Research Paper

Investigating gene signatures associated with immunity in colon adenocarcinoma to predict the immunotherapy effectiveness using NFM and WGCNA algorithms

Weizheng Liang[1,2,*], Xiangyu Yang[3,*], Xiushen Li[4,*], Peng Wang[2], Zhenpeng Zhu[2], Shan Liu[5], Dandan Xu[1], Xuejun Zhi[6], Jun Xue[2]

[1]Central Laboratory, The First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, Hebei, China
[2]Department of General Surgery, The First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, Hebei, China
[3]Department of Gastroenterology and Hepatology, The Second Affiliated Hospital of Chongqing Medical University, Yuzhong 400010, Chongqing, China
[4]Department of Obstetrics and Gynecology, Shenzhen University General Hospital, Shenzhen 518055, Guangdong, China
[5]Bioimaging Core of Shenzhen Bay Laboratory Shenzhen, Shenzhen 518132, Guangdong, China
[6]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, Hebei, China
*Equal contribution

Correspondence to: Weizheng Liang, Xuejun Zhi, Jun Xue; **email:** liangweizheng@hbbfyfy.com, zhixuejun@hbbfyfy.com, xuejun@hebeinu.edu.cn
Keywords: colon adenocarcinoma, immune microenvironment, WGCNA, immunotherapy, prognostic model
Received: November 17, 2023 **Accepted:** March 26, 2024 **Published:** May 13, 2024

ABSTRACT

Colon adenocarcinoma (COAD), a frequently encountered and highly lethal malignancy of the digestive system, has been the focus of intensive research regarding its prognosis. The intricate immune microenvironment plays a pivotal role in the pathological progression of COAD; nevertheless, the underlying molecular mechanisms remain incompletely understood. This study aims to explore the immune gene expression patterns in COAD, construct a robust prognostic model, and delve into the molecular mechanisms and potential therapeutic targets for COAD liver metastasis, thereby providing critical support for individualized treatment strategies and prognostic evaluation. Initially, we curated a comprehensive dataset by screening 2600 immune-related genes (IRGs) from the ImmPort and InnateDB databases, successfully obtaining a rich data resource. Subsequently, the COAD patient cohort was classified using the non-negative matrix factorization (NMF) algorithm, enabling accurate categorization. Continuing on, utilizing the weighted gene co-expression network analysis (WGCNA) method, we analyzed the top 5000 genes with the smallest p-values among the differentially expressed genes (DEGs) between immune subtypes. Through this rigorous screening process, we identified the gene modules with the strongest correlation to the COAD subpopulation, and the intersection of genes in these modules with DEGs (COAD vs COAD vs Normal colon tissue) is referred to as Differentially Expressed Immune Genes Associated with COAD (DEIGRC). Employing diverse bioinformatics methodologies, we successfully developed a prognostic model (DPM) consisting of six genes derived from the DEIGRC, which was further validated across multiple independent datasets. Not only does this predictive model accurately forecast the prognosis of COAD patients, but it also provides valuable insights for formulating personalized treatment regimens. Within the constructed DPM, we observed a downregulation of CALB2 expression levels in COAD tissues, whereas NOXA1,

KDF1, LARS2, GSR, and TIMP1 exhibited upregulated expression levels. These genes likely play indispensable roles in the initiation and progression of COAD and thus represent potential therapeutic targets for patient management. Furthermore, our investigation into the molecular mechanisms and therapeutic targets for COAD liver metastasis revealed associations with relevant processes such as fat digestion and absorption, cancer gene protein polysaccharides, and nitrogen metabolism. Consequently, genes including CAV1, ANXA1, CPS1, EDNRA, and GC emerge as promising candidates as therapeutic targets for COAD liver metastasis, thereby providing crucial insights for future clinical practices and drug development. In summary, this study uncovers the immune gene expression patterns in COAD, establishes a robust prognostic model, and elucidates the molecular mechanisms and potential therapeutic targets for COAD liver metastasis, thereby possessing significant theoretical and clinical implications. These findings are anticipated to offer substantial support for both the treatment and prognosis management of COAD patients.

INTRODUCTION

Colon adenocarcinoma (COAD), one of the most common malignancies, is among the top five in terms of morbidity and death from tumor-related illnesses [1–4]. Predictions indicate that in 2022, colorectal cancer (CRC) diagnoses will number 600,000 in China and 160,000 in the US, with 300,000 and 55,000 cases of CRC-related deaths in each country, respectively [5]. The colon is anatomically more positioned inside than the rectum, which complicates diagnosis and therapy. The lack of accurate COAD indicators, which shows that most colon cancer patients have missed the opportunity for dramatic surgery by the time they are officially diagnosed, is one of the main reasons for the poor prognosis of COAD [6]. CRC is a disease that is well suited for screening since early detection of precancerous lesions greatly reduces the disease's morbidity and death [7–10]. The overall survival of patients with COAD has not increased significantly despite significant breakthroughs in treatment [11]. Therefore, it is necessary to find prognostic biomarkers with high specificity or to create prognostic models with high predictive effect in order to oversee and guide the tailored treatment of COAD patients.

A variety of cytokines secreted during tumorigenesis and progression lead to the reprogramming of its surrounding stromal cells, which in turn promotes the proliferation and survival of tumor cells [12]. A significant factor in the formation, progression, and management of COAD is the interplay among immune cells, stromal cells, and cytokines within the tumor microenvironment [13]. In the tumor microenvironment, cytokines can be secreted by immune cells such as T cells, macrophages, and other immune cells, or produced by the tumor itself or stromal cells. These cytokines influence tumor growth, proliferation, and response to therapy by activating or inhibiting immune cells and controlling the tumor microenvironment [14, 15]. In addition to providing structural support, stromal cells, which include fibroblasts and vascular endothelial cells, control the recruitment and activation of immune cells through the secretion of cytokines [16]. In COAD, regulatory interactions between cytokines and immune cells may lead to the occurrence of immune escape, which promotes tumor cell proliferation and metastasis [17]. In the microenvironment of CRC, IL6 generated by $CD163^{+}$ tumor-associated macrophages stimulate epithelial mesenchymal transition by controlling the STAT3/miR-506-3p/FoxQ1 pathway, which in turn promotes CRC cell invasion and migration [18]. Meanwhile, the accumulation of cytokine IL-6 can also promote the proliferation of CRC cells [19, 20]. Immunogenetic traits are associated with a better prognosis or greater effectiveness of immunotherapy for malignancies [21, 22]. Clinical research indicates that immune checkpoint inhibition combined with divalizumab and trimethoprim may increase overall survival in patients with advanced refractory CRC [23]. Not all colon cancer patients respond to immunotherapy [24]. The specific mechanism of action of the tumor microenvironment remains unclear, despite its major impact on immune efficacy [25, 26]. Therefore, it is an urgent requirement to screen for novel indicators to forecast the effectiveness and post-treatment response of immunotherapy to enhance the individualization of immunotherapy.

In this study, COAD samples were immune clustered using the non-negative matrix clustering (NMF) method based on genes relevant to immunity. Immunological clustering-related gene modules were found using the weighted correlation network analysis (WGCNA) algorithm. The intersection of the gene modules with the differentially expressed genes (DEGs) (COAD vs Normal colon tissue) was defined as differentially expressed immune genes related with COAD (DEIGRC). The DEIGRC prognosis model (DPM) was constructed using a variety of bioinformatics tools, and the expected accuracy of the DPM was confirmed using data from the Gene Expression Omnibus (GEO) database. The capability of the model to forecast treatment outcomes for cancer patients was evaluated, and the distinct

immunological profiles among subgroups were characterized. The development of liver metastases in COAD patients was examined, along with possible processes and important therapeutic genes, using bioinformatics methods.

MATERIALS AND METHODS

Obtaining and processing IRGs and transcriptome sequencing data

The ImmPort database (https://www.immport.org/home) and InnateDB databases (https://www.innatedb.com/) provided the IRGs [27, 28]. Transcriptome sequencing data of 349 healthy colon and 471 COAD tissues were retrieved for this research work using UCSC Xena database (https://xena.ucsc.edu/). The "limma" package in R is used to obtain DEGs (|log (fold change)| > 1, p<0.05). The GEO database (https://www.ncbi.nlm.nih.gov/gds) was used to obtain the GSE17536, GSE39582, and GSE109211 datasets to verify prognostic models' accuracy in predicting outcomes. The development and validation of prognostic models did not include samples that lacked clinical prognostic information. The GSE6988 dataset was used to explore putative biological processes that might underlie the growth of liver metastases in COAD patients.

Immunophenotyping based on IRGs

We retrieved the expression data of IRGs from UCSC Xena database, which consisted of 471 samples from COAD patients. NMF analysis was performed on the screened data utilizing the 'brunet' criterion with the "NMF" package in R [29]. The ideal number of clusters was produced based on the results of the consensus clustering graph and residual sum of squares (rss), dispersion, and cophenetic graphs.

WGCNA analysis

We used the "limma" package of R to compare the expression differences of genes between the immune subtypes, and then used the 5000 genes with the smallest p-value for WGCNA by "WGCNA" package in R [30]. The p-value of different genes between different immune subtypes was calculated through "limma" and 5000 genes with the smallest p-value were used for WGCNA. The steps were as follows: (1) In the data preprocessing stage, genes with standard deviation less than 0.5 were excluded to reduce noise interference and improve the reliability of the subsequent analysis. (2) By calculating the Pearson correlation coefficient between genes, the degree of linear correlation between genes was assessed, thus laying the foundation for the establishment of co-expression network. (3) Construct similarity matrix and neighboring matrix according to Pearson correlation coefficient to measure the correlation and connection strength between genes. (4) Determine the optimal soft threshold using the "sft" function in R, thus determining the topology and module division of the network. (5) Construct a topological overlap matrix (TOM) using the adjacency matrix and cluster genes into different modules. Clarify gene clusters with intrinsic relatedness by 1-TOM similarity transformation. (6) Evaluate the correlation between gene modules and immune subtypes, and obtain gene modules with strong correlation with immune subtypes. (7) Intersect genes from gene modules highly correlated with immune subtypes with DEGs (COAD with normal colon tissue) and define these intersected genes as DEIGRC.

Bioinformatics analysis of DEIGRC

The "clusterProfiler" package and "enrichplot" package in R were employed to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and enrich the potential relationships between the analysis results.

Construction and validation of the DPM

The steps for the construction and validation of DPM were as follows by "survival" and "glmnet" package in R: (1) screen for DEIGRC associated with patient prognosis by univariate Cox regression; (2) reduce the number of genes to solve the multicollinearity problem by least absolute shrinkage and selection operator (LASSO) analysis; (3) construct DPM by multi-factor Cox regression; (4) calculate risk scores for patients with COAD in the UCSC Xena, GSE17536, and GSE39582 datasets; (5) demonstrate the relationship between risk scores and patient prognosis using Kaplan-Meier (KM) survival curves; (6) compare the predictive efficacy of prognostic models with that of constructing prognostic model genes using receiver operating characteristic (ROC) curves and concordance index (C-index).

Construction of nomogram

To verify that DPM was a risk factor independent of the patient's clinical traits, we performed multiple bioinformatics algorithms on risk scores and clinical traits in that order. The C-index was then applied to compare predictive efficacy. Finally, the DPM and clinical features nomogram was produced using the "rms" package in R.

Comparing differences in clinical characteristics between subgroups

The clinical data from the downloaded UCSC Xena database were compiled, and those lacking certain clinical features were eliminated. The distribution and differences in clinical traits between subgroups were visualized.

Comprehensive analysis of immunological profiles between subgroups

The "CIBERSORTx" package in R was used to calculate the proportions of 22 immune cell types in patients with COAD and to analyze the differences in the proportions of immune cells between subgroups. The best "cutoff" values for different immune cell proportions were then obtained using the surv_cutpoint and surv_categorize functions of the "survminer" package in R. The patients were categorized into subgroups with high proportions based on these values. The optimal "cutoff" values for different immune cell ratios were obtained by using the surv_cutpoint and surv_categorize functions in the "survminer" package, based on which the patients were categorized into high- and low-proportions subgroups, and the survival differences between the groups were compared by means of the Kaplan-Meier (KM) survival curve. In addition, the immune cell function scores of COAD patients were estimated with the help of the R software packages "GAVA" and "GSEABase", and patients were categorized into high- and low-scoring sub-groups based on the optimal "cutoff" values, and the survival differences between the subgroups were further evaluated. Finally, the immunomarker expression differences were compared between the subgroups.

Comparison of differences in immunotherapy and targeted therapy between subgroups

Tumor immune dysfunction and rejection (TIDE) scores can be utilized to predict the efficacy of immunotherapy. High TIDE scores predict high immune evasion potential, indicating that patients with tumors may not be suitable for immunotherapy. Transcriptomic data from COAD patients were uploaded to the TIDE database to calculate patients' T-cell-related scores. Compare the predictive efficacy of DPM, TIDE, and tumor inflammatory signature (TIS) models via ROC curves.

Bioinformatics analysis of molecular mechanisms for the development of liver metastases in COAD

Regarding metastasis, the liver is one of the most susceptible organs in COAD patients, and liver metastasis is one of the main causes of the high mortality rate among COAD patients. In this work, we sought to forecast the occurrence of liver metastases in COAD patients using the DPM. However, we could not do so with adequate accuracy. The GSE6988 dataset was employed to explore mechanisms of liver metastases in COAD. We screened DEGs in COAD samples with liver metastases and COAD samples without liver metastases. Build protein-protein interaction (PPI) network and find core genes through the GeneMANIA database and Cytoscape software. Perform GO and KEGG enrichment analysis of DEGs by R.

Acquisition of clinical tumor tissue

The First Affiliated Hospital of Hebei North University provided the COAD tissues and adjacent noncancerous tissue for this research. The tissues were collected within 30 minutes of the surgical specimens being separated. Connective and fatty tissue were removed on the edges of the fresh surgical specimens. Before being covered in RNA protective solution and kept at -80° C, the tissue surface was immediately cleaned of blood and grime with pre-cooled PBS solution or normal saline.

