

Construction of a novel cancer-associated fibroblast-related signature to predict clinical outcome and immune response in colon adenocarcinoma

Lei Zheng^{1,*}, Jiale Zhang^{2,*}, Yingquan Ye^{1,*}, Zhangpeng Shi^{3,4,5}, Yi Huang¹, Mengmeng Zhang¹, Zhongxuan Gui¹, Ping Li¹, Huanlong Qin^{3,4,5}, Weijie Sun⁶, Mei Zhang¹

¹Department of Integrated Chinese and Western Medicine Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

²Department of Gastrointestinal Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China

³Shanghai Clinical College, Anhui Medical University, Shanghai, China

⁴Department of General Surgery, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China

⁵The Fifth Clinical Medical College of Anhui Medical University, Hefei, China

⁶Department of Infectious Diseases, The First Affiliated Hospital of Anhui Medical University, Hefei, China

*Equal contribution

Correspondence to: Mei Zhang, Weijie Sun, Huanlong Qin; **email:** zhangmei@ahmu.edu.cn; swj950128@163.com, <https://orcid.org/0000-0003-0112-0461>; hlq_eduahmu0611@163.com, <https://orcid.org/0009-0007-4525-9509>

Keywords: colon adenocarcinoma, cancer-associated fibroblasts, prognosis, immunotherapy, machine learning

Received: May 8, 2023

Accepted: August 25, 2023

Published: September 16, 2023

Copyright: © 2023 Zheng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The interaction between the tumour and the surrounding microenvironment determines the malignant biological behaviour of the tumour. Cancer-associated fibroblasts (CAFs) coordinate crosstalk between cancer cells in the tumour immune microenvironment (TIME) and are extensively involved in tumour malignant behaviours, such as immune evasion, invasion and drug resistance. Here, we performed differential and prognostic analyses of genes associated with CAFs and constructed CAF-related signatures (CAFRs) to predict clinical outcomes in individuals with colon adenocarcinoma (COAD) based on machine learning algorithms. The CAFRs were further validated in an external independent cohort, GSE17538. Additionally, Cox regression, receiver operating characteristic (ROC) and clinical correlation analysis were utilised to systematically assess the CAFRs. Moreover, CIBERSORT, single sample Gene Set Enrichment Analysis (ssGSEA) and Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) analysis were utilised to characterise the TIME in patients with COAD. Microsatellite instability (MSI) and tumour mutation burden were also analysed. Furthermore, Gene Set Variation Analysis (GSVA), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) elucidated the biological functions and signalling pathways involved in the CAFRs. Consensus clustering analysis was used for the immunological analysis of patients with COAD. Finally, the pRRophic algorithm was used for sensitivity analysis of common drugs. The CAFRs constructed herein can better predict the prognosis in COAD. The cluster analysis based on the CAFRs can effectively differentiate between immune 'hot' and 'cold' tumours, determine the beneficiaries of immune checkpoint inhibitors (ICIs) and provide insight into individualised treatment for COAD.

INTRODUCTION

Colon adenocarcinoma (COAD) is one of the most widespread malignancies globally, with approximately 1.14 million new cases and 570,000 deaths in 2020 [1]. Current treatment options for COAD include endoscopic resection, surgery, radiotherapy, targeted therapy and immunotherapy [2]. Although early screening and diverse treatment options have significantly improved overall survival in COAD, new cases and deaths from colorectal cancer have been estimated to rise significantly in the next decade [3], adding significantly to the public health challenge. The search for novel biomarkers to improve the clinical outcome of patients with COAD is therefore crucial.

Cancer-associated fibroblasts (CAFs) are key components of the tumour microenvironment (TME) [4, 5], promoting not only the malignant phenotype of cancer but also drug resistance and immune rejection by cancer cells [5, 6]. CAFs play a key role in COAD [7–9], and their consideration as a therapeutic target for cancer has gained widespread attention and recognition [10]. Although satisfactory risk models based on CAFs have been developed to predict prognosis and tumour immune microenvironment (TIME) in individuals with certain cancer types [11–13], they are yet to be implemented for COAD. Therefore, it is significant to construct a satisfactory CAFs-based signature in COAD.

The CAF-related signatures (CAFRs) constructed in the present study are excellent biomarkers for predicting clinical outcomes in individuals with COAD and identifying independent risk factors affecting patient prognosis. Additionally, we explored the biological functions and TIME differences in these CAFRs. Microsatellite instability (MSI) status and tumour mutational burden (TMB) were investigated, and a consensus clustering analysis for CAFRs was performed in patients with COAD. The different clusters effectively differentiated patients' TIME characteristics, which not only helps to distinguish immune 'hot' and 'cold' tumours and guides immune checkpoint inhibitors (ICIs) administration but also provides potentially valuable individualised treatment options for patients with cancer.

MATERIALS AND METHODS

Data collection

Transcriptome profiling data, simple nucleotide variation data and clinical parameters of individuals with COAD were downloaded from The Cancer Genome Atlas (TCGA) repository (<https://portal.gdc.cancer.gov/repository>). The downloaded data were collated for follow-up studies using Perl scripts.

Transcriptome and corresponding clinical information from the GSE17538 cohort were downloaded from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/>). Cases in the TCGA and GEO cohorts containing both transcriptomic data and survival data were included in the follow-up study. Immunohistochemical images of CAF-related genes were downloaded from the Human Protein Atlas (HPA, version: 22.0) (<https://www.proteinatlas.org>) [14]. Specific links to all immunohistochemistry images from the Human Protein Atlas used in this study are provided in Supplementary Table 1. The CAF-related gene set was obtained from The Human Gene Database (<https://www.genecards.org/>) [13].

Identification of CAF-related genes in COAD

The mRNA expression matrix of CAF-related genes in the TCGA-COAD cohort was extracted using R (version 4.2.2), and differentially expressed genes (DEGs) between tumour and normal tissues were further identified (fold change (FC) > 1.5, false discovery rate (FDR) < 0.05). The R package 'pheatmap' was utilised to map differential gene volcanoes and mRNA expression heatmaps. Subsequently, the packages 'limma', 'sva' were utilised to obtain the expression data of the DEGs in the TCGA and GEO cohorts and analyse the intersection of the DEGs expression matrix of the two datasets, respectively. The 'survival' and 'survminer' packages performed univariate Cox analysis to obtain the prognosis-related CAF-related genes in the TCGA cohort and draw a forest plot ($P < 0.05$), respectively.

Establishment of CAFRs in COAD

The 'glmnet', and 'survival' packages were utilised to establish CAFRs in COAD. The optimal prognostic genes in the TCGA cohort were screened using univariate regression and least absolute shrinkage and selection operator (LASSO) algorithms and the resultant genes were utilised to construct CAFRs. The risk score of each sample was obtained through the expression of the CAFRs-related genes and the corresponding regression coefficient. The risk equation used was as follows: Risk score = $\sum_{i=1}^n \text{Coef}(i) \times \text{Expr}(i)$. $\text{Coef}(i)$ and $\text{Expr}(i)$ represent the regression coefficient and expression values for each gene in CAFRs, respectively. All individuals were classified into high- and low-risk subgroups based on the median risk score in the TCGA cohort.

Validation of the CAFRs in COAD

The 'pheatmap', 'survival' and 'survminer' packages were utilised to plot risk heatmaps, risk curves, survival

status maps and Kaplan–Meier (K-M) curves for individuals in the TCGA and GEO cohorts. Cox regression evaluated risk scores and clinicopathological parameters to identify independent prognostic variables in the TCGA and GEO cohorts. Additionally, receiver operating characteristic (ROC) curves were drawn utilising the ‘survminer’, ‘survival’ and ‘timeROC’ packages to evaluate the prognostic value of the developed CAFRs based on the size of the area under the curves in the TCGA and GEO cohorts.

Correlation analysis of the CAFRs with clinical parameters in COAD

To stratify and validate the CAFRs, we further divided patients into two groups based on age, gender and tumour stage. Survival differences between high- and low-risk groups across clinical subgroups were analysed using K-M curves to determine the applicability of the constructed CAFRs to the different subgroups of patients with COAD having different clinical parameters. Finally, the ‘ComplexHeatmap’ was utilised to create a heatmap of the status of different clinical parameters in the two risk subgroups.

Correlation analysis of the CAFRs and the TIME in COAD

CIBERSORT, an algorithm, implements a machine learning approach for the high-throughput characterisation of different cell types, such as tumour-infiltrating immune cells (TIICs) [15]. The fraction of 22 TIICs was determined using ‘limma’, ‘CIBERSORT’, ‘preprocessCore’, ‘e1071’ and ‘parallel’ and the differences in TIICs between the two subgroups were further analysed. Gene set enrichment analysis (GSEA) enables the enrichment analysis of gene sets with physiological regulatory roles and biological effects [16, 17]. The single sample GSEA (ssGSEA) was performed utilising the ‘GSEABase’ and ‘GSVA’ to estimate immune cell and immune function scores for each sample. Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) analysis is an expression-based tumour purity determination algorithm [18]. Here, the ‘ESTIMATE’ package was utilised to calculate stromal scores and immune scores in the tumour tissue. Subsequently, the ‘ggpubr’ package was employed to draw box plots of stromal, immune and ESTIMATE scores in the risk subgroups.

Correlation analysis of the CAFRs with MSI and TMB

Genomic hypermutability leads to a molecular tumour phenotype known as MSI [19]. Studies suggest that

MSI has the potential as a viable biomarker for ICIs therapy [20]. The ‘ggplot2’, ‘ggpubr’ and ‘plyr’ were utilised to analyse the proportions of microsatellite-stable (MSS), MSI-High and MSI-Low phenotypes in the different risk groups and plot percentage histograms. Additionally, TMB is defined as the total number of somatic mutations per million bases [21] and is used as a biomarker of response to treatment with ICIs in certain solid tumours [22–24]. We analysed TMB levels in the risk groups and plotted box plots.

GSVA and gene ontology (GO) analysis

The Gene Set Variation Analysis (GSVA) is an algorithm utilised to detect differences in pathway activity among sample populations [25]. GSVA was conducted to obtain the enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the two risk subgroups, and the correlation between KEGG pathways and signature gene expression was analysed. These analyses were implemented using the R ‘limma’, ‘pheatmap’, ‘GSEABase’, ‘reshape2’, ‘ggplot2’ and ‘GSVA’ packages. Additionally, the DEGs ($FC > 2$ and $FDR < 0.05$) between risk groups were determined using ‘limma’. Furthermore, the ‘org.Hs.eg.db’, ‘ggplot2’, ‘enrich’, ‘GOplot’ and ‘clusterProfiler’ were utilised to perform GO and KEGG analysis of DEGs between the risk groups and explore the enrichment of DEGs in cell component, molecular function and biological processes.

Consensus clustering analysis

The package ‘ConsensusClusterPlus’ was utilised to cluster the COAD samples of the TCGA queue according to the established prognostic characteristics. The packages ‘ggplot2’ and ‘Rtsne’ were utilised for principal component analysis (PCA). The relationship between different COAD clusters and patient survival and TIME was further studied using K-M curves, ESTIMATE, MSI and ssGSEA. Additionally, ‘limma’, ‘reshape2’, ‘ggplot2’ and ‘ggpubr’ were utilised to determine the expression of genes related to immune checkpoints in different clusters, and differential box plots were drawn for immune checkpoints with significant differences ($P < 0.05$).

Analysis of clinical therapeutic drug sensitivity

The packages ‘pRRophic’ and ‘ggpubr’ were used to obtain the half maximum inhibitory concentration (IC50) of various drugs in the different clusters and draw a differential box chart for the various drugs ($P < 0.001$). They were also used to explore the potential clinical significance of cluster analysis based on the CAFRs in drug treatment.

COAD tissue samples

Colon tumor tissues and adjacent normal tissues were acquired from the First Affiliated Hospital of Anhui Medical University (Hefei, China). All colon tumors were histologically confirmed as COAD. The study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Anhui Medical University (No. PJ20230861). All enrolled COAD patients provided written informed consent.

Quantitative real-time PCR

RT-qPCR was used to measure the mRNA level in tumor tissues and adjacent normal tissues. RNA extraction and RT-qPCR was performed as previously described [26]. Briefly, the total RNA was isolated using RNA isolation reagent (Takara Bio, Japan), and reverse transcribed into cDNA with PrimeScript™ RT Master Mix (Takara Bio, Japan) following the manufacturer's protocol. Quantitative real-time PCR (qPCR) was performed using SYBR-Green qPCR Master Mix (Vazyme Bio, China). The primer sequences for the CAFRs-related genes used in the experiments were listed in Supplementary Table 2. The GAPDH was used as an internal control for normalization. Relative gene expression was estimated according to the $2^{-\Delta\Delta Ct}$ method.

