



Prognostic signature of colorectal cancer based on uric acid-related genes

Chun Zhuang^{c,1}, Yifan Liu^{b,1}, Ranran Gu^a, Shanqing Du^a, Yin Long^{a,b,*}

^a Department of Clinical Laboratory, Yangpu Hospital, School of Medicine, Tongji University, Shanghai, China

^b Department of Biochemistry and Molecular Cell Biology, Shanghai Key Laboratory for Tumor Microenvironment and Inflammation, Key Laboratory of Cell Differentiation and Apoptosis of the Chinese Ministry of Education, Shanghai Jiao Tong University School of Medicine, Shanghai, China

^c Department of Gastrointestinal Surgery, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

ARTICLE INFO

Keywords:

colorectal cancer
Prognosis
uric acid
immune infiltration
Somatic mutation

ABSTRACT

Colorectal cancer (CRC) is one of the deadliest cancers worldwide. Numerous studies have reported a correlation between uric acid (UA) level and CRC risk. Here, we investigated the role and prognostic value of UA-related genes in CRC progression. CRC-associated gene expression and clinical data were retrieved from The Cancer Genome Atlas (TCGA), and UA-related genes were identified by overlapping the TCGA and GeneCards databases. The Gene Ontology annotation, Kyoto Encyclopedia of Genes and Genomes pathway, and Molecular Signatures Database dataset were subjected to gene set enrichment analysis. A prognostic model was constructed using the univariate and multivariate COX regression and least absolute shrinkage and selection operator (LASSO) analyses and validated using the Gene Expression Omnibus cohort. Competing endogenous RNA network, CellMiner, and Human Protein Atlas were used to detect the signature of 13 UA-related genes in the prediction model. The expression of five potential UA-related genes in CRC cell lines was confirmed via qPCR. CIBERSORT was used to evaluate immune cell infiltration in the TCGA-CRC dataset. Thirteen highly prognostic UA-related genes were used to construct a prognostic model of CRC with risk score accuracy and predictive efficacy. Abundance of activated M0 macrophages, monocytes, CD8⁺ T cells, and natural killer cells positively correlated with the risk score. Five promising UA-related genes showed higher expression levels in CRC than in colonic cell lines. Thus, our model posits a direct relationship between UA-related genes and CRC risk, offering novel insights into diagnosis, prognosis, and treatment.

Abbreviations: CC, cellular components; ceRNA, competing endogenous RNA; CRC, colorectal cancer; BP, biological processes; DEG, Differentially Expressed Genes; DMEM, Dulbecco's modified Eagle's medium; DSS, disease-free survival; FBS, fetal bovine serum; FC, fold change; FDR, false discovery rate; FDA, Food and Drug Administration; GEO, Gene Expression Omnibus; GO, Gene ontology; GSEA, gene set enrichment analysis; GSVA, Gene Set Variation Analysis; HPA, The Human Protein Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; lncRNA, long non-coding RNAs; MF, molecular functions; MSigDB, Molecular Signatures Database; NES, normalized enrichment scores; OS, overall survival; PCA, Principal component analysis; PFS, progression-free survival; ssGSEA, single-sample gene set enrichment analysis; TCGA, The Cancer Genome Atlas; TPM, transcripts per million reads; UA, uric acid.

* Corresponding author. Department of Clinical Laboratory, Yangpu Hospital, School of Medicine, Tongji University, No. 450 Tengyue Road, Shanghai, 200090, China.

E-mail address: longyin8864@163.com (Y. Long).

¹ Contributed equally.

<https://doi.org/10.1016/j.heliyon.2023.e22587>

Received 23 May 2023; Received in revised form 14 November 2023; Accepted 15 November 2023

Available online 22 November 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Colorectal cancer (CRC) is the third most prevalent and deadly cancer worldwide, with approximately 106 180 new diagnosed cases each year [1]. Despite great improvements in diagnosis, molecular targeted therapy, immunotherapy, and other therapeutic strategies, the prognosis and survival of patients with CRC remain poor, highlighting the importance of distinguishing the risk of CRC and identifying specific therapeutic targets [2–4].

Metabolic disorders, somatic mutations, and dysregulated tumor microenvironment are known risk factors for the development of CRC [5–7]. Smoking, alcohol consumption, obesity, low-dietary-fiber diet, and long-term intake of red meat and processed meats can increase the risk of CRC [8]. With the improvement of living standards, the intake of purine-rich foods has increased greatly, and this change in diet structure predisposes individuals to the development of metabolic diseases. Metabolism-related diseases, such as obesity, type 2 diabetes, insulin resistance, and metabolic syndrome, are associated with the incidence of tumors and mortality [5]. Metabolic programming refers to the long-term effects of early-life nutritional and environmental exposures on an individual's metabolism and the subsequent risk of developing various diseases later in life, including cancer. CRC is among the disease that have been associated with metabolic programming [9]. Additionally, metabolic programming can affect epigenetic modifications, which can be influenced by various environmental factors, including nutrition [10]. Altered epigenetic patterns can regulate the expression of genes involved in cancer development and progression, including those related to cell cycle control, DNA repair, and apoptosis [11]. CRC is a complex condition influenced by various genetic, lifestyle, and environmental factors. Therefore, understanding the role of metabolic programming can provide valuable insights into disease prevention and intervention strategies.

Uric acid (UA), the final product of purine catabolism, is mostly found in the liver, gut, muscles, and vascular endothelium [12]. Despite its crucial antioxidative role, high levels of UA are associated with several metabolic and non-metabolic diseases. Additionally, UA induces an inflammatory response by activating NF- κ B, thereby promoting the pathogenesis of metabolic disorders [13]. A recent study reported an association between elevated UA levels and CRC, suggesting UA as an independent predictive risk factor for CRC [14]. Recent studies on the relationship between UA metabolism and CRC have provided novel insights into the prevention and treatment of cancer in patients with metabolic diseases. However, UA-related genes that are associated with the prognosis of patients with CRC remain largely unknown. Therefore, disruptions in important metabolic pathways have therefore been the focus of many cancer studies [15,16].

Based on our findings, a total of 13 UA-related genes with high prognostic value were screened from The Cancer Genome Atlas (TCGA) and GeneCards databases and used to construct a prognostic model, which was validated using the Gene Expression Omnibus (GEO) cohort. The results indicated that the risk score of the candidate genes had accuracy and convenient predictive efficacy. Furthermore, the immune infiltration and somatic mutation analyses of the CRC data from TCGA were performed.

2. Materials and methods

2.1. Cell culture

The human colon cancer cell lines HCT116, SW480, SW620, RKO, HT29, and SW1116 were obtained from National Collection of Authenticated Cell Cultures (Shanghai, China) and cultured in either Dulbecco's modified Eagle's medium (DMED) (Hyclone, USA) supplemented with 10 % fetal bovine serum (FBS) (Gibco, USA) and 1 % penicillin/streptomycin at 37 °C in a humidified atmosphere containing 5 % CO₂. FHC cell line (FuHeng Cell Center, Shanghai, China) was maintained in FHC cultured medium (FuHeng Cell Center, Shanghai, China) at 37 °C, 5 % CO₂ humidified atmosphere.

2.2. Data collection

The "TCGAbiolinks" R package [17] was used to download the raw RNA sequencing (RNA-seq) data profiles (Workflow Type: Count) from the TCGA-COAD project (<https://portal.gdc.cancer.gov/projects/TCGA-COAD>), containing 447 CRC and 41 para-carcinoma tissue samples. Subsequently, the Count data were transformed into transcripts per million reads (TPM), which in turn were transformed by z-score normalization for further analysis. In addition, patient clinical information, including age, gender, survival status, overall survival (OS) time, disease-free survival (DSS) time, progression-free survival (PFS) time, stage, and TNM stage, was downloaded using the GDC software (<https://portal.gdc.cancer.gov/>; Table S1). After excluding the CRC samples with incomplete survival and TNM stage information, 364 CRC and 41 para-carcinoma tissue samples were selected for subsequent analyses. The CRC somatic mutation and Masked Copy Number Segment data were retrieved using the TCGAbiolinks R package and visualized using the "maftools" R package [18]. The validation cohort was obtained from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) using the "GEOquery" R package [19] with the query GSE17536 on the GPL570 platform. The GEO cohort consisted of 175 CRC samples, with a survival time of >30. The data were transferred by z-score and standardized by the "limma" R package [20]. We retrieved 3180 UA-related genes from the GeneCards database [21] (<https://www.genecards.org>) by searching the keyword "Uric Acid."

2.3. Construction and validation of the UA metabolism-associated risk model

Least absolute shrinkage and selection operator (LASSO) regression, a current machine-learning algorithm, was used to construct a diagnosis model. Regularization was utilized to avoid the overfitting problem in the curve-fitting process to increase model accuracy. The "glmnet" Package [22] was used to construct the model, and the parameter was set.seed (2), family = " binomial."

2.4. Screening of significant differentially expressed genes (DEGs)

The “Deseq2” R package was employed to identify the differentially expressed mRNAs between normal and tumor samples. The cutoff value of the DEG criterion was set as log fold change (FC) > 1, along with adjP < 0.05. DEGs with logFC > 1 were considered upregulated genes in the tumor, and those with logFC < -1 were considered downregulated genes. The results of the differential analysis were plotted as heat and volcano maps using the R packages “pheatmap” and “ggplot2,” respectively [23].

2.5. Gene enrichment of DEGs in normal and cancer tissue samples

Gene ontology (GO) annotation analysis [24], a common method for conducting large-scale functional enrichment studies, classifies gene and protein functions into three GO categories, including biological processes (BP), molecular functions (MF), and cellular components (CC). The KEGG [25] is a widely used database that stores information about genomes, biological pathways, diseases, and drugs. The GO annotation and KEGG pathway enrichment analyses of DEGs were performed using the “clusterProfiler” R package [26], and a critical value of < 0.05 for false discovery rate (FDR) was considered statistically significant.

To investigate the differences in biological processes between different subgroups, gene set enrichment analysis (GSEA) [26] was performed for gene enrichment based on the TCGA-COAD gene expression profile dataset with logFC as the alignment. GSEA is a computational method used to analyze statistical differences in a particular gene set between two biological states and is commonly used to estimate changes in pathways and biological process activity in datasets. The gene set “msigdb.v7.0.symbols” was downloaded from the Molecular Signatures Database (MSigDB) database [27] for GSEA, and an FDR of < 0.25 was considered significantly enriched.

In addition, the enrichment scores of relevant pathways in the Molecular Signatures Database (MSigDB) database were calculated based on the gene expression matrix of each sample by the method of Gene Set Variation Analysis (GSVA) using the GSVA R package [28]. Furthermore, the differences were screened by the “limma” package, setting |logFC| > 0.5 and adjP value < 0.05 as the threshold of difference in pathways, and the relevant enrichment entries with statistically significant differences were displayed using heat maps. Enrichment scores for UA-related genes were calculated using the single-sample GSEA (ssGSEA) method [29] based on the gene expression matrix of each sample separately and presented using box plots.

