



OPEN ACCESS

EDITED BY

Xiangshan Fan,
Nanjing University Medical School,
China

REVIEWED BY

Bahman Panahi,
Agricultural Biotechnology Research
Institute of Iran, Iran
Xiaoyi Lv,
Xinjiang University, China

*CORRESPONDENCE

Huaidong Cheng,
chd1975ay@126.com

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

SPECIALTY SECTION

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

RECEIVED 07 April 2022

ACCEPTED 25 July 2022

PUBLISHED 29 August 2022

CITATION

Tang L, Yu S, Zhang Q, Cai Y, Li W, Yao S
and Cheng H (2022), Identification of
hub genes related to CD4⁺ memory
T cell infiltration with gene co-
expression network predicts prognosis
and immunotherapy effect in
colon adenocarcinoma.
Front. Genet. 13:915282.
doi: 10.3389/fgene.2022.915282

COPYRIGHT

© 2022 Tang, Yu, Zhang, Cai, Li, Yao and
Cheng. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Identification of hub genes related to CD4⁺ memory T cell infiltration with gene co-expression network predicts prognosis and immunotherapy effect in colon adenocarcinoma

Lingxue Tang^{1,2†}, Sheng Yu^{1,2†}, Qianqian Zhang^{1,2†}, Yinlian Cai^{1,2},
Wen Li^{1,2}, Senbang Yao^{1,2} and Huaidong Cheng^{1,2*}

¹Department of Oncology, the Second Affiliated Hospital of Anhui Medical University, Hefei, China,

²Department of Oncology, Anhui Medical University, Hefei, China

Background: CD4⁺ memory T cells (CD4⁺ MTCs), as an important part of the microenvironment affecting tumorigenesis and progression, have rarely been systematically analyzed. Our purpose was to comprehensively analyze the effect of CD4⁺ MTC infiltration on the prognosis of colon adenocarcinoma (COAD).

Methods: Based on RNA-Seq data, weighted gene co-expression network analysis (WGCNA) was used to screen the CD4⁺ MTC infiltration genes most associated with colon cancer and then identify hub genes and construct a prognostic model using the least absolute shrinkage and selection operator algorithm (LASSO). Finally, survival analysis, immune efficacy analysis, and drug sensitivity analysis were performed to evaluate the role of the prognostic model in COAD.

Results: We identified 929 differentially expressed genes (DEGs) associated with CD4⁺ MTCs and constructed a prognosis model based on five hub genes (F2RL2, TGFB2, DTNA, S1PR5, and MPP2) to predict overall survival (OS) in COAD. Kaplan–Meier analysis showed poor prognosis in the high-risk group, and the analysis of the hub gene showed that overexpression of TGFB2, DTNA, S1PR5, or MPP2 was associated with poor prognosis. Clinical prediction nomograms combining CD4⁺ MTC-related DEGs and clinical features were constructed to accurately predict OS and had high clinical application value.

Abbreviations: COAD, colon adenocarcinoma; OS, overall survival; TME, tumor microenvironment; CD4⁺ MTCs, CD4⁺ memory T cells; WGCNA, weighted gene co-expression network analysis; LASSO, the least absolute shrinkage, and selection operator; DEGs, differentially expressed genes; TCGA, the Cancer Genome Atlas database; GEO, Gene Expression Omnibus; TCIA, the Cancer Immunome Database; TMB, tumor mutational burden; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; PD-1, programmed cell death protein 1; FDR, false discovery rate, K-M, Kaplan–Meier; ROC, receiver operating characteristic; IC₅₀, half-maximal inhibitory concentration; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; AUC, area under the curve; ICIs, immune checkpoint inhibitors.

Immune efficacy and drug sensitivity analysis provide new insights for individualized treatment.

Conclusion: We constructed a prognostic risk model to predict OS in COAD and analyzed the effects of risk score on immunotherapy efficacy or drug sensitivity. These studies have important clinical significance for individualized targeted therapy and prognosis.

KEYWORDS

colon adenocarcinoma, immunotherapy, CD4 + memory T cell, weighted gene coexpression network analysis, gene

Introduction

Colon cancer is one of the most common gastrointestinal malignancies in humans. GLOBOCAN 2020 estimated 1,148,515 new cases and 5,76,858 deaths due to colon cancer, ranking fifth among 36 cancers globally, accounting for 6 and 5.8%, respectively (Sung et al., 2021). Among all histological subtypes, colorectal adenocarcinoma accounts for more than 90% of colon cancer types (Barresi et al., 2015). Surgery and chemotherapy have improved the overall survival (OS) of colon cancer to a certain extent, but postoperative recurrence and emergence of acquired drug resistance have affected the prognosis of patients. The development of colon cancer is a multistep process caused by gradual accumulation of mutations in tumor suppressor genes, oncogenes, and epigenetic changes (Migheli and Migliore, 2012). In recent years, with in-depth exploration of tumor markers and biomarkers, clinical treatment decisions for colon cancer have changed greatly. A series of markers have been shown to play an important role in the early diagnosis of cancer, monitoring the efficacy of treatment and follow-up of possible recurrence. It is now easier to choose the most appropriate strategy for managing colon cancer (Lech et al., 2016). For example, microsatellite instability-high (MSI-H) was shown to be a predictor of improved overall survival (OS), and chromosome 18q deletion was associated with worse prognosis (Popat and Houlston, 2005; Kim et al., 2007). Patients with mutations in the tumor suppressor gene p53 had better OS when treated with adjuvant chemotherapy than those treated with surgery alone (Russo et al., 2005). KRAS mutation was confirmed to be correlated with non-responsiveness to cetuximab and panitumumab (Di Fiore et al., 2007), and BRAF mutations make patients resistant to anti-EGFR monoclonal antibodies and predict worse prognosis (Roth et al., 2010). Although targeted therapy has been incorporated into the treatment regimen for colon cancer, there is currently no comprehensive drug selection strategy to identify patients who will benefit the most. Therefore, it is of great significance to construct diagnostic and predictive biomarker models to identify the best prognostic biomarkers and help the selection of therapeutic drugs.

The occurrence and development of cancer are closely related to the complex tumor microenvironment (TME). Immune cells in the immune system which contain immune parameters related to survival are an important component of the microenvironment (Galon et al., 2013). Recently, several studies have confirmed that the molecular profile of immune-related genes in TMB may be a promising biomarker for predicting OS in cancer patients. (Huang R et al., 2020). In general, understanding the interaction of cancer and immune cells can help patients assess whether they would benefit from clinical treatment, especially immunotherapy. At present, the choice of treatment options and prognosis evaluation of colon cancer mainly depend on pathological tissue type, TNM stage, and biomarkers (Angell et al., 2020). Although these prediction methods are widely used in clinical practice, they still cannot provide complete prognostic information. For example, patients with the same histological tumor stage may have a significantly different clinical prognosis. Therefore, individualized treatment can maximize the benefits and minimize the harm to patients, resulting in optimal survival status and relatively long survival time.

T cell immunity is a hot research topic in recent years. CD4⁺ memory T cells (CD4⁺ MTCs) are closely associated with the prognosis in breast cancer (Deng et al., 2019), gastric cancer (Ning et al., 2020), lung adenocarcinoma (Choi and Na, 2018), and pancreatic cancer (Gu et al., 2020), but the role in colon adenocarcinoma (COAD) is unclear. Common detection methods for immune infiltration include flow cytometry and immunohistochemistry, but they cannot comprehensively measure the immune effects of different immune cell types. The wide application of high-throughput sequencing makes transcriptomic data more accessible and provides large amounts of resources for the analysis of immune cell infiltration (Lv et al., 2022). Predictive biomarker screening based on a database has been widely used in various diseases and achieved good results, especially in cancer-related fields. Weighted gene co-expression network analysis (WGCNA) is a comprehensive biological analysis method used to describe the correlation pattern between genes in microarray samples and pairwise relationships between gene transcripts. WGCNA can also be used to find clusters or modules of highly related genes,

analyze correlations between modules and clinical characteristics, and identify biomarkers or therapeutic targets (Langfelder and Horvath, 2008; Yang et al., 2021a). This method has been successfully applied to analyze gene expression data from various types of cancers, such as breast cancer (Tian et al., 2020), lung cancer (Ding et al., 2019), melanoma (Wan et al., 2018), hepatocellular carcinoma (Yang et al., 2021a), glioblastoma (Zhou J et al., 2021), oral squamous cell carcinoma (Yang et al., 2021b), and ovarian cancers (Su et al., 2021). Compared with other signature construction methods, WGCNA pays more attention to the strong associations between genes and can more accurately identify the prognostic-related hub genes, which provides a novel method for us to construct a higher-resolution prognostic model and a new idea for predicting disease prognosis (Panahi and Hejazi, 2021). As a feature selection method, the least absolute shrinkage and selection operator (LASSO) is increasingly used in colon. We usually use LASSO regression analysis to mitigate the over-fitting of genes with prognostic value (Narala et al., 2021). To improve the accuracy of prediction and the generalization of statistical models, LASSO eliminates unnecessary covariates in a combined nonlinear and interactive manner (Obermeyer and Emanuel, 2016; Pavlou et al., 2016). Compared to traditional statistical models, LASSO has a better ability to identify key predictors of clinical features.

In this study, we screened the CD4⁺ MTC-related differentially expressed genes (DEGs) by WGCNA and then used LASSO-Cox regression analysis to identify hub genes and construct a prognostic model. We used the prognostic model to predict OS and established nomograms to improve prediction capacity. We also discussed the specific role of CD4⁺ MTC-related hub genes in colon adenocarcinoma (COAD) and predicted the effectiveness of immunotherapy and potential therapeutic drugs. Finally, we explored the function and biological signaling pathway of CD4⁺ MTC-related genes. These results will provide more precise treatment strategies for prognosis of colon cancer.

Materials and methods

Data collection and preprocessing

437 transcriptome data files (including 398 COAD and 39 normal samples), 385 clinical data files, and 399 gene mutation data files of the colon cancer training set were downloaded from the Cancer Genome Atlas database (TCGA, <https://portal.gdc.cancer.gov/>). The retrieval strategy is shown in Supplementary Table S1. Public microarray data and clinical data of the testing set (GSE40967-GPL570) were downloaded from the Gene Expression Omnibus database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) by using the keywords “colon cancer,” “survival,” and “*Homo sapiens*”. Then, we use Perl software for preliminary processing of the TCGA data. We first extracted gene expression data from transcriptome files of 398 tumor samples. These data were then analyzed to identify DEGs between normal and COAD samples.

Second, the clinical features of each sample were extracted from 385 clinical samples, including survival status, OS, age, gender, and TNM stage. Subsequently, we merged the gene expression data and clinical data based on the sample IDs and finally obtained a total of 379 training set samples (Supplementary Table S2). Similarly, we screened 579 samples with complete clinical features from 585 GEO samples and merged them with the microarray data (Supplementary Table S3). In addition, we also obtained tumor mutational burden (TMB) data from 399 samples (Supplementary Table S4). TMB is defined as the total number of somatic gene coding errors, base substitutions, gene insertions, or deletion errors detected per million bases (Yarchoan et al., 2017). Finally, the cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) immunotherapy score data were downloaded from the Cancer Immunome Database for further analysis (TCIA, <https://tcia.at/>).

Immune cell infiltration and co-expression network construction

The “CIBERSORT” (Newman et al., 2015) algorithm was used to quantify the tumor-infiltrating immune cells from cancer RNA-seq data, and the relative proportions of 22 types of immune infiltrating cells were determined from the reference data set (the gene expression characteristic set of 22 immune cell subtypes) (Chen et al., 2018). Also, we searched the genes in normal samples and COAD circularly to identify the DEGs between the two groups. Both $|\log FC| \geq 1.0$ and adjusted $p < 0.05$ were used as the thresholds for DEGs. Based on the results of immune cell infiltration in DEGs, a co-expression network was constructed by R package “WGCNA” (18) and Pearson correlations (Botía et al., 2017) to understand correlation patterns between genes and identify important modules associated with COAD. In order to meet the requirements of scale-free topology, the soft threshold method was used to evaluate the correlation coefficient and noise filtering capability. The optimal soft threshold power β is determined by the function in WGCNA. The topological overlap matrix (TOM) and corresponding dissimilarity matrix (1-TOM) visualize the network graph for module detection. Subsequently, a scale-free topology plot was generated under the optimal soft threshold power, and a clustering tree of co-expressed gene modules was established, with the main parameters as $\text{cutHeight} = 10,000$, $\text{minSize} = 10$. DEGs closely related to 22 types of immune infiltrating cells were obtained by WGCNA to represent the expression profiles of module genes.