Statistical analysis

All statistical analyses were performed using R software (version 4.1.2) and the corresponding R packages. K-M method was utilised to plot the survival curves of different subgroups. The correlation between different continuous variables was assessed by Pearson correlation test. The Wilcoxon test was utilised for comparing two groups. $P < 0.05$ was considered as statistically significant for a difference.

Data availability statement

All data presented in this study are available from the corresponding author upon reasonable request.

RESULTS

Identification of CAF-related genes in COAD

The study flow chart is illustrated in Figure 1. A total of 473 COAD tumour samples and 41 normal samples were acquired from the TCGA database with relevant data. Overall, 431 CAF-related genes were acquired from the Genecards, all with relevance scores greater than 5 (Appendix 1). A total of 244 CAF-related genes were differentially expressed in COAD tumours and

normal tissues, of which 172 were upregulated and the remaining downregulated (Figure 2A). The Cox regression indicated that 16 CAF-related genes were associated with the overall survival (OS) of COAD (Figure 2B). The expression patterns of the 50 CAF-related genes with the highest up- and down-regulation folds among the DEGs are presented as a heat map (Figure 2C).

Construction of the CAFRs in COAD

To avoid overfitting, the LASSO algorithm was utilised (Figure 3A, 3B), identifying 15 CAF-related genes for CAFRs construction (Table 1). The K-M curves of 15 signature-related genes in the TCGA-COAD cohort further confirmed the relationship between the expression of these genes and the survival of patients with COAD (Supplementary Figure 1). Additionally, box line plots demonstrate the differential expression status of CAFRs-related genes in COAD tumor tissues and normal tissues (Supplementary Figure 2).

The risk score for each case was obtained by calculating the risk regression coefficient and expression of each gene in the CAFRs. Risk score = $FGF9 \times (0.0787497) + PCAT6 \times (0.3741071) + CD36 \times (0.1114249) + TIMP1 \times (0.3006318) + TERT \times (0.5012208) + CDKN2A \times (0.1204453) + CYP19A1 \times (0.5903825) + IL13 \times (-1.4691038) + SNAI1 \times (0.1950326) + BDNF \times (0.4283392) + GPC1 \times (0.0962728) + NRG1 \times (-0.4469785) + SERPINH1 \times (-0.2432318) + AGER \times (0.0734801) + ENO2 \times (0.0958250)$.

We further validated the CAFRs in the TCGA and GSE17538 cohorts. The expression status of the 15 signature genes in the two cohorts is shown in heat maps (Figure 3C, 3E). K-M analyses of the two cohorts revealed a significantly poorer clinical outcome for individuals with COAD in the high-risk subgroup (Figure 3D, 3F). In addition, immunohistochemical images of the HPA database showed the expression of proteins encoded by some of the signature-related genes in COAD normal and tumor tissues (Figure 4A, 4B).

Assessment of the CAFRs in COAD

Risk scores based on the CAFRs were identified as an independent prognostic indicator for the TCGA-COAD cohort using univariate and multivariate Cox regression, with hazard ratio values of 3.014 (2.240–4.055; $P < 0.001$) and 2.716 (1.966–3.752; $P < 0.001$) (Figure 5A, 5B). The tumour stage was also an independent factor ($P < 0.001$). The ROC curves were utilised to evaluate the specificity and sensitivity of the CAFRs for COAD prognosis. The area under the curve values for the CAFRs predicting OS at 1-, 3- and 5- years were

Table 1. LASSO regression coefficients for CAFs.

CAFs-related genes	Coefficient
FGF9	0.0787497
PCAT6	0.3741071
CD36	0.1114249
TIMP1	0.3006318
TERT	0.5012208
CDKN2A	0.1204453
CYP19A1	0.5903825
IL13	-1.4691038
SNAI1	0.1950326
BDNF	0.4283392
GPC1	0.0962728
NRG1	-0.4469785
SERPINH1	-0.2432318
AGER	0.0734801
ENO2	0.0958250

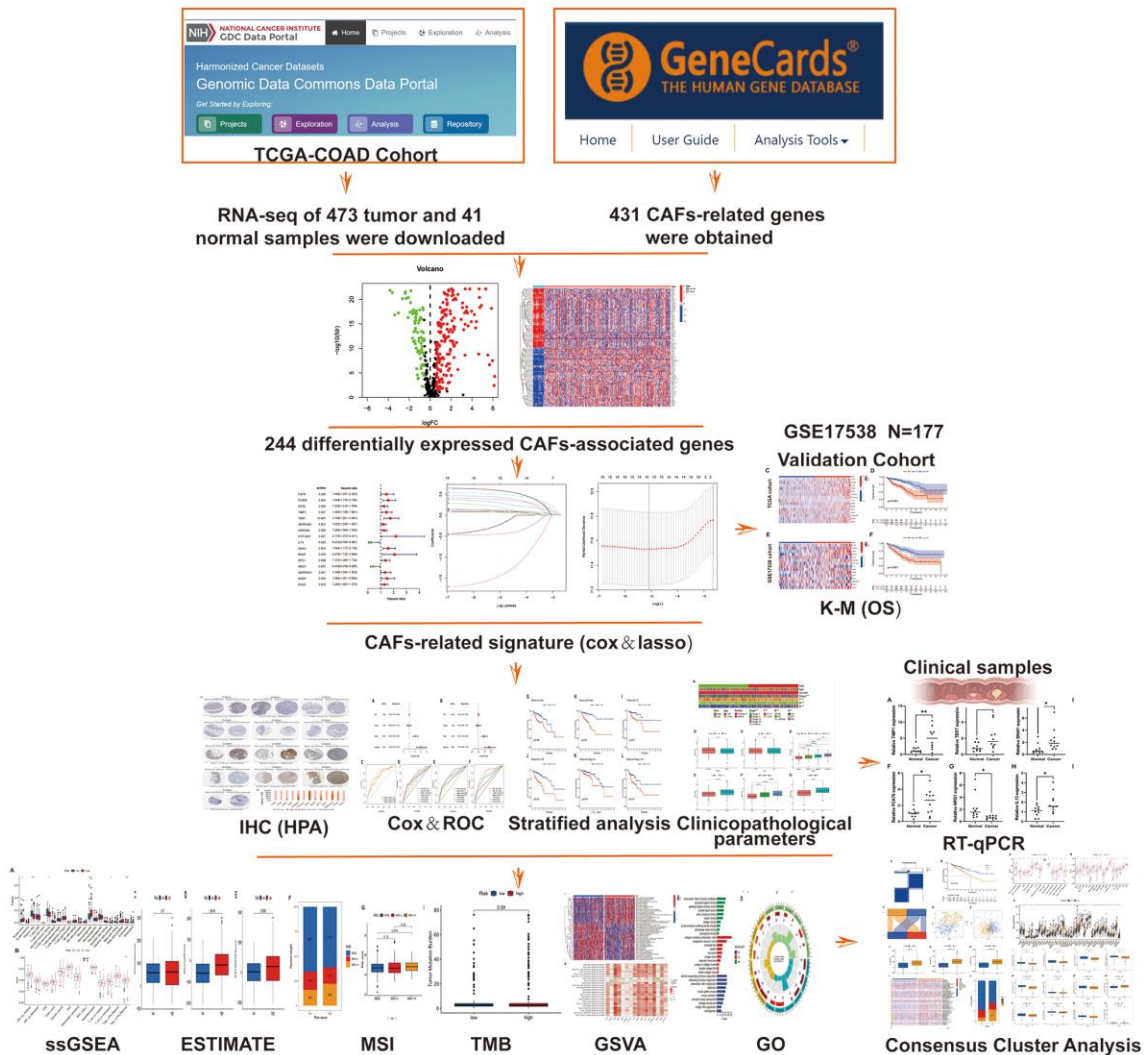


Figure 1. Flow chart.

0.711, 0.749 and 0.788 (Figure 5C–5F). Additionally, Cox regression analysis and ROC curves of the GEO validation cohort further validated that CAFRs is an independent prognostic factor for COAD with good prognostic predictive efficacy (Figure 5G–5L).

Correlation of the CAFRs with clinical parameters in COAD

We further analysed the correlation of the CAFRs with the clinical parameters. The heat map shows the status

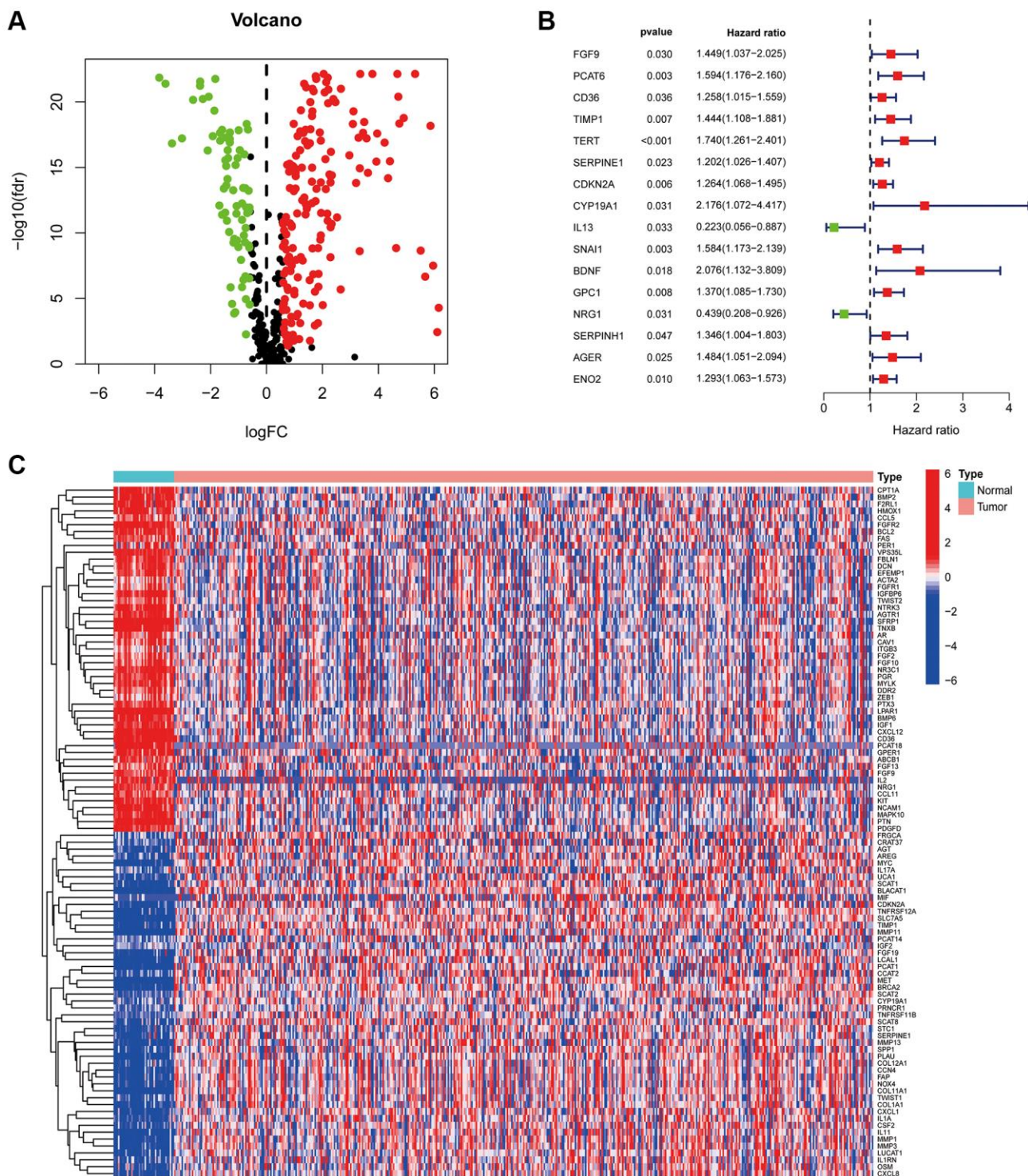


Figure 2. CAF-related genes in COAD. (A) The volcano plot of 244 CAFs-associated genes shows differential expression. (B) The risk forest plot showed that 16 CAF-related genes were associated with COAD prognosis (C) Heat map of differentially expressed CAF-related genes.