2.6. Estimation of immune infiltration into CRC tumor microenvironment

Based on the principle of linear support vector regression, CIBERSORT [30] is an analytical tool for deconvoluting the transcriptome expression matrix to estimate the composition and abundance of immune cells in a mixture of cells. The gene expression matrix data (TPM) was uploaded to CIBERSORT combined with the LM22 eigengene matrix and filtered samples with an output of p < 0.05 to derive the immune cell infiltration matrix. The distribution of the 22 immune cell infiltrates in each sample was displayed by histograms using the “ggplot2” R package.

2.7. Protein–protein interaction analysis of UA-related genes

The STRING database [31] is used for predicting protein–protein interactions. It contains 9.6 million proteins and 138 million protein–protein interactions from 2031 species. It contains data obtained from experiments, text mining results from PubMed abstracts, other databases, and bioinformatic predictions. Protein–protein interaction networks were constructed for UA-related differential genes using the STRING database, setting 0.9 as the correlation coefficient. The protein–protein Interaction Network (PPI) results were exported from the STRING database and visualized using Cytoscape [32].

2.8. Construction of the competing endogenous RNA (ceRNA) network for DEGs

The ceRNA network elucidates the mechanisms of long non-coding RNAs (lncRNA), microRNA, and mRNA interactions. The data associated with the lncRNA-miRNA data (starBaseV3 hg19 CLIP-seq lncRNA all) were downloaded from the StarBase database to screen for lncRNAs with pancancerNum > 5 and clipExpNum > 3. Using the multiMiR package [33], miRNAs were predicted using the miRTarBase database, with dual luciferase validation as a condition.

2.9. Chemotherapy sensitivity analysis of prognostic genes based on CellMiner

The mRNA expression profiles and drug activity data of 13 UA-related DEGs in the prognostic model were downloaded from the CellMiner database (<https://discover.nci.nih.gov/cellminer/>) [34], which is a web-based tool containing genomic and pharmacological information for researchers to use NCI-60 cell line transcript and drug response data compiled by the National Cancer Institute. The transcript expression levels of 22 379 genes, 360 miRNAs, and drug responses for 20 503 compounds are available on the CellMiner website. We calculated correlations between the expression of the 13 prognostic model genes and compound sensitivity using Pearson’s correlation analysis. Statistical significance was set at P < 0.05.

2.10. Immunohistology analysis from HPA database

The Human Protein Analysis (HPA; <https://www.proteinatlas.org/>) [35] presents information on the tissue and cellular distribution of 24 000 human proteins obtained from 48 human normal tissues, 20 tumor tissues, 47 cell lines, and 12 blood cell types. All results are expressed in at least 576 immunohistochemical images, which are read and cited by professionals. Protein expression levels and differences in 13 model genes in normal and colon cancer tissues were retrieved from the HPA database (<https://www.proteinatlas.org/>).

2.11. RNA isolation and qPCR

Total cell RNA was isolated using TRIzol (Invitrogen, USA) according to the protocol previously reported [36]. The first-strand cDNAs were generated utilizing HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, China) and analyzing using ChamQ SYBR qPCR Master Mix (Vazyme, China) by Quantitative Real-time Polymerase Chain Reaction (PCR) in LightCycler 480 II (Roche, Switzerland) according to manufacture protocol. The sequences of primers are listed as followed: *PLAC1*: forward primer: 5'-GGAGTCTGTCAAG-GAGTGCC-3' and reverse primer: 5'- TGGTGAAGGCAGTTGTGCTA -3'; *TERT*: forward primer: 5'- GGCACGGCTTTTGTTCAGATG -3' and reverse primer: 5'- GGAGGCTGTTACCTGCAAAT -3'; *TIMP1*: forward primer: 5'- ATTCGACCTCGTCATCAGG -3' and reverse primer: 5'- GGACCTGTGGAAGTATCCGC -3'; *ACTB*: forward primer: 5'-AGAGCTACGAGCTGCCTGAC-3' and reverse primer: 5'-AGCACTGTGTTGGCGTACAG-3'; *TRIP6*: forward primer: 5'-CGACCAGACCGACAGGCATA -3' and reverse primer: 5'-GCTGGCGGTAGTGTAAGAGG -3'; *WDR72*: forward primer: 5'- CAGCAACTCAAACCTGGCAA -3' and reverse primer: 5'-GATGCCTTGCTCACCTGG -3'.

2.12. Statistical analysis

All calculations and statistical analyses were performed using the R software (<https://www.r-project.org/>, version 4.0.2). For comparing the two groups of continuous variables, the independent Student's *t*-test was employed to estimate the statistical significance of normally distributed variables, and the Mann-Whitney *U* test (i.e., Wilcoxon rank-sum test) was used to analyze the differences between non-normally distributed variables. All statistical *P*-values were two-sided and were statistically significant at < 0.05 . The overall flow diagram is shown in Fig. S1.

3. Results

3.1. Identification of differential UA metabolism-associated genes and biological function analysis

Principal component analysis (PCA) was used to analyze the CRC data from the TCGA database (TCGA-COAD). As illustrated in Fig. S2 A, the expression profiles of cancer tissues were significantly different from those of normal tissues. To explore the potential mechanisms of CRC development and to identify diagnostic and prognostic markers, we performed differential analysis of the expression profiles of CRC and normal tissues. Screening with the threshold of $|\log_{2}FC| > 1$ and adjusted $P < 0.05$, it showed that 4786 genes were upregulated and 3480 genes were downregulated, as visualized using the volcano plot and heat map. Brown represents upregulated expression and dark green represents downregulated expression (Figure S 2B–C). To further understand the alteration of pathways in CRC, we analyzed the enrichment scores of relevant datasets in MSigDB in all samples using GSEA and screened the differentially enriched pathways with $|\log_{2}FC| > 0.5$, adjusted $P < 0.05$, using the “limma” package. Multiple pathways were suggested to differ between the groups. The most significantly different pathways were visualized using a heat map that included multiple cancer-related pathways.

Additionally, GSEA was conducted to enrich the GO, KEGG, and MSigDB pathway-related datasets with the list of genes ranked from the highest to the lowest difference multiplicity. The results demonstrated that the gene sets were significantly enriched by the GO enrichment analysis and were ranked by normalized enrichment scores (NES) in descending order: condensed chromosome centromeric and condensed chromosome kinetochore; however, the significantly suppressed ones were the cation channel complex sarcolemma (Fig. S3A). The GO enrichment analysis network indicated that the pathway was focused on chromosomes, centromeric regions, spindle poles, and spliceosomal complexes (Fig. S3B) (Table S1). We visualized the top 10 functional entries that were significantly enriched according to the descending order of the absolute NES values. The main functions were associated with the extracellular matrix and extracellular vesicles.

The KEGG enrichment analysis revealed that the cell cycle and cGMP-PKG signaling pathways were significantly enriched, whereas the ribosome and calcium signaling pathways were significantly inhibited (Fig. S3D) (Table S2). The KEGG enrichment network analysis showed that the pathways were focused on signaling pathways regulating stem cell pluripotency, cell cycle, and the cGMP-PKG signaling pathway (Fig. S3E). We also visualized the significant enrichment of related pathways in descending order of $|NES|$, including signaling pathways regulating stem cell pluripotency, the Hippo signaling pathway, cell adhesion molecules, and other related pathways, among which purine metabolism was significantly inhibited CRC (Fig. S3F).

The MSigDB dataset enrichment analysis showed the enrichment of several pathways, including *sabates_colorectal_adenoma* and *vecchi_gastric_cancer_early*, which were significantly suppressed (Fig. S3G). The enrichment analysis network diagram shows that the pathways focus on the Anderson cholangiocarcinoma class2 and Basaki YBX1 targets up, and Benporath proliferation. (Fig. S3H). The enrichment of apparently related pathways were also visualized in the descending order of $|NES|$ (Fig. S3I).

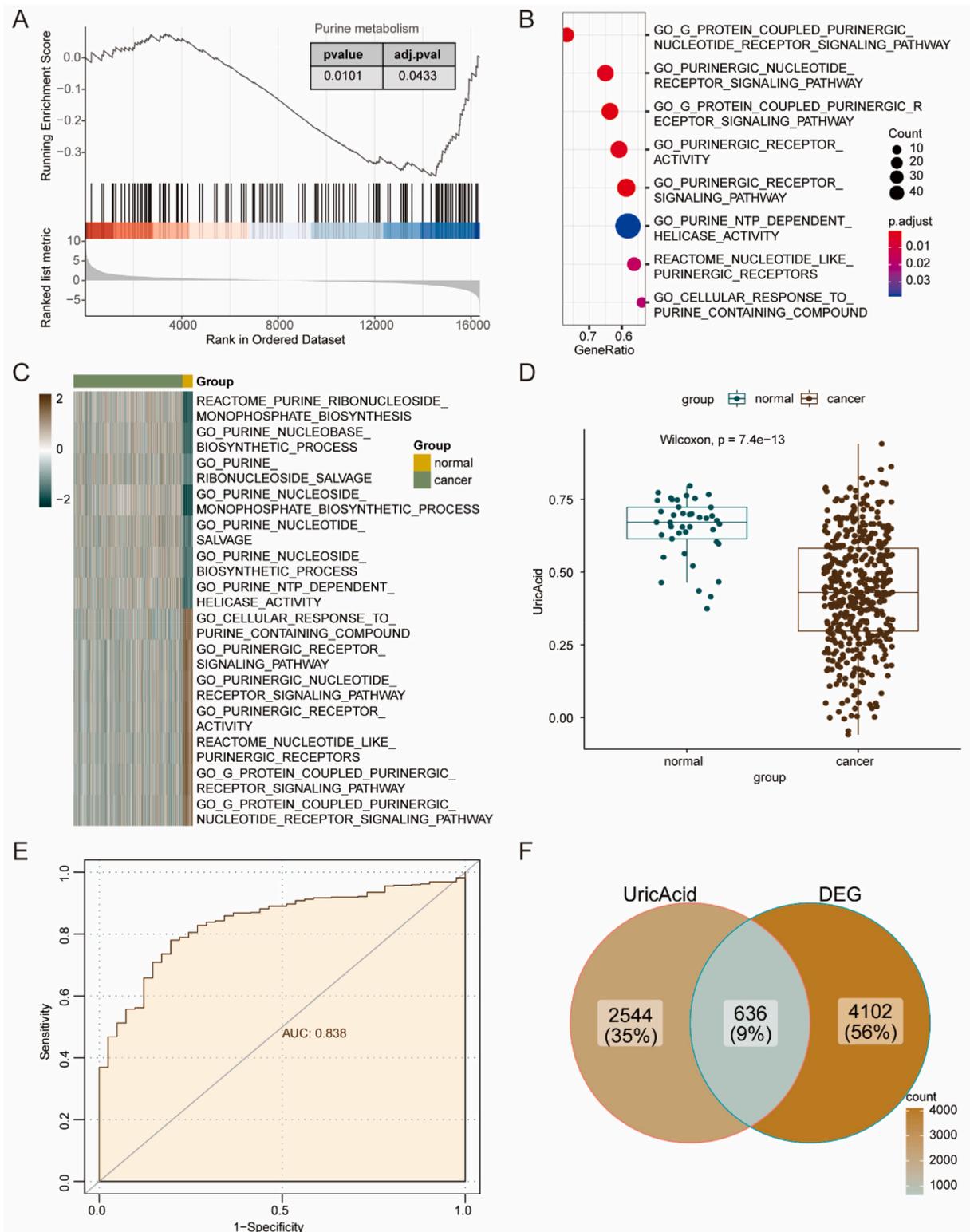


Fig. 1. Diagnosis value of uric acid-related genes

(A)Enrichment plots of purine metabolism-associated pathway from GSEA analysis. (B) Bubble plots of purine metabolism-associated pathway from GSEA analysis. (C) Heatmap of significantly differential pathway between CRC and normal tissues, yellow: normal, green: CRC. (D) Box plots of uric acid-related DEGs, brown: CRC, green: normal. (E)Area under the curve (AUC) of ROC curve displaying high diagnosis efficacy of uric acid-related genes (AUC = 0.838).(F)The Venn diagram of uric acid-related DEGs.