RT-qPCR

A sterile, enzyme-free EP tube with 300 μl of Solution R1 was filled with about 30 mg of tissue, which was then ground for 1-2 min. Centrifuge at 13,000 rpm for 15 seconds once there were no longer any visible tissue fragments. The supernatant was transferred to a new EP tube and mixed thoroughly with 500 μl of Solution R2. Centrifuge at 13,000 rpm for 30 seconds after adding the mixed liquid to the adsorption column to remove the waste product. Add 500 μl of RNA Wash Buffer to the adsorbent column, centrifuge at 13,000 rpm for 15 seconds, and repeat once. The cDNA Reverse Transcription Kit's instructions were followed to build the reverse transcription system, and reverse transcription was performed at 37° C for 15 min and then at 85° C for 5 seconds. Set up the reaction schedule after preparing the amplification system following the amplification kit's instructions. Cyclic reaction: 95° C, 10 s, then 60° C, 30 s, 40 cycles; pre-denaturation: 95° C, 30 s, 1 cycle.

Availability of data and material

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

RESULTS

Download results of IRGs and transcriptome datasets

1793 and 1226 IRGs were obtained from the ImmPort database (Supplementary Table 1) and InnateDB database (Supplementary Table 2), respectively, and 2660 IRGs were obtained after merging (Figure 1A). We obtained sequencing data from 349 normal samples and 471 COAD tissue samples from the UCSC Xena database and identified DEGs using this dataset. Furthermore, we acquired two COAD datasets, GSE17536 and

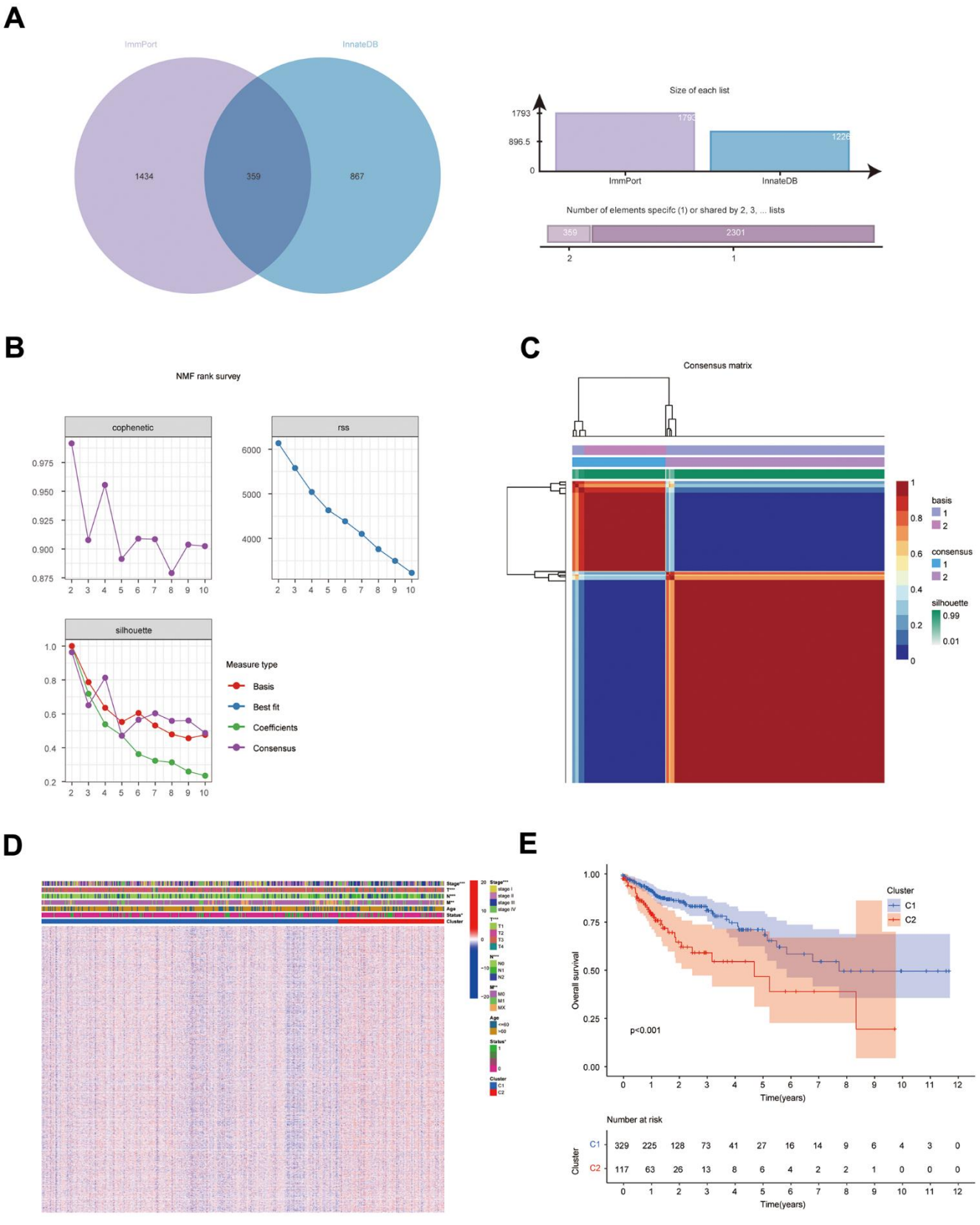


Figure 1. Results of NMF clustering analysis. (**A**) Venn diagram of IRGs in ImmPort database and InnateDB database; (**B**) Rss, dispersion, and cophenetic plots of the number of screened NMF clusters; (**C**) Consensus diagram of NMF clustering of IRGs; (**D**) Distribution of IRGs and clinical traits in immune subgroups; (**E**) KM survival curves for immune subgroups.

GSE39582, from the GEO database, consisting of chip data for 177 and 573 COAD samples, respectively. These datasets were utilized to validate the prognostic impact of the DEIGRC Prognostic Model (DPM) across different datasets. Additionally, we curated the GSE109211 dataset from the GEO database, encompassing chip data for 67 tumor samples subjected to targeted therapy. This dataset served to authenticate the accuracy and efficacy of our developed DPM in forecasting clinical benefits of targeted therapy.

Identification of immune subtypes in COAD

The NFM technique was employed to cluster the expression data of IRG from the COAD sample (UCSC Xena database). Cluster consensus maps from 2 to 10 (Supplementary Figure 1), rss maps, dispersion maps, and cophenetic maps (Figure 1B) were used to determine the optimal number of clusters 2 (Figure 1C), which resulted in the division of patients into cluster 1 (C1) and cluster 2 subgroups (C2). Heat maps were applied to display the expression and clinical traits of IRGs in the subgroups (Figure 1D). The KM survival analysis revealed that subgroup C1 had a better prognosis (Figure 1E).

Detection of important gene modules in subtypes

The DEGs (C1 *vs.* C2) obtained with the top 5000 *p*-values were then used for WGCNA. The "sft" function in R helped to obtain the best soft threshold 7 measurements (Figure 2A). Seven gene modules were created by calculating correlations between gene modules and subgroups and merging substantially identical gene modules (Figure 2B). With correlation coefficients above 0.6, the brown and green gene modules had the strongest connection with immunophenotyping. Volcano and heat maps demonstrate DEGs (normal colon tissues *vs.* COAD tissues) in the COAD patients from UCSC Xena database (Supplementary Figure 2A, 2B). Intersecting genes of gene modules with the highest subgroup correlation (green and brown) and DEGs (COAD) were defined as DEIGRC (Figure 2C).

Functional and pathway enrichment analysis of DEIGRC

Figure 3A–3D illustrates the enrichment results for GO and KEGG. Results of KEGG enrichment analysis demonstrated that DEIGRC was mainly enriched in the Calcium signaling pathway, Glutathione signaling pathway, cGMP-PKG signaling pathway, etc.

Construction and validation of DPM

58 DEIGRC associated with prognosis in COAD patients were obtained by univariate Cox regression analysis (Figure 4A, $p<0.05$). LASSO and multi-factor Cox regression analyses were performed sequentially on the 58 DEIGRC, resulting in the construction of the DPM consisting of 6 genes (Figure 4B–4D). CALB2, NOXA1, and TIMP1 were positively related to great poor prognosis, whereas KDF1, LARS2, and GSR were positively related to excellent prognosis (Supplementary Figure 3A–3F). In the COAD patients from UCSC Xena database, GSE17536 and GSE39582 datasets, the DEIGRC model displayed a strong ability to predict prognosis (Figure 4E–4G). The DEIGRC model had a greater prediction performance than the genes used to build it, according to the ROC curve and the C-index (Figure 4H, 4I).

Construction of nomogram

The results of KM survival curves indicated that Age was not associated with patient prognosis (Supplementary Figure 4A, $P=0.089$), while tumor (T), metastasis (M), node (N), and Stage were associated with patient prognosis (Supplementary Figure 4B–4E, $p<0.001$). DPM and clinical traits were sequentially analyzed by bioinformatics algorithms, and results revealed that DPM could be used as a prognostic indicator independent of clinical traits (Figure 5A–5D). Results of the C-index demonstrated that risk score had better predictive efficacy than clinical features in COAD patients (Figure 5E). DPM and clinical traits were further used to construct a nomogram for forecasting prognosis (Figure 5F).

Clinical characteristics of different subgroups of patients

After collecting and filtering the clinical information of COAD patients from the UCSC Xena database, heat maps were created to show the clinical traits of the subgroups. The distribution of T, M, N, and Stage between subgroups was substantially different (Figure 6A, $p<0.01$). Using the "ComplexHeatmap" package in R, it is possible to more clearly demonstrate the distribution of T, M, N, and Stage between subgroups (Figure 6B–6E).

Immunologic characteristics and therapeutic treatments of patients between different subgroups

Results of the 22 immune cell type ratios are displayed in Supplementary Figure 5A. The immune cell proportions and immune cell function scores for the various subgroups are shown in Figure 7A, 7B. The relationship between immune cell ratios and the prognosis of COAD patients was demonstrated through KM survival curves (Figure 7C–7I). The prognosis of high-score and low-score subgroups was connected with functional scores of

aDCs, APC co-inhibition, APC co-stimulation, and other immune cell function (Supplementary Figure 5B–5P). According to box plots of immunological marker expression, HLA-B, HLA-C, HLA-F, CD70, and TGFB1 were strongly upregulated and CD160, HAVCR-1, and ICOS were strongly downregulated in the high-risk subgroup (Figure 8A–8D).

Low-risk subgroup had larger microsatellite instability (MSI) scores, while the high-risk subgroup had larger TIDE, T-cell rejection, and T-cell dysfunction scores (Figure 9A–9D). We, therefore, considered that the DEIGRC model may be used to forecast which patients might benefit from immunotherapy. ROC curve results showed DPM had higher predictive efficacy than the TIDE and TIS models (Figure 9E). Patients who responded to sorafenib had lower risk scores than those who did not, so the DEIGRC model might be able to be used to predict the effectiveness of sorafenib treatment (Figure 9F).

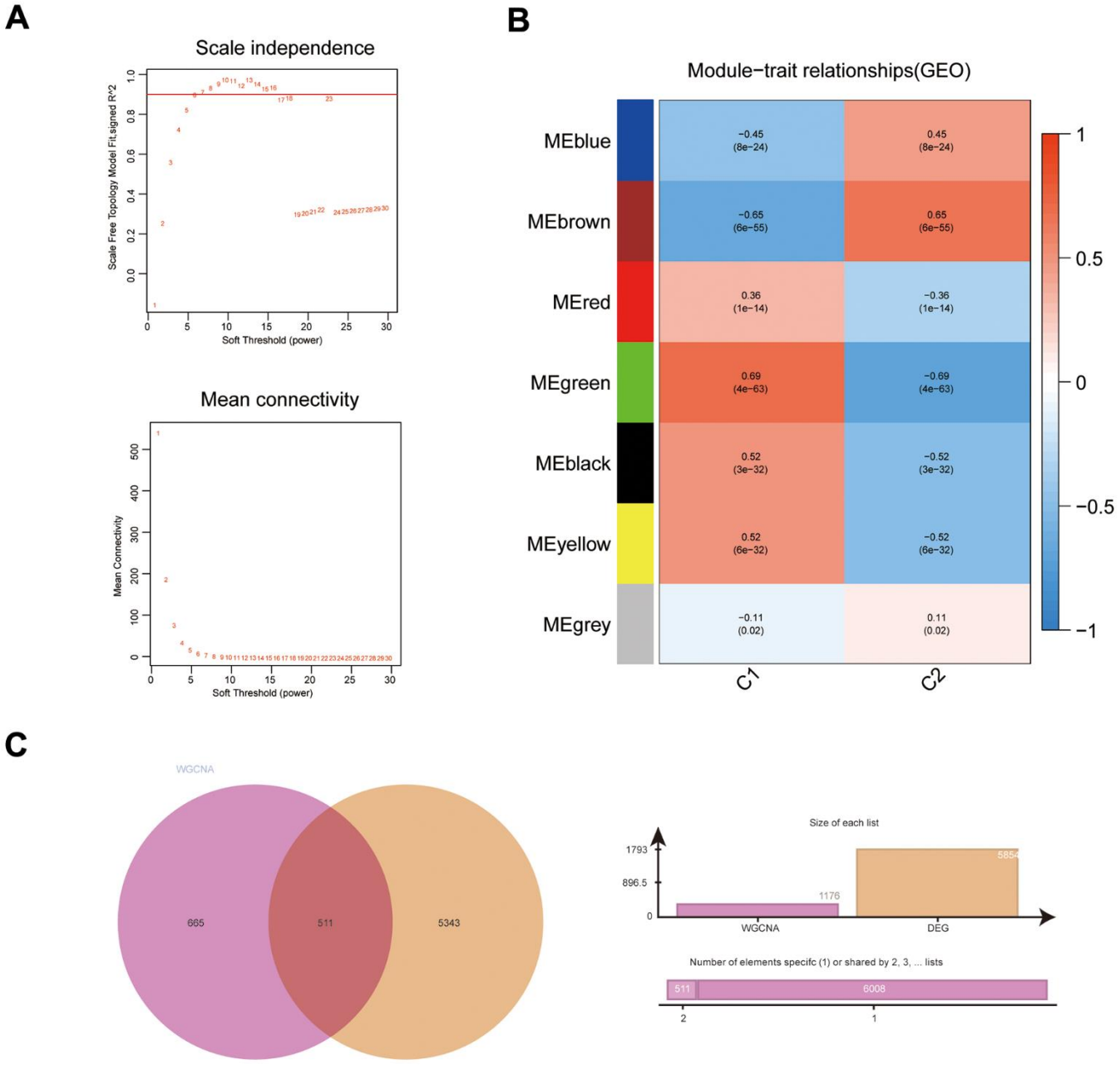


Figure 2. WGCNA results for immune subgroups. (**A**) Results of screening for soft threshold power; (**B**) correlation analysis of gene modules with immune subgroups; (**C**) Venn diagram of gene modules with high correlation to immunophenotyping and DEGs.