of different clinical factors in the risk subgroups (Figure 6A). To further analyse whether the CAFRs apply to individuals with different clinicopathological factors, survival analyses were performed on different clinical subgroups of patients. Stratified K-M curves indicated

that individuals with different gender, age and tumour stages had worse prognoses in the high-risk group (Figure 6B–6G), demonstrating the stability of the CAFRs. Additionally, box plots of risk scores for different clinical subgroups showed that as the tumour

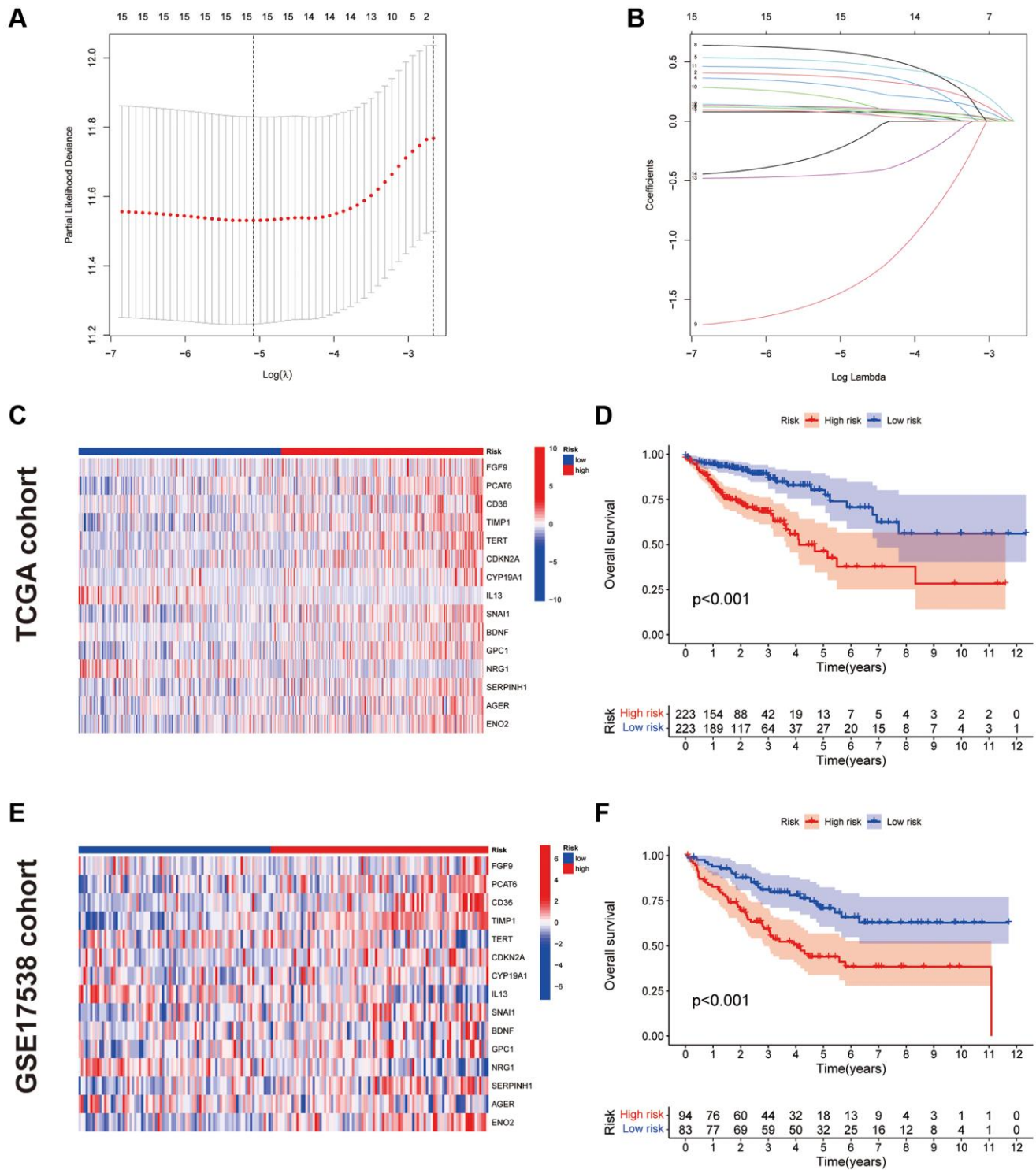


Figure 3. Construction and validation of CAFRs. (A, B) The coefficient and partial likelihood deviance of the prognostic signature. (C) Heat map of the expression of the 15 CAFs-associated genes in the TCGA cohort. (D) K–M curve for OS in the TCGA cohort. (E) Heat map of the expression of the 15 CAFs-associated genes in the GEO cohort. (F) K–M curve for OS in the GEO cohort.

stage increased (stages I-IV), the risk score also increased significantly. Moreover, the same results were observed for the T-, N- and M-stage. However, the risk scores did not differ significantly in the age and gender subgroups (Figure 6H–6M).

Correlation of the CAFs with TIME in COAD

CIBERSORT algorithm revealed that naive B cells, plasma cells, resting CD4+ T cells, M0 macrophages, activated dendritic cells and eosinophils differed

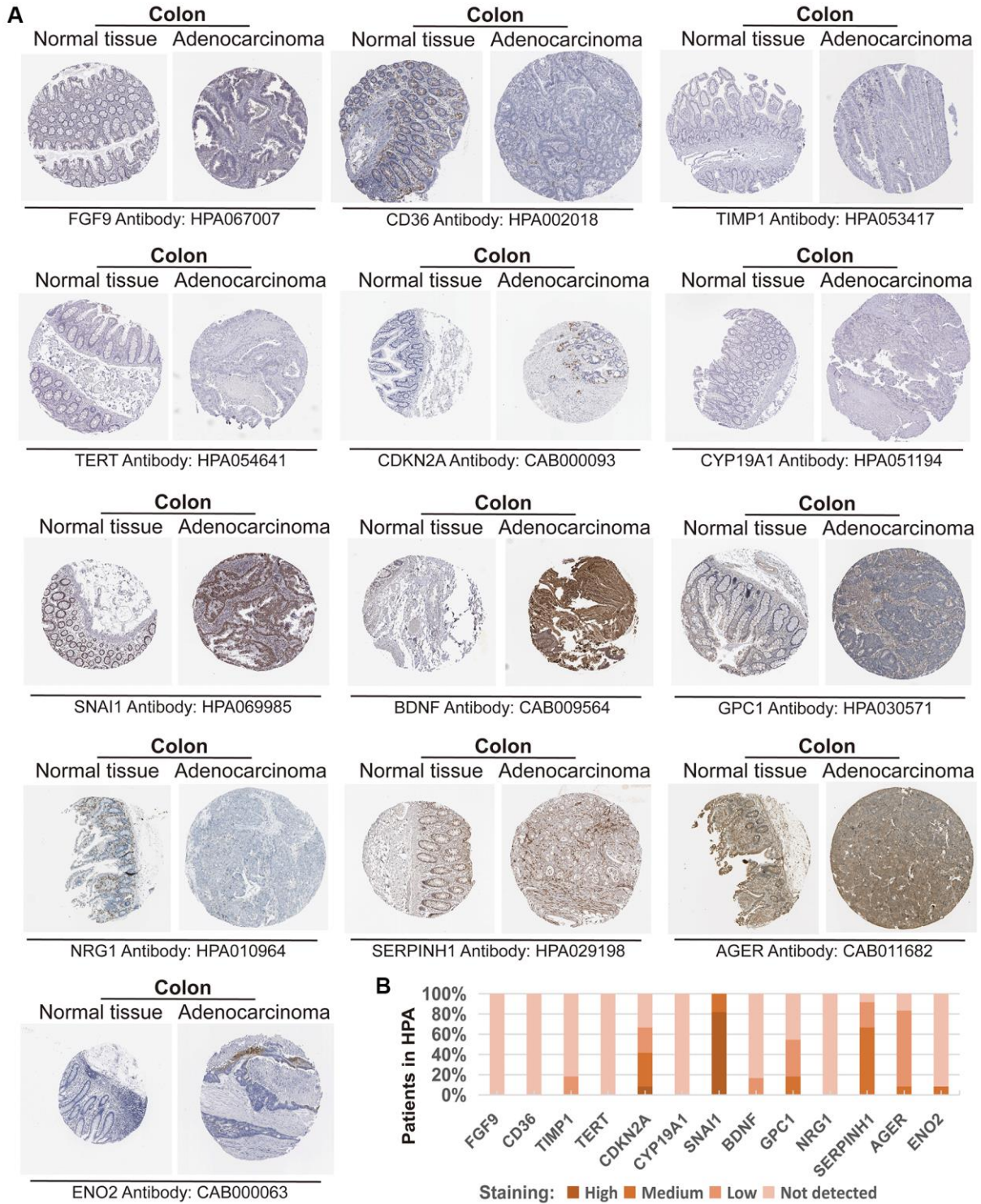


Figure 4. Expression of the protein encoded by the CAFs-related gene in the Human Protein Atlas (HPA). (A) Immunohistochemical images of the protein encoded by some CAFs-related genes in COAD normal and tumour tissue in the HPA. (B) The proportion of the protein encoded by some CAFs-related genes that is expressed in the COAD of the HPA.

significantly between the high- and low-risk subgroups (Figure 7A). However, most of the other immune infiltrating cells did not differ significantly in the risk groups. ssGSEA also showed no significant difference

in most immune-related functions between the high- and low-risk groups, with the exception of type II IFN response (Figure 7B). ESTIMATE analysis revealed that the stromal and ESTIMATE scores were

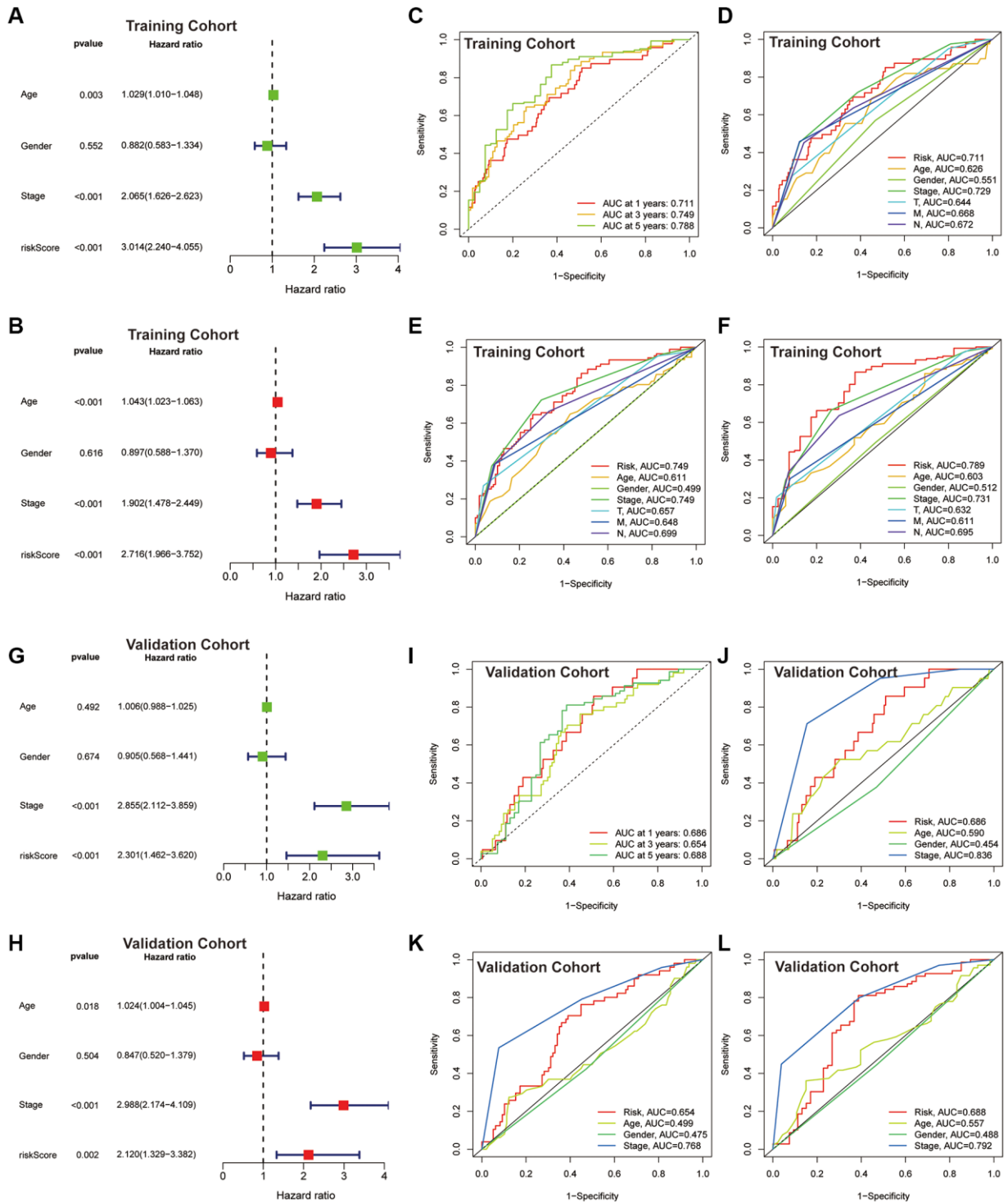


Figure 5. Assessment of the CAFrs. (A, B) Forest plot for univariate and multivariate Cox regression analyses in the TCGA-COAD cohort. (C) ROC curves of 1-, 3- and 5-year survival for the CAFrs in the TCGA-COAD cohort. (D-F) Comparison of the prediction accuracy of the CAFrs with age, gender, TNM-stage, T-stage, N-stage and M-stage at 1-, 3- and 5 years in the TCGA-COAD cohort. (G, H) Forest plot for univariate and multivariate Cox regression analyses in the GEO cohort. (I) ROC curves of 1-, 3- and 5-year survival for the CAFrs in the GEO cohort. (J-L) Comparison of the prediction accuracy of the CAFrs with age, gender and stage in the GEO cohort.

significantly higher in the high-risk group but the immune score did not differ between the two subgroups (Figure 7C–7E).