3.2. Diagnostic signature of UA-related genes in CRC

In enrichment analysis, we found that purine metabolism was significantly reduced in CRC (Fig. 1A). Moreover, the results of the GSEA and GSVA enrichment analyses showed that several purine metabolism-related pathways in the MSigDB dataset were significantly altered in CRC and normal tissues (Fig. 1B and C). Disorders in purine metabolism may be involved in CRC development, and several clinical studies have suggested that serum UA may be associated with CRC risk. To further explore the role of UA-related genes in CRC, we searched the GeneCards database for UA-related genes using the keyword “Uric Acid,” and finally obtained 3180 UA-related genes. The ssGSEA algorithm was used to calculate the enrichment scores of UA-related genes in each TCGA-COAD sample. The enrichment levels of UA-related genes were significantly different between colon cancer and normal tissues (Fig. 1D). The results of the receiver operating characteristics (ROC) curve analysis also showed high accuracy (AUC = 83.8 %) of using the UA-related gene enrichment score as a diagnostic marker (Fig. 1E).

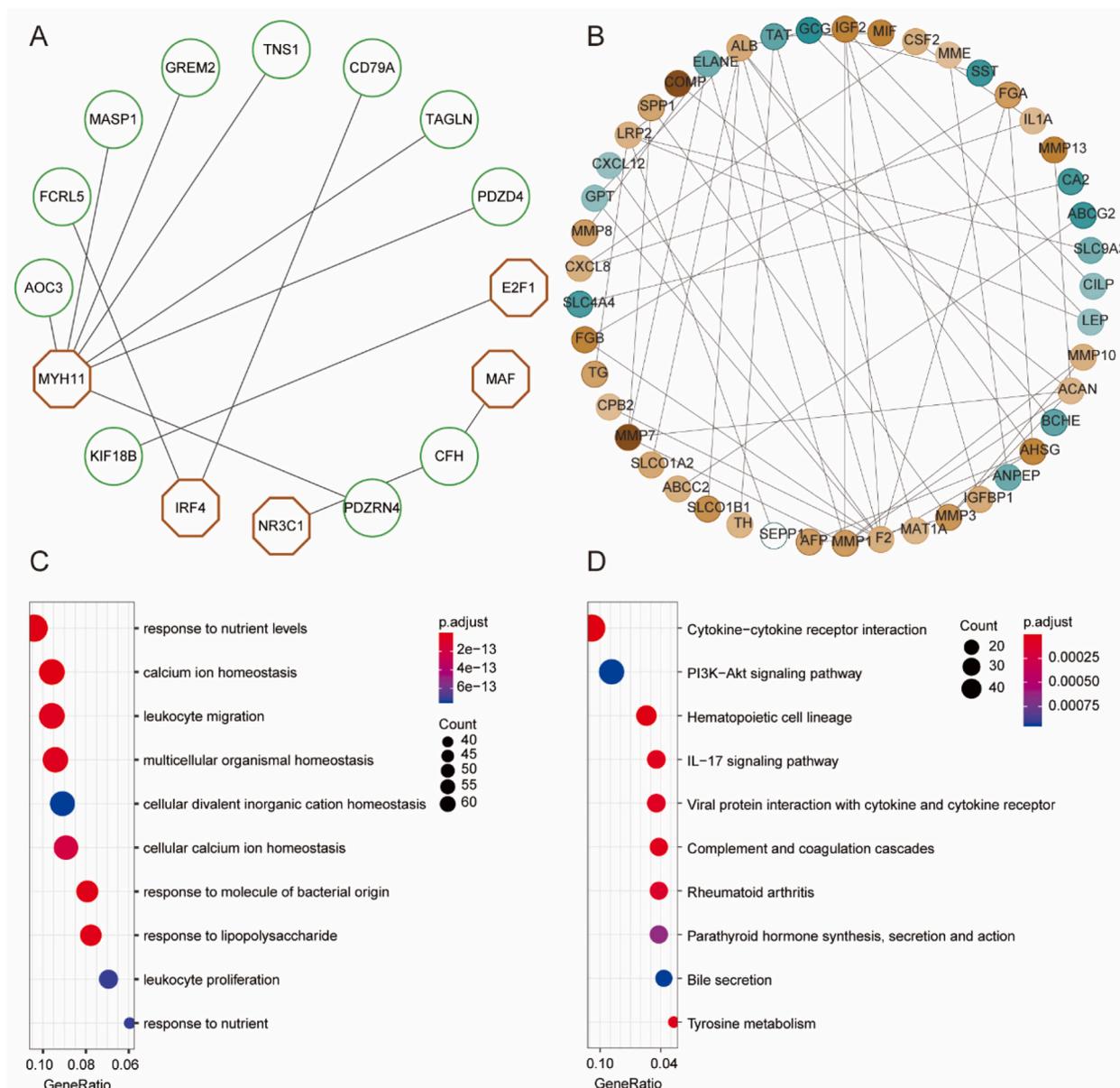


Fig. 2. The signature of uric acid-related DEGs (A)Regulation network based on transcript factors of MYH11, IRF4, NR3C1, MAF, E2F1, brown: Transcript factors, green: correlated genes (correlation coefficient>0.8).(B) PPI network constructed by the top 150 uric acid-related DEGs.(C) Bubble plots displaying GO annotation analysis of uric acid-related DEGs.(d) Bubble plots displaying KEGG pathway analysis of uric acid-related DEGs.

3.3. Identification of the signature of UA metabolism-associated DEGs

By intersecting the UA-related genes with DEGs, we obtained 636 UA-related DEGs, which were visualized using a Venn diagram (Fig. 1F). Among the UA-related DEGs, abnormally expressed transcript factors (*MYH1*, *NR3C1*, *IRF4*, *MAF*, and *E2F1*) were identified.

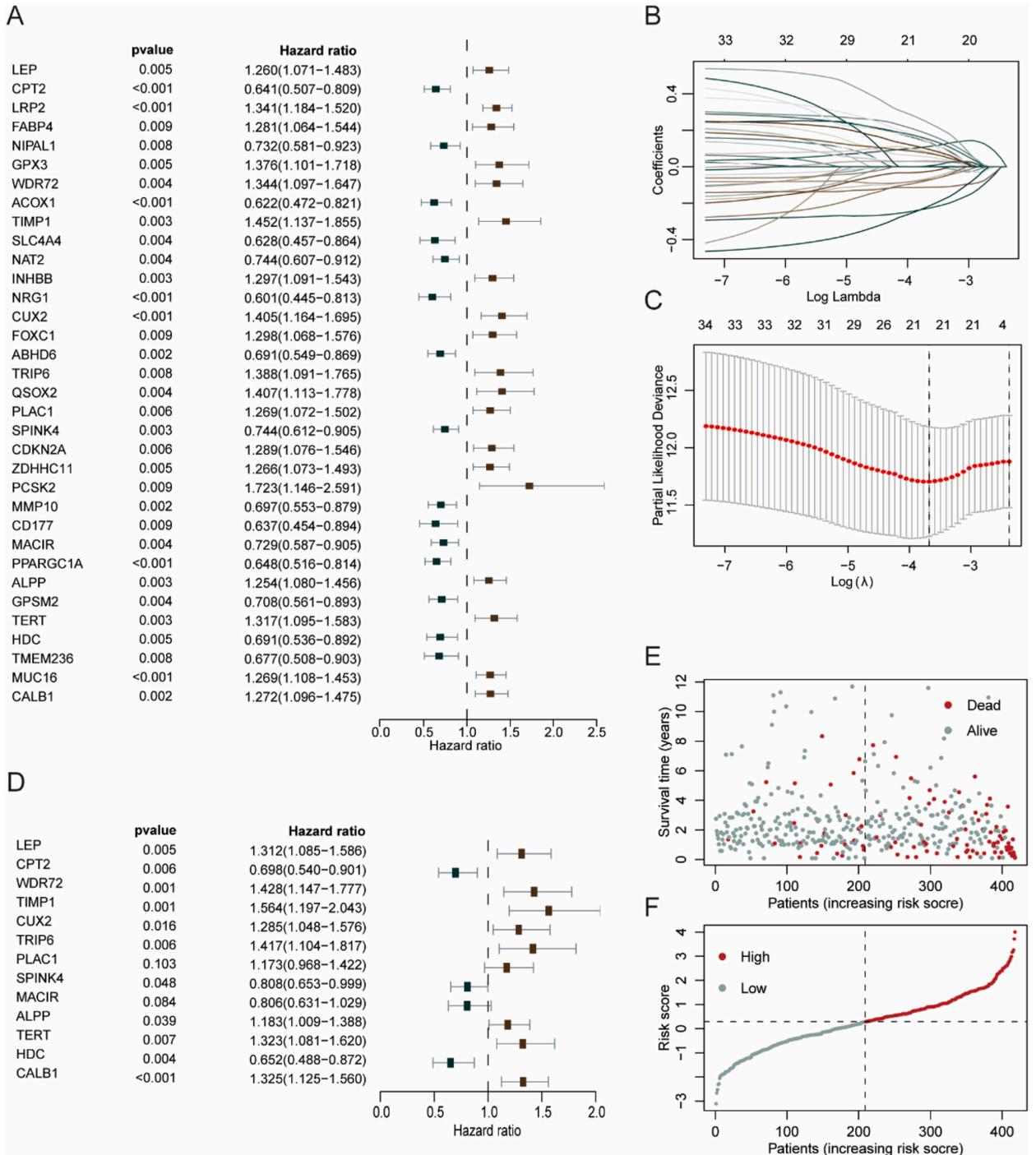
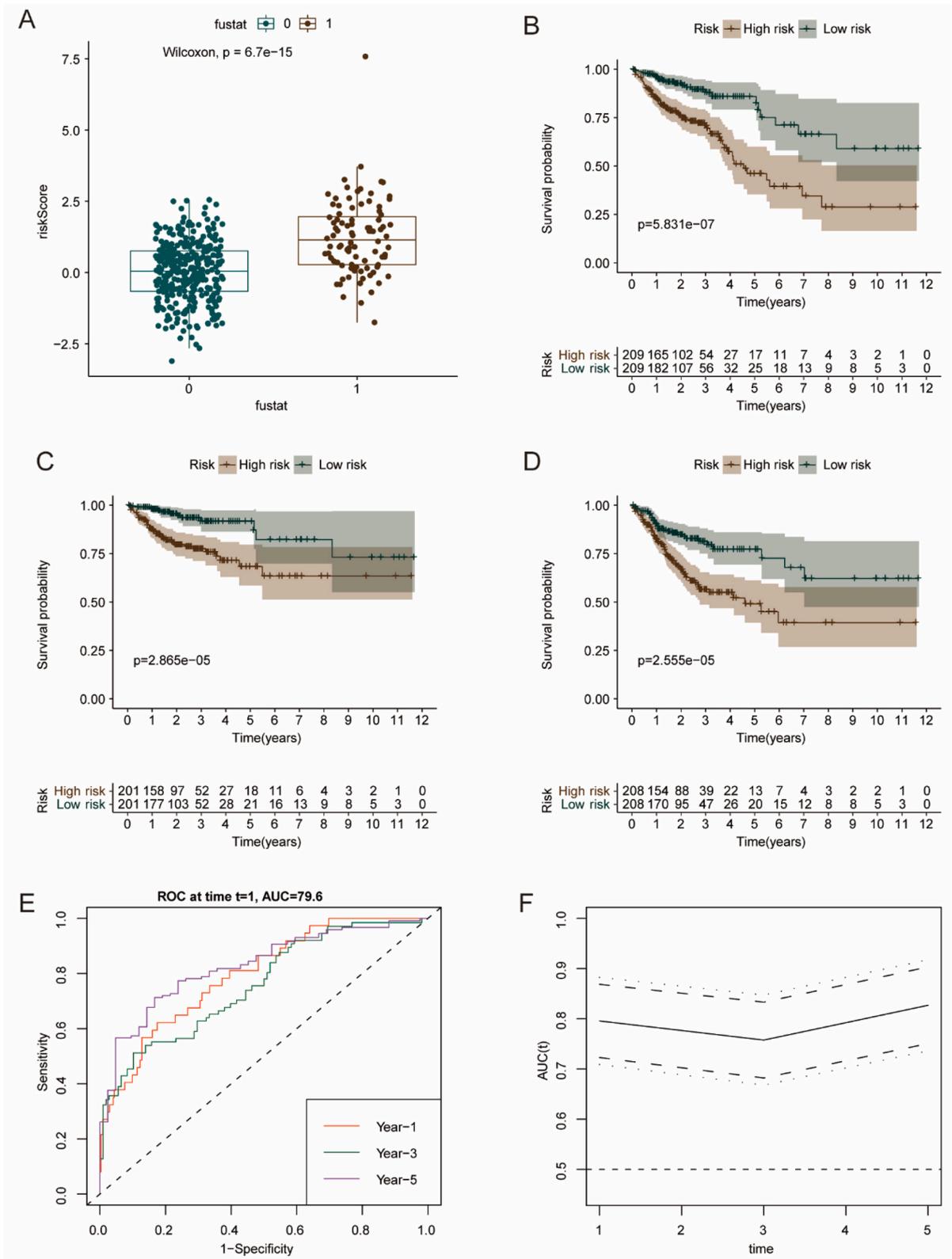