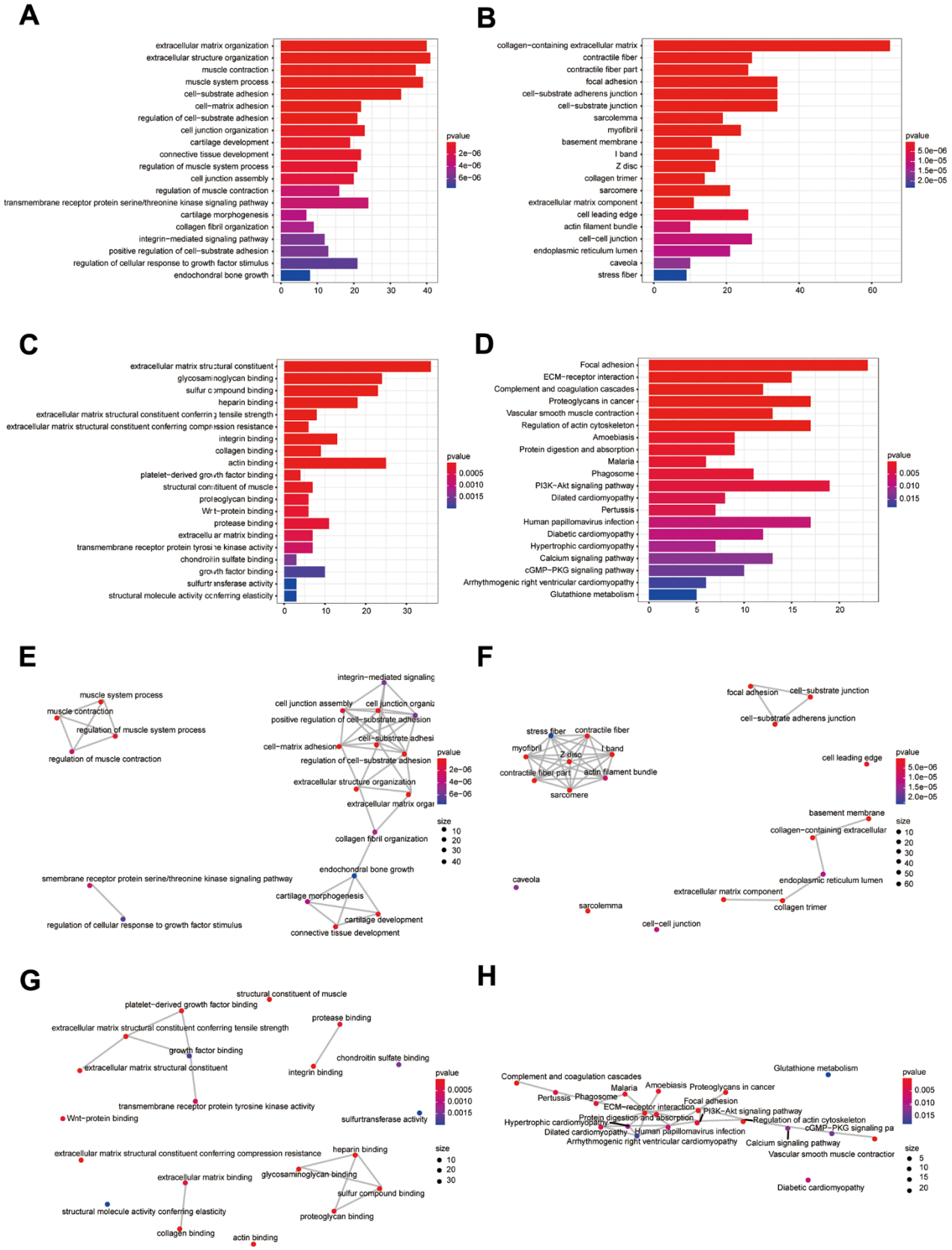


Figure 3. Bioinformatics analysis of DEIGRC. (**A**–**D**) Bar charts of BP, CC, MF, and KEGG enrichment analysis results; (**E**–**H**) Correlation analysis of the results from BP, CC, MF, and KEGG enrichment analysis.

Potential mechanisms for the development of liver metastases from COAD

DEGs in COAD samples with liver metastasis and COAD samples without liver metastasis were presented in Figure 10A. The heat map demonstrated expression distribution for the 25 most highly expressed genes and the 25 most lowly expressed genes in COAD with liver metastases samples (Figure 10B). The DEGs were applied to construct a PPI network, and core genes were further screened based on the degree of nodes using Cytoscape software (Figure 10C, 10D). The main genes contributing to the development of liver metastases in COAD patients include CAV1, ANXA1, CPS1,

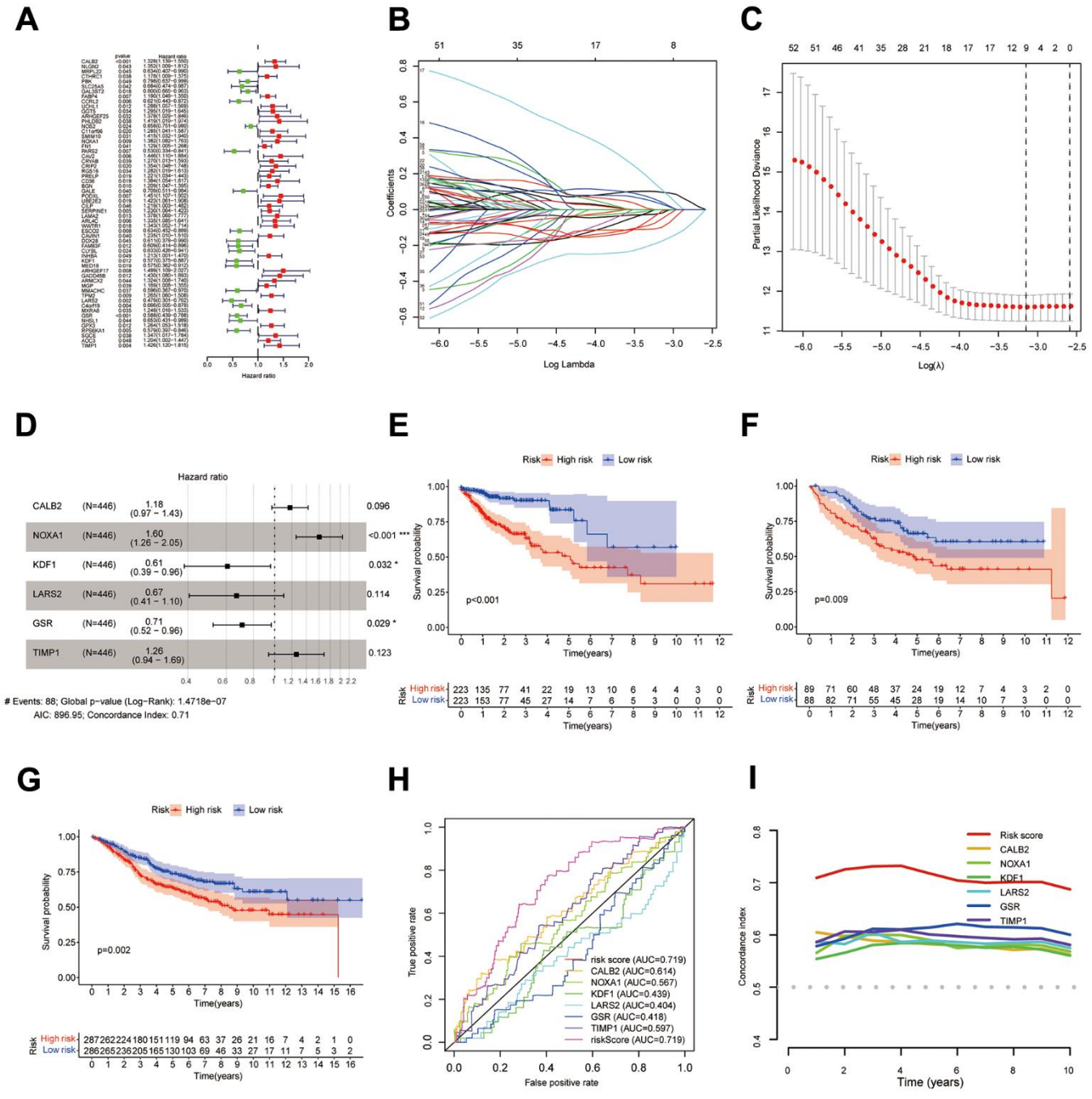


Figure 4. Construction of DPM. (**A**–**D**) Results of univariate Cox, LASSO, and multivariate Cox analyses of DEIGRC; (**E**–**G**) KM survival curves for the COAD patients from UCSC Xena database, GSE17536 and GSE39582 datasets; (**H**, **I**) ROC curves and C-indexes compare predictive efficacy of DPM with those of the genes used to construct the model.

EDNRA, and GC. Figure 10E, 10F show the findings of the GO and KEGG enrichment analyses, respectively. According to KEGG enrichment analysis, potential mechanisms underlying the development of liver metastases in COAD patients may be Fat digestion and absorption, Proteoglycans in cancer and Nitrogen metabolism.

Expression levels of 6 DPM genes in clinical tissue samples

This study obtained surgical samples from 10 pairs of colon cancer patients. Using bar graphs to depict the individual mRNA expression from each of the 6 genes in the COAD dataset (TCGA and GTEx), we discovered that CALB2 was only lowly expressed in COAD tissues, whereas the other 5 genes were highly expressed (Figure 11A–11F). The mRNA expression of these 6 genes was then further investigated by reverse transcription quantitative polymerase chain reaction (RT-qPCR) in the collected adjacent and COAD tissues (Table 1), and the outcomes were found to agree with the gene expression in database (Figure 11G–11L).

The Human Protein Atlas database (https://www.proteinatlas.org/) was used to retrieve the expression of the 6 genes at the protein level in healthy colon tissues and colon cancer tissues, and the results were consistent with RT-qPCR (Figure 11M–11X).

DISCUSSION

COAD accounts for approximately 70% of patients with CRC [31]. The highly heterogeneous features of COAD make its clinical treatment more complex and difficult. Immunotherapy, as one of the most likely therapeutic cures for tumors, is classified into two main categories: immune cell therapy and immune checkpoint therapy, and has been applied in the clinical treatment of a variety of tumors with encouraging therapeutic results [32–35]. For treating COAD patients, the immune checkpoint-related medications nabolutumab and pablizumab are now being used in clinical trials [36]. Nevertheless, due to a lack of biomarkers and models to predict patient response and prognosis to immunotherapy, low response rates to immunotherapy, and the fact that only some patients indicate treatment,

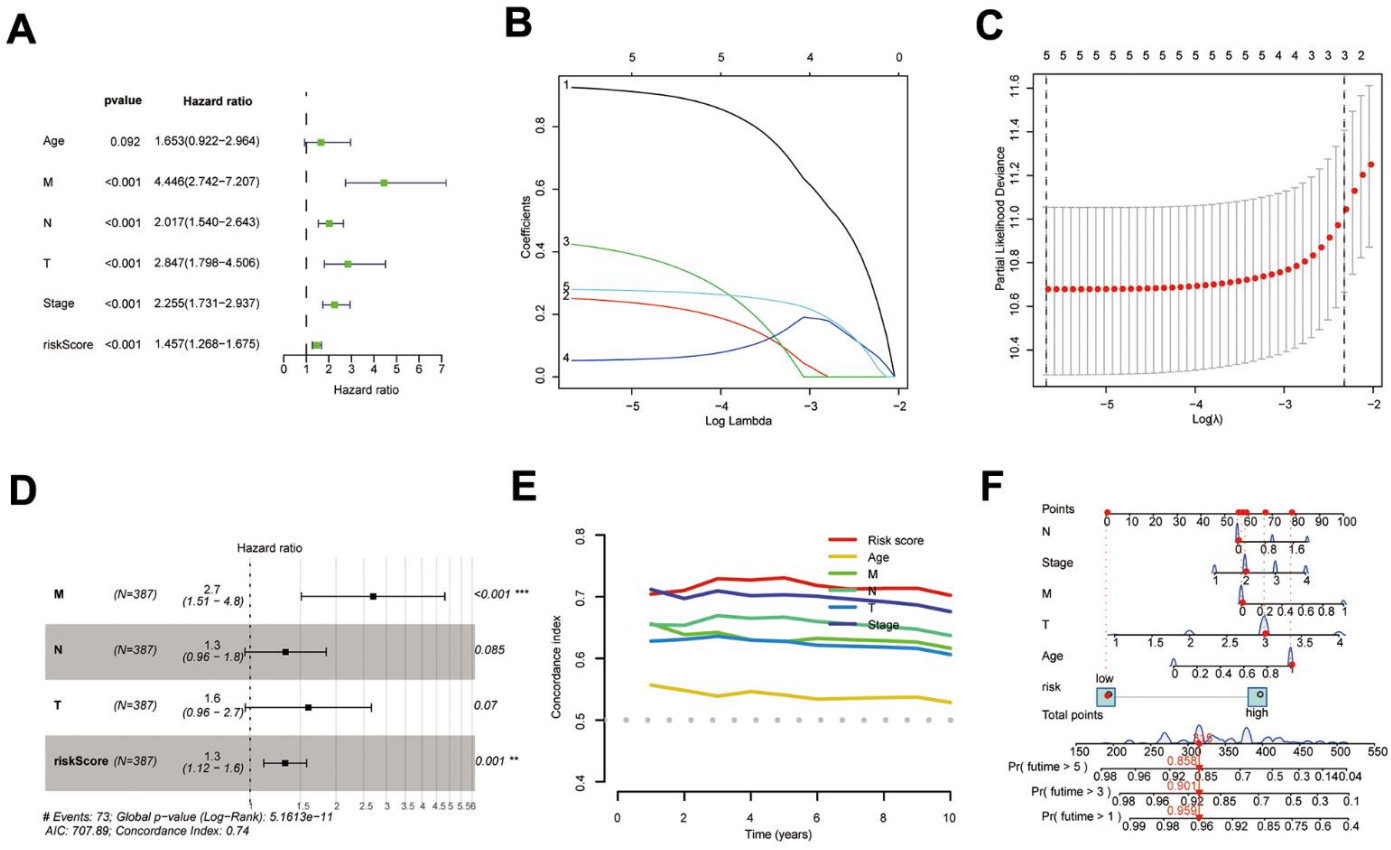


Figure 5. Construction of the COAD nomogram. (**A**–**D**) Results of univariate Cox, LASSO, and multi-factor Cox analyses of prognostic models and clinical traits; (**E**) C-index comparing the predictive efficacy of the prognostic model with that of clinical traits; (**F**) The COAD nomogram was constructed using prognostic models and clinical traits.

immunotherapy has experienced several challenges in clinical practice [37, 38]. IRG model has been built to evaluate the prognosis of patients with tumors based on transcriptome sequencing data [39, 40]. However, the potential mechanisms between IRGs, COAD, and immune features remain underdeveloped. In the current investigation, NFM and WGCNA algorithms were employed to filter DEIGRC, and prognostic models were constructed using one-factor Cox, LASSO, and multi-factor Cox algorithms. The DPM demonstrated superior predictive efficacy as a prognostic factor independent of clinical features and could be employed for predicting clinical outcomes, according to a further bioinformatics investigation.