Correlation of the CAFRs with MSI and TMB in COAD

MSI status closely correlates with immunotherapy response in gastroenterology tumours. The histogram of proportions shows that the proportions of MSS, MSI-L and MSI-H in the low-risk subgroup were 66%, 18% and 16%, respectively, while the values were 61%, 17% and 22%, respectively, in the high-risk subgroup

(Figure 7F). Furthermore, there was no significant difference in the risk scores of individuals in the MSS, MSI-L and MSI-H groups (Figure 7G). Additionally, there was also no significant difference in TMB status between the risk groups (Figure 7H).

GSA and GO analysis of the CAFRs in COAD

GSA investigated the biological differences between the risk groups and revealed that the high-risk subgroup was enriched in pathways such as circadian rhythm, Notch signalling pathway, MAPK signalling pathway, actin cytoskeleton regulation, calcium signalling

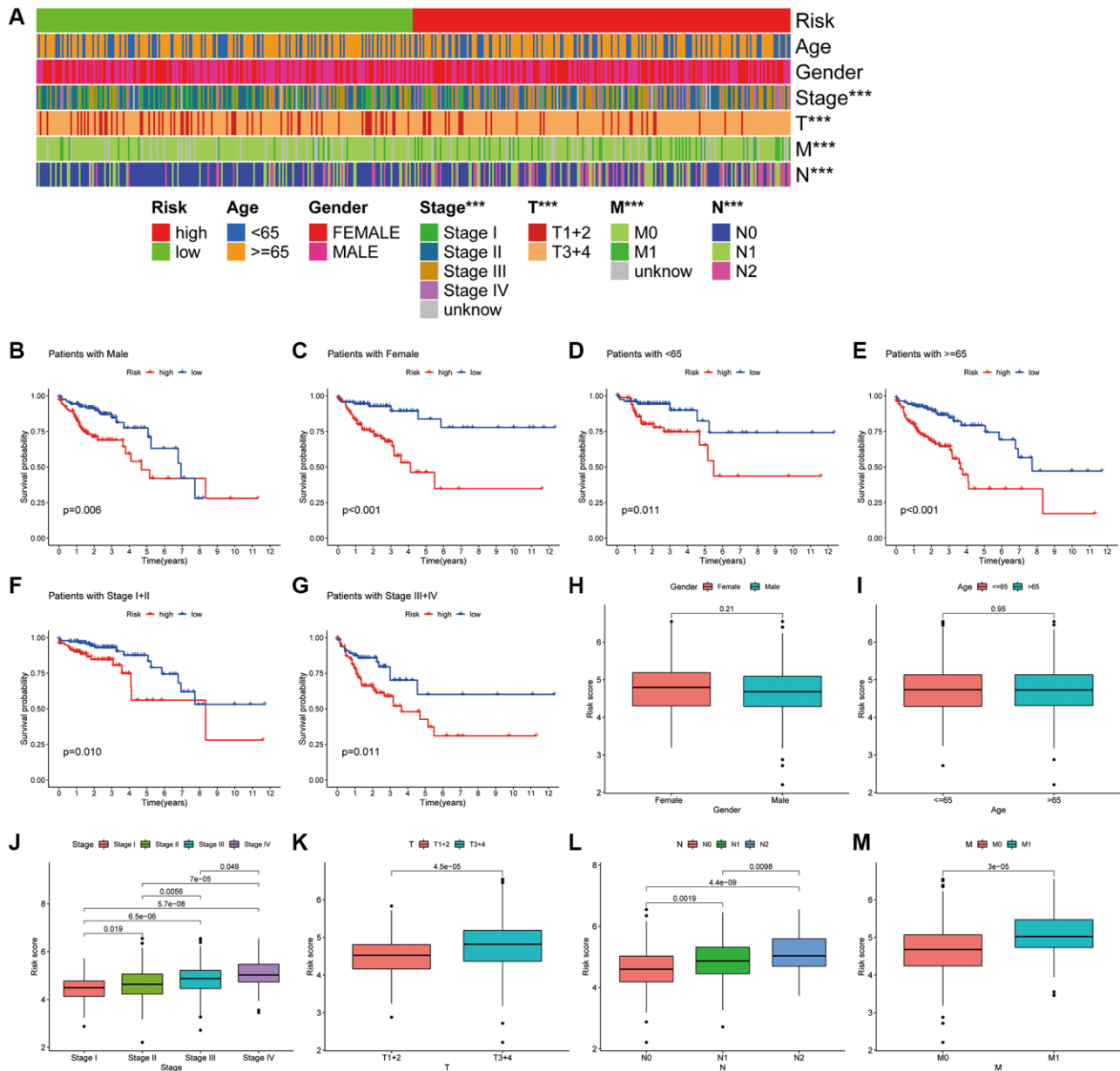


Figure 6. Association of CAFRs with clinicopathological parameters in COAD. (A) A strip chart of the associations between risk status and clinical parameters. (B–G) K–M curves of low- and high-risk subgroups sorted by gender, age and TNM stage. (H–M) Box plot of the difference in risk scores by gender, age, TNM-stage, T-stage, N-stage and M stage.

pathway, extracellular matrix receptor interactions, basal cell carcinoma and Hedgehog signalling pathway, which are associated with tumour malignancies. Additionally, the metabolism of retinol; toxic metabolism of cytochrome P450; interconversion of

pentose and glucuronide, ascorbic acid and aldehyde; metabolism of drugs; metabolism of glutathione; metabolism of fatty acids; mismatch repair and DNA replication were enriched in the low-risk subgroup (Figure 8A). Furthermore, spearman correlation

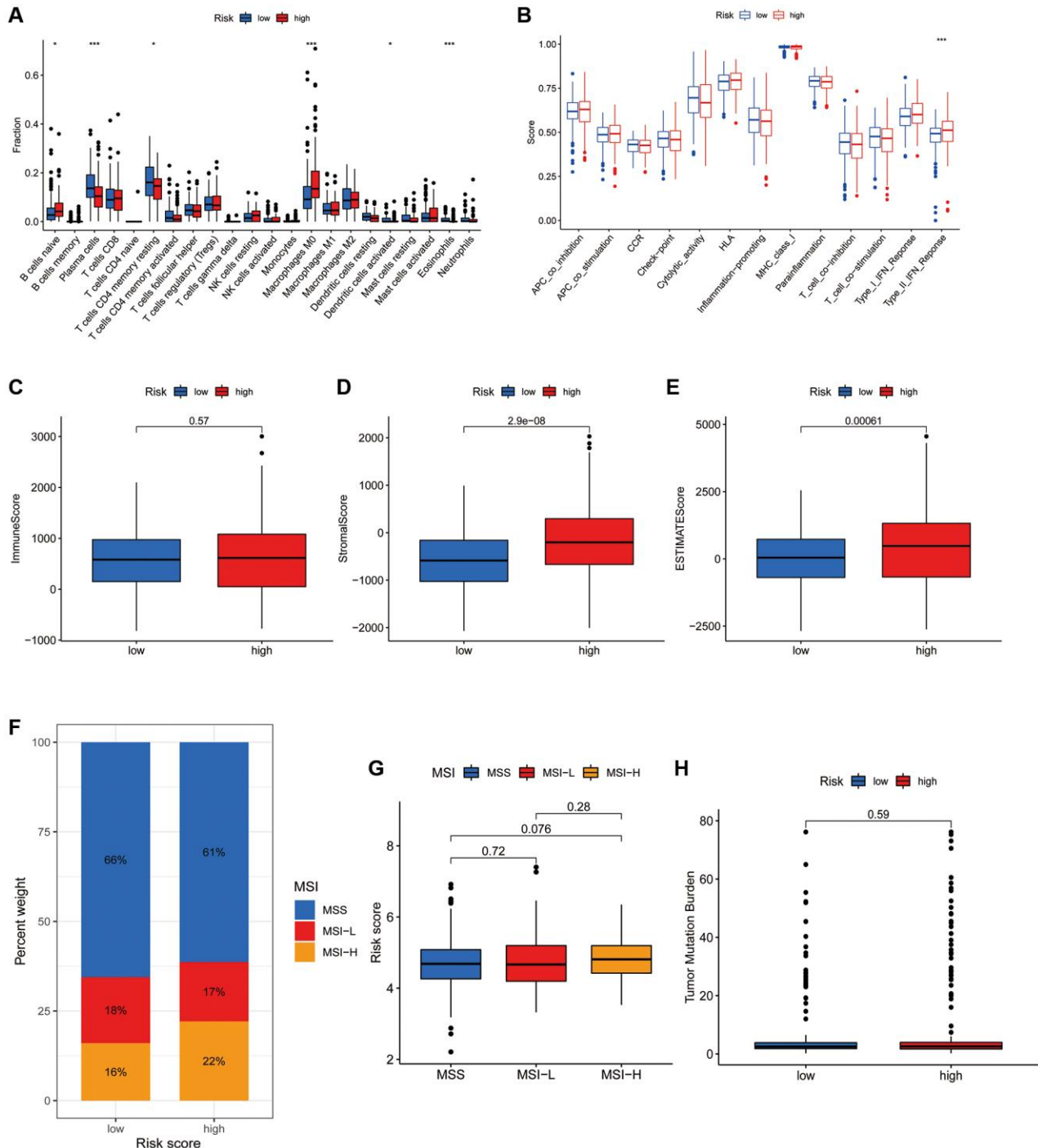


Figure 7. Association of the CRFRs with the immune microenvironment of COAD. (A) Box plot showing differences in immune cells between the high- and low-risk subgroups using the CIBERSORT algorithm. (B) Box plot showing differences in immune-related functions between the high- and low-risk groups using the ssGSEA algorithm. (C–E) Stromal score, immunity score and ESTIMATE score in the two risk subgroups. (F) Histogram of proportions showing the proportion of patients with MSS, MSI-L and MSI-H in the high- and low-risk subgroups. (G) Box plot of differences in risk scores for patients in the MSS, MSI-L and MSI-H subgroups. (H) Box plot of TMB difference for the high- and low-risk subgroups. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

analysis showed a strong correlation between the expression of the 15 genes in the CAFs and signalling pathways related to tumour evolution (Figure 8B).

Additionally, we investigated the biological functions of DEGs in the different risk groups. In terms of biological processes, the DEGs were enriched in extracellular matrix structural constituent, signalling receptor activator activity, glycosaminoglycan binding, receptor-

ligand activity, sulfur compound binding and extracellular matrix binding. Regarding molecular function, the DEGs were enriched in the external encapsulating structure organisation, extracellular structure organisation, extracellular matrix organisation, ossification, connective tissue development and other functions. Furthermore, the DEGs were enriched in cellular components such as the endoplasmic reticulum lumen, contractile fibre and myofibril (Figure 8C).