Identification of hub genes and construction of a prognostic model

We extracted the target module of CD4⁺ MTC infiltration DEGs from the co-expression network according to the p -value of the correlation coefficient. The correlation was deemed

significant when the false discovery rate (FDR) was $p < 0.05$ (Chen et al., 2021). Based on the gene expression in TCGA and GEO preprocessing results, the expression levels of target module DEGs in each sample were obtained. Then, we performed a univariate Cox regression analysis on the TCGA training set to screen out DEGs which were significantly associated with OS (Supplementary Table S5). Subsequently, LASSO regression analysis was performed on these DEGs to remove genes that were highly correlated and prevent the overfitting of the model. Next, DEGs with the least error in LASSO regression were determined by cross-validation. Based on the results of LASSO, we constructed a Cox model to obtain the prognostic hub genes and model formula. Then, the risk score was calculated using the model formula, and the training set was divided into high- and low-risk groups according to the median value of the risk score. Similarly, the GEO validation set was also divided into high- and low-risk groups to verify the accuracy of the prognostic model.

Predictive ability of the model

We validated the prognostic value of the model by Kaplan–Meier (K–M) analysis. The accuracy of the prognostic model in predicting 1-, 3-, and 5-year OS rates was assessed by time-dependent receiver operating characteristic (ROC) curve. In addition, multivariate and univariate Cox regression analyses were performed to verify whether the risk score could be used as a prognostic indicator independent of other clinical features. We also performed survival analysis for each hub gene to determine the impact of their differential expression on the prognosis of COAD. Furthermore, based on the results of independent prognostic factors, we used “regplot” and “rms” packages to draw nomograms and calibration curves. Each prognostic factor corresponds to a score, and the scores of all prognostic factors were added to obtain a total score. Then, the total score was used to predict the 1-, 3-, and 5-year survival of COAD, and the calibration curve was used to verify the accuracy of the nomogram.

Correlation analysis of risk score with immune cell infiltration, immunotherapy, and drug sensitivity

Based on pan-cancer immune cell infiltration data from the TCGA database (Supplementary Table S6), the correlation of immune cells with risk scores was analyzed. We also analyzed the correlation of each hub gene with immune cell infiltration and immune checkpoint. After downloading the CTLA-4 and PD-1 immunotherapy scores of COAD patients downloaded from the TCIA website, we compared the difference in immunotherapy efficacy between high- and low-risk groups. The higher the immunity score, the more patients benefit from immunotherapy. Finally, the pre-prepared installation package

“pRRophetic” was used to predict drug sensitivity for COAD. The “pRRophetic” package is mainly used to predict the phenotype and drug sensitivity of external cell lines by using gene expression data. It can also be used to predict clinical data. (i.e., predicting clinical outcomes based on the cancer genome project cell coefficient). The half-maximal inhibitory concentration (IC_{50}) value is used to represent the sensitivity of drugs. The smaller the IC_{50} value, the more sensitive the patient is to the drug.

Functional and pathway enrichment analysis

We used the R package “limma” package and logFC function to analyze the genes in the prognostic model and screened out DEGs between high- and low-risk groups. LogFC >0 means the gene is upregulated in the high-risk group, and conversely, it is upregulated in the low-risk group. Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Dennis et al., 2003) enrichment analyses were performed based on the data obtained from risk analysis. The FDR (< 0.05) correction was used to determine the statistical significance of GO and KEGG terms.

Tumor mutational burden analysis

We analyzed the sorted gene mutation data with the R package “maftools” and generated the corresponding waterfall diagram according to the high- and low-risk groups, showing the 30 genes with the highest mutation frequency. R software was used to verify whether there was a difference in TMB between high- and low-risk groups, and the association between TMB and OS was verified by K–M analysis.

Statistical analysis

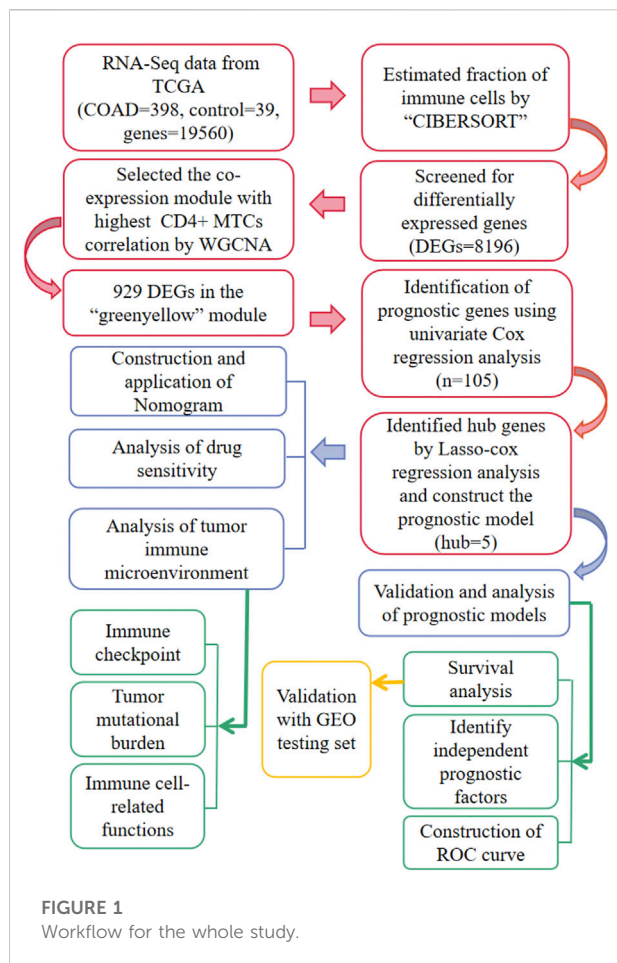
R (v 4.1.3) (<https://www.r-project.org/>) and SPSS 23 were used for statistical analysis, and Strawberryperl (v 5.30.0) (<https://www.perl.org/>) was used to sort and merge the downloaded data.

Results

Evaluation of tumor-infiltrating immune cells

The flowchart of the whole study is summarized in Figure 1.

The RNA-Seq data of TCGA contained a total of 398 COAD and 39 normal samples, including 19,560 genes. After obtaining the expression matrix file, the proportion of 22 types of immune cell infiltration was determined by the “CIBERSORT” algorithm. Among the 22 types of immune cells, “Macrophages M0,” “T cells



follicular helper,” “Macrophages M1,” and “T cells CD4 memory activated” were more actively expressed in COAD than in normal samples (Figure 2A). The heatmap shows the relationship between 22 immune cells in COAD (Figure 2B). The number in the small square is the Pearson product-moment correlation coefficient, which is used to measure the degree of linear correlation between variable X and variable Y. Its value is between -1 and 1 , greater than 0 means positively correlated, and less than 0 means negatively correlated. 0 means that there is no linear correlation between the two variables. (For example, “B cells memory” and “Macrophages M0” are positively correlated, correlation coefficient = 0.08). Also, the bar plot shows the proportion of immune cell infiltration in each COAD. We can conclude that there was a higher proportion of T cells and macrophages (Figure 2C).

Construction of the gene co-expression network

We first analyzed all TCGA genes and extracted 8,196 DEGs between normal samples and COAD. Then, R software was used

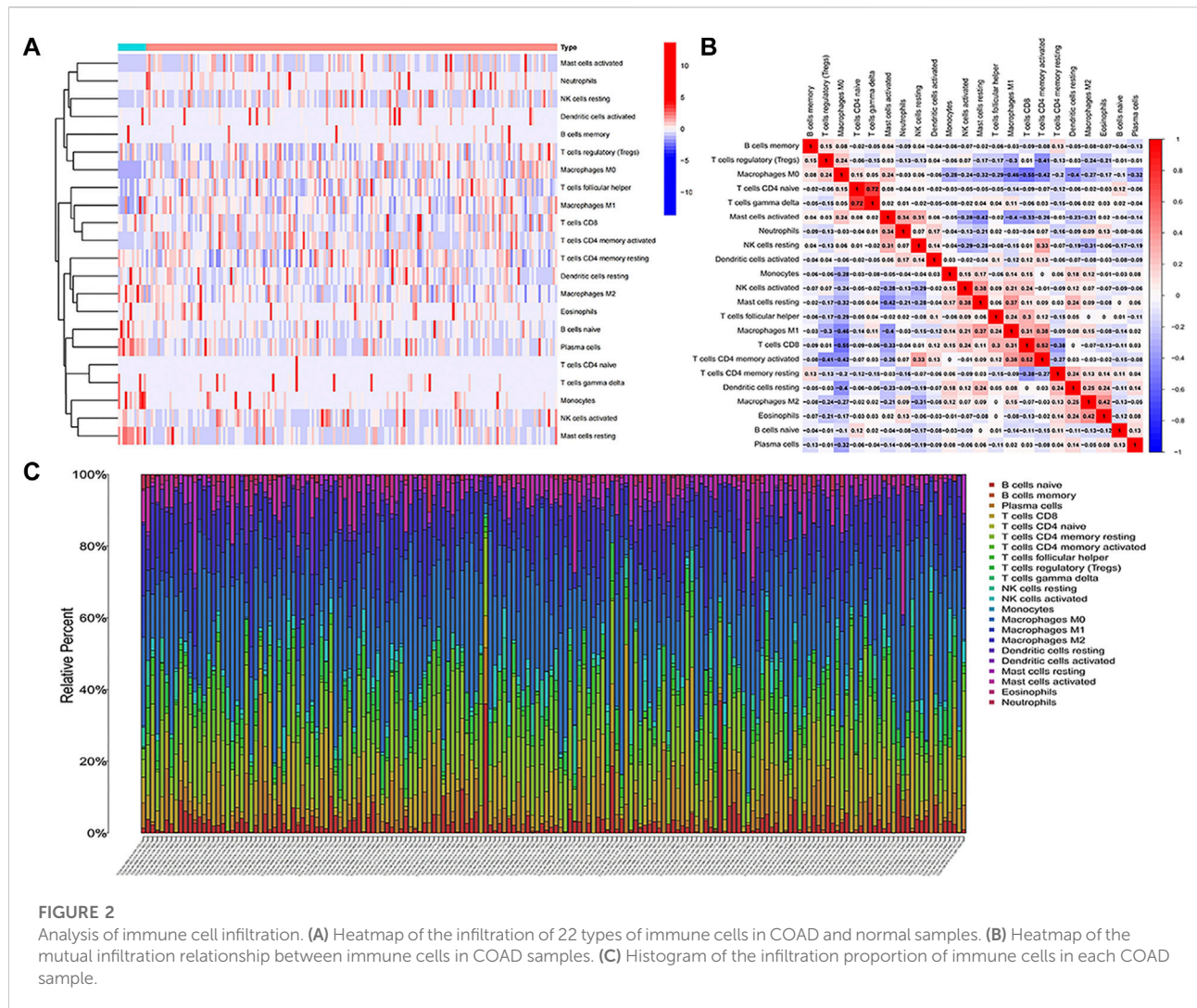
to eliminate the normal samples and genes with small fluctuation values in the data set and checked whether there is any deletion in the sample data and then removed the offending genes and samples from the data. Next, we detected outliers through sample clustering and eliminated them (Figure 3A). “`sft$powerEstimate`” function was used to determine the optimal soft-power threshold ($\beta = 5$), and the scale-free fit index of network topology was obtained by soft-thresholding power analysis (Figure 3B). Furthermore, we constructed a hierarchical clustering tree by using the dynamic shearing method and searched genes with similar expression data for modular clustering to generate a new hierarchical clustering tree (Figure 3C), and eight modules were generated. Finally, we draw the correlation plot between WGCNA modules and immune cells for further analysis (Figure 3D).