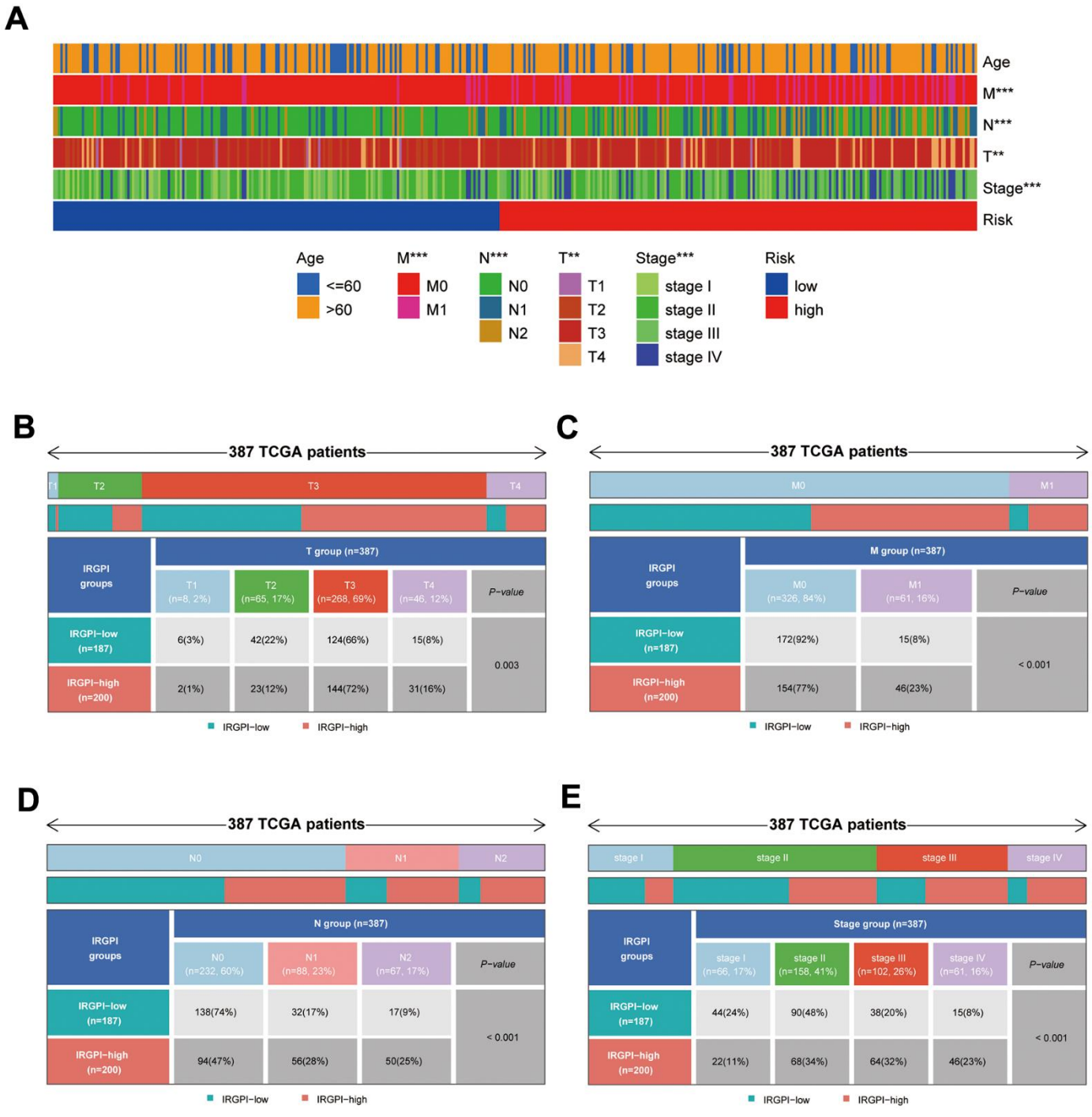


Figure 6. Results of differential analysis of clinical traits between subgroups. (**A**) Distribution of clinical traits between subgroups; (**B**–**E**) Results of differential analysis of T-stage, M-stage, N-stage, and stage between subgroups.

Using the NFM and WGCNA algorithms, we finally identified 511 DEIGRCs. The prognostic model constructed in this study consisted of 6 DEIGRCs, of which CALB2, NOXA1, and TIMP1 were risk factors for COAD patients, while KDF1, LARS2, and GSR were protective factors for COAD patients. GSE17536 and GSE39582 datasets also validated the model's prognostic value. A calcium receptor protein called CALB2 has the ability to bind Ca^{2+} [41, 42]. In colon cancer tissues, the expression level of CALB2 was significantly higher than that of normal colon epithelial cells, and at the same time, the expression level of CALB2 was positively correlated with the metastasis of local lymph nodes and other organs [43, 44]. Zhang et al. discovered that CALB2 up-regulated MMM9 and down-regulated E-cadherin, which in turn encouraged colon cancer cells to invade and migrate [45]. NoxA1, a homolog of $p67^{phox}$, is thought to be an activator of Nox1 [46, 47]. Nox1 and its regulators NoxO1 and NoxA1 are expressed greater in human gastric and intestinal adenocarcinomas than in normal gastric mucosa, suggesting that Nox protein activation could be a sign of tumor transformation [48]. In COAD cells, NoxA1 is crucial for the functional and reactive oxygen species-dependent development of endocryptal fossas [49]. TIMP1, as a soluble protein, is a member of the

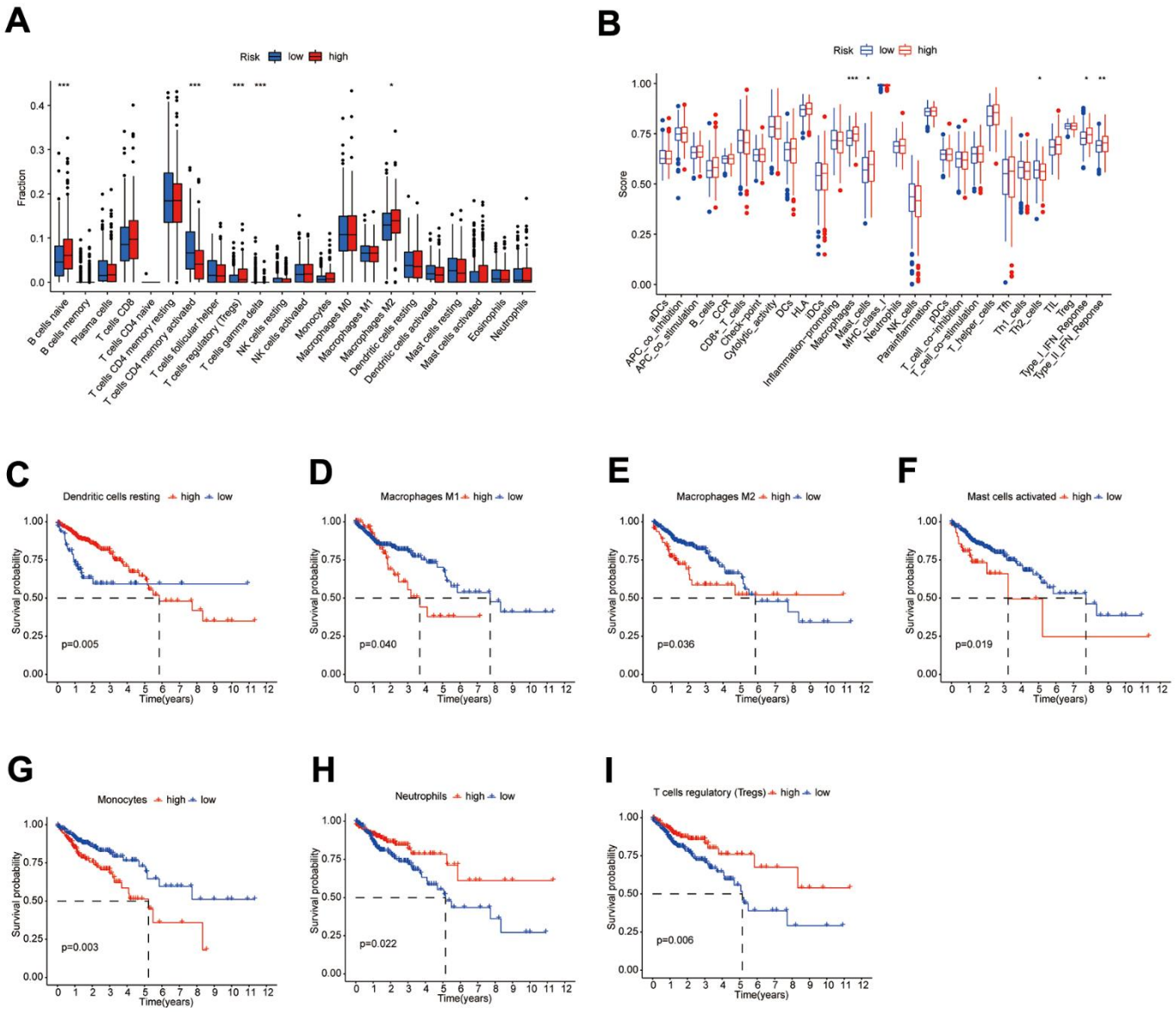


Figure 7. Differences in immune characteristics between subgroups. (**A**, **B**) Differences in the proportion of immune cells and Immune cell function scores between high- and low-risk subgroups; (**C**–**I**) KM survival curves for the proportion of 7 immune cells in high- and low-risk subgroups.

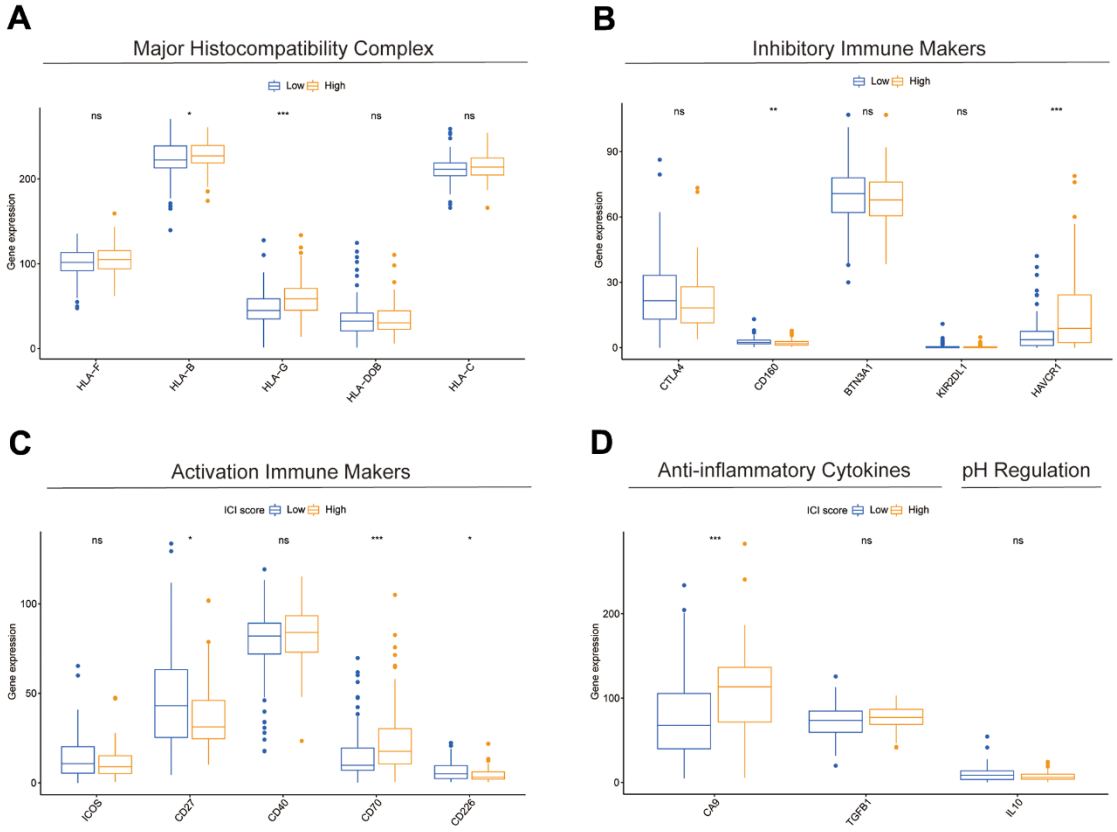


Figure 8. Differences in immune-related markers between subgroups. (**A**–**D**) Differences in expression of multiple types of immune marker genes between subgroups.