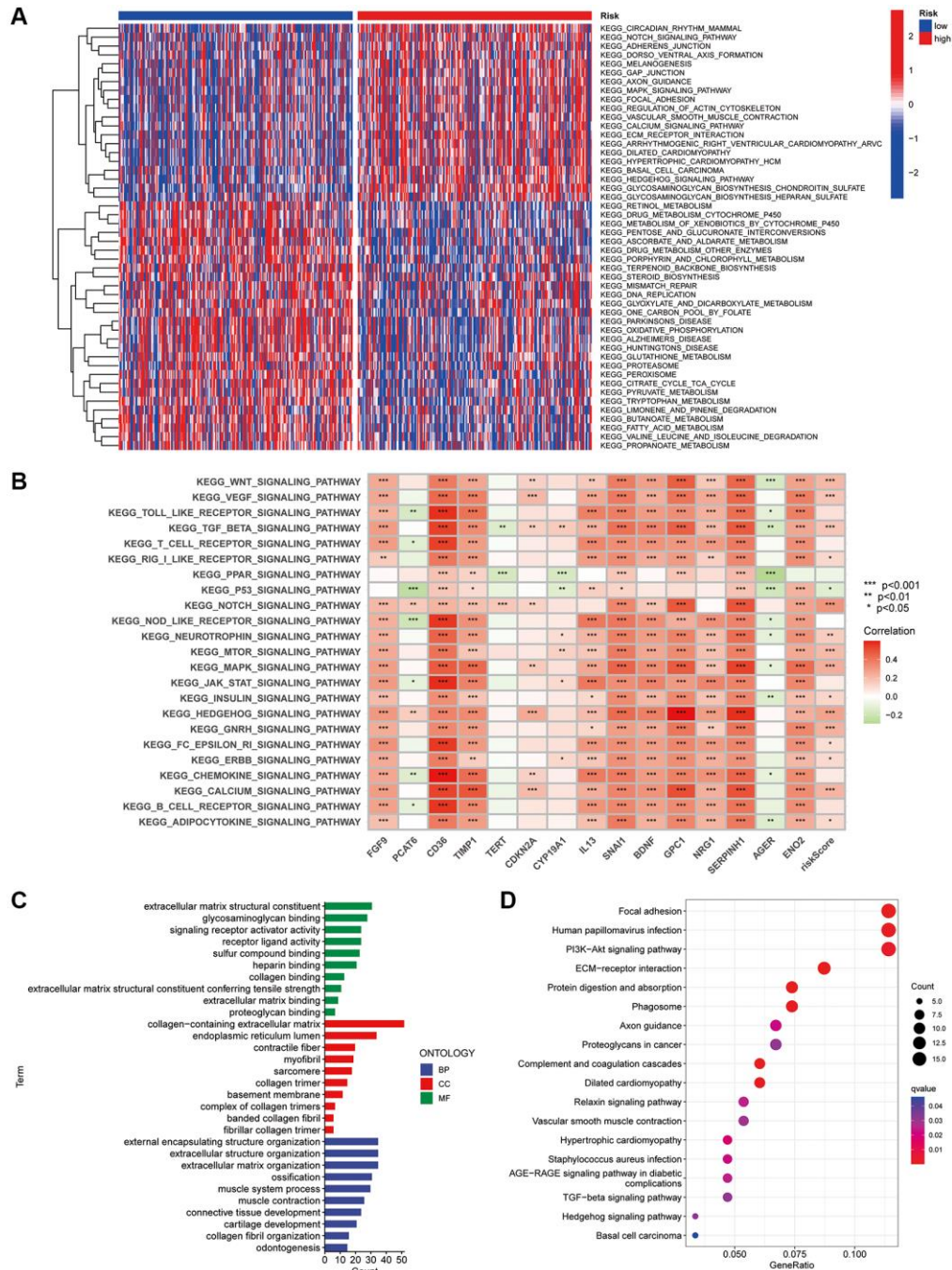


Figure 8. GSEA and GO analysis. (A) Heat map of functional pathway enrichment differences between the two risk groups. (B) Heat map of the correlation between the expression of signature genes and signalling pathways. (C) GO analysis shows the enrichment of DEGs between the high- and low-risk subgroups. (D) KEGG analysis shows the enrichment of DEGs between the two risk subgroups.

Finally, KEGG analysis revealed that DEGs were enriched in pathways including focal adhesion, PI3K-Akt signaling pathway, ECM-receptor interactions, and protein digestion and absorption (Figure 8D).

Consensus clustering based on the CAFRs

There is growing evidence that tumour subgroups derived from consensus clustering analysis have different TIME landscapes and influence the response to tumour immunotherapy [27, 28]. All patients in the TCGA-COAD cohort were divided into k ($k = 2-9$) clusters using ConsensusClusterPlus. According to the

cumulative distribution function curve of the consensus scores, the best classification occurs when $k = 2$. Therefore, all patients in the TCGA-COAD cohort were classified into cluster 1 ($n = 223$) and cluster 2 ($n = 223$), when the variability was lowest within clusters and highest between clusters (Figure 9A–9D). The K-M curves revealed that individuals in cluster 2 have worse survival rates than those in cluster 1 ($P = 0.003$) (Figure 9E). The Sankey plots revealed that the majority of individuals in cluster 1 were in the low-risk group, while the majority of patients in cluster 2 were in the high-risk group (Figure 9F). The findings indicate that the cluster typing developed can help determine the

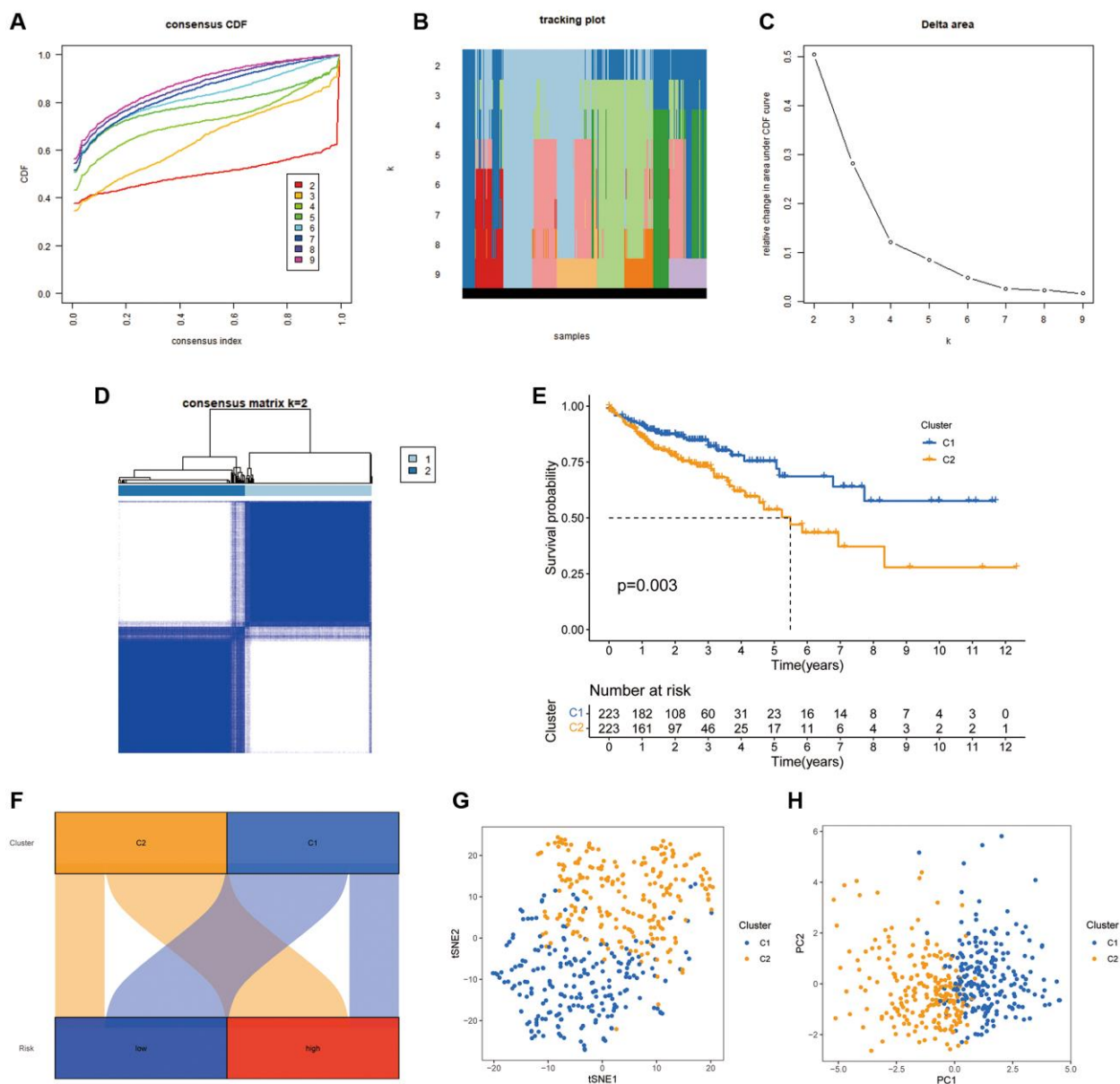


Figure 9. COAD classification based on the CAFRs. (A) The cumulative distribution function curves for $k = 2-9$. (B) The tracking plot of consistent clustering. (C) The elbow plot showing relative change in area under the cumulative distribution function curve. (D) Consensus clustering matrix for $k = 2$. (E) K-M curves of the two clusters. (F) Sankey diagram of the association between the risk groups and clusters. (G, H) PCA and tSNE analyses of the clusters.

prognosis of patients with COAD. Additionally, the PCA and tSNE significantly distinguished the distributional features of the two clusters (Figure 9G, 9H).

We further explored the impact of cluster analysis on the TIME of COAD tumours using ESTIMATE

analysis, which revealed that the immune, stromal and ESTIMATE scores were significantly higher in cluster 2 (Figure 10A–10C). The heat map showed that the majority of immune infiltrating cells were significantly less abundant in cluster 1 than in cluster 2 (Figure 10D). Additionally, ssGSEA validated these findings,

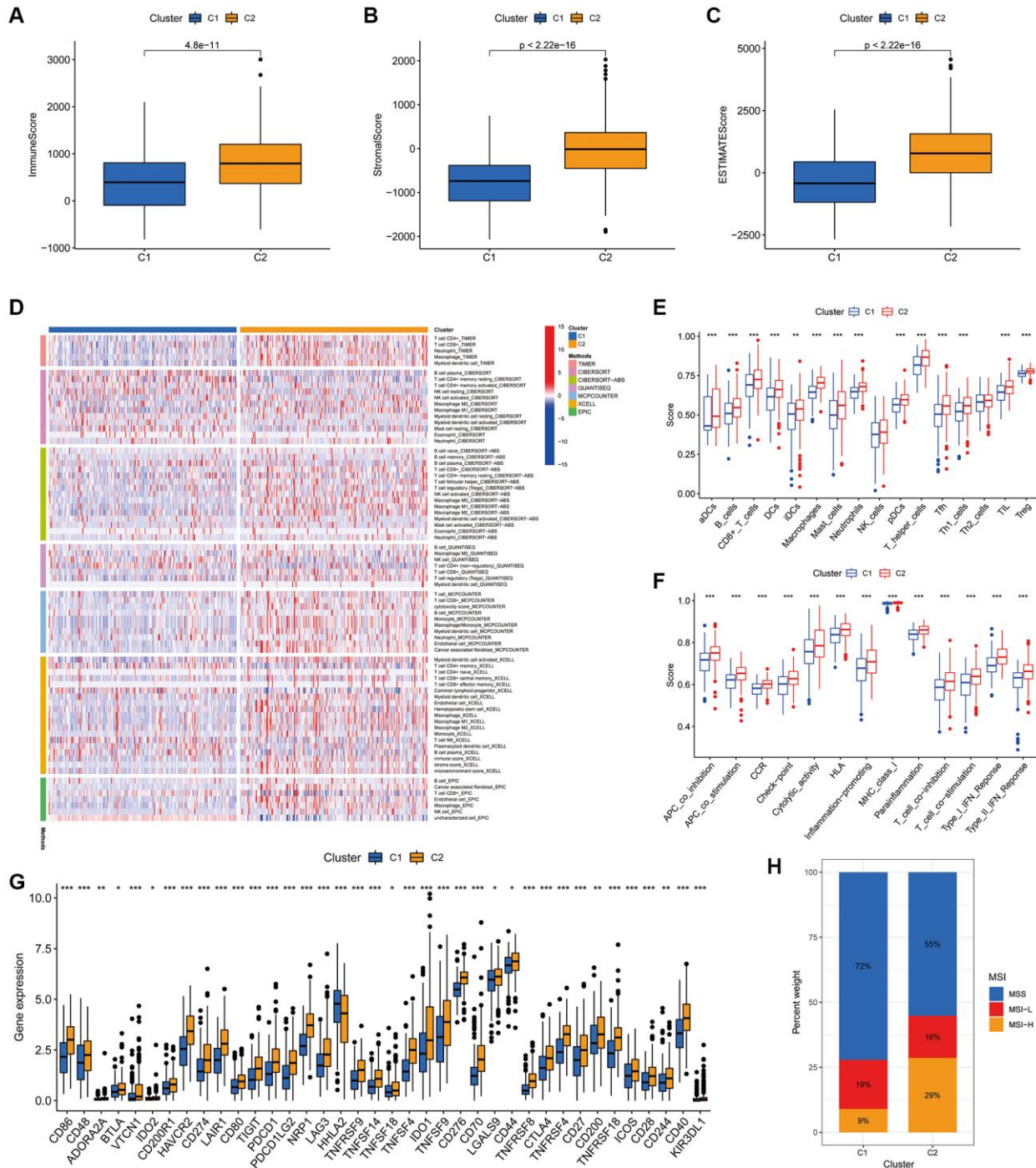


Figure 10. Association of the two clusters with the TIME. (A–C) Immune, stromal and ESTIMATE scores in the two clusters. (D) Heat map of the proportion of different types of immune cells. (E) Box plot showing differences in immune cells between the two clusters using the ssGSEA algorithm. (F) Box plot showing differences in immune-related functions between the two clusters using the ssGSEA algorithm. (G) Expression of immune checkpoint markers in the two clusters. (H) Histogram of proportions showing the proportion of patients with MSS, MSI-L and MSI-H in the two clusters. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