Identification of hub genes and establishment of a prognostic model

The highest correlation with the CD4⁺ MTC-related gene was found in the “greenyellow” module ($R^2 = -0.35$, $p < 0.0001$), including 929 DEGs in total. According to the gene names in the “greenyellow” module, we searched in the TCGA and GEO sets respectively to extract expression levels of each gene. Also, gene expression files were merged with clinical data (Supplementary Table S2 and Supplementary Table S3) to prepare for further analysis. Univariate Cox analysis was performed on the “greenyellow” module genes and identified 105 CD4⁺ MTC-related DEGs associated with the OS of COAD (Supplementary Table S5), and we preferentially showed 27 DEGs with $p < 0.005$ in the forest plot (Figure 4A). We further identified 12 genes by LASSO regression analysis (Figure 4B). Then, a Cox model (Gill, 1982) was constructed based on these 12 genes, and finally, five hub genes (TGFB2, DTNA, S1PR5, F2RL2, and MPP2) and a model formula were obtained. The model formula was used to calculate the risk score of the training set: risk score = $[TGFB2 \times 0.320,583] + [F2RL2 \times (-0.612,387)] + [DTNA \times 0.411,655] + [S1PR5 \times 0.447,946] + [MPP2 \times 0.718,418]$ (Table 1). According to the median value of the risk score, the samples of the training set were divided into high- and low-risk groups to predict the OS of COAD. Similarly, the samples of the testing set were also divided into two groups to verify the prediction accuracy of the prognostic model.

Predictive ability assessment of the prognostic model

We performed a K-M analysis between high- and low-risk groups on the prognostic model, and the result showed that patients with lower risk scores had a better outcome ($p <$

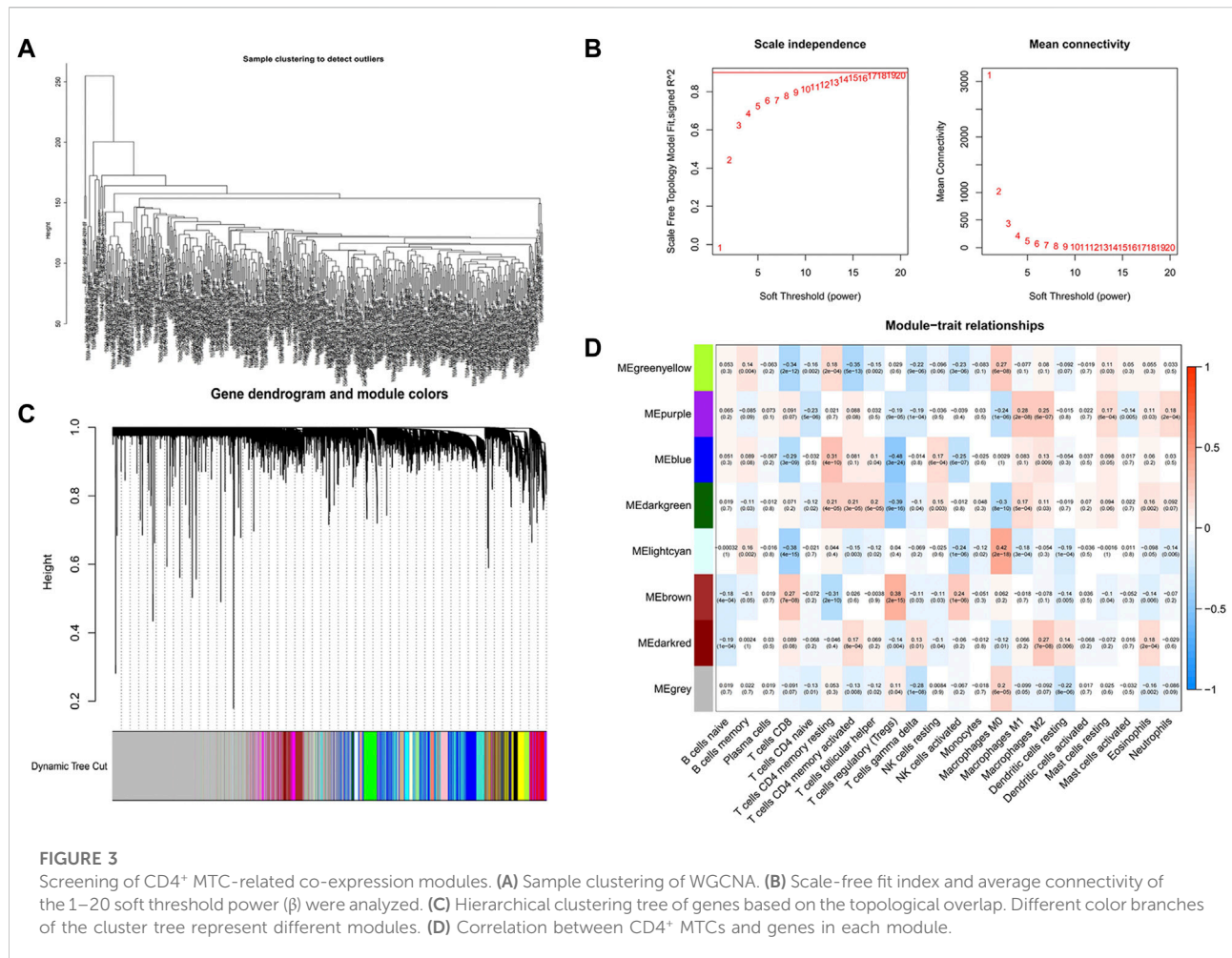


0.001, Figure 4C). This conclusion was also confirmed by the GEO testing set ($p = 0.002$, Figure 4D). As the risk score increased, the number of COAD patients who died increased accordingly, and patients at low-risk generally live longer than those at high risk (Figures 5A,B). In addition, in our prognostic model, F2RL2 was confirmed to be a low-risk gene, while TGFB2, DTNA, S1PR5, and MPP2 were high-risk genes (Figures 5C,D). K–M analysis proved that all five hub genes were closely related to OS of COAD, and patients with F2RL2 overexpression or SLC35G2, DTNA, S1PR5, and MPP2 low-expression had a better OS outcome (Figures 5E–I).

Meanwhile, univariate and multivariate Cox analyses showed that age, stage, and risk score were independent prognostic factors associated with the OS of COAD (Figures 6A,B). Considering the application of the prognostic model in clinical practice, we developed nomograms to predict the OS at 1, 3, and 5 years based on the baseline characteristics and pathological parameters of COAD. Each COAD was individually

matched with a nomogram, which fully embodied the individualization of clinical application. In the presented nomogram, the predicted 1-, 3-, and 5-year OS for this patient was 85.9%, 73.5%, and 54.7%, respectively (Figure 6C). Also, age, stage, and risk score were significantly associated with OS, which was consistent with the multivariate Cox analysis result. Subsequently, we use the calibration curve to verify the prediction accuracy of the nomogram. We can see that the prediction lines are very close to the diagonal dotted line, indicating that the nomogram has high accuracy (Figure 6D).

Furthermore, time-dependent-ROC analysis was performed to assess the predictive power of the risk score. Factors with an area under the curve (AUC) > 0.5 indicates that the prognostic model has a predictive value, and the greater the AUC, the higher the accuracy of prediction. Our analysis results showed that the prognostic model had good predictive value in predicting 1-, 3- and 5- year OS, and the prediction accuracy was 3- year (AUC: 0.720) >1- year (AUC:



0.718) >5- year (AUC: 0.692) (Figure 6E). Likewise, the predictive power of the prognostic model was superior to that of clinical predictors such as age and gender (Figure 6F).

Finally, we analyzed the clinical features between high- and low-risk groups in the prognostic model. The results showed that there were significant differences in gender and TNM stage between the two groups. The risk score of females was higher than that of males, and patients in stage IV had the highest risk score, while patients in stage I had the lowest risk score. However, it was worth noting that the risk scores between stage II and III ($p = 0.17$) and stage III and stage IV ($p = 0.16$) were not statistically significant (Figures 6G–I).

Analysis of the tumor microenvironment and immune cell function

We compared the prognostic model data with the pan-cancer immune cell infiltration data and analyzed the

correlation between risk score, hub genes, and immune cell infiltration by R package “limma”. The bubble plot shows the results of the analysis performed by different software (Figure 7A). A correlation coefficient greater than zero is considered a positive correlation and that less than zero is a negative correlation. Similarly, the scatter plot suggested that CD4⁺ MTC-related DEGs were negatively correlated with immune scores (Figure 7B). In addition, we analyzed the association of the TME with risk scores. The COAD TME score was calculated by the ESTIMATE algorithm (estimation of stromal and immune cells in malignant tumor tissues using expression data). The violin plot reported significant differences in stromal cell scores between the high- and low-risk groups ($p < 0.01$), with higher scores in the high-risk group. But, there were no significant differences in immune cell score and ESTIMATE score between the two groups (Figure 7C).

Analysis of 22 types of immune cells showed that a total of seven types had significant differences in infiltration between the

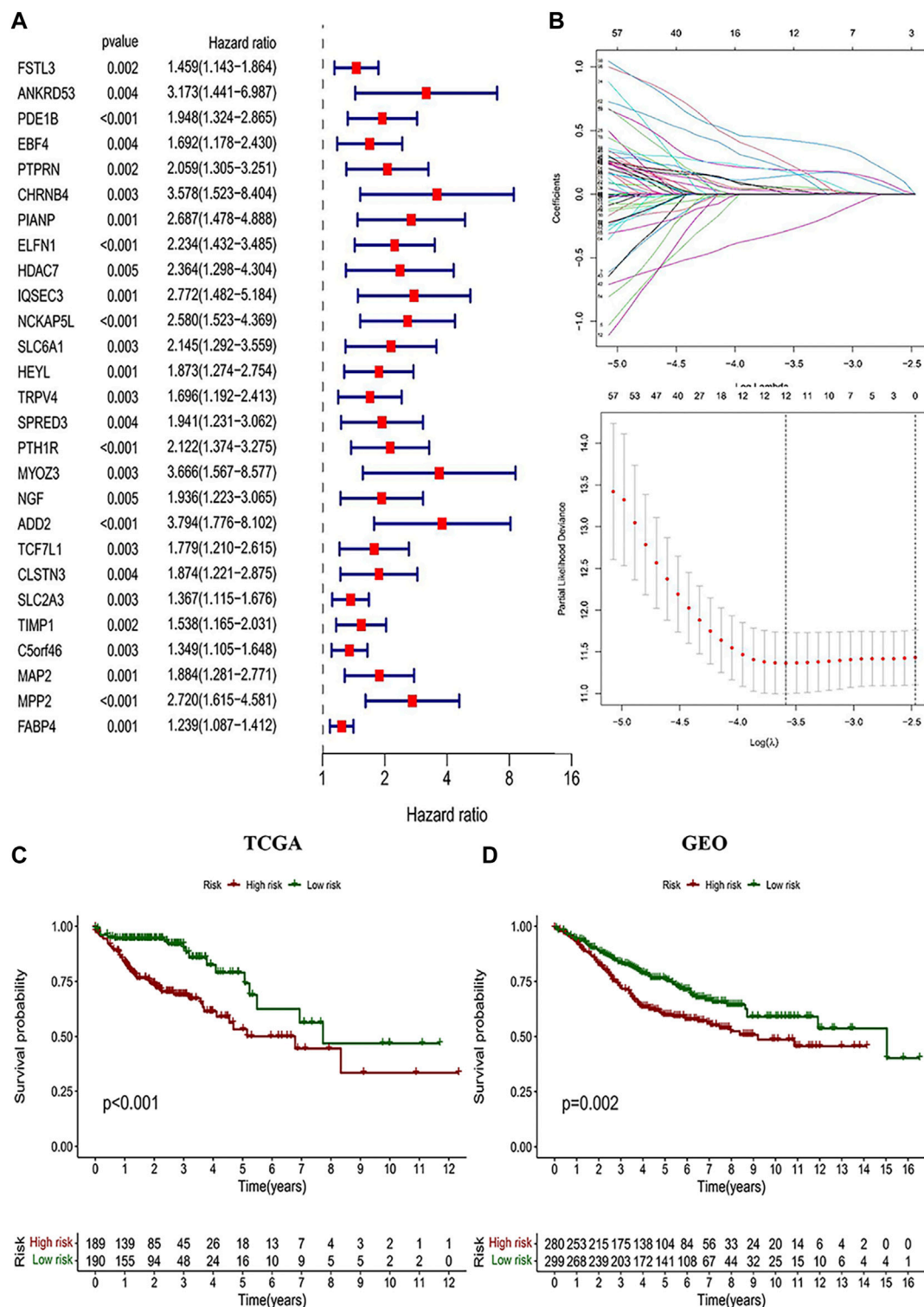


FIGURE 4 Identification of CD4⁺ MTC-related DEGs and construction of the prognosis model. (A) Univariate Cox regression analysis result of 103 prognosis-related genes in the study (27 genes with $p < 0.005$ were preferentially displayed). (B) Lasso regression and cross-validation showed that the number of genes corresponding to the point with the smallest error was 13. (C) COAD patients with low risk in the prognostic model predicted better OS outcomes. (D) GEO testing set was used for validation, and the results were consistent with those of the prognostic model.