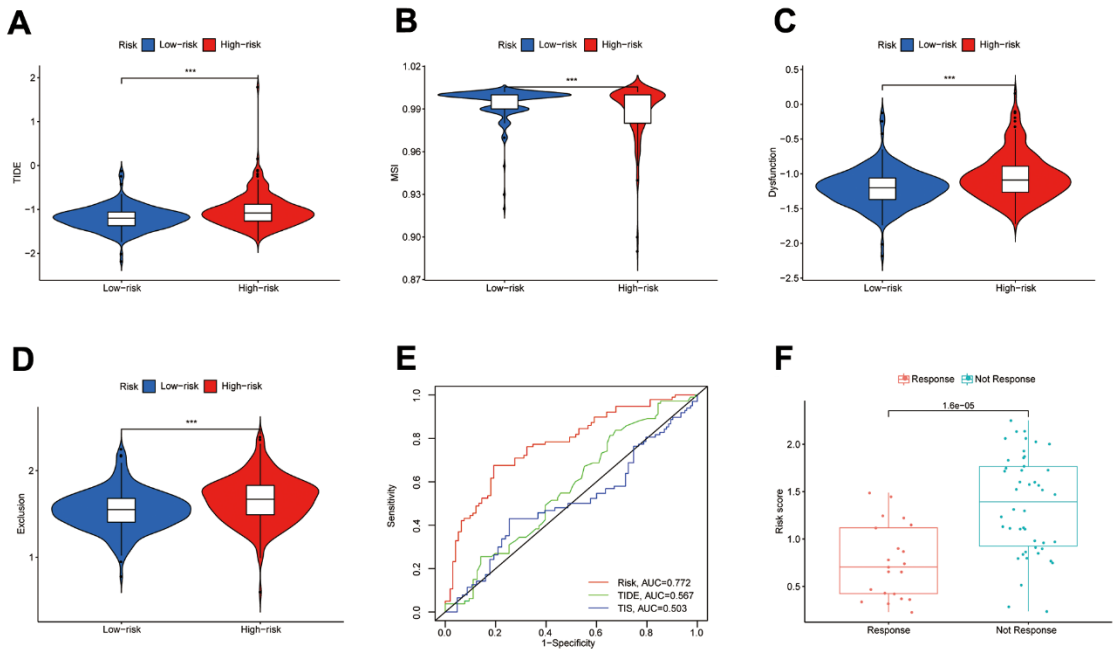


Figure 9. Value of the DPM in predicting immunotherapy outcomes. (**A**–**D**) Differences in TIDE, MSI, and T-cell dysfunction and exclusion scores between subgroups; (**E**) ROC curves comparing prognostic efficacy of prognostic models with those of TIDE and TIS; (**F**) Differences in risk scores between immune responsive and non-responsive patients (GSE109211 dataset).

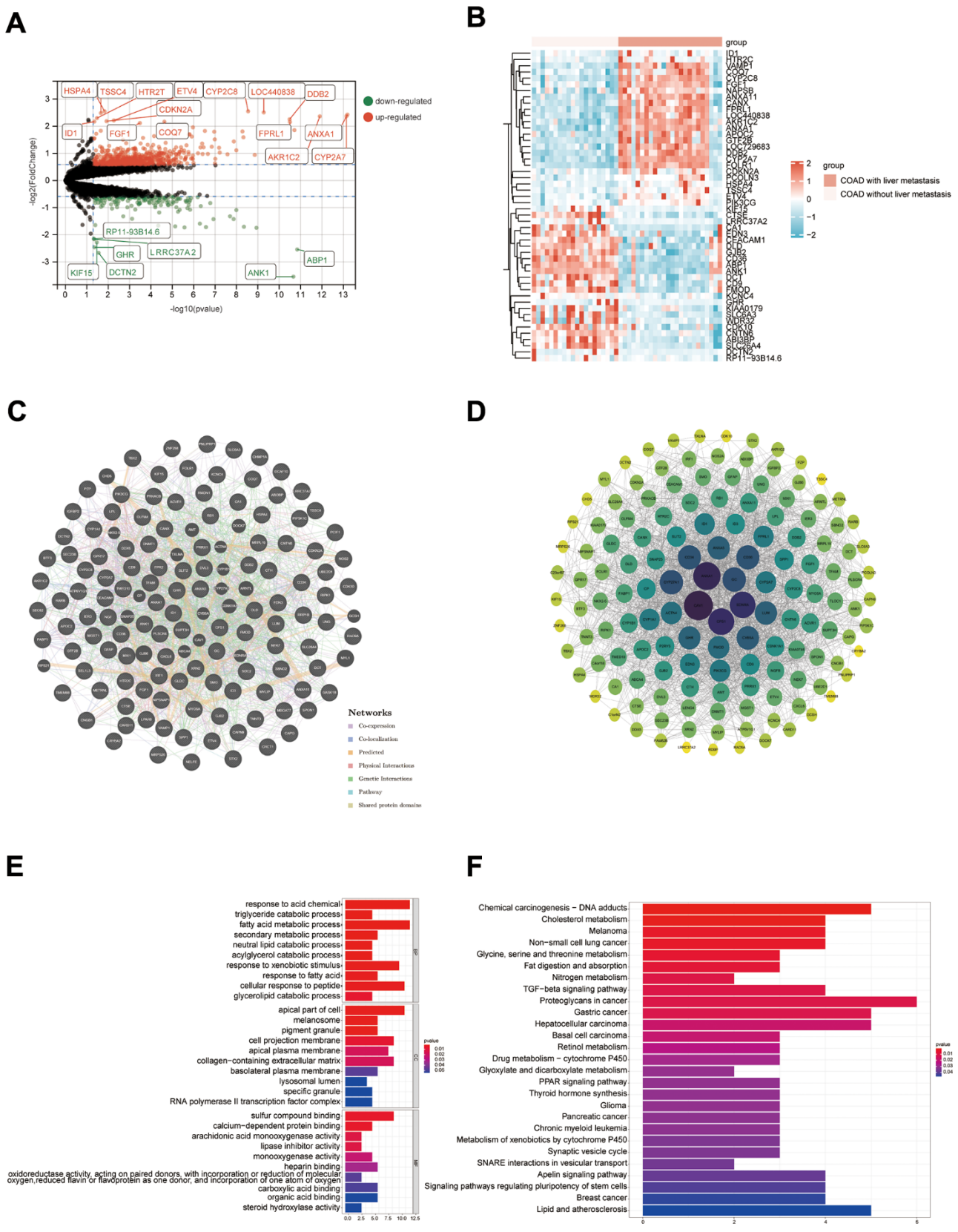


Figure 10. Results of bioinformatics analysis of potential mechanisms for developing liver metastasis in COAD. (**A**) Volcano map; (**B**) Heat map (only the top 25 highly and lowly expressed genes are shown); (**C**, **D**) PPI networks were constructed from the GeneMANIA database and Cytoscape software, respectively; (**E**, **F**) Results of GO and KEGG enrichment analysis, respectively.

tissue inhibitor family of metalloproteinases [50]. TIMP1 is secreted by cancer cells, fibroblasts, and endometrial cells, and has been associated with a poor prognosis in a variety of malignancies [51–53]. TIMP1 is regarded as a novel predictive biomarker for colon cancer due to its close relationships to processes and functions pertaining to the metastasis, proliferation, and apoptosis of cancer cells [54]. TIMP1 is also thought to be a viable target for the treatment of colon cancer since it was shown that it stimulates the growth and invasiveness of right-sided colon cancer cells via the FAK/Akt signaling pathway [55]. By encoding a precursor to mitochondrial leucine-tRNA synthetase, amino-tRNA synthetase LARS2 regulates the translation of mitochondria-encoded genes [56]. Breast cancer tumor growth and proliferative capacity were increased in mouse mammary glands with single allele LARS gene deletion [57]. LARS2-secreting B-cell subsets are highly correlated with the prognosis of CRC patients and promote immune escape of colorectal cells [58]. KDF1 is a crucial regulator of epidermal differentiation and an inhibitor of cell proliferation [59]. For tissue homeostasis and cancer prevention, KDF1 plays a crucial function in preserving the right balance between cell division and differentiation. Reduced KDF1 expression has been discovered in cancer cells, and it has been demonstrated to correlate with patient survival positively and negatively correlate with tumor grade [60]. Chromosome 8p12, where GSR is located, is frequently deleted in CRC [61]. There is growing evidence that the deletion of chromosome 8p lowers the survival rates of cancer patients and enhances the aggressiveness and metastatic potential of CRC [62–64]. Glutathione peroxidase utilizes the reducing capacity of GSR to scavenge excess reactive oxygen species in the cytoplasm, thereby preventing oxidative

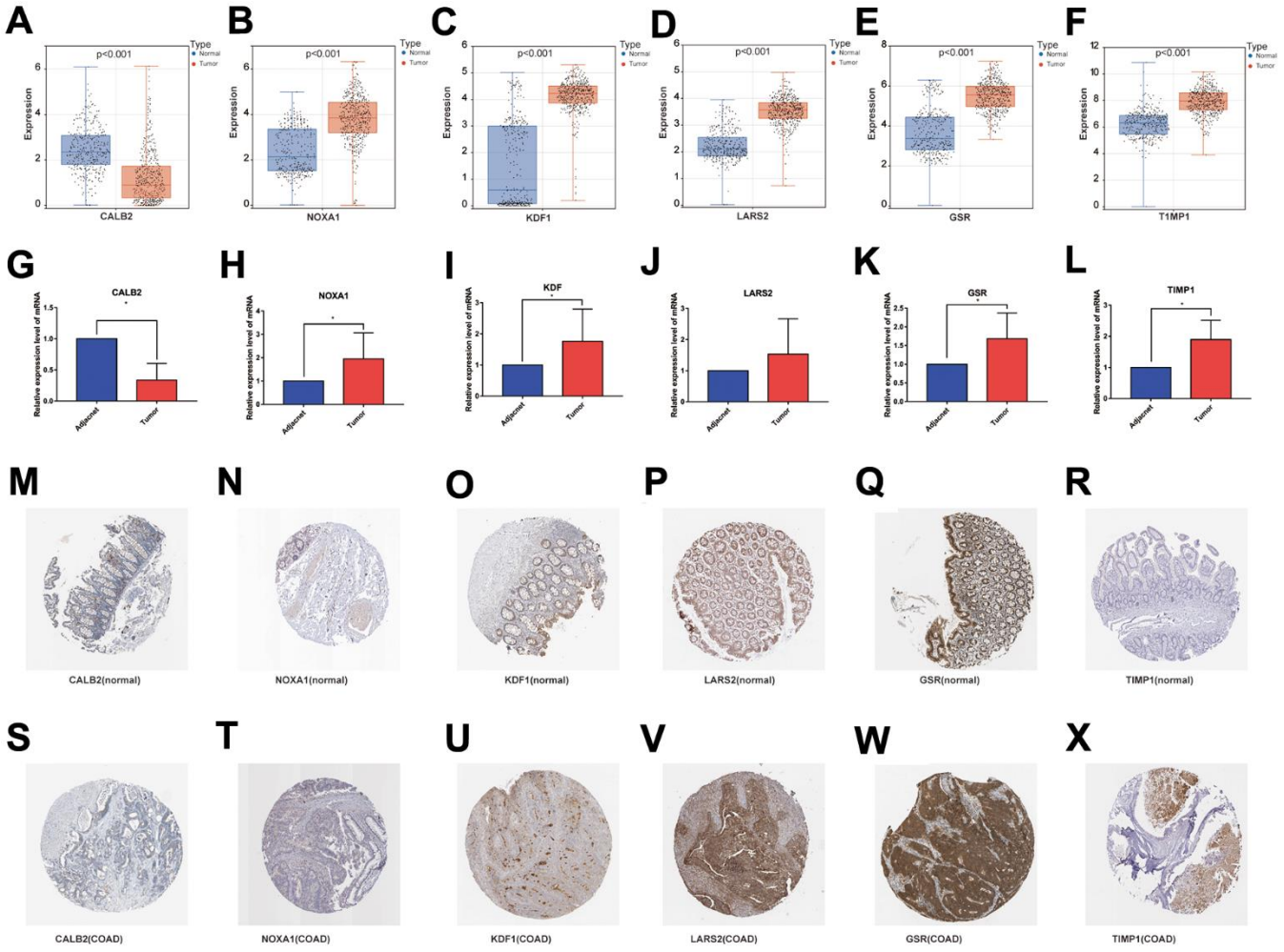


Figure 11. (**A**–**F**) mRNA expression levels of the 6 genes modeled in the databases (TCGA and GTEx); (**G**–**L**) RT-qPCR results of the 6 genes modeled in the collected clinical samples; (**M**–**R**) and (**S**–**X**) immunohistochemical results of the 6 genes modeled in normal colon tissue and COAD tissue.

Table 1. Primer sequences for the 6 genes of DMP.

Primer name	Specific sequence	Length
CALB2	Forward 5'-TTCCATCCACCACCTTGCCAATG-3'	24
	Reverse 5'-AAAGGAGCACGCCGAGTAAAGAAG-3'	23
NOXA1	Forward 5'-CCGCCAGGCTGTGCTTCAAC-3'	22
	Reverse 5'-TGGTCACGGCTTGGTCAAATGC-3'	20
KDF1	Forward 5'-CAGCAGCATCACGCAGGACTAC-3'	21
	Reverse 5'-CAGCAGCCCGAGTTGAACGAC-3'	22
LARS2	Forward 5'-CTACACCATCAGCGACACCATAGC-3'	22
	Reverse 5'-GCGGCATTTTCAGCAGGCAATC-3'	24
GSR	Forward 5'-CTGGAGTGCGGTGGTGCTATTTC-3'	23
	Reverse 5'-ATGGTGGTGCGTGCCTGTAATTC-3'	23
TIMP1	Forward 5'-ATCCTGTTGTTGCTGTGGCTGATAG-3'	24
	Reverse5'-CGCTGGTATAAGGTGGTCTGGTTG-3'	25

stress-driven cancer progression [65]. The mechanism of COAD cell differentiation and proliferation is intimately linked to GSR [66].