suggesting that both immune-related functions and immune cell infiltration were significantly stronger in cluster 2 (Figure 10E, 10F). Furthermore, most immune checkpoints were significantly more highly expressed in cluster 2 (Figure 10G). This suggests that patients in cluster 2 were more likely to benefit from ICIs compared to the cluster 1 population. Furthermore, the histogram of proportions revealed that the proportions

of MSS, MSI-L and MSI-H in cluster 1 were 72%, 19% and 9%, respectively, while the values were 55%, 16% and 29%, respectively, in cluster 2 (Figure 10H).

Drug sensitivity analysis of the two clusters revealed variations in IC50 for numerous chemical and targeted anti-cancer agents between the clusters ($P < 0.001$) (Figure 11A–11T). These findings imply that our

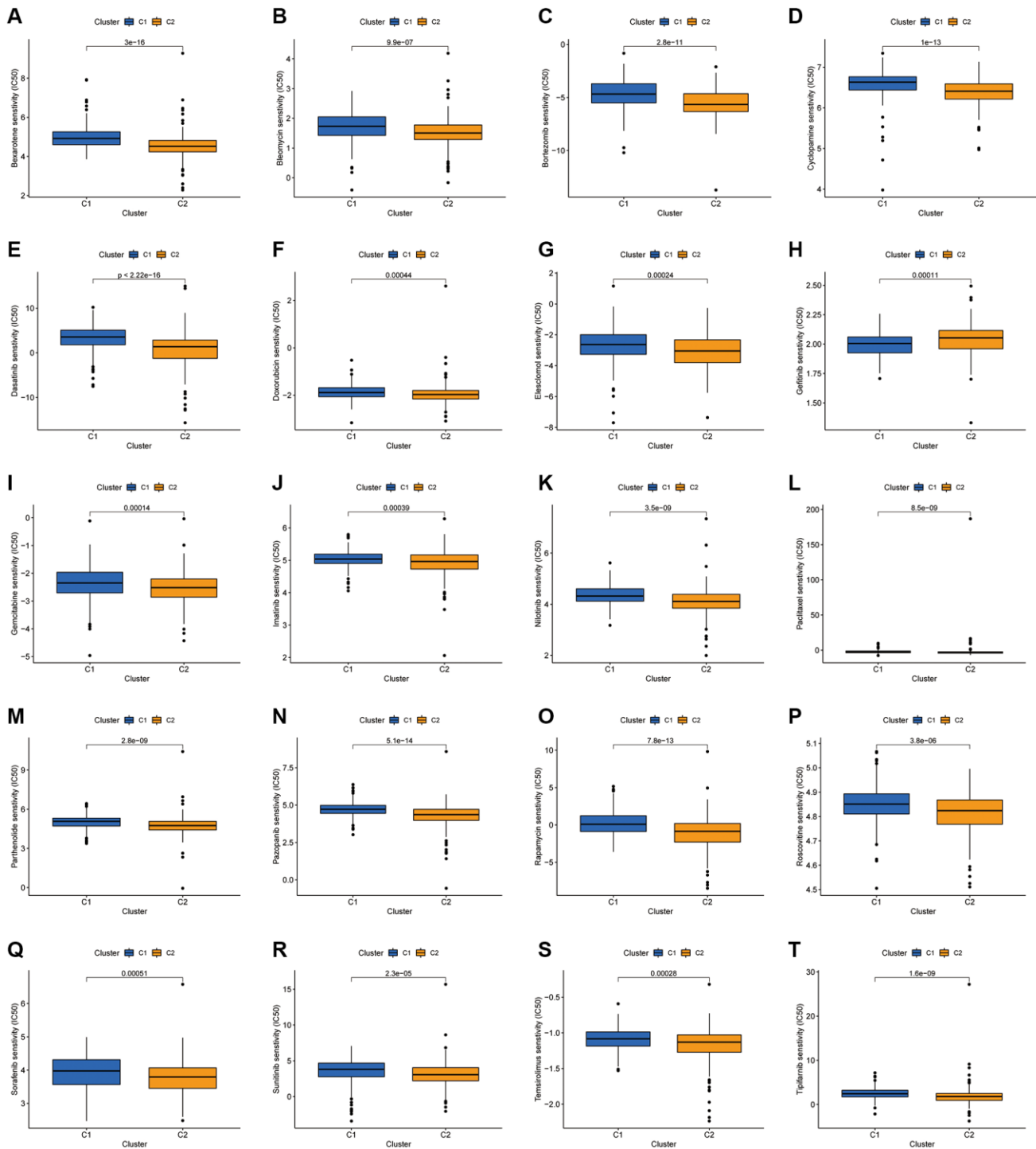


Figure 11. Investigation of drug sensitivity. (A–T) Boxplots of IC50 values for different agents in the two clusters.

clustering analysis could offer a basis for the selection of targeted therapeutic regimens and chemotherapeutic agents for patients with COAD.

Validation of CAFRs genes expression levels in COAD tissues

To further investigate the expression levels of CAFRs genes in COAD clinical tissues, we examined the mRNA expression levels of CAFRs genes in COAD tumor tissues and adjacent normal tissues. The qRT-PCR results showed that the mRNAs of all CAFRs genes were differentially expressed in COAD tumor tissues and adjacent normal tissues, among which CD36, NRG1 and FGF9 were highly expressed in adjacent normal tissues, whereas TIMP1, TERT, CDKN2A, PCAT6, CYP19A1, IL13, SNAI1, BDNF, GPC1, SERPINH1, AGER, and ENO2 were highly expressed in the tumor tissues (Figure 12A–12O).

DISCUSSION

Malignant tumours remain one of the major diseases that pose a serious threat to human health. The

evolution of tumours is determined by a combination of the intrinsic properties of the tumour cells and the external environment consisting of various other components in the TME [29]. CAFs are a major component of the TME, interacting extensively with tumour cells and influencing other components of the TME [30]. CAFs not only play a role in promoting tumour proliferation, metastasis and invasion but also in inducing anti-tumour drug resistance and immunosuppression [31, 32]. Additionally, growing evidence suggests that CAFs are strongly associated with the efficacy of tumour immunotherapy [13, 33, 34]. Zheng et al. reported that CAFs correlated with CD8+ T cells in the TME and also that the CD8+ T cell/CAFs ratio influenced the response to immunotherapy [35]. Furthermore, targeted therapy against CAFs is considered an effective strategy to improve the efficacy of immunotherapy [36]. Therefore, it is vital to understand the role of CAFs in assessing the prognosis and immunotherapy efficacy of patients with tumours.

In this study, we constructed CAFRs to predict the prognosis of patients with COAD. Patients with COAD were categorised into high- and low-risk subgroups

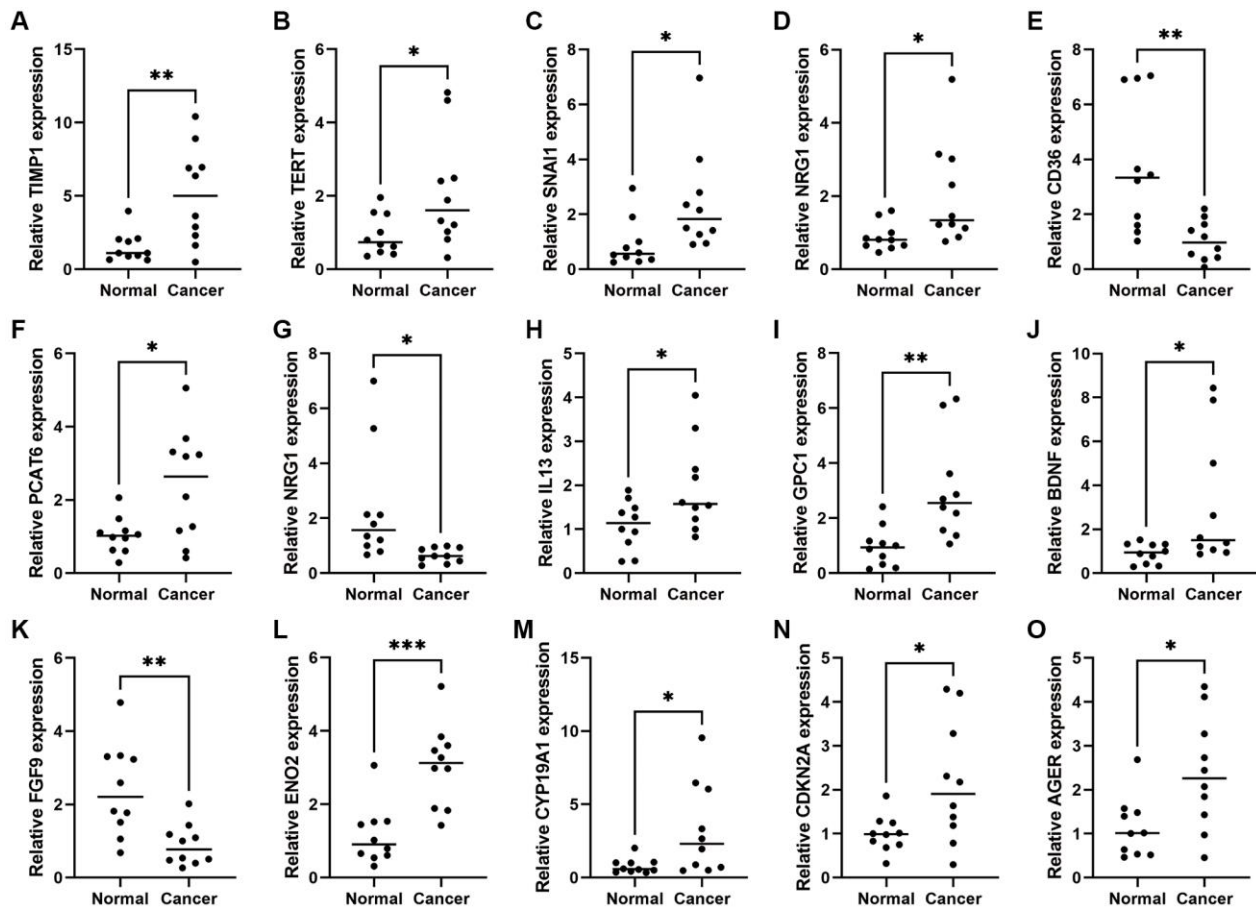


Figure 12. Expression of CAFRs-related genes in COAD and adjacent normal tissues. (A–O) mRNA expression levels of the 15 CAFRs-related genes in COAD and adjacent normal tissues ($n = 10$). * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

based on risk scores in the CAFRs. Additionally, we validated the prognostic predictive value of CAFRs in COAD cohorts and further assessed the signature using a range of methods, including univariate and multivariate Cox regression analysis and time-dependent ROC curves. The findings suggest that the constructed CAFRs have reliable and excellent prognostic predictive power for COAD.