TABLE 1 Information and the corresponding coefficients of hub genes.

Gene symbol	Full name	Coefficient
F2RL2	Coagulation factor II (thrombin) receptor-like 2	-0.612,387
TGFB2	Transforming growth factor- β 2	0.320,583
DTNA	Dystrobrevin-alpha	0.411,655
S1PR5	Sphingosine 1-phosphate receptor 5	0.447,946
MPP2	Membrane palmitoylated protein 2	0.718,418

high- and low-risk groups (Figure 7D, $p < 0.05$). “Plasma cells”, “T cells CD4 memory resting”, “T cells CD4 memory activated”, “dendritic cells resting” and “mast cells resting” infiltration was upregulated in the low-risk group, while “macrophages M0” and “neutrophil” infiltration was upregulated in the high-risk group. In addition, analysis of immune cell-related functions showed that “T cell co-stimulation” and “MHC_class_I” functions were more active in the low-risk group (Figure 7E, $p < 0.05$).

Immune checkpoint and immunotherapy

Comparing the hub genes and risk score with immune checkpoints to analyze their correlation, it is noteworthy to observe that immune checkpoints VTCN1, TNFSF4, TNFSF14, TNFRSF8, TNFRSF4, NRP1, LAIR1, CD40, CD70, CD276, and CD200 were positively correlated with risk scores, while TNFSF18, TMIGD2, ICOS, and HHLA2 were negatively correlated (Figure 8A). In the prediction of the therapeutic effect of immune checkpoint inhibitors (ICIs) in COAD, anti-CTLA-4 or anti-PD-1 or anti-CTLA-4 plus anti-PD-1 treatment was achieved better OS with a low-risk score ($p < 0.005$; Figures 8B–E). Meanwhile, we conducted a drug sensitivity analysis and concluded that bosutinib and cetuximab were more sensitive in COAD patients with low-risk scores (Figures 9A,B), while bleomycin (50 μ M), dasatinib, foretinib, midostaurin, pazopanib, saracatinib, shikonin, and talazoparib were more sensitive in the high-risk score (Figures 9C–I). The specific mechanism of action and targeted pathway of these drugs can be searched in Genomics of Drug Sensitivity in Cancer (<https://www.cancerrxgene.org/>). Finally, we analyzed the association of TME with risk scores.

Functional and pathway enrichment in the prognostic model

We conducted GO and KEGG enrichment according to the analysis results of DEGs between high- and low-risk groups. In the GO histogram (Figure 10A), we showed the number of risk-DEGs enriched in the three categories of GO (BP: Biological Process; CC: Cellular Component; MF: Molecular Function). In the GO bubble plot (Figure 10B), we showed the number and

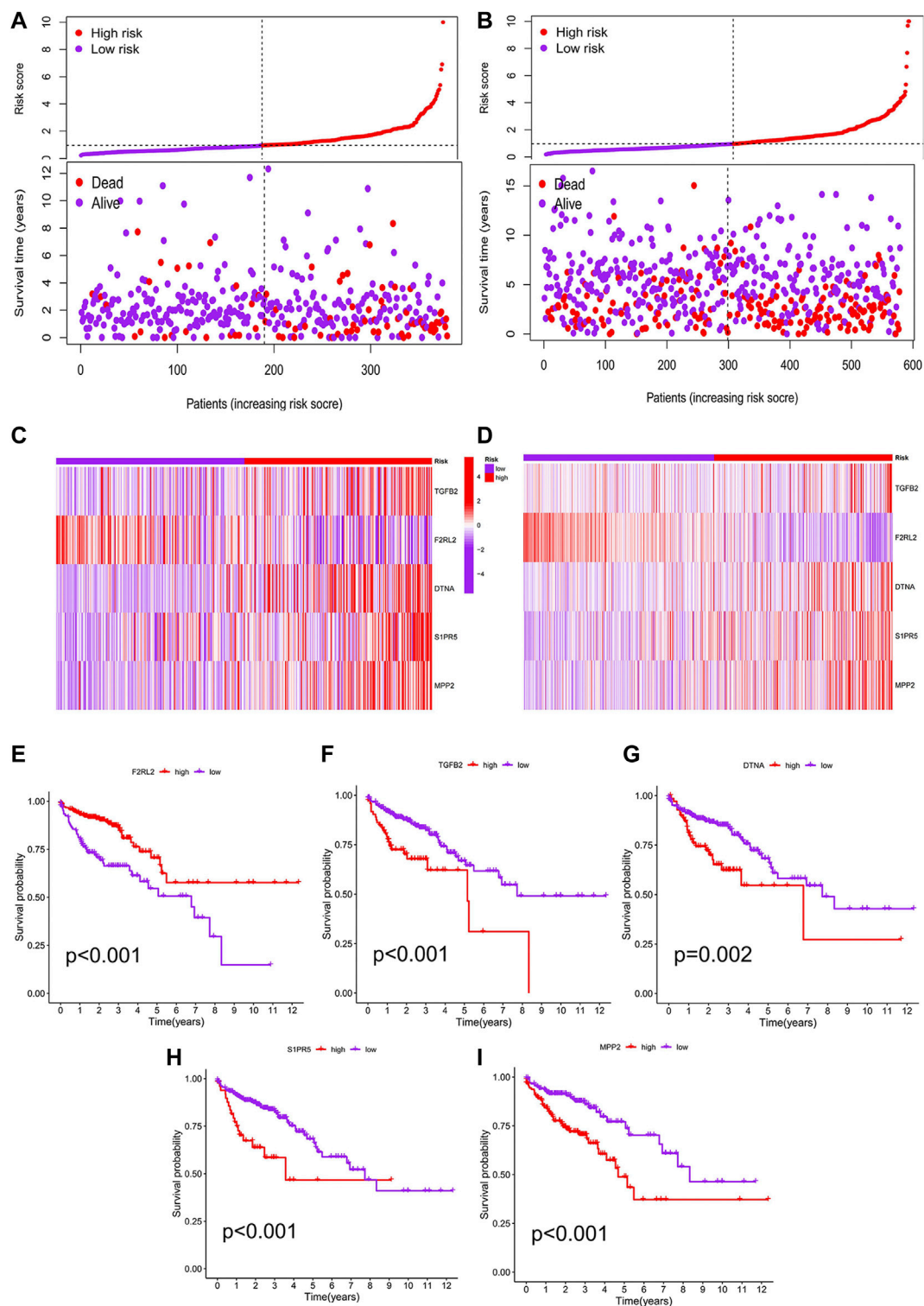
difference significance of the functional enrichment of risk-DEGs. In the circle plot (Figure 10C), the outermost circle represented the ID of GO, the second circle represented the number of genes enriched on each GO term, the third circle represented the number of DEGs enriched on each GO term, and the inner circle represented the proportion of genes. In the abovementioned plots, we showed 30 GO functions that were significantly associated with risk-DEGs and the number of genes enriched for each function. However, this was different from the results of GO enrichment analysis. In addition, we performed KEGG analyses of hub genes and risk scores to assess the association of these factors with the KEGG pathway and visualized the results with heat maps (Figure 10D).

Association of tumor mutational burden with risk score

The TMB analysis was performed through R packages “BiocManager”, “ggpubr” and “maftools”. In the waterfall plot, we can observe that COAD in the low-risk group (100%) has higher gene mutation frequencies than that in the high-risk group (99.39%) and show the proportion of the top 30 genes with the highest mutation frequency (Figures 11A,B). However, the correlation analysis of TMB with risk score showed no significant difference between high- and low-risk groups (Figures 11C,D). Furthermore, K-M analysis was used to predict the OS of patients with TMB, and we observed that patients with low TMB have better OS ($p = 0.04$; Figure 11E). Also, we found that TMB combined with risk score had a significant difference in evaluating the prognosis of patients. Patients with high-TMB/low-risk had the best prognosis, followed by low-TMB/low-risk, low-TMB/high-risk, and high-TMB/high-risk ($p < 0.001$; Figure 11F).

Discussion

Colorectal cancer is one of the most common types of malignancies, with the third highest morbidity and mortality in both males and females (Siegel et al., 2021). Although the

**FIGURE 5**

Correlation analysis between risk score and prognosis. **(A)** Patients in the prognostic model were divided into high- and low-risk groups according to the median value of risk score. With the increase in risk score, the number of deaths increased correspondingly, and the survival prognosis of patients with low risk was better. **(B)** Conclusions of the prognostic model were validated by the GEO testing set. **(C)** F2RL2 was confirmed to be a low-risk gene, while TGFB2, DTNA, S1PR5, and MPP2 were high-risk genes. **(D)** Conclusions of hub genes were validated by the GEO testing set. **(E–I)** K-M analysis results of five hub genes: F2RL2, TGFB2, DTNA, S1PR5, and MPP2.