We discovered statistically significant variations in the proportion of B cells naive, T cells regulatory, Macrophages M2, and T cells CD4 memory activated between subgroups. B cells play an important role in the tumor microenvironment. By secreting cytokines, B cells naive can prevent lung cancer cells from proliferating, and the presence of B cells naive is favorably correlated with a positive prognosis for lung cancer patients [67]. Through cell interaction and bodily fluids, T cells govern various immune cells, including macrophages [68]. Eliminating T cell regulation increases T cells' capacity to attack tumor cells and boosts the patient's immunological response to tumors [69, 70]. Macrophages are part of the immune system and contain M0 macrophages, M1 macrophages, and M2 macrophages, with MO macrophages polarized into M1 or M2 types. M1 macrophages can boost the inflammatory response and destroy tumor cells, while M2 macrophages have the efficacy to suppress the inflammatory response and promote tumor cell proliferation and metastasis [71]. T cells CD4 memory activated was positively correlated with good prognosis in breast and bladder cancer [72, 73].

KM survival curve indicated low-risk subgroup had a better prognosis. The TIEDE database, which was created to score the T-cell function of the samples by computing the transcriptome sequencing data of the samples, could be applied to forecast the immunotherapeutic outcome of patients [74]. In the high-risk subgroup, TIDE, T-cell Dysfunction and Exclusion scores were higher than in the low-risk subgroup. The higher these scores, the greater the likelihood of immunological escape and the worse the patient's immunotherapeutic outcome. Additionally, the ROC curve's predictions of patient survival time were more accurate than those of the TIDE and TIS models. Additionally, since the DEIGRC model only includes 6 genes, it is easier for clinical prediction in COAD patients.

CONCLUSIONS

We constructed DPM consisting of an immune-related model that can predict the prognosis for COAD patients. Further studies revealed that this model may be used to predict the suitability of immunotherapy and targeted therapy for oncology patients, which may help in the clinical management of oncology patients. We also identified potential molecular mechanisms and therapeutic targets for developing liver metastases in COAD patients, which may contribute to fundamental research related to liver metastases in COAD and the development of related new drugs.

Abbreviations

COAD: colon adenocarcinoma; IRGs: immune-related genes; NMF: non-negative matrix clustering; DEGs: differentially expressed genes; WGCNA: weighted correlation network analysis; DEIGRC: differentially expressed immune genes related with COAD; DPM: DEIGRC prognostic model; CRC: colorectal cancer; GEO: Gene Expression Omnibus; rss: residual sum of squares; TOM: topological overlap matrix; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LASSO: least absolute shrinkage and selection operator; KM: Kaplan-Meier; ROC: receiver operating characteristic; C-index: concordance index; TIDE: tumor immune dysfunction and rejection; TIS: tumor inflammatory signature; PPI: protein-protein interaction; C1: cluster1; C2: cluster2; RT-qPCR: reverse

transcription quantitative polymerase chain reaction; T: tumor; M: metastasis; N: node; MSI: microsatellite instability.

AUTHOR CONTRIBUTIONS

Weizheng Liang, Jun Xue and Xuejun Zhi developed the project concept. The data were collected and evaluated by Weizheng Liang, and Xiushen Li with the other authors' help. The manuscript was written and revised by Weizheng Liang and Xiangyu Yang. The experiments were performed by Xiangyu Yang and Peng Wang. All authors reviewed and discussed the results and contributed to the paper preparation. All authors have read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICAL STATEMENT AND CONSENT

This study was approved by the Ethics Committee of the First Affiliated Hospital of Hebei North University (Approval No. K2023040). The written informed consent was obtained from 10 surgical patients of First Affiliated Hospital of Hebei North University.

FUNDING

The study was supported by Hebei Provincial Natural Science Foundation (No. H2022405033), Zhangiiakou City Key R&D Plan Project (No.2322088D and 2311038D), Hebei Province Key R&D Plan Project (No.22377784D), Medical Science Research Subject Plan Project of Hebei Provincial Health Commission (No.20240805 and 20240782), and the High-level talent Starting Research Fund of The First Affiliated Hospital of Hebei North (No.20221107).

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022; 72:7–33. https://doi.org/10.3322/caac.21708 PMID:35020204

2. Li X, Liang W, Yu C, Meng Q, Zhang W, Wu X, Xue J, Deng S, Wang H. Potential therapeutic strategies for quercetin targeting critical pathological mechanisms associated with colon adenocarcinoma and COVID-19. Front Pharmacol. 2022; 13:988153. https://doi.org/10.3389/fphar.2022.988153 PMID:36249762

3. Liang W, Li X, Yao Y, Meng Q, Wu X, Wang H, Xue J. Puerarin: A Potential Therapeutic for Colon Adenocarcinoma (COAD) Patients Suffering From SARS-CoV-2 Infection. Front Pharmacol. 2022; 13:921517. https://doi.org/10.3389/fphar.2022.921517 PMID:35677450

4. Liang W, Yi H, Mao C, Meng Q, Wu X, Li S, Xue J. Research Progress of RNA Methylation Modification in Colorectal Cancer. Front Pharmacol. 2022; 13:903699. https://doi.org/10.3389/fphar.2022.903699 PMID:35614935

5. Xia C, Dong X, Li H, Cao M, Sun D, He S, Yang F, Yan X, Zhang S, Li N, Chen W. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J (Engl). 2022; 135:584–90. https://doi.org/10.1097/CM9.0000000000002108 PMID:35143424

6. He Z, Lin J, Chen C, Chen Y, Yang S, Cai X, He Y, Liu S. Identification of BGN and THBS2 as metastasis-specific biomarkers and poor survival key regulators in human colon cancer by integrated analysis. Clin Transl Med. 2022; 12:e973. https://doi.org/10.1002/ctm2.973 PMID:36377223

7. Inadomi JM, Vijan S, Janz NK, Fagerlin A, Thomas JP, Lin YV, Muñoz R, Lau C, Somsouk M, El-Nachef N, Hayward RA. Adherence to colorectal cancer screening: a randomized clinical trial of competing strategies. Arch Intern Med. 2012; 172:575–82. https://doi.org/10.1001/archinternmed.2012.332 PMID:22493463

8. García Sánchez J. [Colonoscopic polypectomy and long-term prevention of colorectal cancer deaths]. Rev Clin Esp. 2012; 212:408. https://doi.org/10.1016/j.rce.2012.04.012 PMID:22937540

9. Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JM, Parkin DM, Wardle J, Duffy SW, Cuzick J, and UK Flexible Sigmoidoscopy Trial Investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet. 2010; 375:1624–33. https://doi.org/10.1016/S0140-6736(10)60551-X PMID:20430429

10. Espey DK, Wu XC, Swan J, Wiggins C, Jim MA, Ward E, Wingo PA, Howe HL, Ries LA, Miller BA, Jemal A, Ahmed F, Cobb N, et al. Annual report to the nation on the status of cancer, 1975-2004, featuring cancer in American Indians and Alaska Natives. Cancer. 2007; 110:2119–52. https://doi.org/10.1002/cncr.23044 PMID:17939129

11. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity.

Nat Rev Clin Oncol. 2020; 17:11–32. https://doi.org/10.1038/s41571-019-0241-1 PMID:31289352

12. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. Cancer Res. 2019; 79:4557–66. https://doi.org/10.1158/0008-5472.CAN-18-3962 PMID:31350295

13. Kasprzak A. The Role of Tumor Microenvironment Cells in Colorectal Cancer (CRC) Cachexia. Int J Mol Sci. 2021; 22:1565. https://doi.org/10.3390/ijms22041565 PMID:33557173

14. Yamamoto T, Tsunedomi R, Nakajima M, Suzuki N, Yoshida S, Tomochika S, Xu M, Nakagami Y, Matsui H, Tokumitsu Y, Shindo Y, Watanabe Y, Iida M, et al. IL-6 Levels Correlate with Prognosis and Immunosuppressive Stromal Cells in Patients with Colorectal Cancer. Ann Surg Oncol. 2023; 30:5267–77. https://doi.org/10.1245/s10434-023-13527-y PMID:37222942

15. Yue T, Cai Y, Zhu J, Liu Y, Chen S, Wang P, Rong L. Autophagy-related IFNG is a prognostic and immunochemotherapeutic biomarker of COAD patients. Front Immunol. 2023; 14:1064704. https://doi.org/10.3389/fimmu.2023.1064704 PMID:36756126

16. Pearce OMT, Delaine-Smith RM, Maniati E, Nichols S, Wang J, Böhm S, Rajeeve V, Ullah D, Chakravarty P, Jones RR, Montfort A, Dowe T, Gribben J, et al. Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers. Cancer Discov. 2018; 8:304–19. https://doi.org/10.1158/2159-8290.CD-17-0284 PMID:29196464

17. Leonard NA, Reidy E, Thompson K, McDermott E, Peerani E, Tomas Bort E, Balkwill FR, Loessner D, Ryan AE. Stromal Cells Promote Matrix Deposition, Remodelling and an Immunosuppressive Tumour Microenvironment in a 3D Model of Colon Cancer. Cancers (Basel). 2021; 13:5998. https://doi.org/10.3390/cancers13235998 PMID:34885111

18. Wei C, Yang C, Wang S, Shi D, Zhang C, Lin X, Liu Q, Dou R, Xiong B. Crosstalk between cancer cells and tumor associated macrophages is required for mesenchymal circulating tumor cell-mediated colorectal cancer metastasis. Mol Cancer. 2019; 18:64. https://doi.org/10.1186/s12943-019-0976-4 PMID:30927925

19. De Simone V, Franzè E, Ronchetti G, Colantoni A, Fantini MC, Di Fusco D, Sica GS, Sileri P, MacDonald TT, Pallone F, Monteleone G, Stolfi C. Th17-type cytokines, IL-6 and TNF-α synergistically activate STAT3 and NF-kB to promote colorectal cancer cell growth. Oncogene. 2015; 34:3493–503. https://doi.org/10.1038/onc.2014.286 PMID:25174402

20. Franzè E, Marafini I, Troncone E, Salvatori S, Monteleone G. Interleukin-34 promotes tumorigenic signals for colon cancer cells. Cell Death Discov. 2021; 7:245. https://doi.org/10.1038/s41420-021-00636-4 PMID:34535634

21. Bedognetti D, Hendrickx W, Marincola FM, Miller LD. Prognostic and predictive immune gene signatures in breast cancer. Curr Opin Oncol. 2015; 27:433–44. https://doi.org/10.1097/CCO.0000000000000234 PMID:26418235

22. Galon J, Angell HK, Bedognetti D, Marincola FM. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. Immunity. 2013; 39:11–26. https://doi.org/10.1016/j.immuni.2013.07.008 PMID:23890060

23. Chen EX, Jonker DJ, Loree JM, Kennecke HF, Berry SR, Couture F, Ahmad CE, Goffin JR, Kavan P, Harb M, Colwell B, Samimi S, Samson B, et al. Effect of Combined Immune Checkpoint Inhibition vs Best Supportive Care Alone in Patients With Advanced Colorectal Cancer: The Canadian Cancer Trials Group CO.26 Study. JAMA Oncol. 2020; 6:831–8. https://doi.org/10.1001/jamaoncol.2020.0910 PMID:32379280

24. Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K, Lopez-Yurda M, Grootscholten C, Beets GL, Snaebjornsson P, Maas M, Mertz M, Veninga V, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat Med. 2020; 26:566–76. https://doi.org/10.1038/s41591-020-0805-8 PMID:32251400

25. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, Lugli A, Zlobec I, Rau TT, Berger MD, Nagtegaal ID, Vink-Börger E, Hartmann A, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018; 391:2128–39. https://doi.org/10.1016/S0140-6736(18)30789-X PMID:29754777

26. Angell HK, Bruni D, Barrett JC, Herbst R, Galon J. The Immunoscore: Colon Cancer and Beyond. Clin Cancer Res. 2020; 26:332–9.

https://doi.org/10.1158/1078-0432.CCR-18-1851 PMID:31413009

27. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, Thomson E, Wiser J, Butte AJ. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data. 2018; 5:180015. https://doi.org/10.1038/sdata.2018.15 PMID:29485622

28. Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan THW, Shah N, Lo R, Naseer M, Que J, et al. InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol. 2008; 4:218. https://doi.org/10.1038/msb.2008.55 PMID:18766178

29. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010; 11:367. https://doi.org/10.1186/1471-2105-11-367 PMID:20598126

30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559 PMID:19114008

31. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021; 71:7–33. https://doi.org/10.3322/caac.21654 PMID:33433946

32. Lee DW, Kochenderfer JN, Stetler-Stevenson M, Cui YK, Delbrook C, Feldman SA, Fry TJ, Orentas R, Sabatino M, Shah NN, Steinberg SM, Stroncek D, Tschernia N, et al. T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. Lancet. 2015; 385:517–28. https://doi.org/10.1016/S0140-6736(14)61403-3 PMID:25319501

33. Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, Wunderlich JR, Nahvi AV, Helman LJ, Mackall CL, Kammula US, Hughes MS, Restifo NP, et al. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. J Clin Oncol. 2011; 29:917–24. https://doi.org/10.1200/JCO.2010.32.2537 PMID:21282551

34. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, Schadendorf D, Hassel JC, Akerley W, van den Eertwegh AJM, Lutzky J, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med. 2010; 363:711–23. https://doi.org/10.1056/NEJMoa1003466 PMID:20525992

35. Brahmer JR, Tykodi SS, Chow LQM, 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

36. Xu M, Chang J, Wang W, Wang X, Wang X, Weng W, Tan C, Zhang M, Ni S, Wang L, Huang Z, Deng Z, Li W, et al. Classification of colon adenocarcinoma based on immunological characterizations: Implications for prognosis and immunotherapy. Front Immunol. 2022; 13:934083. https://doi.org/10.3389/fimmu.2022.934083 PMID:35967414

37. Chung KY, Gore I, Fong L, Venook A, Beck SB, Dorazio P, Criscitiello PJ, Healey DI, Huang B, Gomez-Navarro J, Saltz LB. Phase II study of the anti-cytotoxic T-lymphocyte-associated antigen 4 monoclonal antibody, tremelimumab, in patients with refractory metastatic colorectal cancer. J Clin Oncol. 2010; 28:3485–90. https://doi.org/10.1200/JCO.2010.28.3994 PMID:20498386

38. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature. 2017; 541:321–30. https://doi.org/10.1038/nature21349 PMID:28102259