Immune checkpoints can send ‘off’ signals to suppress immune function, thereby allowing tumour cells to evade immune killing [37, 38]. Recently, ICIs have been increasingly used in the treatment of patients with gastrointestinal tumours, displaying promising efficacy. In particular, ICIs represented by the anti-PD1/PD-L1 pathway have been most effective in patients with mismatch repair deficiency (dMMR)/MSI-H and have been approved for the second-line treatment of metastatic COAD with dMMR/MSI-H [39]. However, the majority of patients represented by pMMR/MSS not only responded poorly to ICIs but also caused disease progression in some patients. It is therefore important to explore biomarkers for predicting and evaluating the response to treatment with ICIs in COAD, thereby benefitting personalised treatment regimens.

Previous studies have confirmed that the molecular subtype of the tumour correlates with the clinical outcome and immune microenvironment characteristics of patients [40, 41]. To analyse the differences in survival and immune landscape of patients with different subtypes of COAD, we performed consensus clustering analysis based on the constructed CAFRs and divided the patients into two clusters. Further analysis revealed that most immune effector cells were more infiltrated in cluster 2 compared to cluster 1. Additionally, our study shows that most immune checkpoints were significantly highly expressed in cluster 2, suggesting that cluster 2 has a highly immunosuppressive microenvironment that promotes the immune escape of tumour cells, which is also corroborated with the poor clinical outcomes of this population.

Despite that the promising effects of ICIs, their low overall efficiency is an urgent issue for clinical immunotherapy. The sparse infiltration of effector immune cells in tumour tissues, known as ‘immune cold tumours’, is considered to be the main factor for the low efficiency of ICIs [42]. Contrastingly, ‘immune hot tumours’ are characterised by a high infiltration of effector immune cells and the activation of immune checkpoints, and respond better to ICIs [43]. Taken together, patients in cluster 2 were more consistent with the characteristics of an ‘immune hot

tumour’. Furthermore, in terms of MSI, cluster 2 had up to 29% of patients with MSI-H, which was higher than cluster 1 (9%). This further validates that cluster 2 could be a beneficial population for treatment with ICIs. Thus, our cluster analysis not only facilitates the prediction of prognosis and immune microenvironmental characteristics of different subtypes of COAD but also provides a basis for the identification of a population that is advantageous for the treatment of ICIs.

Patients with advanced COAD usually choose chemotherapy to control the progression of their disease; however, in some patients, efficacy is reduced after conventional first- and second-line standard chemotherapy. With rapid advances in drug development, molecularly targeted drugs with different mechanisms of action have been developed and are commonly used to treat patients with advanced COAD who have failed second-line therapy. Meanwhile, patients who show progress after standard treatment are often treated clinically with a combination of drugs with different mechanisms of action. Therefore, the rational arrangement of different drugs in combination with personalised treatment regimens is crucial for patients with advanced COAD. Notably, most of the targeted drugs, including Dasatinib, Imatinib, Nilotinib, Pazopanib, Sorafenib, Sunitinib and Tipifarnib, had lower IC50s in cluster 2, suggesting that cluster 2 could also be a beneficiary population for small molecule tyrosine kinase inhibitors.

CONCLUSION

The CAFRs and CAFRs-based clusters established can effectively predict the prognosis of patients with COAD and differentiate TIME characteristics in patients. This aids in distinguishing immune ‘hot’ and ‘cold’ tumours and guides ICI administration. Furthermore, the constructed signature provides valuable individualised treatment options for patients with cancer. Nevertheless, the therapeutic potential of CAFRs in clinical settings requires further validation in the future using prospective clinical trials with large samples.

AUTHOR CONTRIBUTIONS

Conceptualization, L.Z. and J.Z.; methodology, L.Z. and Y.Y.; software, Y.Y. and Z.S.; validation, Y.H. and M.Z.; formal analysis, Z.G.; data curation, L.Z.; writing—original draft preparation, L.Z.; writing—review and editing, J.Z. and P.L.; supervision, H.Q. and M.Z.; project administration, W.S. and M.Z. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We acknowledge the contributions from TCGA, HPA, GEO and Genecards databases.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT AND CONSENT

The study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Anhui Medical University (No. PJ20230861). All enrolled COAD patients provided written informed consent.

REFERENCES

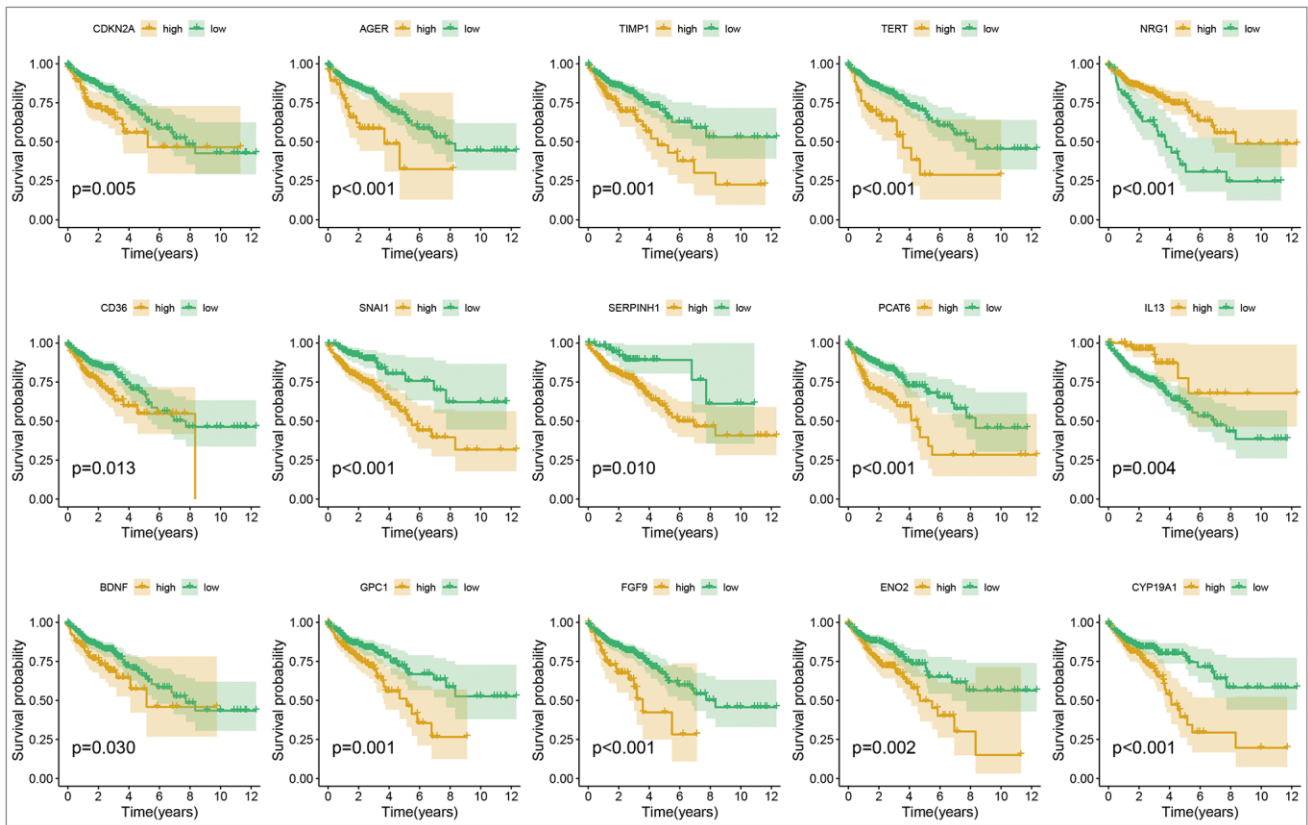
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
2. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol.* 2021; 14:101174. <https://doi.org/10.1016/j.tranon.2021.101174> PMID:34243011
3. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, Bray F. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut.* 2023; 72:338–44. <https://doi.org/10.1136/gutjnl-2022-327736> PMID:36604116
4. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T, Hynes RO, Jain RK, Janowitz T, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer.* 2020; 20:174–86. <https://doi.org/10.1038/s41568-019-0238-1> PMID:31980749
5. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov.* 2019; 18:99–115. <https://doi.org/10.1038/s41573-018-0004-1> PMID:30470818
6. Biffi G, Tuveson DA. Diversity and Biology of Cancer-Associated Fibroblasts. *Physiol Rev.* 2021; 101:147–76. <https://doi.org/10.1152/physrev.00048.2019> PMID:32466724
7. Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol.* 2019; 16:282–95. <https://doi.org/10.1038/s41575-019-0115-0> PMID:30778141
8. Kamali Zonouzi S, Pezeshki PS, Razi S, Rezaei N. Cancer-associated fibroblasts in colorectal cancer. *Clin Transl Oncol.* 2022; 24:757–69. <https://doi.org/10.1007/s12094-021-02734-2> PMID:34839457
9. Sun Y, Wang R, Qiao M, Xu Y, Guan W, Wang L. Cancer associated fibroblasts tailored tumor microenvironment of therapy resistance in gastrointestinal cancers. *J Cell Physiol.* 2018; 233:6359–69. <https://doi.org/10.1002/jcp.26433> PMID:29334123
10. Chen Y, McAndrews KM, Kalluri R. Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat Rev Clin Oncol.* 2021; 18:792–804. <https://doi.org/10.1038/s41571-021-00546-5> PMID:34489603
11. Dong W, Xie Y, Huang H. Prognostic Value of Cancer-Associated Fibroblast-Related Gene Signatures in Hepatocellular Carcinoma. *Front Endocrinol (Lausanne).* 2022; 13:884777. <https://doi.org/10.3389/fendo.2022.884777> PMID:35733776
12. Zheng H, Liu H, Li H, Dou W, Wang X. Weighted Gene Co-expression Network Analysis Identifies a Cancer-Associated Fibroblast Signature for Predicting Prognosis and Therapeutic Responses in Gastric Cancer. *Front Mol Biosci.* 2021; 8:744677. <https://doi.org/10.3389/fmolb.2021.744677> PMID:34692770
13. Ye Y, Zhao Q, Wu Y, Wang G, Huang Y, Sun W, Zhang M. Construction of a cancer-associated fibroblasts-related long non-coding RNA signature to predict prognosis and immune landscape in pancreatic adenocarcinoma. *Front Genet.* 2022; 13:989719. <https://doi.org/10.3389/fgene.2022.989719> PMID:36212154
14. Pontén F, Schwenk JM, Asplund A, Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. *J Intern Med.* 2011; 270:428–46. <https://doi.org/10.1111/j.1365-2796.2011.02427.x> PMID:21752111
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust

- enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–7.
<https://doi.org/10.1038/nmeth.3337>
PMID:25822800
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID:16199517
17. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009; 462:108–12.
<https://doi.org/10.1038/nature08460>
PMID:19847166
18. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612.
<https://doi.org/10.1038/ncomms3612>
PMID:24113773
19. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med*. 2016; 22:1342–50.
<https://doi.org/10.1038/nm.4191>
PMID:27694933
20. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*. 2015; 372:2509–20.
<https://doi.org/10.1056/NEJMoa1500596>
PMID:26028255
21. McNamara MG, Jacobs T, Lamarca A, Hubner RA, Valle JW, Amir E. Impact of high tumor mutational burden in solid tumors and challenges for biomarker application. *Cancer Treat Rev*. 2020; 89:102084.
<https://doi.org/10.1016/j.ctrv.2020.102084>
PMID:32738738
22. Jiang T, Chen J, Xu X, Cheng Y, Chen G, Pan Y, Fang Y, Wang Q, Huang Y, Yao W, Wang R, Li X, Zhang W, et al. On-treatment blood TMB as predictors for camrelizumab plus chemotherapy in advanced lung squamous cell carcinoma: biomarker analysis of a phase III trial. *Mol Cancer*. 2022; 21:4.
<https://doi.org/10.1186/s12943-021-01479-4>
PMID:34980131
23. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC, Kindler HL, Lopez-Martin JA, Miller WH Jr, Italiano A, Kao S, Piha-Paul SA, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol*. 2020; 21:1353–65.
[https://doi.org/10.1016/S1470-2045\(20\)30445-9](https://doi.org/10.1016/S1470-2045(20)30445-9)
PMID:32919526
24. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019; 51:202–6.
<https://doi.org/10.1038/s41588-018-0312-8>
PMID:30643254
25. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 14:7.
<https://doi.org/10.1186/1471-2105-14-7>
PMID:23323831
26. Sun W, Shen J, Liu J, Han K, Liang L, Gao Y. Gene Signature and Prognostic Value of Ubiquitin-Specific Proteases Members in Hepatocellular Carcinoma and Explored the Immunological Role of *USP36*. *Front Biosci (Landmark Ed)*. 2022; 27:190.
<https://doi.org/10.31083/j.fbl2706190>
PMID:35748266
27. DeBerardinis RJ. Tumor Microenvironment, Metabolism, and Immunotherapy. *N Engl J Med*. 2020; 382:869–71.
<https://doi.org/10.1056/NEJMcibr1914890>
PMID:32101671
28. Das S, Camphausen K, Shankavaram U. Cancer-Specific Immune Prognostic Signature in Solid Tumors and Its Relation to Immune Checkpoint Therapies. *Cancers (Basel)*. 2020; 12:2476.
<https://doi.org/10.3390/cancers12092476>
PMID:32882873
29. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022; 12:31–46.
<https://doi.org/10.1158/2159-8290.CD-21-1059>
PMID:35022204
30. Park D, Sahai E, Rullan A. SnapShot: Cancer-Associated Fibroblasts. *Cell*. 2020; 181:486–486.e1.
<https://doi.org/10.1016/j.cell.2020.03.013>
PMID:32302576