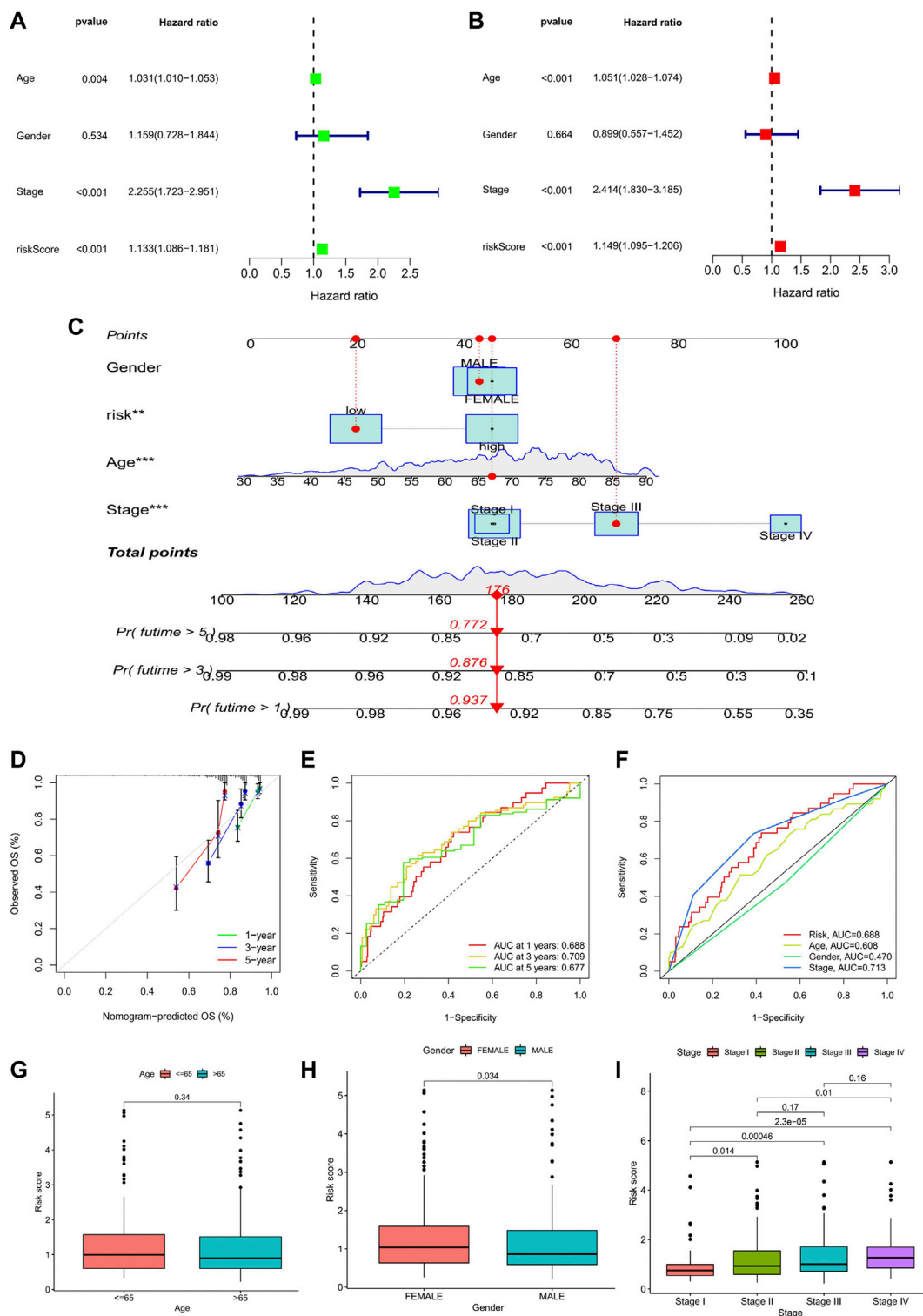
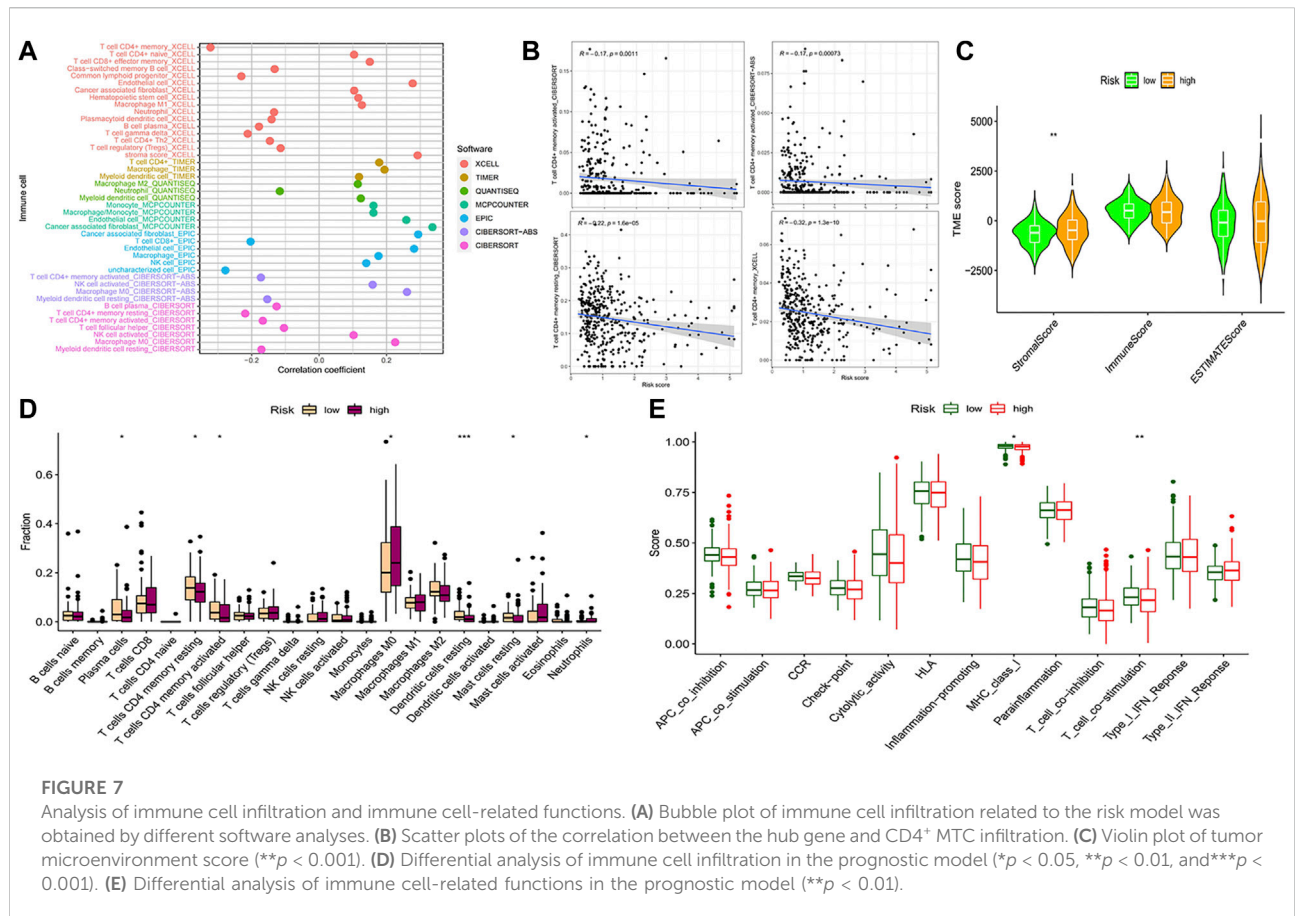


FIGURE 6

Analysis of prognostic factors and construction of nomogram. (A) Univariate analysis identified prognostic factors, and (B) multivariate analysis identified prognostic factors as independent predictors. (C) Nomogram used to predict the 1-, 3-, and 5-year survival rate in clinical medicine. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (D) Calibration curve to verify the accuracy of nomogram prediction. (E) ROC curves for OS prediction accuracy. (F) ROC curves for risk score and clinical features. (G–I) Analysis of clinical features between high- and low-risk groups in the prognostic model: age; gender; TNM stage.

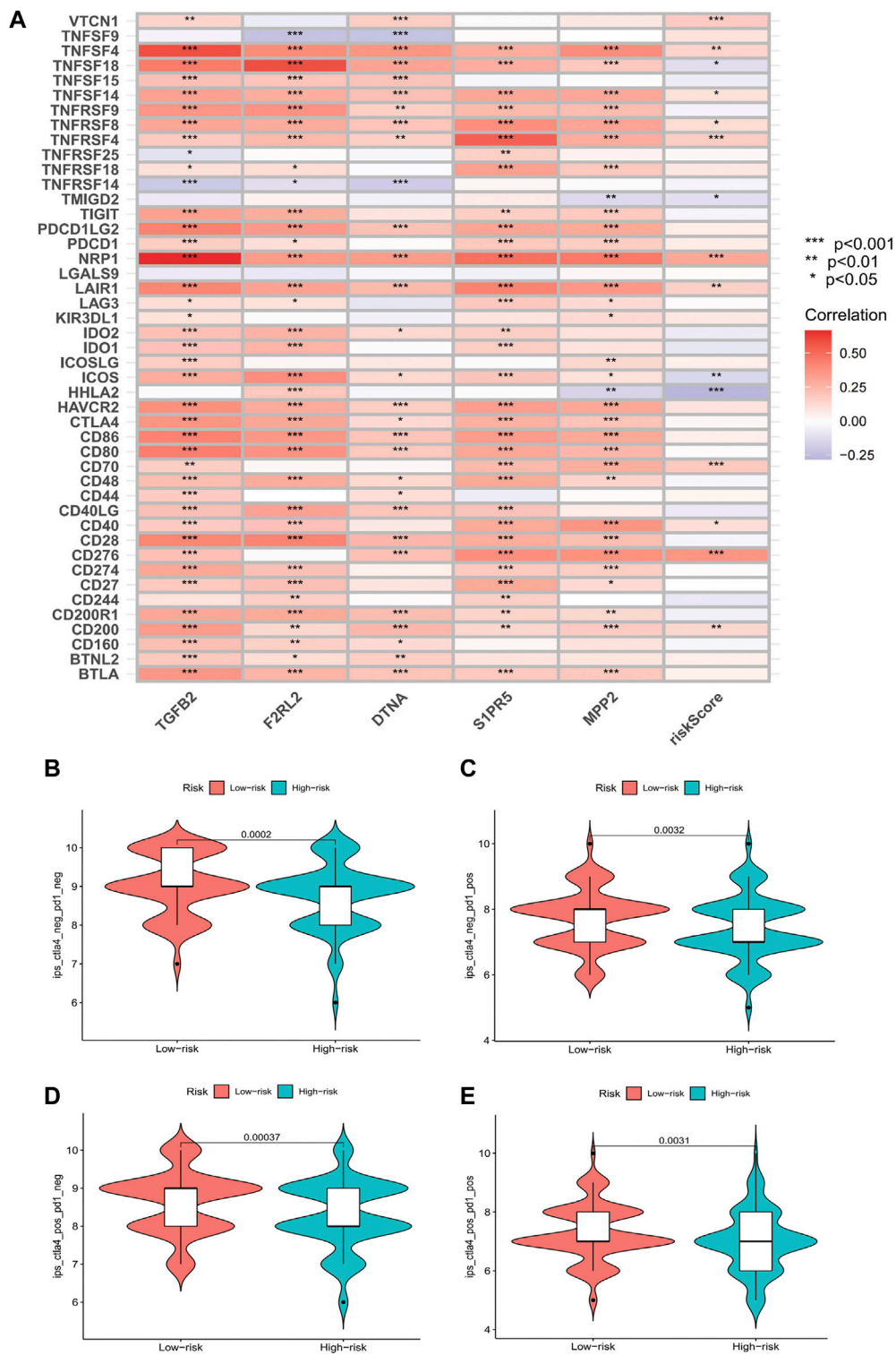


overall morbidity and mortality have decreased year by year, the incidence has shown an upward trend in patients < 50 years of age (Cheng et al., 2011; Bailey et al., 2015; Siegel et al., 2017). Surgery, adjuvant chemotherapy, targeted therapy, and immunotherapy are the main options for treatment of colon cancer (Benson et al., 2018). With the younger age of colon cancer incidence population, the individualization and precision of treatment strategies are promising. Recent studies have pointed out that the TME plays a key role in cancer proliferation, invasion, and metastasis and helps predict the prognosis of the disease (Wang et al., 2021). CD4⁺ MTCs in the microenvironment can make a rapid and direct immune response to protect the host against the invasion of cancer cells (Hirahara et al., 2021). These findings suggest that it is a potential therapeutic target (Quail and Joyce, 2013).

In our study, we constructed a co-expression network based on gene expression in 398 TCGA-COAD samples. Through the “CIBERSORT” algorithm and WGCNA analysis, we preliminarily determined that the “greenyellow” module containing 929 DEGs was most significantly associated with CD4⁺ MTC infiltration. Subsequently, the optimal five hub genes (F2RL2, TGFB2, DNMTA, S1PR5, and MPP2) were identified by univariate Cox-LASSO regression analysis, and

model formulas were obtained. According to the model formula, TCGA-COAD patients were divided into high- and low-risk groups, and the prognostic risk model was constructed. In addition, K-M analysis was performed on the prognostic risk model to assess whether there was a significant difference in OS between the high- and low-risk groups. The result showed that patients in the low-risk group benefited more from OS and had a better prognosis than those in the high-risk group. More importantly, we validated this result with the GEO testing set, and K-M analysis also showed that patients in the low-risk group predicted better OS. This conclusion is consistent with the analysis results of our prognostic risk model, indicating that the TCGA prognostic risk model constructed by us has high accuracy in predicting the OS of COAD patients, which provides an important reference value for our clinical application.