Fig. 3. Construction of risk predictive model

(A)Cox regression analysis of uric acid-related DEGs, brown: HR > 1, green: HR < 1. (B)The selection of LASSO model tuning parameters by 10-fold cross validation. (C)LASSO coefficient of uric acid-related DEGs gene set. Dashed line showing the value of 10-fold cross validation.(D)13 independent risk factors in risk predictive model constructed by multivariate COX regression analysis.(E)Distribution of patients in TCGA cohort based on survival status with risk value in model. (F)Distribution of patients in low- and high-risk group based on the median risk score.



(caption on next page)

Fig. 4. The predictive value of uric acid-related DEGs risk score model in survival status of patients from TCGA cohort

(A) The box plot of differential risk value between survival patients and dead patients, 0: survival, 1: dead. (B) The comparison of overall survival between low- and high-risk group, brown: high-risk group, green: low-risk group. (C) The comparison of disease specific survival (DSS) between low- and high-risk groups, brown: high-risk group, green: low-risk group. (D) The comparison of progression free survival (PFS) between low- and high-risk group, brown: high-risk group, green: low-risk group.(E) AUC of time-dependent ROC curve suggesting the good performance of the risk scores (AUC>0.7). (F) Diagnosis efficacy of AUC in time progression based on risk scores.

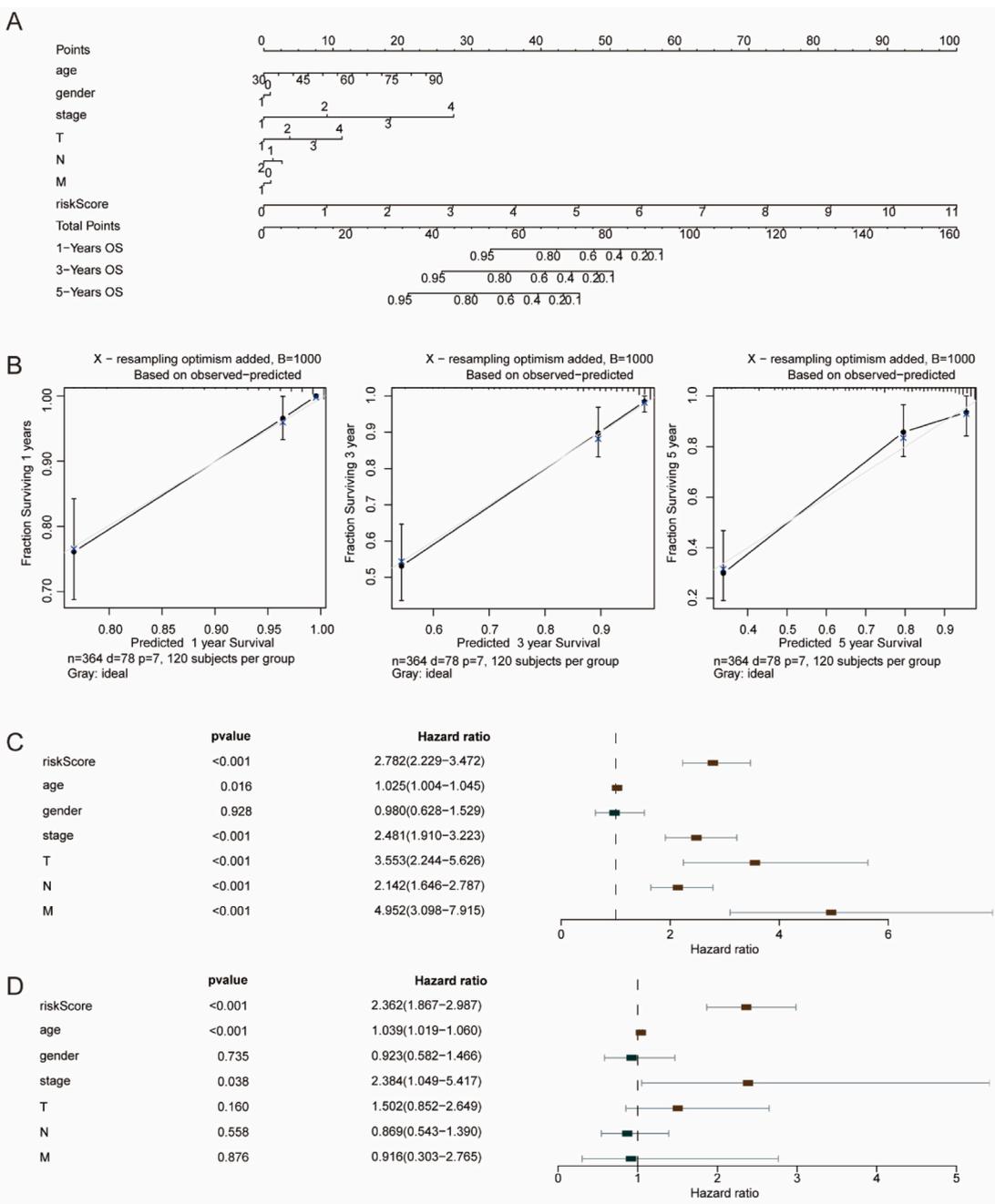
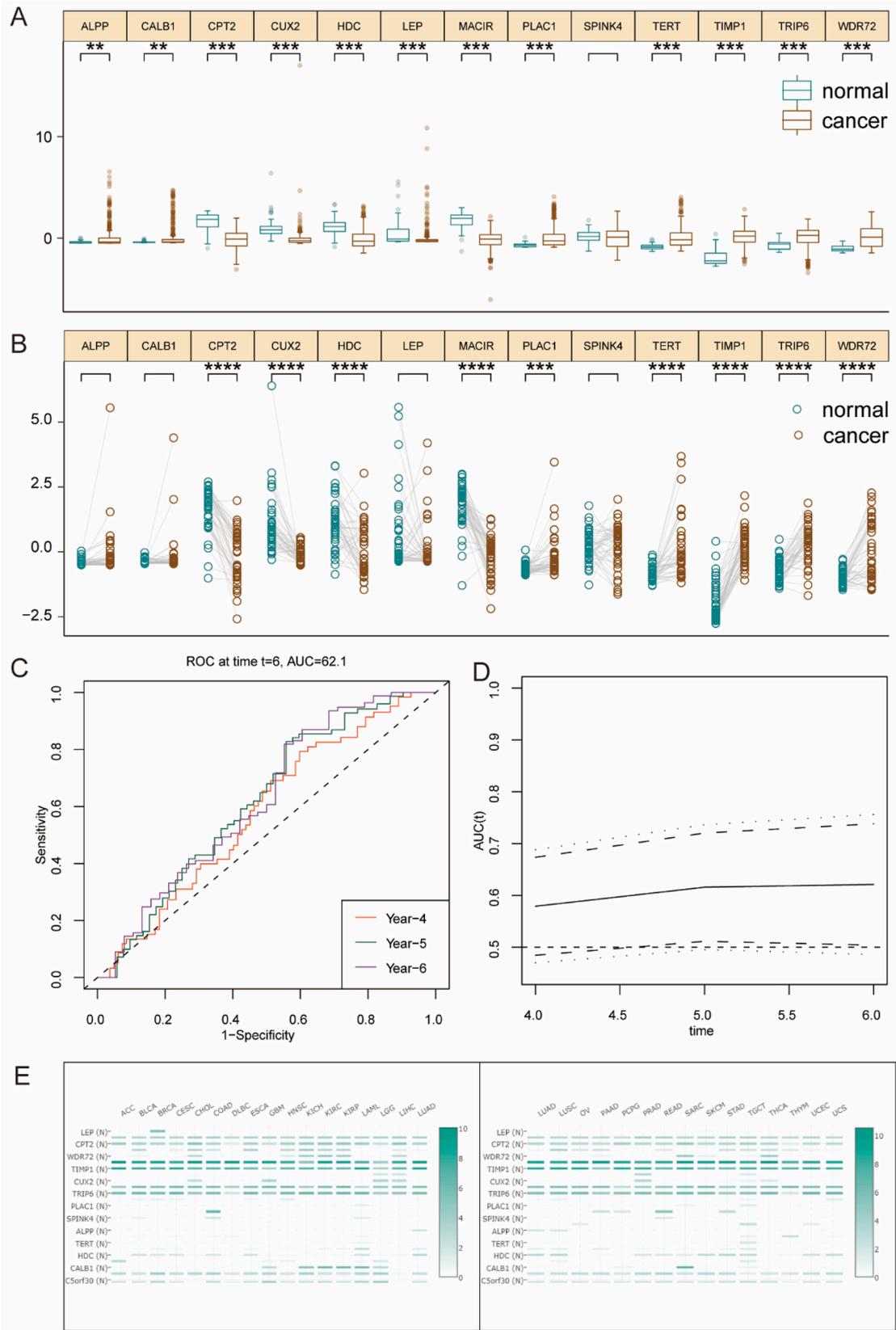


Fig. 5. The predictive value of uric acid-related DEGs risk score model in combination with clinical pathological characteristics of patients from TCGA cohort

(A) Nomogram elucidating higher contribution value than other clinical pathological characteristics. (B) The calibration plots of the nomogram. (C) Univariate Cox regression analysis of the nomogram. (E) Multivariate Cox regression analysis of the nomogram.