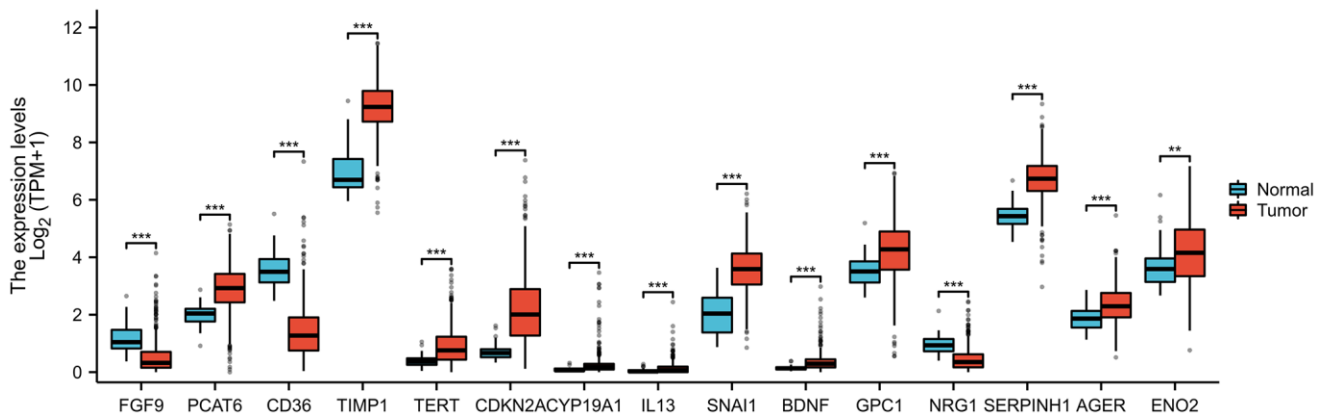
31. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer*. 2016; 16:582–98.
<https://doi.org/10.1038/nrc.2016.73>
PMID:[27550820](https://pubmed.ncbi.nlm.nih.gov/27550820/)
32. D'Arcangelo E, Wu NC, Cadavid JL, McGuigan AP. The life cycle of cancer-associated fibroblasts within the tumour stroma and its importance in disease outcome. *Br J Cancer*. 2020; 122:931–42.
<https://doi.org/10.1038/s41416-019-0705-1>
PMID:[31992854](https://pubmed.ncbi.nlm.nih.gov/31992854/)
33. Wang Q, Zhang X, Du K, Wu X, Zhou Y, Chen D, Zeng L. Machine learning identifies characteristics molecules of cancer associated fibroblasts significantly correlated with the prognosis, immunotherapy response and immune microenvironment in lung adenocarcinoma. *Front Oncol*. 2022; 12:1059253.
<https://doi.org/10.3389/fonc.2022.1059253>
PMID:[36439484](https://pubmed.ncbi.nlm.nih.gov/36439484/)
34. Zou R, Jiang Q, Jin T, Chen M, Yao L, Ding H. Pan-cancer analyses and molecular subtypes based on the cancer-associated fibroblast landscape and tumor microenvironment infiltration characterization reveal clinical outcome and immunotherapy response in epithelial ovarian cancer. *Front Immunol*. 2022; 13:956224.
<https://doi.org/10.3389/fimmu.2022.956224>
PMID:[36032075](https://pubmed.ncbi.nlm.nih.gov/36032075/)
35. Zheng X, Jiang K, Xiao W, Zeng D, Peng W, Bai J, Chen X, Li P, Zhang L, Zheng X, Miao Q, Wang H, Wu S, et al. CD8⁺ T cell/cancer-associated fibroblast ratio stratifies prognostic and predictive responses to immunotherapy across multiple cancer types. *Front Immunol*. 2022; 13:974265.
<https://doi.org/10.3389/fimmu.2022.974265>
PMID:[36439099](https://pubmed.ncbi.nlm.nih.gov/36439099/)
36. Hanley CJ, Thomas GJ. Targeting cancer associated fibroblasts to enhance immunotherapy: emerging strategies and future perspectives. *Oncotarget*. 2021; 12:1427–33.
<https://doi.org/10.18632/oncotarget.27936>
PMID:[34262652](https://pubmed.ncbi.nlm.nih.gov/34262652/)
37. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, Drake CG, Camacho LH, Kauh J, Odunsi K, Pitot HC, Hamid O, Bhatia S, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*. 2012; 366:2455–65.
<https://doi.org/10.1056/NEJMoa1200694>
PMID:[22658128](https://pubmed.ncbi.nlm.nih.gov/22658128/)
38. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012; 12:252–64.
<https://doi.org/10.1038/nrc3239>
PMID:[22437870](https://pubmed.ncbi.nlm.nih.gov/22437870/)
39. Kather JN, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer. *Semin Cancer Biol*. 2018; 52:189–97.
<https://doi.org/10.1016/j.semcancer.2018.02.010>
PMID:[29501787](https://pubmed.ncbi.nlm.nih.gov/29501787/)
40. Wang Y, Peng B, Ning C, He S, Yang H, Mao Y, Sun L. Characterization of immune features and immunotherapy response in subtypes of hepatocellular carcinoma based on mitophagy. *Front Immunol*. 2022; 13:966167.
<https://doi.org/10.3389/fimmu.2022.966167>
PMID:[36304466](https://pubmed.ncbi.nlm.nih.gov/36304466/)
41. Wang L, Qu H, Ma X, Liu X. Identification of Oxidative Stress-Associated Molecular Subtypes and Signature for Predicting Survival Outcome of Cervical Squamous Cell Carcinoma. *Oxid Med Cell Longev*. 2022; 2022:1056825.
<https://doi.org/10.1155/2022/1056825>
PMID:[36225179](https://pubmed.ncbi.nlm.nih.gov/36225179/). Retraction in: *Oxid Med Cell Longev*. 2023; 2023:9795318.
<https://doi.org/10.1155/2023/9795318>
PMID:[37565171](https://pubmed.ncbi.nlm.nih.gov/37565171/)
42. Bonaventura P, Shekarian T, Alcazer V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, Caux C, Depil S. Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol*. 2019; 10:168.
<https://doi.org/10.3389/fimmu.2019.00168>
PMID:[30800125](https://pubmed.ncbi.nlm.nih.gov/30800125/)
43. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov*. 2019; 18:197–218.
<https://doi.org/10.1038/s41573-018-0007-y>
PMID:[30610226](https://pubmed.ncbi.nlm.nih.gov/30610226/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. K-M curves of signature-related genes in the TCGA-COAD cohort.



Supplementary Figure 2. The differential expression status of CAFs-related genes in COAD tumor tissues and normal tissues in TCGA-COAD cohort.

Supplementary Tables

Supplementary Table 1. Specific links to all immunohistochemistry images from the Human Protein Atlas.

Protein	Specific link
FGF9	Normal: https://www.proteinatlas.org/ENSG00000102678-FGF9/tissue/colon
	Tumor: https://www.proteinatlas.org/ENSG00000102678-FGF9/pathology/colorectal+cancer#ihc
CD36	Normal: https://www.proteinatlas.org/ENSG00000135218-CD36/tissue/colon
	Tumor: https://www.proteinatlas.org/ENSG00000135218-CD36/pathology/colorectal+cancer#img
TIMP1	Normal: https://www.proteinatlas.org/ENSG00000102265-TIMP1/tissue/colon
	Tumor: https://www.proteinatlas.org/ENSG00000102265-TIMP1/pathology/colorectal+cancer#img
TERT	Normal: https://www.proteinatlas.org/ENSG00000164362-TERT/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000164362-TERT/pathology/colorectal+cancer#img
CDKN2A	Normal: https://www.proteinatlas.org/ENSG00000147889-CDKN2A/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000147889-CDKN2A/pathology/colorectal+cancer#img
CYP19A1	Normal: https://www.proteinatlas.org/ENSG00000137869-CYP19A1/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000137869-CYP19A1/pathology/colorectal+cancer#img
SNAI1	Normal: https://www.proteinatlas.org/ENSG00000124216-SNAI1/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000124216-SNAI1/pathology/colorectal+cancer#img
BDNF	Normal: https://www.proteinatlas.org/ENSG00000176697-BDNF/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000176697-BDNF/pathology/colorectal+cancer#img
GPC1	Normal: https://www.proteinatlas.org/ENSG00000063660-GPC1/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000063660-GPC1/pathology/colorectal+cancer#img
NRG1	Normal: https://www.proteinatlas.org/ENSG00000157168-NRG1/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000157168-NRG1/pathology/colorectal+cancer#img
SERPINH1	Normal: https://www.proteinatlas.org/ENSG00000149257-SERPINH1/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000149257-SERPINH1/pathology/colorectal+cancer#img
AGER	Normal: https://www.proteinatlas.org/ENSG00000204305-AGER/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000204305-AGER/pathology/colorectal+cancer#img
ENO2	Normal: https://www.proteinatlas.org/ENSG00000111674-ENO2/tissue/colon#img
	Tumor: https://www.proteinatlas.org/ENSG00000111674-ENO2/pathology/colorectal+cancer#img

Supplementary Table 2. Primer sequences used for RT-qPCR.

Gene	Primer sequence
GAPDH	F: ACAACTTTGGTATCGTGGAAGG
	R: GCCATCACGCCACAGTTTC
ENO2	F: CCGGGAAGCTCAGACCTCATC
	R: CTCTGCACCTAGTCGCATGG
FGF9	F: ATGGCTCCCTTAGGTGAAGTT
	R: CCCAGGTGGTCACTTAACAAAAC
GPC1	F: TGAAGCTGGTCTACTGTGCTC
	R: CCCAGAACTTGTCGGTGATGA
CD36	F: CTTTGGCTTAATGAGACTGGGAC
	R: GCAACAAACATCACCACACCA
TIMP1	F: ACCACCTTATACCAGCGTTATGA
	R: GGTGTAGACGAACCGGATGTC
TERT	F: TCACGGAGACCACGTTTCAA
	R: TTCAAGTGCTGTCTGATTCCAAT

CDKN2A F: ATGGAGCCTTCGGCTGACT
R: GTAACCTATTCGGTGC GTTGGG
CYP19A1 F: ACTACAACCGGGTATATGGAGAA
R: TCGAGAGCTGTAATGATTGTGC
IL-13 F: CCTCATGGCGCTTTTGTGAC
R: TCTGGTTCTGGGTGATGTTGA
SNAI1 F: TCGGAAGCCTAACTACAGCGA
R: AGATGAGCATTGGCAGCGAG
BDNF F: TAACGGCGGCAGACAAAAAGA
R: TGCACTTGGTCTCGTAGAAGTAT
PCAT6 F: CCCCTCCTTACTCTTGGACAAC
R: GACCGAATGAGGATGGAGACAC
NRG1 F: TCCATTAGAATATCAGTATCCACAG
R: CATAAGCGACACACAGGATTTC
SERPINH1 F: GCGGGCTAAGAGTAGAATCG
R: ATGGCCAGGAAGTGGTTTG
AGER F: ACTACCGAGTCCGTGTCTACC
R: GGAACACCAGCCGTGAGTT