F2RL2 is a G protein-coupled receptor that regulates protease-activated receptor-3 involved in inflammatory and immune responses (Zhou et al., 2019). It has been reported as a prognostic marker for oral squamous cell carcinoma (Huang SN et al., 2020), metastatic breast cancer (Liu et al., 2021), pancreatic cancer (Chen et al., 2022), and glioma (Lvu et al., 2020), and down-regulated F2RL2 expression has been detected in rectal cancer (Supiot et al., 2013). TGFB2 is a protein-coding



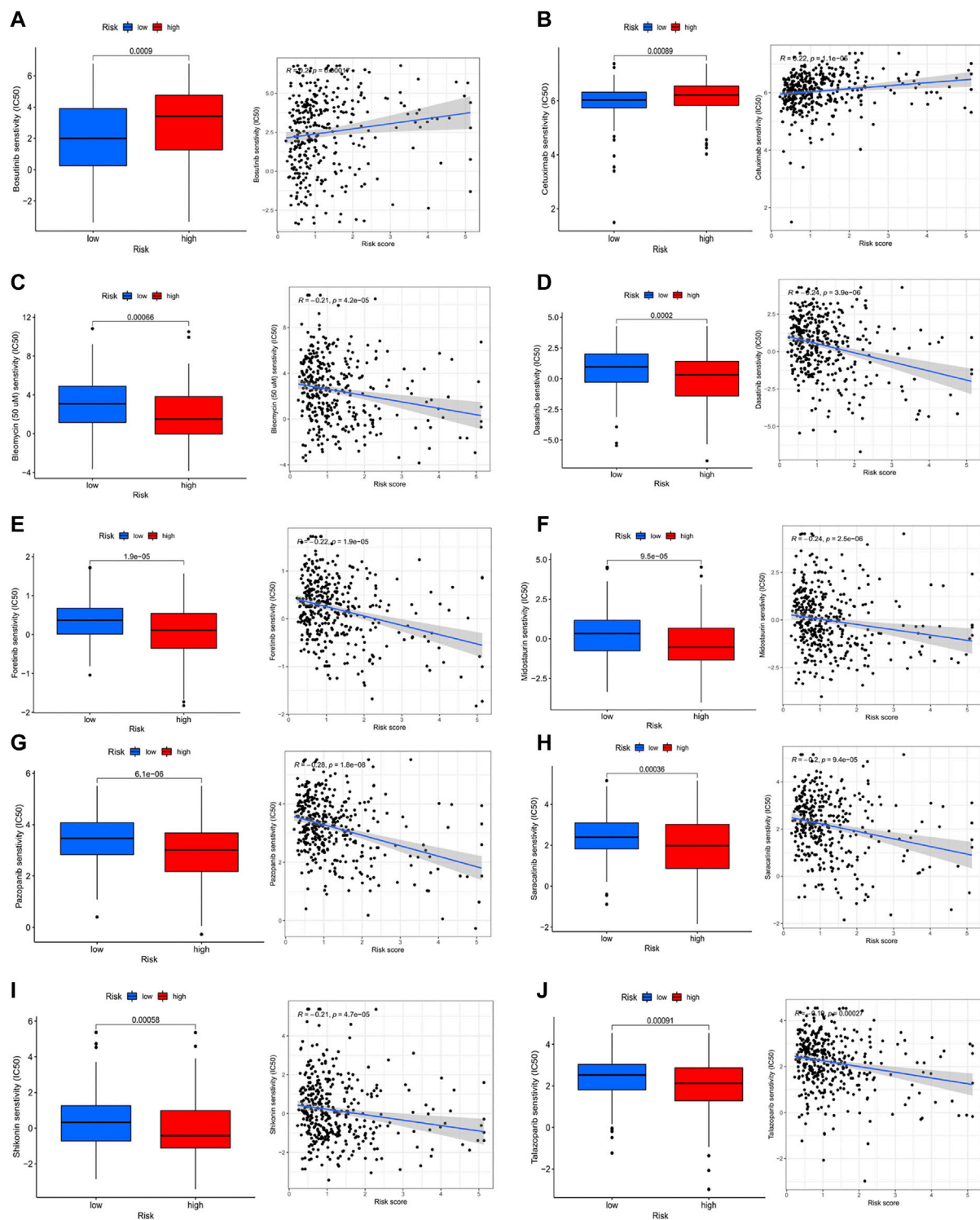
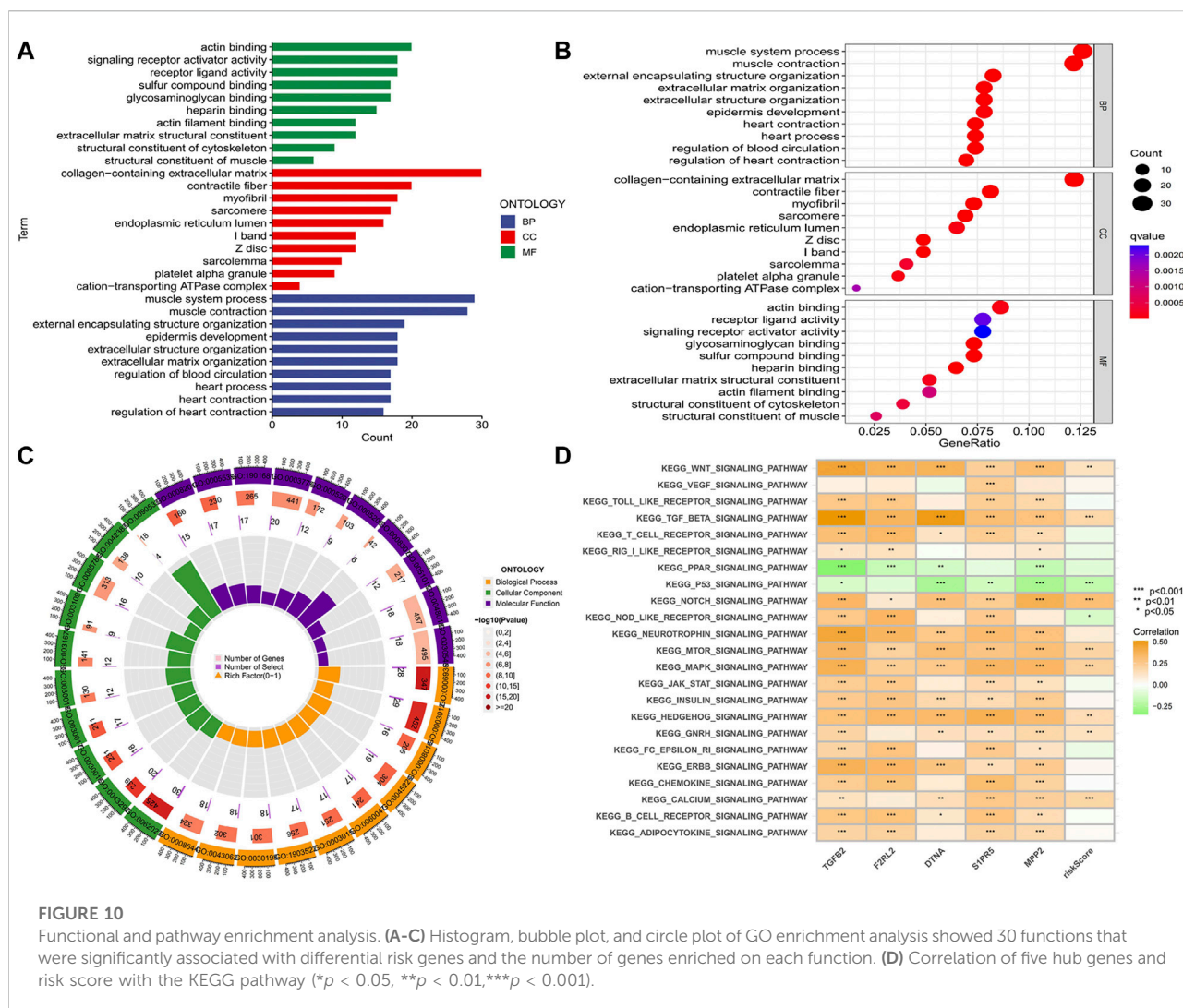


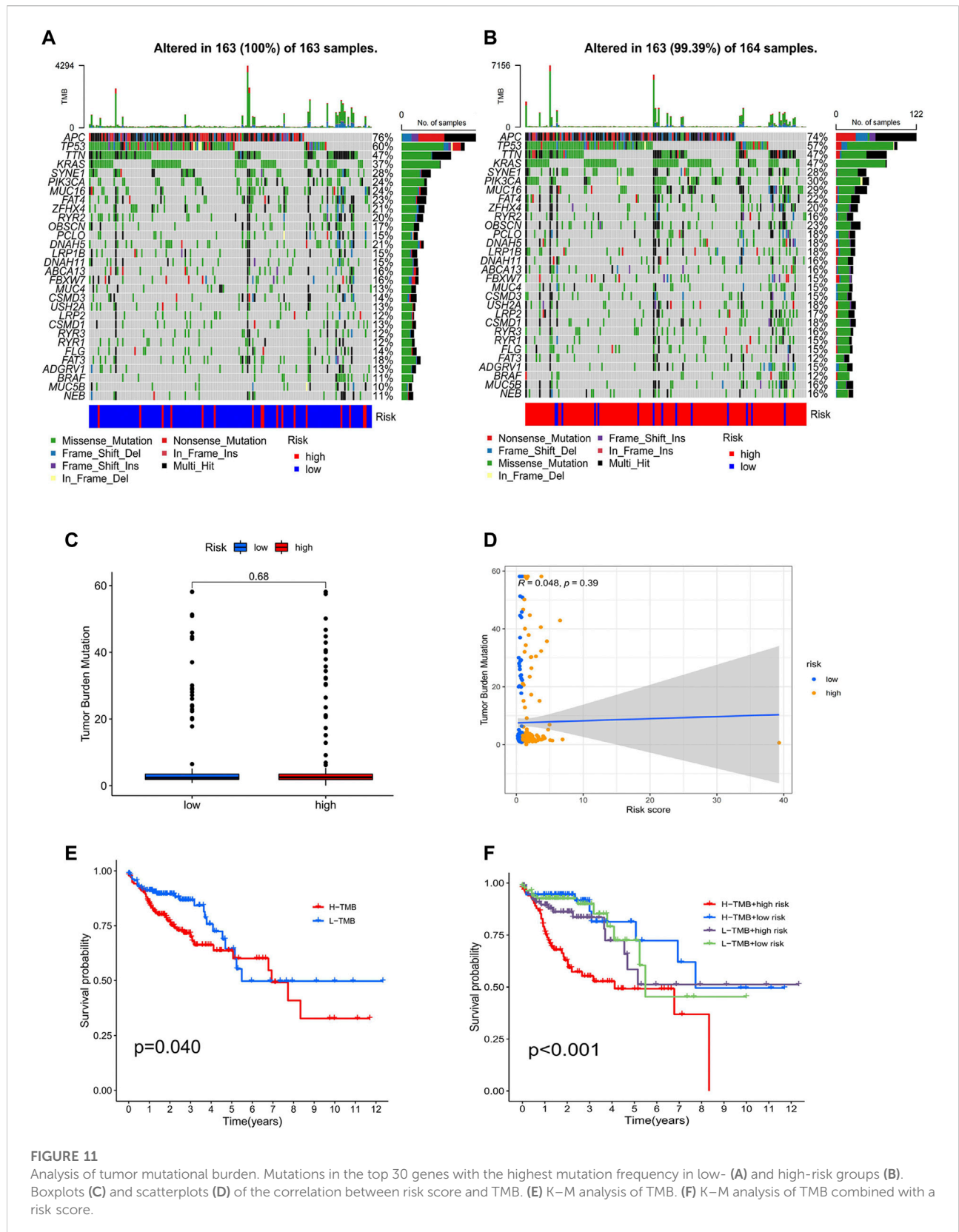
FIGURE 9

Analysis of drug sensitivity. (A–J) bosutinib; cetuximab; bleomycin (50 uM); dasatinib; foretinib; midostaurin; pazopanib; saracatinib; shikonin; talazoparib.