39. Wang Q, Tang H, Luo X, Chen J, Zhang X, Li X, Li Y, Chen Y, Xu Y, Han S. Immune-Associated Gene Signatures Serve as a Promising Biomarker of Immunotherapeutic Prognosis for Renal Clear Cell Carcinoma. Front Immunol. 2022; 13:890150. https://doi.org/10.3389/fimmu.2022.890150 PMID:35686121

40. Yao Y, Kong X, Liu R, Xu F, Liu G, Sun C. Development of a Novel Immune-Related Gene Prognostic Index for Breast Cancer. Front Immunol. 2022; 13:845093. https://doi.org/10.3389/fimmu.2022.845093 PMID:35558081

41. Bertschy S, Genton CY, Gotzos V. Selective immunocytochemical localisation of calretinin in the human ovary. Histochem Cell Biol. 1998; 109:59–66. https://doi.org/10.1007/s004180050202 PMID:9452956

42. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, et al., and International Human Genome Sequencing Consortium. Initial

sequencing and analysis of the human genome. Nature. 2001; 409:860–921. https://doi.org/10.1038/35057062 PMID:11237011

43. Gotzos V, Wintergerst ES, Musy JP, Spichtin HP, Genton CY. Selective distribution of calretinin in adenocarcinomas of the human colon and adjacent tissues. Am J Surg Pathol. 1999; 23:701–11. https://doi.org/10.1097/00000478-199906000-00010 PMID:10366153

44. Gotzos V, Schwaller B, Gander JC, Bustos-Castillo M, Celio MR. Heterogeneity of expression of the calcium-binding protein calretinin in human colonic cancer cell lines. Anticancer Res. 1996; 16:3491–8. PMID:9042211

45. Zhang D, Zhao Y, Wang S, Wang X, Sun Y. A Prognostic Model of Angiogenesis and Neutrophil Extracellular Traps Related Genes Manipulating Tumor Microenvironment in Colon Cancer. J Cancer. 2023; 14:2109–27. https://doi.org/10.7150/jca.85778 PMID:37497410

46. Lyle AN, Deshpande NN, Taniyama Y, Seidel-Rogol B, Pounkova L, Du P, Papaharalambus C, Lassègue B, Griendling KK. Poldip2, a novel regulator of Nox4 and cytoskeletal integrity in vascular smooth muscle cells. Circ Res. 2009; 105:249–59. https://doi.org/10.1161/CIRCRESAHA.109.193722 PMID:19574552

47. Bánfi B, Clark RA, Steger K, Krause KH. Two novel proteins activate superoxide generation by the NADPH oxidase NOX1. J Biol Chem. 2003; 278:3510–3. https://doi.org/10.1074/jbc.C200613200 PMID:12473664

48. Tominaga K, Kawahara T, Sano T, Toida K, Kuwano Y, Sasaki H, Kawai T, Teshima-Kondo S, Rokutan K. Evidence for cancer-associated expression of NADPH oxidase 1 (Nox1)-based oxidase system in the human stomach. Free Radic Biol Med. 2007; 43:1627–38. https://doi.org/10.1016/j.freeradbiomed.2007.08.029 PMID:18037128

49. Gianni D, Taulet N, DerMardirossian C, Bokoch GM. c-Src-mediated phosphorylation of NoxA1 and Tks4 induces the reactive oxygen species (ROS)-dependent formation of functional invadopodia in human colon cancer cells. Mol Biol Cell. 2010; 21:4287–98. https://doi.org/10.1091/mbc.E10-08-0685 PMID:20943948

50. Batra J, Robinson J, Soares AS, Fields AP, Radisky DC, Radisky ES. Matrix metalloproteinase-10 (MMP-10) interaction with tissue inhibitors of metalloproteinases TIMP-1 and TIMP-2: binding studies and crystal structure. J Biol Chem. 2012; 287:15935–46. https://doi.org/10.1074/jbc.M112.341156 PMID:22427646

51. Wang YY, Li L, Zhao ZS, Wang HJ. Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. World J Surg Oncol. 2013; 11:81. https://doi.org/10.1186/1477-7819-11-81 PMID:23548070

52. Bjerre C, Vinther L, Belling KC, Würtz SØ, Yadav R, Lademann U, Rigina O, Do KN, Ditzel HJ, Lykkesfeldt AE, Wang J, Nielsen HB, Brünner N, et al. TIMP1 overexpression mediates resistance of MCF-7 human breast cancer cells to fulvestrant and down-regulates progesterone receptor expression. Tumour Biol. 2013; 34:3839–51. https://doi.org/10.1007/s13277-013-0969-7 PMID:23881388

53. Peng L, Yanjiao M, Ai-guo W, Pengtao G, Jianhua L, Ju Y, Hongsheng O, Xichen Z. A fine balance between CCNL1 and TIMP1 contributes to the development of breast cancer cells. Biochem Biophys Res Commun. 2011; 409:344–9. https://doi.org/10.1016/j.bbrc.2011.05.021 PMID:21586274

54. Zhang Y, Li Y, Zuo Z, Li T, An Y, Zhang W. An epithelial-mesenchymal transition-related mRNA signature associated with the prognosis, immune infiltration and therapeutic response of colon adenocarcinoma. Pathol Oncol Res. 2023; 29:1611016. https://doi.org/10.3389/pore.2023.1611016 PMID:36910014

55. Ma B, Ueda H, Okamoto K, Bando M, Fujimoto S, Okada Y, Kawaguchi T, Wada H, Miyamoto H, Shimada M, Sato Y, Takayama T. TIMP1 promotes cell proliferation and invasion capability of right-sided colon cancers via the FAK/Akt signaling pathway. Cancer Sci. 2022; 113:4244–57. https://doi.org/10.1111/cas.15567 PMID:36073574

56. Feng S, Wan S, Liu S, Wang W, Tang M, Bai L, Zhu Y. LARS2 Regulates Apoptosis via ROS-Mediated Mitochondrial Dysfunction and Endoplasmic Reticulum Stress in Ovarian Granulosa Cells. Oxid Med Cell Longev. 2022; 2022:5501346. https://doi.org/10.1155/2022/5501346 PMID:35585880

57. Passarelli MC, Pinzaru AM, Asgharian H, Liberti MV, Heissel S, Molina H, Goodarzi H, Tavazoie SF. Leucyl-tRNA synthetase is a tumour suppressor in breast cancer and regulates codon-dependent translation dynamics. Nat Cell Biol. 2022; 24:307–15. https://doi.org/10.1038/s41556-022-00856-5 PMID:35288656

58. Wang Z, Lu Z, Lin S, Xia J, Zhong Z, Xie Z, Xing Y, Qie J, Jiao M, Li Y, Wen H, Zhao P, Zhang D, et al. Leucine-tRNA-synthase-2-expressing B cells contribute to colorectal cancer immunoevasion. Immunity. 2022; 55:1067–81.e8. https://doi.org/10.1016/j.immuni.2022.04.017 PMID:35659337

59. Lee S, Kong Y, Weatherbee SD. Forward genetics identifies Kdf1/1810019J16Rik as an essential regulator of the proliferation-differentiation decision in epidermal progenitor cells. Dev Biol. 2013; 383:201–13. https://doi.org/10.1016/j.ydbio.2013.09.022 PMID:24075906

60. Zheng JM, Gan MF, Yu HY, Ye LX, Yu QX, Xia YH, Zhou HX, Bao JQ, Guo YQ. KDF1, a Novel Tumor Suppressor in Clear Cell Renal Cell Carcinoma. Front Oncol. 2021; 11:686678. https://doi.org/10.3389/fonc.2021.686678 PMID:34136411

61. Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y. Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. Cancer Res. 1992; 52:5368–72. PMID:1356616

62. Diep CB, Kleivi K, Ribeiro FR, Teixeira MR, Lindgjaerde OC, Lothe RA. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. Genes Chromosomes Cancer. 2006; 45:31–41. https://doi.org/10.1002/gcc.20261 PMID:16145679

63. Macartney-Coxson DP, Hood KA, Shi HJ, Ward T, Wiles A, O'Connor R, Hall DA, Lea RA, Royds JA, Stubbs RS, Rooker S. Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level. BMC Cancer. 2008; 8:187. https://doi.org/10.1186/1471-2407-8-187 PMID:18590575

64. Cai Y, Crowther J, Pastor T, Abbasi Asbagh L, Baietti MF, De Troyer M, Vazquez I, Talebi A, Renzi F, Dehairs J, Swinnen JV, Sablina AA. Loss of Chromosome 8p Governs Tumor Progression and Drug Response by Altering Lipid Metabolism. Cancer Cell. 2016; 29:751–66. https://doi.org/10.1016/j.ccell.2016.04.003 PMID:27165746

65. Sies H. Role of metabolic H2O2 generation: redox signaling and oxidative stress. J Biol Chem. 2014; 289:8735–41. https://doi.org/10.1074/jbc.R113.544635 PMID:24515117

66. Bravard A, Beaumatin J, Dussaulx E, Lesuffleur T, Zweibaum A, Luccioni C. Modifications of the anti-oxidant metabolism during proliferation and differentiation of colon tumor cell lines. Int J Cancer. 1994; 59:843–7. https://doi.org/10.1002/ijc.2910590622 PMID:7989127

67. Chen J, Tan Y, Sun F, Hou L, Zhang C, Ge T, Yu H, Wu C, Zhu Y, Duan L, Wu L, Song N, Zhang L, et al. Single-cell transcriptome and antigen-immunoglobin analysis reveals the diversity of B cells in non-small cell lung cancer. Genome Biol. 2020; 21:152. https://doi.org/10.1186/s13059-020-02064-6 PMID:32580738

68. Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T cells and immune tolerance. Cell. 2008; 133:775–87. https://doi.org/10.1016/j.cell.2008.05.009 PMID:18510923

69. Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. Cell Res. 2017; 27:109–118. https://doi.org/10.1038/cr.2016.151 PMID:27995907

70. Zou W. Regulatory T cells, tumour immunity and immunotherapy. Nat Rev Immunol. 2006; 6:295–307. https://doi.org/10.1038/nri1806 PMID:16557261

71. Xia Y, Rao L, Yao H, Wang Z, Ning P, Chen X. Engineering Macrophages for Cancer Immunotherapy and Drug Delivery. Adv Mater. 2020; 32:e2002054. https://doi.org/10.1002/adma.202002054 PMID:32856350

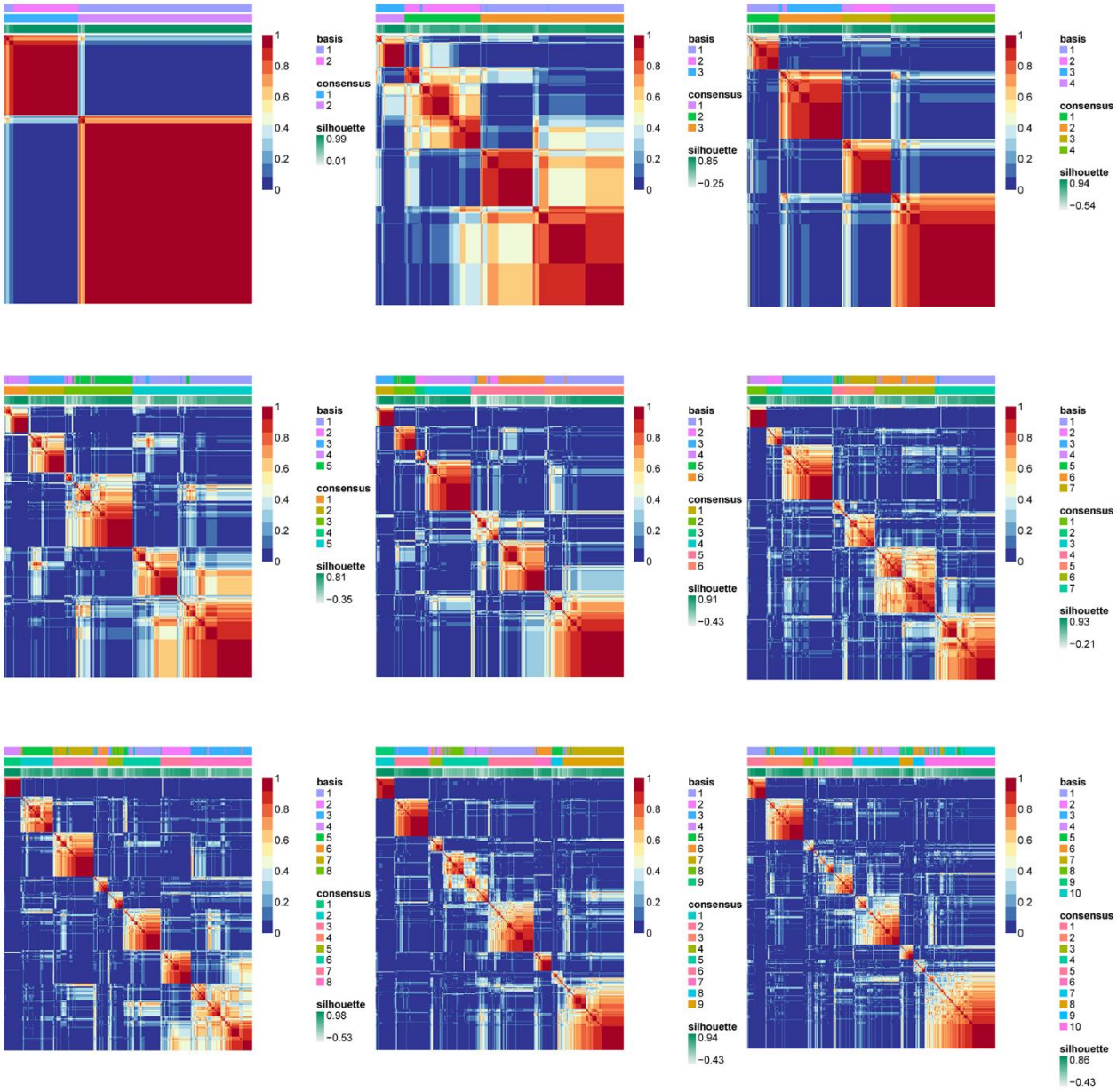
72. Li W, Zeng J, Luo B, Mao Y, Liang Y, Zhao W, Hu N, Chen G, Zheng X. [High expression of activated CD4+ memory T cells and CD8+ T cells and low expression of M0 macrophage are associated with better clinical prognosis in bladder cancer patients]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi. 2020; 36:97–103. Chinese. PMID:32314705

