, M., de Reyniès, A., Sautès-Fridman, C. & Fridman, W. H. The tumor microenvironment in the response to immune checkpoint Blockade therapies. *Front. Immunol.* **11**, 784 (2020).

13. Spill, F., Reynolds, D. S., Kamm, R. D. & Zaman, M. H. Impact of the physical microenvironment on tumor progression and metastasis. *Curr. Opin. Biotechnol.* **40**, 41–48 (2016).
14. Turley, S. J., Cremasco, V. & Astarita, J. L. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat. Rev. Immunol. Nov.* **15** (11), 669–682 (2015).
15. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell. Mar.* **20** (3), 309–322 (2012).
16. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Sci. 3*. **348** (6230), 74–80 (2015).
17. Sorokin, L. The impact of the extracellular matrix on inflammation. *Nat. Rev. Immunol. Oct.* **10** (10), 712–723 (2010).
18. Fiori, M. E. et al. Cancer-associated fibroblasts as abettors of tumor progression at the crossroads of EMT and therapy resistance. *Mol. Cancer*. **30** (1), 70 (2019).
19. Gieniec, K. A., Butler, L. M., Worthley, D. L. & Woods, S. L. Cancer-associated fibroblasts-heroes or villains? *Br. J. Cancer*. **121** (4), 293–302 (2019).
20. De Jaeghere, E. A., Denys, H. G. & De Wever, O. Fibroblasts fuel immune escape in the tumor microenvironment. *Trends Cancer* **5** (11), 704–723 (2019).
21. Park, D., Sahai, E., Rullan, A. & SnapShot Cancer-associated fibroblasts. *Cell. Apr*. **16** (2), 486–486e481 (2020).
22. Zhao, Z., Li, T., Sun, L., Yuan, Y. & Zhu, Y. Potential mechanisms of cancer-associated fibroblasts in therapeutic resistance. *Biomed. Pharmacother 1*. **166**, 115425 (2023).
23. Bu, L., Baba, H., Yasuda, T., Uchihara, T. & Ishimoto, T. Functional diversity of cancer-associated fibroblasts in modulating drug resistance. *Cancer Sci.* **111** (10), 3468–3477 (2020).
24. Liu, T. et al. Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J. Hematol. Oncol.* **28** (1), 86 (2019).
25. Mhaidly, R. & Mechta-Grigoriou, F. Fibroblast heterogeneity in tumor micro-environment: role in immunosuppression and new therapies. *Semin Immunol.* **48**, 101417 (2020).
26. Li, T. et al. Colorectal carcinoma-derived fibroblasts modulate natural killer cell phenotype and antitumor cytotoxicity. *Med. Oncol.* **30** (3), 663 (2013).
27. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife Nov*. **13**, 6 (2017).
28. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol. 20*. **17** (1), 218 (2016).
29. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **29**, 9:559 (2008).
30. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16** (5), 284–287 (2012).
31. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **6** (D1), D587–d592 (2023).
32. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
33. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res. 1*. **28** (1), 27–30 (2000).
34. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44** (D1), D457–462 (2016).
35. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41** (Database issue), D955–961 (2013).
36. Maeser, D., Gruener, R. F. & Huang, R. S. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22** (6), 5 (2021).
37. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
38. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*. **12** (5), 453–457 (2015).
39. Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **26** (9927), 1490–1502 (2014).
40. Zhang, Q., Li, J., Shen, L., Li, Y. & Wang, X. Opportunities and challenges of immunotherapy for dMMR/MSI-H colorectal cancer. *Cancer Biol. Med.* 20 (2023).
41. Gong, J., Wang, C., Lee, P. P., Chu, P. & Fakih, M. Response to PD-1 Blockade in microsatellite stable metastatic colorectal cancer harboring a POLE mutation. *J. Natl. Compr. Canc Netw.* **15** (2), 142–147 (2017).
42. Hanley, C. J. & Thomas, G. J. T-cell tumour exclusion and immunotherapy resistance: a role for CAF targeting. *Br. J. Cancer* **123** (9), 1353–1355 (2020).
43. He, L. et al. Identification of four immune subtypes in locally advanced rectal cancer treated with neoadjuvant chemotherapy for predicting the efficacy of subsequent immune checkpoint Blockade. *Front. Immunol.* **13**, 955187 (2022).
44. Zhao, Z. et al. Integrative analysis of cancer-associated fibroblast signature in gastric cancer. *Heliyon* **9** (9), e19217 (2023).
45. Ren, Q. et al. A fibroblast-associated signature predicts prognosis and immunotherapy in esophageal squamous cell cancer. *Front. Immunol.* **14**, 1199040 (2023).
46. Zhang, R. & Liu, F. Cancer-associated fibroblast-derived gene signatures predict radiotherapeutic survival in prostate cancer patients. *J. Transl. Med.* **4** (1), 453 (2022).
47. Ford, K. et al. NOX4 Inhibition potentiates immunotherapy by overcoming cancer-associated fibroblast-mediated CD8 T-cell exclusion from tumors. *Cancer Res.* **80** (9), 1846–1860 (2020).
48. Kieffer, Y. et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov*. **10** (9), 1330–1351 (2020).
49. Dominguez, C. X. et al. Single-cell RNA sequencing reveals stromal evolution into LRRC15(+) myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov.* **10** (2), 232–253 (2020).
50. Zhang, Q., Yang, J., Bai, J. & Ren, J. Reverse of non-small cell lung cancer drug resistance induced by cancer-associated fibroblasts via a paracrine pathway. *Cancer Sci.* **109** (4), 944–955 (2018).
51. Deying, W. et al. CAF-derived HGF promotes cell proliferation and drug resistance by up-regulating the c-Met/PI3K/Akt and GRP78 signalling in ovarian cancer cells. *Biosci. Rep.* **37** (2) (2017).
52. Zhang, H. et al. CAF secreted miR-522 suppresses ferroptosis and promotes acquired chemo-resistance in gastric cancer. *Mol. Cancer* **19** (1), 43 (2020).
53. Zhu, S. et al. CAF-derived Exosomal LncRNA FAL1 promotes chemoresistance to oxaliplatin by regulating autophagy in colorectal cancer. *Dig. Liver Dis.* 1 (2023).
54. Galindo-Pumariño, C. et al. SNAI1-expressing fibroblasts and derived-extracellular matrix as mediators of drug resistance in colorectal cancer patients. *Toxicol. Appl. Pharmacol.* **1** (450), 116171 (2022).
55. De, P., Aske, J. & Dey, N. Cancer-associated fibroblast functions as a road-block in cancer therapy. *Cancers (Basel)* **13** (20) (2021).
56. Ganguly, D. et al. Cancer-associated fibroblasts: versatile players in the tumor microenvironment. *Cancers (Basel)* **12** (9), 17 (2020).

Author contributions

Daoyang Zou was responsible for the bioinformatics analysis, data collating, and manuscript writing of this study, Xi Xin was responsible for data collating and reviewing the manuscript, Huangzhen Xu was responsible

for picture processing and reviewing the manuscript, Yunxian Xu was responsible for picture processing and reviewing the manuscript, Tianwen Xu was responsible for obtaining funding support, designing the project idea, and reviewing the final article.

Funding
This study was supported by the Bureau of Science and Technology of Ganzhou Municipality, No.GZ2021ZSF341; Jiangxi Provincial Administration of Traditional Chinese Medicine, No. SZYYB20224071.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.