gene of TGF- β 2. It is considered to be the most critical factor in epithelial–mesenchymal transition and is involved in the key biological processes of growth, proliferation, migration, and invasion of malignant cells (Dave et al., 2011; Abraham et al., 2018; Yang et al., 2020), especially in suppressing antitumor immune response. Recent studies have shown that TGFB2 expression is upregulated in gastric cancer, non–small cell lung cancer, gallbladder cancer, colorectal carcinoma, and high-grade glioma, which is associated with poor prognosis (Bogdahn et al., 2011; Zhang et al., 2020; Liao et al., 2021; Song and Zhou, 2021). Therefore, targeted TGFB2 therapy for cancer patients may be a promising strategy. Currently, several therapies that specifically inhibit TGFB2, such as antisense phosphorothioate oligodeoxynucleotide trabedersen (AP12009), have entered clinical development in patients with advanced cancers (Jaschinski et al., 2011). DTNA is a scaffold protein that maintains the structural integrity of the heart and skeletal

muscle (Cao et al., 2017) and has been proven to predict the survival prognosis of bladder cancer (Zhang et al., 2021), hepatocellular carcinoma (Huang SN et al., 2020), gastric adenocarcinoma (Qin et al., 2019) and esophageal cancer (Fu et al., 2021). Liu et al. (Liu et al., 2017) also found that DTNA had a reference value in early colon cancer screening. S1PR5 is a G protein-coupled receptor, belonging to one of the five subtypes of S1PRs, which is distributed in many tissues and cells of the human body, especially in immune cells (Takabe and Spiegel, 2014; Patmanathan et al., 2017). Also, d Zhou et al. showed that up-regulation of S1PR5 could activate the NF- κ B/IDO1 signaling pathway and promote the progression of colon cancer (Zhou H et al., 2021). MPP2 is a scaffold protein belonging to the membrane-associated guanylate kinase-P55 (MAGUK) subfamily. MPP2 is closely associated with cell adhesion and is critical for the formation of multiprotein complexes involved in cell–cell communication. MPP2 is known for its junctional function



in epithelial cells (Baumgartner et al., 2009; Rademacher et al., 2016). At present, the role and mechanism of MPP2 in cancer are rarely reported. In the only two *in vitro* studies on MPP2, Li et al. (2021) confirmed that MPP2 can be up-regulated through miR-34a demethylation, promoted liver cancer cell apoptosis, and reduced proliferation, migration, and invasion. Similarly, Maschietto et al. reported that the MPP2 gene was downregulated in relapse Wilms tumors (Maschietto et al., 2011). In our prognostic risk model, overexpression of MPP2 was associated with a higher risk score and worse OS. Although this conclusion is different from the abovementioned results of liver cancer or Wilms tumors, there is no relevant report on the role of MPP2 in the occurrence and development of colon cancer. Therefore, the relationship between MPP2 and colon cancer and its specific mechanism remains to be explored.

Our findings showed that F2RL2 was downregulated in high-risk patients and considered to be a protective gene, while TGFB2, DTNA, S1PR5, and MPP2 were up-regulated in high-risk patients, suggesting that they were associated with poor prognosis. In our risk model, age, stage, and risk score were independent prognostic factors for COAD. The older the age, the higher the risk score, the higher the cancer stage, and the worse the prognosis. According to the gene expression level in the selected panel, we drew a nomogram to evaluate the survival time of patients. Each sample can get the corresponding nomogram through software, which is almost cost-free, portable, and intuitive in clinical applications.

Current treatments for colon cancer include surgery, chemotherapy, and targeted therapy. However, existing clinical studies have reported that EGFR inhibitor treatment is not conducive to long-term survival and disease remission (Cunningham et al., 2004; Lièvre et al., 2006; Van Emburgh et al., 2014). Therefore, the exploration of immunotherapy is highly expected. ICI therapy has shown promising results in melanoma and lung cancer (Lichtenstern et al., 2020). In 2019, ICI drugs were approved for breast cancer (Schmid et al., 2020). However, there are few reports of other cancers benefiting from ICI therapy. Nowadays, there is no large-scale transcriptome data for colon cancer immunotherapy, so we use RNA-seq in the TCGA database to calculate Immunophenoscore and then verify the effectiveness of ICI treatment. Our analysis found that anti-CTLA-4 or anti-PD-1 or anti-CTLA-4/anti-PD-1 therapy was more effective in low-risk patients. In addition, we screened for immune checkpoints associated with risk scores. These results provide an idea for future immunotherapy studies and drug selection.

Although our CD4⁺ MTC infiltration prognosis model has achieved some important results, it must be admitted that there are still shortcomings. First, this is a retrospective analysis using the public database, and we still lack

validation of prospective studies. In addition, our study considered the influence of immune infiltration and gene mutation on the progression of colon cancer, but there are many other epigenetic modifications in the pathology of the disease. An analysis combining these factors may be of greater reference value. Furthermore, in future studies, the potential mechanisms of action associated with hub genes in our prognostic risk model need to be further validated *in vivo* and *in vitro*. It is worth mentioning that when using the GEO dataset to validate the TCGA prognostic risk model, the selection of different datasets may bring some deviation to our validation results. For example, the sample size of the GEO dataset is insufficient or some GEO datasets may study a specific race or different disease stages (AJCC stage or TNM stage) may be used among different GEO datasets. Therefore, in order to make the validation results more reliable, we need to consider the clinical characteristics and sample size of the samples comprehensively when selecting the GEO dataset.

Conclusion

In summary, we constructed a risk assessment model of five CD4⁺ MTC-related gene markers for COAD and drew the nomogram of the hub gene and clinical independent risk factors to assess immunotherapy efficacy, disease prognosis, and survival time of the patient. These results provide a reference for target selection and individualized immunotherapy of COAD.

Data availability statement

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

Author contributions

All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by QZ, YC, SYu, SYa, and WL. The first draft of the manuscript was written by LT, HC revised it critically for important intellectual content, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81872504).

Acknowledgments

We thank HC for providing technical guidance and all authors who participated in this study and thank the TCGA, GEO, and TCIA databases for providing valuable data sets. We also thank the National Natural Science Foundation of China for its financial support.

Conflict of interest

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

References

- Abraham, C. G., Ludwig, M. P., Andryskiv, Z., Pandey, A., Joshi, M., Galbraith, M. D., et al. (2018). Δ Np63 α suppresses TGFB2 expression and RHOA activity to drive cell proliferation in squamous cell carcinomas. *Cell Rep.* 24 (12), 3224–3236. doi:10.1016/j.celrep.2018.08.058
- Angell, H. K., Bruni, D., Barrett, J. C., Herbst, R., and Galon, J. (2020). The immunoscore: colon cancer and beyond. *Clin. Cancer Res.* 26 (2), 332–339. doi:10.1158/1078-0432.CCR-18-1851
- Bailey, C. E., Hu, C. Y., You, Y. N., Bednarski, B. K., Rodriguez-Bigas, M. A., Skibber, J. M., et al. (2015). Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975–2010. *JAMA Surg.* 150 (1), 17–22. doi:10.1001/jamasurg.2014.1756
- Barresi, V., Reggiani Bonetti, L., Ieni, A., Caruso, R. A., and Tuccari, G. (2015). Histological grading in colorectal cancer: new insights and perspectives. *Histol. Histopathol.* 30 (9), 1059–1067. doi:10.14670/HH-11-633
- Baumgartner, M., Weiss, A., Fritzius, T., Heinrich, J., and Moelling, K. (2009). The PDZ protein MPP2 interacts with c-Src in epithelial cells. *Exp. Cell Res.* 315 (17), 2888–2898. doi:10.1016/j.yexcr.2009.07.028
- Benson, A. B., Venook, A. P., Al-Hawary, M. M., Cederquist, L., Chen, Y. J., Ciombor, K. K., et al. (2018). NCCN guidelines insights: colon cancer, version 2.2018. *J. Natl. Compr. Canc. Netw.* 16 (4), 359–369. doi:10.6004/jnccn.2018.0021
- Bogdahn, U., Hau, P., Stockhammer, G., Venkataramana, N. K., Mahapatra, A. K., Suri, A., et al. (2011). Targeted therapy for high-grade glioma with the TGF- β 2 inhibitor trabedersen: results of a randomized and controlled phase IIb study. *Neuro. Oncol.* 13 (1), 132–142. doi:10.1093/neuonc/nuq142
- Botía, J. A., Vandrovцова, J., Forabosco, P., Guelfi, S., D'Sa, K., Hardy, J., et al. (2017). An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* 11 (1), 47. doi:10.1186/s12918-017-0420-6
- Cao, Q., Shen, Y., Liu, X., Yu, X., Yuan, P., Wan, R., et al. (2017). Phenotype and functional analyses in a transgenic mouse model of left ventricular noncompaction caused by a DTNA mutation. *Int. Heart J.* 58 (6), 939–947. doi:10.1536/ihj.16-019
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12
- Chen, X., Robinson, D. G., and Storey, J. D. (2021). The functional false discovery rate with applications to genomics. *Biostatistics* 22 (1), 68–81. doi:10.1093/biostatistics/kxz010
- Chen, X., Yang, J., Lu, Z., and Ding, Y. (2022). A 70-RNA model based on SVR and RFE for predicting the pancreatic cancer clinical prognosis. *Methods* 204, 278–285. doi:10.1016/j.ymeth.2022.02.011
- Cheng, L., Eng, C., Nieman, L. Z., Kapadia, A. S., and Du, X. L. (2011). Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. *Am. J. Clin. Oncol.* 34 (6), 573–580. doi:10.1097/COC.0b013e3181fe41ed
- Choi, H., and Na, K. J. (2018). Integrative analysis of imaging and transcriptomic data of the immune landscape associated with tumor metabolism in lung

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.915282/full#supplementary-material>

- adenocarcinoma: clinical and prognostic implications. *Theranostics* 8 (7), 1956–1965. doi:10.7150/thno.23767
- Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., et al. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* 351 (4), 337–345. doi:10.1056/NEJMoa033025
- Dave, H., Trivedi, S., Shah, M., and Shukla, S. (2011). Transforming growth factor beta 2: a predictive marker for breast cancer. *Indian J. Exp. Biol.* 49 (11), 879–887.
- Deng, L., Lu, D., Bai, Y., Wang, Y., Bu, H., and Zheng, H. (2019). Immune profiles of tumor microenvironment and clinical prognosis among women with triple-negative breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 28 (12), 1977–1985. doi:10.1158/1055-9965.EPI-19-0469
- Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (5), R60. doi:10.1186/gb-2003-4-9-r60
- Di Fiore, F., Blanchard, F., Charbonnier, F., Le Pessot, F., Lamy, A., Galais, M. P., et al. (2007). Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. *Br. J. Cancer* 96 (8), 1166–1169. doi:10.1038/sj.bjc.6603685
- Ding, M., Li, F., Wang, B., Chi, G., and Liu, H. (2019). A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. *J. Cell. Biochem.* 120 (6), 10855–10863. doi:10.1002/jcb.28377
- Fu, G., Pei, Z., and Song, N. (2021). Oncogenic microRNA-301b regulates tumor repressor dystrobrevin alpha to facilitate cell growth, invasion and migration in esophageal cancer. *Esophagus* 18 (2), 315–325. doi:10.1007/s10388-020-00764-3
- Galon, J., Angell, H. K., Bedognetti, D., and Marincola, F. M. (2013). The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity* 39 (1), 11–26. doi:10.1016/j.immuni.2013.07.008
- Gill, R. (1982). Understanding Cox's regression model. *Exp. Suppl.* 41, 187–199.
- Gu, J., Zhang, J., Huang, W., Tao, T., Huang, Y., Yang, L., et al. (2020). Activating miRNA-mRNA network in gemcitabine-resistant pancreatic cancer cell associates with alteration of memory CD4(+) T cells. *Ann. Transl. Med.* 8 (6), 279. doi:10.21037/atm.2020.03.53
- Hirahara, K., Kokubo, K., Aoki, A., Kiuchi, M., and Nakayama, T. (2021). The role of CD4(+) resident memory T cells in local immunity in the mucosal tissue - protection versus pathology. *Front. Immunol.* 12, 616309. doi:10.3389/fimmu.2021.616309
- Huang R, R., Mao, M., Lu, Y., Yu, Q., and Liao, L. (2020). A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. *Aging (Albany NY)* 12 (8), 6966–6980. doi:10.18632/aging.103054
- Huang SN, S. N., Li, G. S., Zhou, X. G., Chen, X. Y., Yao, Y. X., Zhang, X. G., et al. (2020). Identification of an immune score-based gene panel with prognostic power for oral squamous cell carcinoma. *Med. Sci. Monit.* 26, e922854. doi:10.12659/MSM.922854