73. Liu D, Vadgama J, Wu Y. Basal-like breast cancer with low TGFβ and high TNFα pathway activity is rich in activated memory CD4 T cells and has a good prognosis. Int J Biol Sci. 2021; 17:670–82. https://doi.org/10.7150/ijbs.56128 PMID:33767579

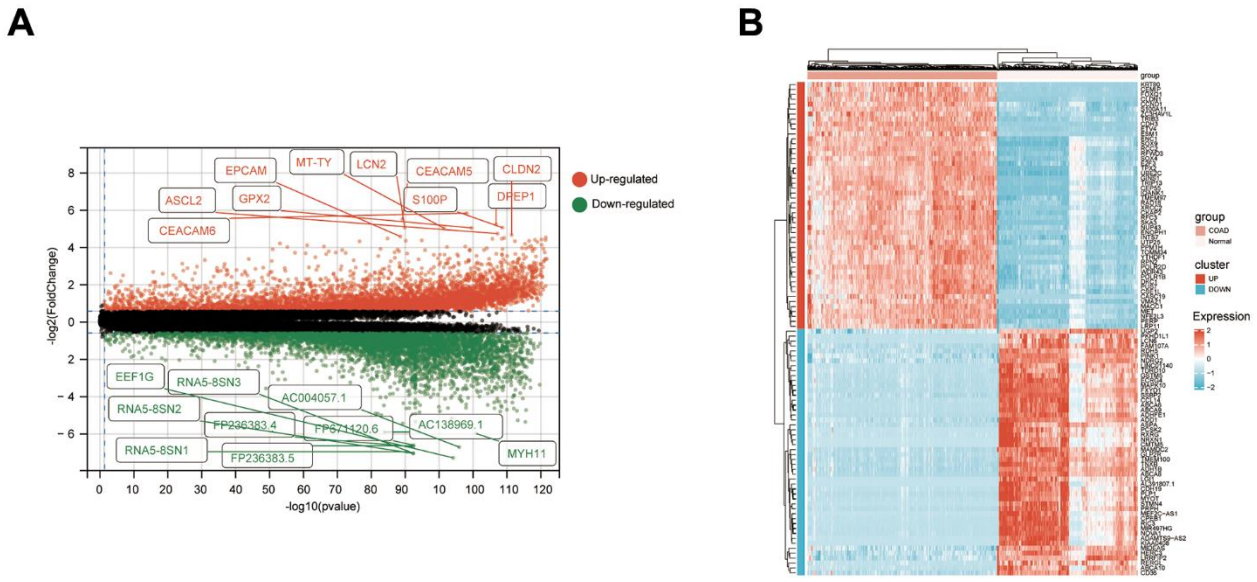
74. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018; 24:1550–8. https://doi.org/10.1038/s41591-018-0136-1 PMID:30127393

SUPPLEMENTARY MATERIALS

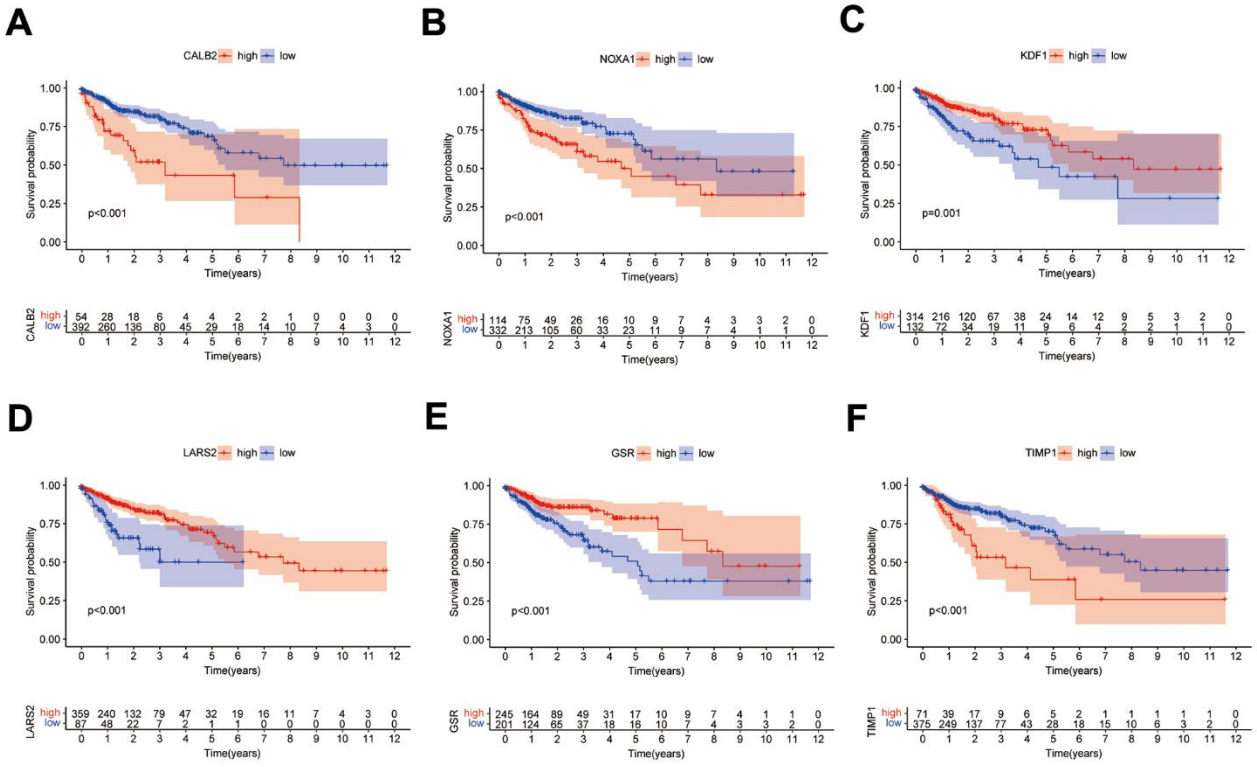
Supplementary Figures



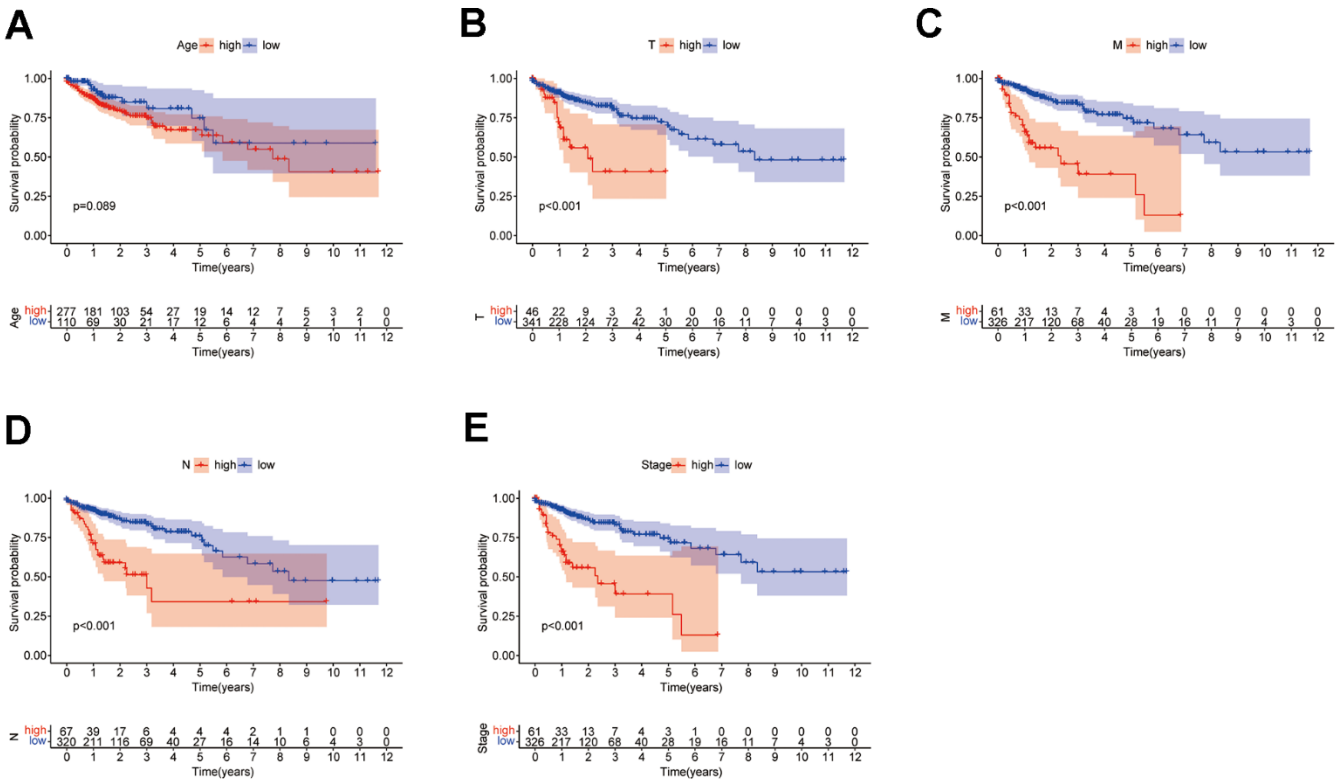
Supplementary Figure 1. Results of NFM analysis for classification numbers from 2 to 10.



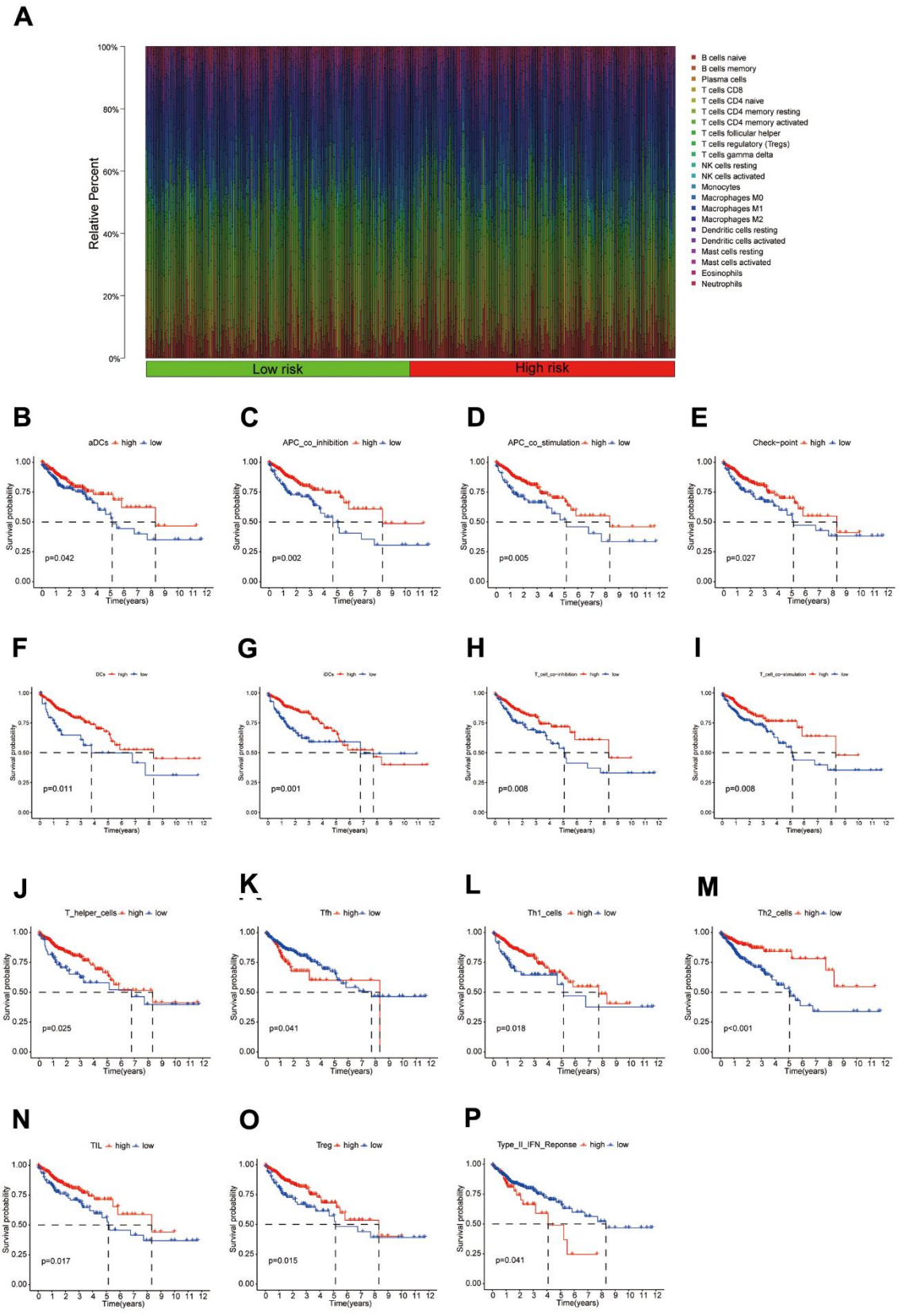
Supplementary Figure 2. DEGs in COAD patients. (**A**) Volcano plot; (**B**) Heat map (only the top 50 highly and lowly expressed genes are shown).



Supplementary Figure 3. KM curves for genes used in the construction of DPM. (**A**–**F**) KM survival curves for CALB2, NOXA1, KDF1, LARS2, GSR, and TIMP1.



Supplementary Figure 4. Relationship between clinical traits and prognosis of patients with COAD. (**A**–**E**) KM survival curves for Age, T-stage, M-stage, N-stage, and stage.



Supplementary Figure 5. Relationship between immune cell function scores and the prognosis of patients with COAD. (**A**) The proportion of immune cell distribution between subgroups; (**B**–**P**) KM survival curves for 15 immune cell function scores significantly associated with prognosis in COAD patients (p<0.05).

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 2.

Supplementary Table 1. Immune-related genes obtained from ImmPort.

Supplementary Table 2. Immune-related genes obtained from InnateDB.