(caption on next page)

Fig. 6. The verification of model gene set expression profile and predictive risk model

(A) The box plots of the expression difference of 13 genes in risk model between CRC and normal tissues. (B) The paired dot plots of the expression difference of 13 genes in risk model between CRC and cancer adjacent tissues. (C) The validation of predictive ability of risk model in GSE17536 cohort. (D) Diagnosis efficacy of AUC in time progression based on risk scores in GSE17536 cohort. (E) The heatmap of differential expression of 13 model genes in pan-cancer. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Abnormal expression of transcription factors often causes extensive abnormal expression of their downstream target genes. We constructed a transcription factor network by performing Pearson's correlation analysis with a correlation coefficient of 0.8 and a threshold of P -value < 0.001 (Fig. 2A). Subsequently, we took the top 150 UA-related differential genes in the descending order of ploidy. We further constructed a protein–protein interaction network using STRING (correlation coefficient 0.9) and visualized it using Cytoscape (Fig. 2B). The top three genes with the highest protein interactions were *ALB* (7), *ACAN* (5), and *F2* (5). To further explore the functions and pathways of UA-related genes, we performed a functional analysis of the 636 differentially expressed UA-related genes. The GO functional analysis showed that these genes are involved in the body's response to nutrients, homeostasis, and leukocyte migration (Fig. 2C), and the KEGG functional analysis suggested that these genes are involved in tumor signaling pathways, such as Phosphatidylinositide 3 kinases-protein kinase B (PI3K-Akt) immune-related pathways, and tyrosine metabolism (Fig. 2D).

3.4. Construction of the prognostic model of UA-related genes

To explore the role of UA-related genes in the prognosis of CRC, we used a univariate COX regression analysis to identify 35 UA-related genes that might be associated with CRC prognosis from the 636 UA-related DEGs (Fig. 3A). Subsequently, the accuracy of the prediction model was improved by the least absolute shrinkage and selection operator (LASSO) regression analysis. The model had the best predictive accuracy when 21 genes were selected (*LEP*, *CPT2*, *LRP2*, *WDR72*, *TIMP1*, *INHBB*, *CUX2*, *TRIP6*, *QSOX2*, *PLAC1*, *SPINK4*, *ZDHHC11*, *MMP10*, *MACIR*, *PPARGC1A*, *ALPP*, *GPSM2*, *TERT*, *HDC*, *MUC16*, and *CALB1*) (Fig. 3B and C). Finally, a multivariable Cox regression analysis was used to identify 13 independent predictors (*LEP*, *CPT2*, *WDR72*, *TIMP1*, *CUX2*, *TRIP6*, *PLAC1*, *SPINK4*, *MACIR*, *ALPP*, *TERT*, *HDC*, and *CALB1*) (Fig. 3D). We also constructed a prognostic model for CRC based on 13 UA-related genes and calculated the risk values for each sample based on the correlation coefficients. The model showed that the number of deaths significantly increased, and the survival time significantly decreased with an increase in risk values (Fig. 3E and F). We visualized the differences in risk values across survival states, and found that the risk values were significantly higher in the death group than in the survival group (Fig. 4A). The samples were divided into high- and low-risk groups with median values before performing a survival analysis, which showed that overall survival (OS), disease-specific survival (DSS), and progression-free survival (PFS) were significantly lower in the high-risk group (Fig. 4B–D). To verify the efficacy of the risk model, we constructed time-dependent ROC curves using risk values as predictors. The results showed that the 1-, 3-, and 5-year AUC values were 79.57 %, 75.76 %, and 82.69 %, respectively, indicating high predictive accuracy of the model (Fig. 4E and F). Using a nomogram, we found that the risk score contributed most to the survival risk compared with age, gender, stage, and TNM stage (Fig. 5A). Subsequently, calibration curves were used to test the prediction accuracy of the model and found that the prediction accuracies of 1-, 3- and 5-year survival rates were high (Fig. 5B). The clinicopathological characteristics of patients in the TCGA-COAD cohort are shown in Table S1.

To determine the effect of clinicopathological characteristics on risk scores, we performed univariate and multivariable Cox analysis by combining clinical characteristics (age, sex, stage, T, N, M), and the results showed that the risk score of the prediction model remained an independent risk factor after combining other clinical characteristics (multifactorial COX regression HR = 2.363, 95 % CI [1.867–2.987]; $P < 0.001$) (Fig. 5C and D). We visualized the expression of 13 model genes in cancer versus normal tissues, among which *CPT2*, *CUX2*, *HDC*, *LEP*, and *MACR* were poorly expressed in cancer, whereas *ALPP*, *CALB1*, *PLAC1*, *TERT*, *TIMP1*, *TRIP6*, and *WDR72* were highly expressed in cancer. The cancer versus paracancer expression profiles of *CPT2*, *CUX2*, *HDC*, *MACR*, *ALPP*, *CALB1*, *PLAC1*, *TERT*, *TIMP1*, *TRIP6*, and *WDR72* were significantly different, whereas those of *ALPP*, *CALB1*, *LEP*, and *SPINK4* were not significantly different, and this was attributed to the small paired-sample size (Fig. 6A and B). We then downloaded the CRC-related dataset (GSE17536) from the GEO database for prognostic model validation, and the TimeROC results indicated that our prognostic model had some predictive power (Fig. 6C and D). We explored the expression of 13 predictive genes in pan-cancer using the Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/>), and the results suggested that these 13 predictive genes were significantly differentially expressed in multiple cancers and normal tissues.

3.5. Comparison of somatic mutation and tumor immune infiltration

We visualized the mutation gene profiles of TCGA-COAD samples using the “maftools” R package (Fig. 7A). A horizontal histogram shows the mutation characteristics of the genetic loci with high mutation frequencies, and the top 10 genes with the highest mutation frequencies were *APC* (71 %), *TP53* (54 %), *TTN* (49 %), *KRAS* (42 %), *SYNE1* (30 %), *MUC16* (30 %), *PIK3CA* (30 %), *FAT4* (25 %), *RYR2* (22 %), and *CSMD3* (21 %). Similarly, horizontal histograms were used to show mutations in 13 prognostic model genes (Fig. 7B and C), with the top three being *CUX2* (9 %), *WDR72* (4 %), and *TRIP6* (3 %). Detailed mutation information for each sample is shown in the waterfall plot, with different mutation types annotated in different colors at the bottom. Lollipop plots were applied to present the amino acid mutation sites of *CUX2*, *WDR72*, and *TRIP6* (Fig. 7D). Furthermore, we analyzed the mutation correlation among the selected 13 genes (Fig. 7E).

Functional analysis of the UA-related DEGs suggested a close association with immunity. We computed the level of infiltration of 22

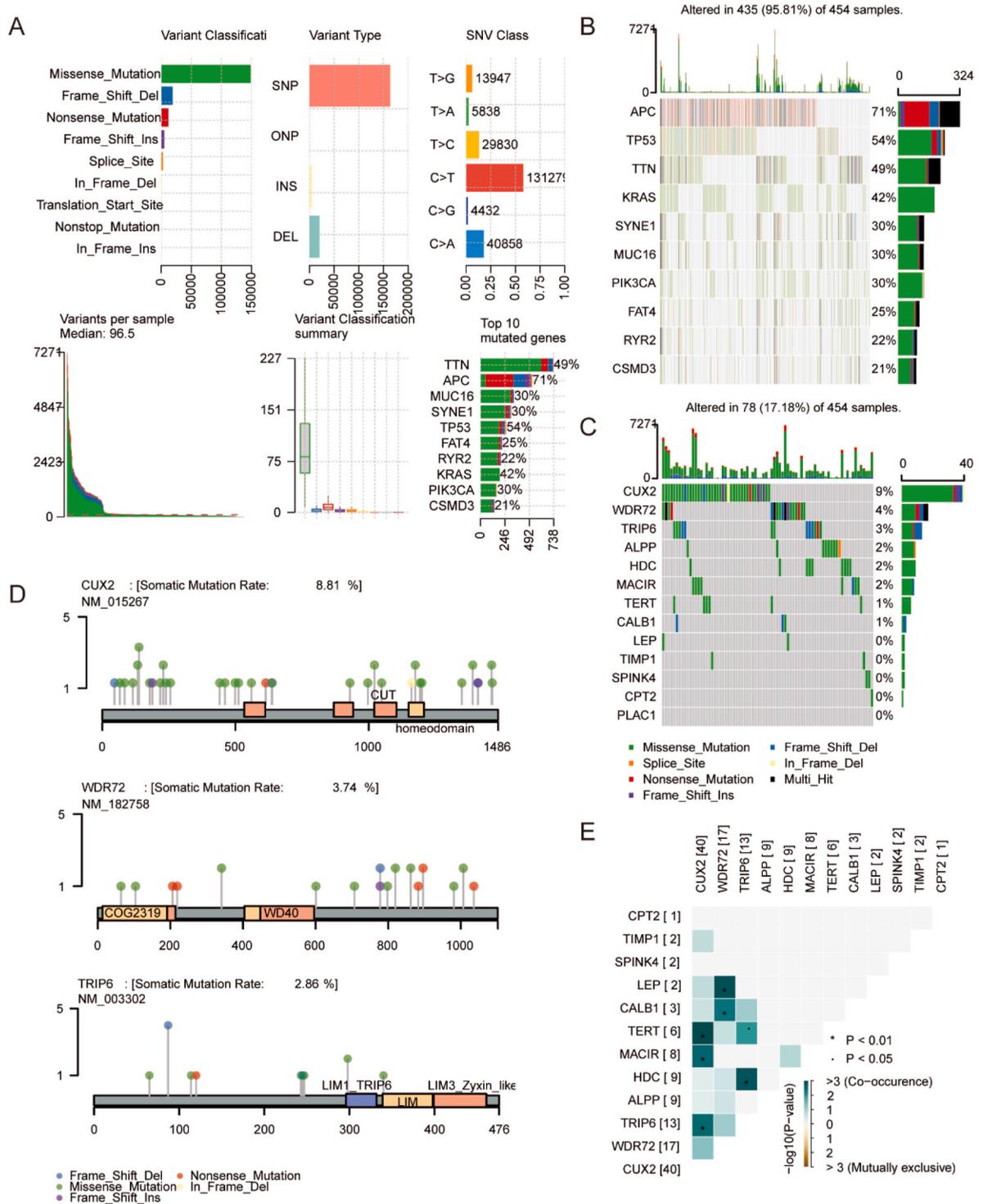


Fig. 7. The mutation signature of data from TCGA-COAD (A) The mutation profile of data from TCGA-COAD. (B) The waterfall plot depicting the top 10 mutation genes in TCGA-COAD samples. (C) The waterfall plot depicting the mutation frequency of 13 uric acid-related DEGs of risk model in TCGA-COAD samples. (D) The somatic mutation rate of CUX2, WDR72 and TRIP6. (E) The mutation co-occurrence and exclusion analyses of 13 uric acid-related DEGs of risk model.