- Jaschinski, F., Rothhammer, T., Jachimczak, P., Seitz, C., Schneider, A., and Schlingensiepen, K. H. (2011). The antisense oligonucleotide trabedersen (AP 12009) for the targeted inhibition of TGF- β 2. *Curr. Pharm. Biotechnol.* 12 (12), 2203–2213. doi:10.2174/138920111798808266
- Kim, G. P., Colangelo, L. H., Wieand, H. S., Paik, S., Kirsch, I. R., Wolmark, N., et al. (2007). Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a national cancer institute-national surgical adjuvant breast and bowel project collaborative study. *J. Clin. Oncol.* 25 (7), 767–772. doi:10.1200/JCO.2006.05.8172
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Lech, G., Slotwiński, R., Słodkowski, M., and Krasnodębski, I. W. (2016). Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J. Gastroenterol.* 22 (5), 1745–1755. doi:10.3748/wjg.v22.i5.1745
- Li, F. Y., Fan, T. Y., Zhang, H., and Sun, Y. M. (2021). Demethylation of miR-34a upregulates expression of membrane palmitoylated proteins and promotes the apoptosis of liver cancer cells. *World J. Gastroenterol.* 27 (6), 470–486. doi:10.3748/wjg.v27.i6.470
- Liao, H., Liang, Y., Kang, L., Xiao, Y., Yu, T., and Wan, R. (2021). miR-454-3p inhibits non-small cell lung cancer cell proliferation and metastasis by targeting TGF β 2. *Oncol. Rep.* 45 (5), 67. doi:10.3892/or.2021.8018
- Lichtenstern, C. R., Ngu, R. K., Shalpour, S., and Karin, M. (2020). Immunotherapy, inflammation and colorectal cancer. *Cells* 9 (3), E618. doi:10.3390/cells9030618
- Lièvre, A., Bachet, J. B., Le Corre, D., Boige, V., Landi, B., Emile, J. F., et al. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* 66 (8), 3992–3995. doi:10.1158/0008-5472.CAN-06-0191
- Liu, J., Liu, F., Li, X., Song, X., Zhou, L., and Jie, J. (2017). Screening key genes and miRNAs in early-stage colon adenocarcinoma by RNA-sequencing. *Tumour Biol.* 39 (7), 1010428317714899. doi:10.1177/1010428317714899
- Liu, S., Song, A., Wu, Y., Yao, S., Wang, M., Niu, T., et al. (2021). Analysis of genomics and immune infiltration patterns of epithelial-mesenchymal transition related to metastatic breast cancer to bone. *Transl. Oncol.* 14 (2), 100993. doi:10.1016/j.tranon.2020.100993
- Lv, L. H., Lu, J. R., Zhao, T., Liu, J. L., and Liang, H. Q. (2022). A CD8(+) T cell-related genes expression signature predicts prognosis and the efficacy of immunotherapy in breast cancer. *J. Mammary Gland. Biol. Neoplasia* 27, 53–65. doi:10.1007/s10911-022-09510-0
- Lvu, W., Fei, X., Chen, C., and Zhang, B. (2020). *In silico* identification of the prognostic biomarkers and therapeutic targets associated with cancer stem cell characteristics of glioma. *Biosci. Rep.* 40 (8), BSR20201037. doi:10.1042/BSR20201037
- Maschietto, M., Piccoli, F. S., Costa, C. M., Camargo, L. P., Neves, J. I., Grundy, P. E., et al. (2011). Gene expression analysis of blastemal component reveals genes associated with relapse mechanism in Wilms tumour. *Eur. J. Cancer* 47 (18), 2715–2722. doi:10.1016/j.ejca.2011.05.024
- Migheli, F., and Migliore, L. (2012). Epigenetics of colorectal cancer. *Clin. Genet.* 81 (4), 312–318. doi:10.1111/j.1399-0004.2011.01829.x
- Narala, S., Li, S. Q., Klimas, N. K., and Patel, A. B. (2021). Application of least absolute shrinkage and selection operator logistic regression for the histopathological comparison of chondrodermatitis nodularis helices and hyperplastic actinic keratosis. *J. Cutan. Pathol.* 48 (6), 739–744. doi:10.1111/cup.13931
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337
- Ning, Z. K., Hu, C. G., Huang, C., Liu, J., Zhou, T. C., and Zong, Z. (2020). Molecular subtypes and CD4(+) memory T cell-based signature associated with clinical outcomes in gastric cancer. *Front. Oncol.* 10, 626912. doi:10.3389/fonc.2020.626912
- Obermeyer, Z., and Emanuel, E. J. (2016). Predicting the future - big data, machine learning, and clinical medicine. *N. Engl. J. Med.* 375 (13), 1216–1219. doi:10.1056/NEJMp1606181
- Panahi, B., and Hejazi, M. A. (2021). Weighted gene co-expression network analysis of the salt-responsive transcriptomes reveals novel hub genes in green halophytic microalgae *Dunaliella salina*. *Sci. Rep.* 11 (1), 1607. doi:10.1038/s41598-020-80945-3
- Patmanathan, S. N., Wang, W., Yap, L. F., Herr, D. R., and Paterson, I. C. (2017). Mechanisms of sphingosine 1-phosphate receptor signalling in cancer. *Cell. Signal.* 34, 66–75. doi:10.1016/j.cellsig.2017.03.002
- Pavlou, M., Ambler, G., Seaman, S., De Iorio, M., and Omar, R. Z. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat. Med.* 35 (7), 1159–1177. doi:10.1002/sim.6782
- Popat, S., and Houlston, R. S. (2005). A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *Eur. J. Cancer* 41 (14), 2060–2070. doi:10.1016/j.ejca.2005.04.039
- Qin, G., Yang, L., Ma, Y., Liu, J., and Huo, Q. (2019). The exploration of disease-specific gene regulatory networks in esophageal carcinoma and stomach adenocarcinoma. *BMC Bioinforma.* 20 (22), 717. doi:10.1186/s12859-019-3230-6
- Quail, D. F., and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394
- Rademacher, N., Schmerl, B., Lardong, J. A., Wahl, M. C., and Shoichet, S. A. (2016). MPP2 is a postsynaptic MAGUK scaffold protein that links SynCAM1 cell adhesion molecules to core components of the postsynaptic density. *Sci. Rep.* 6, 35283. doi:10.1038/srep35283
- Roth, A. D., Tejpar, S., Delorenzi, M., Yan, P., Fiocca, R., Klingbiel, D., et al. (2010). Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J. Clin. Oncol.* 28 (3), 466–474. doi:10.1200/JCO.2009.23.3452
- Russo, A., Bazan, V., Iacopetta, B., Kerr, D., Soussi, T., Gebbia, N., et al. (2005). The TP53 colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *J. Clin. Oncol.* 23 (30), 7518–7528. doi:10.1200/JCO.2005.00.471
- Schmid, P., Rugo, H. S., Adams, S., Schneeweiss, A., Barrios, C. H., Iwata, H., et al. (2020). Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet. Oncol.* 21 (1), 44–59. doi:10.1016/S1470-2045(19)30689-8
- Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A., et al. (2017). Colorectal cancer statistics, 2017. *CA Cancer J. Clin.* 67 (3), 177–193. doi:10.3322/caac.21395
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654
- Song, C., and Zhou, C. (2021). HOXA10 mediates epithelial-mesenchymal transition to promote gastric cancer metastasis partly via modulation of TGF β 2/Smad/METTL3 signaling axis. *J. Exp. Clin. Cancer Res.* 40 (1), 62. doi:10.1186/s13046-021-01859-0
- Su, R., Jin, C., Zhou, L., Cao, Y., Kuang, M., Li, L., et al. (2021). Construction of a ceRNA network of hub genes affecting immune infiltration in ovarian cancer identified by WGCNA. *BMC Cancer* 21 (1), 970. doi:10.1186/s12885-021-08711-w
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). 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. doi:10.3322/caac.21660
- Supiot, S., Gouraud, W., Campion, L., Jezéquel, P., Buecher, B., Charrier, J., et al. (2013). Early dynamic transcriptomic changes during preoperative radiotherapy in patients with rectal cancer: a feasibility study. *World J. Gastroenterol.* 19 (21), 3249–3254. doi:10.3748/wjg.v19.i21.3249
- Takabe, K., and Spiegel, S. (2014). Export of sphingosine-1-phosphate and cancer progression. *J. Lipid Res.* 55 (9), 1839–1846. doi:10.1194/jlr.R046656
- Tian, Z., He, W., Tang, J., Liao, X., Yang, Q., Wu, Y., et al. (2020). Identification of important modules and biomarkers in breast cancer based on WGCNA. *Oncotargets Ther.* 13, 6805–6817. doi:10.2147/OTT.S258439
- Van Emburgh, B. O., Sartore-Bianchi, A., Di Nicolantonio, F., Siena, S., and Bardelli, A. (2014). Acquired resistance to EGFR-targeted therapies in colorectal cancer. *Mol. Oncol.* 8 (6), 1084–1094. doi:10.1016/j.molonc.2014.05.003
- Wan, Q., Tang, J., Han, Y., and Wang, D. (2018). Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp. Eye Res.* 166, 13–20. doi:10.1016/j.exer.2017.10.007
- Wang, H., Yung, M. M. H., Ngan, H. Y. S., Chan, K. K. L., and Chan, D. W. (2021). The impact of the tumor microenvironment on macrophage polarization in cancer metastatic progression. *Int. J. Mol. Sci.* 22 (12), 6560. doi:10.3390/ijms22126560
- Yang, B., Bai, J., Shi, R., Shao, X., Yang, Y., Jin, Y., et al. (2020). TGF β 2 serves as a link between epithelial-mesenchymal transition and tumor mutation burden in gastric cancer. *Int. Immunopharmacol.* 84, 106532. doi:10.1016/j.intimp.2020.106532
- Yang, Z., Zi, Q., Xu, K., Wang, C., and Chi, Q. (2021). Development of a macrophages-related 4-gene signature and nomogram for the overall survival prediction of hepatocellular carcinoma based on WGCNA and LASSO algorithm. *Int. Immunopharmacol.* 90, 107238. doi:10.1016/j.intimp.2020.107238

Yang, Z., Yan, G., Zheng, L., Gu, W., Liu, F., Chen, W., et al. (2021). YKT6, as a potential predictor of prognosis and immunotherapy response for oral squamous cell carcinoma, is related to cell invasion, metastasis, and CD8+ T cell infiltration. *Oncoimmunology* 10 (1), 1938890. doi:10.1080/2162402X.2021.1938890

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* 377 (25), 2500–2501. doi:10.1056/NEJMc1713444

Zhang, X., Zhang, L., Chen, M., and Liu, D. (2020). miR-324-5p inhibits gallbladder carcinoma cell metastatic behaviours by downregulation of transforming growth factor beta 2 expression. *Artif. Cells Nanomed. Biotechnol.* 48 (1), 315–324. doi:10.1080/21691401.2019.1703724

Zhang, F., Wang, X., Bai, Y., Hu, H., Yang, Y., Wang, J., et al. (2021). Development and validation of a hypoxia-related signature for predicting survival outcomes in

patients with bladder cancer. *Front. Genet.* 12, 670384. doi:10.3389/fgene.2021.670384

Zhou, P. J., Wang, X., An, N., Wei, L., Zhang, L., Huang, X., et al. (2019). Loss of Par3 promotes prostatic tumorigenesis by enhancing cell growth and changing cell division modes. *Oncogene* 38 (12), 2192–2205. doi:10.1038/s41388-018-0580-x

Zhou, J., Guo, H., Liu, L., Hao, S., Guo, Z., Zhang, F., et al. (2021). Construction of co-expression modules related to survival by WGCNA and identification of potential prognostic biomarkers in glioblastoma. *J. Cell. Mol. Med.* 25 (3), 1633–1644. doi:10.1111/jcmm.16264

Zhou, H., Yin, X., Bai, F., Liu, W., Jiang, S., and Zhao, J. (2021). The role and mechanism of S1PR5 in colon cancer [retraction]. *Cancer Manag. Res.* 13, 5723–5724. doi:10.2147/cmar.s329184