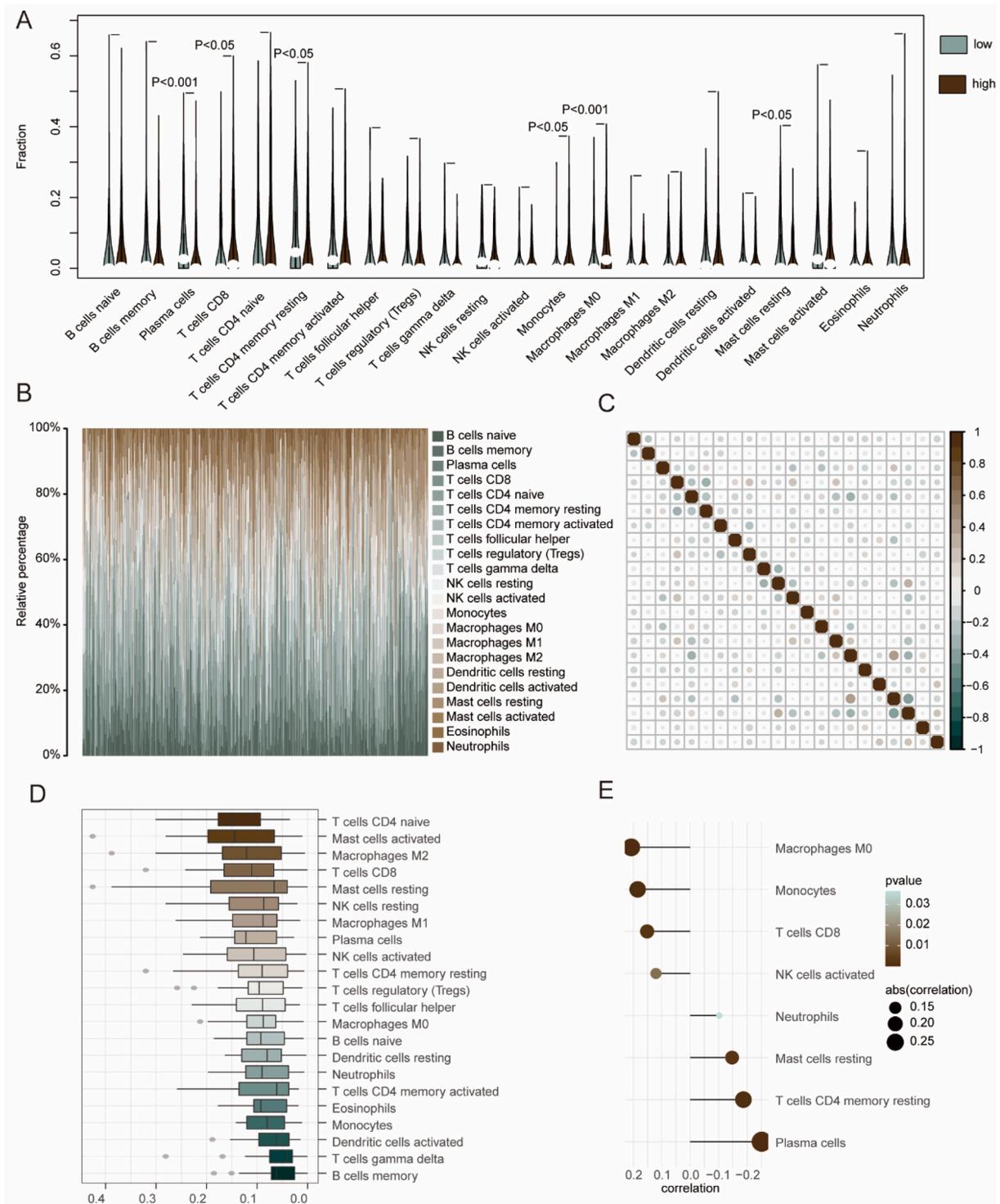


Fig. 8. The analyses of immune infiltration in TCGA cohort (A) The violin plots displaying the distribution difference of 22 immune cell types in low- and high-risk group estimated by CIBERSORT. (B) The bar graphs showing the proportion of 22 immune cells types. (C) The heatmap of correlation between immune cell types.(D) Friend analysis of the connectivity of 22 immune cell types.(E) The lollipop diagrams for the correlation between immune cell types and hazard ratio.

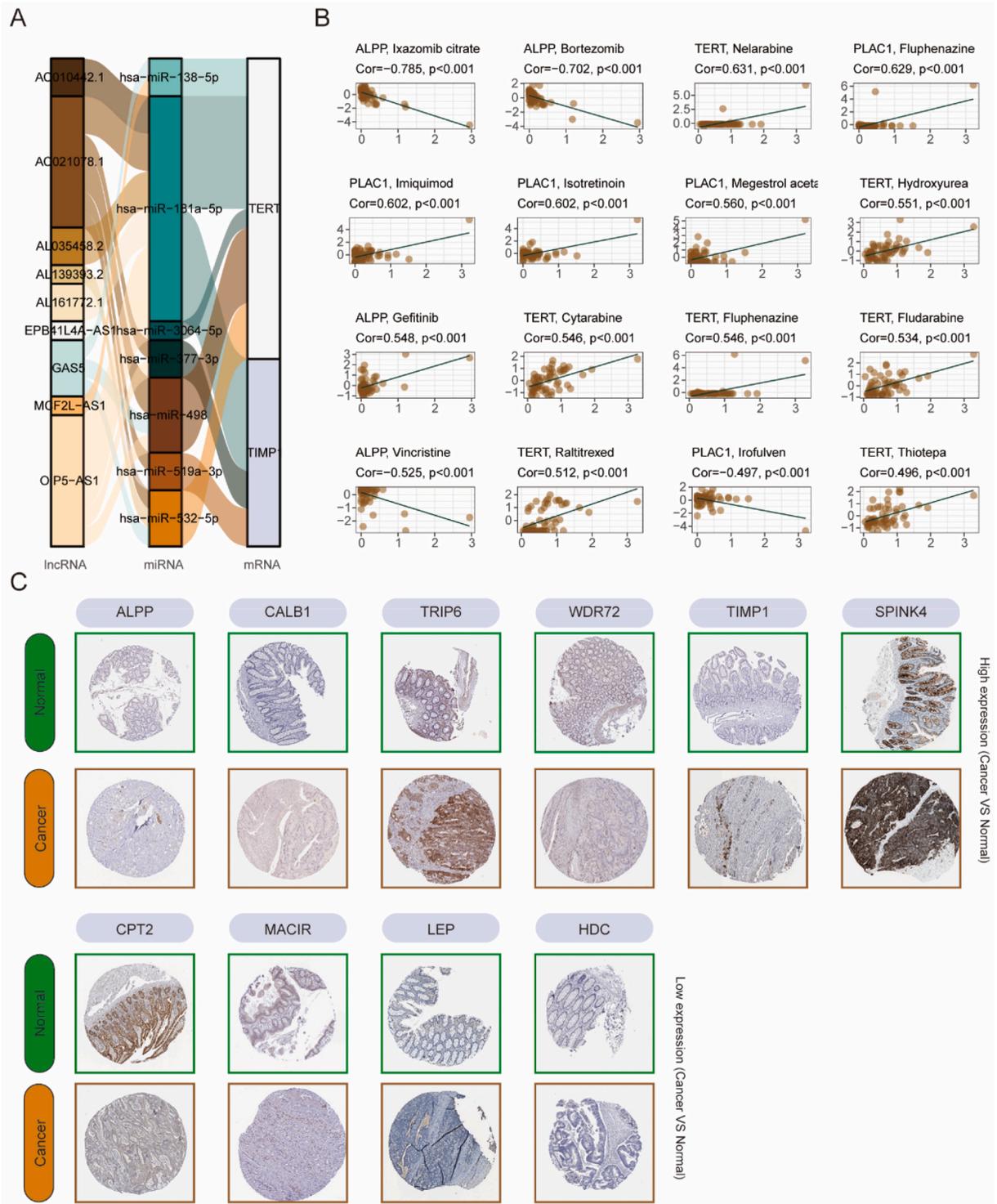


Fig. 9. The construction of ceRNA network, the sensitivity of chemotherapy and immunohistochemistry (A) The alluvial diagram displaying the ceRNA network of TERT and TIMP1. (B) The top 16 susceptibility results based on analysis of chemotherapy sensitivity of 13 uric acid-related DEGs of risk model. (C) The immunohistochemistry results of 13 uric acid-related DEGs of risk model from The Human Protein Atlas database.

immune cells in CRC samples using the CIBERSORT algorithm. We found significant differences in some immune cells in high- and low-risk groups, where CD8⁺ T cells, monocytes, and M0 macrophages were significantly increased in CRC while plasma cells, resting CD4 memory T cells, and resting mast cells were significantly decreased (Fig. 8A). To further understand immune cell infiltration, we visualized the distribution of immune cells in each sample and performed a correlation analysis of 22 immune cells (Fig. 8B and C) and found that naïve CD4⁺ T cells and activated mast cells had the highest association with other immune cells according to Friend analysis (Fig. 8D). Further, the correlation were explored between immune cell expression and risk value using a correlation analysis with a significance threshold of $P < 0.05$. The results showed that the abundance levels of M0 macrophages, monocytes, CD8⁺ T cells, and NK cells were positively correlated with the risk value, whereas those of plasma cells, resting CD4⁺ memory T cells, resting mast cells, and neutrophils were negatively correlated with the risk value (Fig. 8E).

3.6. Analysis of ceRNA network and chemotherapy sensitivity

To further reveal the regulatory network of UA-related gene sets, we constructed ceRNA networks. Based on the miRtarBase database, we screened the miRNAs of predicted gene sets using the threshold of validation with a dual-luciferase reporter assay and predicted lncRNAs based on the starBase database with the threshold of pancancerNum >5 and clipExpNum >3 . The results were visualized using the Cytoscape software. Finally, a ceRNA network of *TERT* and *TIMP1* was constructed (Fig. 9A). To further explore UA-related predicted gene-sensitive drugs, we determined the chemotherapy sensitivity of the predicted gene sets using the CellMiner database. The chemotherapeutic drugs were screened using clinical trials and U.S. Food and Drug Administration (FDA) approval as the benchmark. The Pearson's correlation coefficient of the predicted gene set and chemotherapeutic drugs was calculated, and the top 16 significantly associated genes and drugs were visualized (Fig. 9B). We further confirmed the expression of the predicted genes in cancer versus normal tissues using The Human Protein Atlas (<https://www.proteinatlas.org/>) database, where all 10 genes were found to have significant protein expression in cancer versus normal tissues. Immunohistochemically, *PLAC1* and *TERT* levels showed no significant differences between cancer and normal tissues, while *CUX2* expression level remained unchanged (Fig. 9C).

3.7. Expression of *TRIP6*, *TERT*, *WDR72*, *PLAC1*, and *TIMP1*

To verify the results of the above analysis, we tested the mRNA expression levels of *TRIP6*, *TERT*, *WDR72*, *PLAC1*, and *TIMP1* in different CRC cell lines (HCT116, SW480, SW620, RKO, HT29, and SW1116) and in the normal colonic epithelial cell line FHC. The results of the qPCR assay showed that the expression levels of *TRIP6* and *PLAC1* levels were significantly higher in CRC cell lines than in

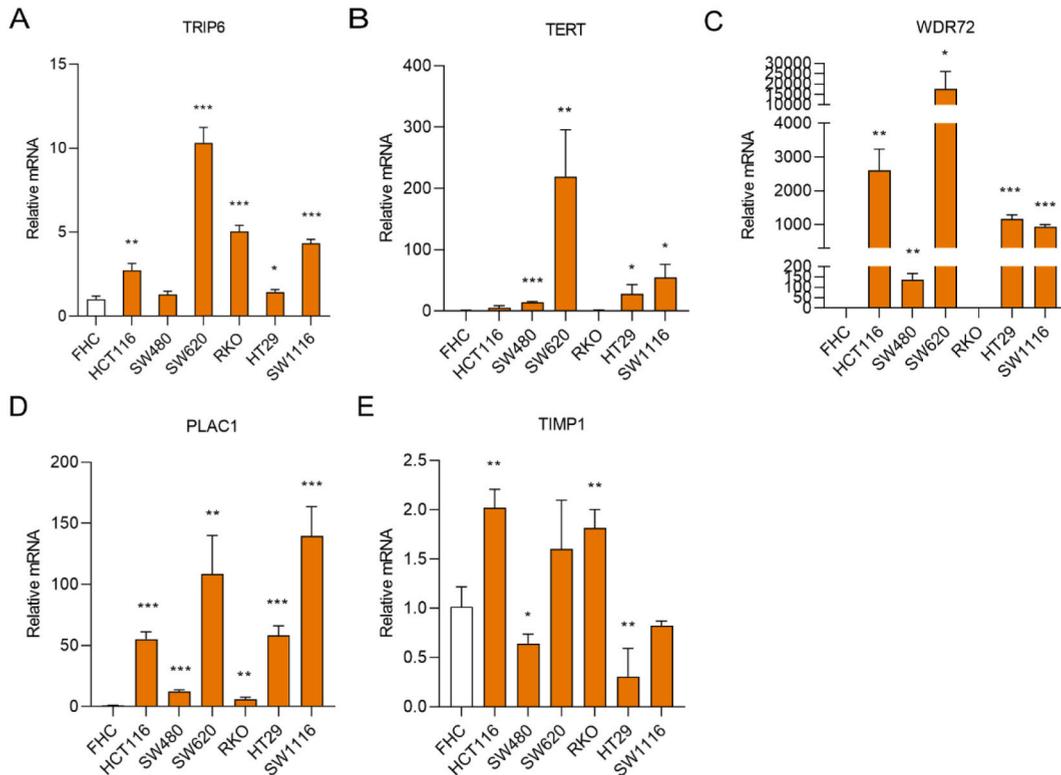


Fig. 10. The expression of *TRIP6*, *TERT*, *WDR72*, *PLAC1* and *TIMP1* in CRC cell lines and normal colonic cell line. The mRNA expression levels of *TRIP6*(A), *TERT*(B), *WDR72*(C), *PLAC1*(D) and *TIMP1*(E) measured by q-pcr. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

FHC. Excluding the RKO cell line, the mRNA levels of *TERT* and *WDR72* were significantly elevated in the HCT116, SW480, SW620, HT29, and SW1116 cell lines. In SW480 and HT29 cells, the mRNA expression level of *TIMP1* was significantly lower than that in FHC, while it was significantly higher in the HCT116 and RKO cell lines than in FHC (Fig. 10).

4. Discussion

CRC is a multifactorial digestive system disease with a high risk of metastasis and recurrence, posing a significant global health burden [37]. Despite extensive research on CRC development, the exact mechanism of CRC -microenvironment interactions has not been elucidated. Recently, cancer cell metabolism reprogramming has become a research hotspot in cancer initiation and progression [38]. Dysregulation of metabolic pathways is a hallmark of cancer [39]. Aberrant changes in metabolism trigger uncontrolled CRC cell proliferation and an unbalanced tumor microenvironment. Although changes in gene and protein expression have been profiled in CRC, little is known about metabolism-related gene alterations that characterize colorectal neoplastic progression.

UA, a catabolic product of purine derived from RNA and DNA, is a well-known antioxidant and proinflammatory compound. UA disorder is a common cause of obesity, diabetes, and metabolic syndrome [40], and has been reported to induce the progression of CRC. However, the integrated roles of multiple UA-related genes in CRC are unknown. Our analysis revealed a novel perspective on the potential significance of UA-related genes in CRC prognosis, thus providing an effective therapeutic strategy.

In this study, we aimed to explore the potential association between UA-related genes and CRC risk. We screened 4786 upregulated DEGs and 3480 downregulated DEGs in CRC tissues compared to normal tissues from the TCGA database. The GSEA, GSVA, and MsigDB analyses were applied for further functional and pathway enrichment, and the identified DEGs were found to be enriched in purine metabolism-related pathways. Additionally, ROC curves were used to test the diagnostic efficacy of the UA enrichment score of the samples in the TCGA-COAD cohort, and the results showed that the UA-related enrichment score is a highly accurate diagnostic marker, suggesting that UA-related metabolic problems play a vital role in CRC tumorigenesis. Of the 8266 DEGs, 636 UA-related DEGs were identified by analyzing the data from the GeneCards database, and a gene signature was constructed in the TF and STRING networks. The GO and KEGG enrichment results showed that UA-related DEGs participate in CRC progression by playing an important role in tumor cell homeostasis and immune response signaling pathways, in accordance with the antitumor immunity function of UA in previous studies [41]. Moreover, we identified 13 UA-related genes (*LEP*, *CPT2*, *WDR72*, *TIMP1*, *CUX2*, *TRIP6*, *PLAC1*, *SPINK4*, *MACIR*, *ALPP*, *TERT*, *HDC*, and *CALB1*) from 636 UA-related DEGs as independent risk predictive factors using univariate COX and LASSO regression analyses and established a prognostic risk score model for CRC in the TCGA cohort. The model could predict the OS, DSS, and PFS of patients with CRC with good accuracy by calculating the risk value of the tumor samples, and the risk score contributed more than the age, sex, stage, and TNM stage. Subsequently, the prognostic risk model was validated in a GEO cohort. While previous studies focused on the predictive function of UA level in CRC prognosis [14,42], this study offers a novel insight into the prognostic role of UA-related genes in CRC. Our data highlight the value of combining biomarkers to predict the prognosis of CRC development.

Notably, because there were significant differences between the low- and high-risk score groups, the UA-related DEGs in the prognostic risk models were further explored. We detected elevated expression levels of *TRIP6*, *PLAC1*, *WDR72*, *TERT*, and *TIMP1* in CRC tissues compared to normal and para-cancerous tissues in the TCGA database. In addition, we detected the mRNA expression of these candidate genes in the colonic epithelial cell line FHC and the CRC cell lines HCT116, SW480, SW620, RKO, HT29, and SW1116. Elevation in the expression levels of these five UA-related genes in CRC cell lines compared to FHC indicated that they are closely related to the progression and prognosis of CRC. Thyroid Hormone Receptor Interactor 6 (*TRIP6*) is a member of the zyxin family that localizes to focal adhesion sites and serves as an adaptor protein to regulate diverse biological processes [43]. *TRIP6*, functioning as an oncogene, promotes Ewing's sarcoma cell migration, invasion, and growth [44] and accelerates cervical cancer cell proliferation and invasion by upregulating oncogenic YAP signaling [44]. In addition, *TRIP6* has been reported to serve as a positive regulator of inflammatory damage during the pathogenesis of colitis via *TRAF6* activation [45]. We also calculated the tumor mutation burden in our TCGA-COAD samples. A total of 2.86 % of *TRIP6* somatic mutations were identified in CRC samples, which suggested its potential role in CRC progression.

WD repeat domain 72 (*WDR72*), a member of the WD 40 repeat domain superfamily, plays a major role in tooth enamel formation [46] and calcium vesicle turnover [47]. However, its function in cancer has rarely been reported. Recently, Ouyang et al. reported that *WDR72* enhances the stemness of lung cancer cells by regulating the AKT/HIF-1 (protein kinase B/Hypoxia-inducible factor 1) signaling pathway. In this study, compared with that in normal tissues and colonic cells, the expression level of *WDR72* was significantly elevated in CRC samples and CRC cell lines, with a 3.74 % somatic mutation rate in CRC samples in the TCGA-COAD database, indicating that abnormal expression and mutation of *WDR72* might increase CRC risk.

Placenta-specific 1 (*PLAC1*) is involved in placental development and is frequently activated and expressed in various cancers. *PLAC1* is a specific antigen in tumor immune response and may play an important role in cancer immunotherapy [48]. Furthermore, *PLAC1* expression was found to be positively correlated with sensitivity to fluphenazine, imiquimod, and isotretinoin, but negatively correlated with sensitivity to irifolven, suggesting that *PLAC1* is a prospective chemotherapy target.

Telomerase reverse transcriptase (*TERT*) is a ribonucleoprotein polymerase that maintains telomere ends by adding the telomere repeat TTAGGG. It engages in tumorigenesis via multiple genetic and epigenetic alterations by affecting telomerase activity [49]. It showed higher expression levels in SW480, SW620, HT29, and SW1116 cell lines than in FHC, indicating that the upregulation of its expression might promote CRC development. Moreover, we predicted that *TERT* expression was positively associated with the sensitivity to nelarabine, hydroxyurea, cytarabine, fluphenazine, fludarabine, and raltitrexed, suggesting that *TERT* is a chemotherapeutic target.

Tissue inhibitor matrix metalloproteinase 1 (*TIMP1*) is a natural inhibitor of matrix metalloproteinases in extracellular matrix

degradation. *TIMP1* possesses anti-apoptotic functionality and is involved in cellular response to several cytokines and hormones. *TIMP1* is correlated with poor prognosis in a wide variety of cancers. It has been reported to induce neutrophil extracellular trap formation in pancreatic cancer [50]. Song et al. reported that *TIMP1* participates in colon cancer progression and metastasis via the Focal Adhesion Kinase-Phosphatidylinositide 3 kinases/protein kinase B(FAK-PI3K/AKT) and mitogen-activated protein kinase (MAPK) pathways [51]. *TIMP1* is closely associated with tumors and inflammation, which is in line with the pro-inflammatory function of UA. Therefore, these five UA-related DEGs are promising diagnostic and prognostic markers for CRC progression, and their functions and mechanisms should be further explored.

An imbalance in the antioxidant and pro-inflammatory patterns of UA can also trigger tumor progression [51]. In the current study, we attempted to elucidate the relationship between the risk model for UA-related genes and immune infiltration. The altered expression profile from the CIBERSORT analysis in our risk score model is highly consistent with current studies on immune cell infiltration. The high-risk CRC patients had increased CD8⁺ T cells, monocytes, and M0 macrophages, and the risk score exhibited a positive correlation with the infiltration of these three immune cell types. CD8⁺ T cells are an important cluster of cells that perform immune surveillance and antitumor immune responses, which play a key role in balancing the tumor immune microenvironment [52, 53]. The progression of monocyte differentiation into tumor-associated macrophages (TAMs) has been reported to contribute to tumor infiltration and metastasis; thus, an increase in monocyte levels could reflect the activity of TAMs [54].

Although we constructed an accurate CRC prognostic model and validated the candidate genes in a separate cohort, there are still some limitations to our study, including the lack of more samples. The retrieval of all data and samples from the public database might have conferred a lack of CRC cases covering all relevant regions. Further *in vivo* and *in vitro* experiments are needed to examine the role of the five screened genes in CRC progression. Finally, further functional studies are required to elucidate the molecular mechanisms underlying the role of UA-related DEGs in CRC.

In conclusion, our results support the notion that UA-related genes might be utilized as a noninvasive approach to assess the prognosis of CRC. We identified 13 UA-related genes with high prognostic value to construct a prognostic model. The risk score of candidate genes can be a new assessment tool for CRC prognosis that shows plausible biological and clinical significance. Immune infiltration of M0 macrophages, monocytes, CD8⁺ T cells, and activated NK cells was positively correlated with the risk score. Five significant UA-related genes were expressed at higher levels in CRC cell lines than in colonic cell lines, and they were associated with chemotherapy treatment. To the best of our knowledge, this is the first study to explore the relationship between UA-related genes and CRC and may provide a significant basis for future mechanistic studies.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81903056), and the Youth Foundation of Shanghai Municipal Commission of Health and Family Planning (No. 20184Y0095).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found in the article.

CRedit authorship contribution statement

Chun Zhuang: Writing – original draft, Supervision, Methodology, Formal analysis, Data curation. **Yifan Liu:** Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Ranran Gu:** Resources. **Shanqing Du:** Resources. **Yin Long:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e22587>.

References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J Clin* 72 (1) (2022) 7–33, <https://doi.org/10.3322/caac.21708>.
- [2] L.H. Biller, D. Schrag, Diagnosis and treatment of metastatic colorectal cancer: a review, *JAMA* 325 (7) (2021) 669–685, <https://doi.org/10.1001/jama.2021.0106>.

- [3] L. Klimeck, T. Heisser, M. Hoffmeister, H. Brenner, Colorectal cancer: a health and economic problem, *Best Pract. Res. Clin. Gastroenterol.* 66 (2023), 101839, [10.1016/j.bpg.2023.101839](https://doi.org/10.1016/j.bpg.2023.101839).
- [4] B. Medici, B. Riccò, E. Caffari, et al., Early onset metastatic colorectal cancer: current insights and clinical management of a rising condition, *Cancers* 15 (13) (2023), <https://doi.org/10.3390/cancers15133509>.
- [5] E.H. Jin, K. Han, D.H. Lee, et al., Association between metabolic syndrome and the risk of colorectal cancer diagnosed before age 50 Years according to tumor location, *Gastroenterology* 163 (3) (2022) 637–648.e632, [10.1053/j.gastro.2022.05.032](https://doi.org/10.1053/j.gastro.2022.05.032).
- [6] A.S. Aghabozorgi, A. Bahreyni, A. Soleimani, et al., Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives, *Biochimie* 157 (2019) 64–71, [10.1016/j.biochi.2018.11.003](https://doi.org/10.1016/j.biochi.2018.11.003).
- [7] M. Nenkov, Y. Ma, N. Gaßler, Y. Chen, Metabolic reprogramming of colorectal cancer cells and the microenvironment: implication for therapy, *Int. J. Mol. Sci.* 22 (12) (2021), <https://doi.org/10.3390/ijms22126262>.
- [8] N. Keum, E. Giovannucci, Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies, *Nat. Rev. Gastroenterol. Hepatol.* 16 (12) (2019) 713–732, <https://doi.org/10.1038/s41575-019-0189-8>.
- [9] L. Xia, L. Oyang, J. Lin, et al., The cancer metabolic reprogramming and immune response, *Mol. Cancer* 20 (1) (2021) 28, <https://doi.org/10.1186/s12943-021-01316-8>.
- [10] S.W. Waldrop, S. Niemiec, C. Wood, et al., Cord blood DNA methylation of immune and lipid metabolism genes is associated with maternal triglycerides and child adiposity, *Obesity* (2023), <https://doi.org/10.1002/oby.23915>.
- [11] Q. Hu, Y. Li, D. Li, et al., Amino acid metabolism regulated by lncRNAs: the propellant behind cancer metabolic reprogramming, *Cell Commun. Signal. : CCS* 21 (1) (2023) 87, <https://doi.org/10.1186/s12964-023-01116-1>.
- [12] H. Yanai, H. Adachi, M. Hakoshima, H. Katsuyama, Molecular biological and clinical understanding of the pathophysiology and treatments of hyperuricemia and its association with metabolic syndrome, cardiovascular diseases and chronic kidney disease, *Int. J. Mol. Sci.* 22 (17) (2021), <https://doi.org/10.3390/ijms22179221>.
- [13] W. Lu, Y. Xu, X. Shao, et al., Uric acid produces an inflammatory response through activation of NF- κ B in the hypothalamus: implications for the pathogenesis of metabolic disorders, *Sci. Rep.* 5 (2015), 12144, <https://doi.org/10.1038/srep12144>.
- [14] N. Mi, J. Huang, C. Huang, et al., High serum uric acid may associate with the increased risk of colorectal cancer in females: a prospective cohort study, *Int. J. Cancer* 150 (2) (2022) 263–272, [10.1002/ijc.33807](https://doi.org/10.1002/ijc.33807).
- [15] R.J. Kishton, J.C. Rathmell, Novel therapeutic targets of tumor metabolism, *Cancer Journal (Sudbury, Mass)* 21 (2) (2015) 62–69, <https://doi.org/10.1097/ppo.000000000000099>.
- [16] J. Yang, Y. Wang, Q. Zhao, et al., Association of serum uric acid with increased risk of cancer among hypertensive Chinese, *Int. J. Cancer* 141 (1) (2017) 112–120, <https://doi.org/10.1002/ijc.30731>.
- [17] A. Colaprico, T.C. Silva, C. Olsen, et al., TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, *Nucleic Acids Res.* 44 (8) (2016) e71, <https://doi.org/10.1093/nar/gkv1507>.
- [18] A. Mayakonda, D.C. Lin, Y. Assenov, C. Plass, H.P. Koeffler, Maftools: efficient and comprehensive analysis of somatic variants in cancer, *Genome Res.* 28 (11) (2018) 1747–1756, [10.1101/gr.239244.118](https://doi.org/10.1101/gr.239244.118).
- [19] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression Omnibus (GEO) and BioConductor, *Bioinformatics* 23 (14) (2007) 1846–1847, [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254).
- [20] M.E. Ritchie, B. Phipson, D. Wu, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47, <https://doi.org/10.1093/nar/gkv007>.
- [21] G. Stelzer, N. Rosen, I. Plaschkes, et al., The GeneCards suite: from gene data mining to disease genome sequence analyses, *Current protocols in bioinformatics* 54 (2016) 1, [30.31-31.3033](https://doi.org/10.3131-31.3033), doi.org/10.1002/cpbi.5.
- [22] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Software* 33 (1) (2010) 1–22.
- [23] E.K. Gustavsson, D. Zhang, R.H. Reynolds, S. Garcia-Ruiz, M. Ryten, ggrtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2, *Bioinformatics* 38 (15) (2022) 3844–3846, <https://doi.org/10.1093/bioinformatics/btac409>.
- [24] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (1) (2000) 25–29, [10.1038/75556](https://doi.org/10.1038/75556).
- [25] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (1) (2000) 27–30, [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27).
- [26] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS A J. Integr. Biol.* 16 (5) (2012) 284–287, <https://doi.org/10.1089/omi.2011.0118>.
- [27] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J.P. Mesirov, Molecular signatures database (MSigDB) 3.0, *Bioinformatics* 27 (12) (2011) 1739–1740, <https://doi.org/10.1093/bioinformatics/btr260>.
- [28] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 14 (2013) 7, <https://doi.org/10.1186/1471-2105-14-7>.
- [29] D.A. Barbie, P. Tamayo, J.S. Boehm, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, *Nature* 462 (7269) (2009) 108–112, <https://doi.org/10.1038/nature08460>.
- [30] A.M. Newman, C.L. Liu, M.R. Green, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457, [10.1038/nmeth.3337](https://doi.org/10.1038/nmeth.3337).
- [31] D. Szklarczyk, A.L. Gable, K.C. Nastou, et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res.* 49 (D1) (2021) D605–d612, [10.1093/nar/gkaa1074](https://doi.org/10.1093/nar/gkaa1074).
- [32] P. Shannon, A. Markiel, O. Ozier, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (11) (2003) 2498–2504, [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).
- [33] Y. Ru, K.J. Kechris, B. Tabakoff, et al., The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations, *Nucleic Acids Res.* 42 (17) (2014) e133, <https://doi.org/10.1093/nar/gku631>.
- [34] U.T. Shankavaram, S. Varma, D. Kane, et al., CellMiner: a relational database and query tool for the NCI-60 cancer cell lines, *BMC Genom.* 10 (2009) 277, <https://doi.org/10.1186/1471-2164-10-277>.
- [35] M. Uhlén, L. Fagerberg, B.M. Hallström, et al., Proteomics. Tissue-based map of the human proteome, *Science (New York, NY)* 347 (6220) (2015), 1260419, [10.1126/science.1260419](https://doi.org/10.1126/science.1260419).
- [36] R. Mahabir, M. Tanino, A. Elmansuri, et al., Sustained elevation of Snail promotes glial-mesenchymal transition after irradiation in malignant glioma, *Neuro Oncol.* 16 (5) (2014) 671–685, [10.1093/neuonc/not239](https://doi.org/10.1093/neuonc/not239).
- [37] J. Gong, Y. Lin, H. Zhang, et al., Reprogramming of lipid metabolism in cancer-associated fibroblasts potentiates migration of colorectal cancer cells, *Cell Death Dis.* 11 (4) (2020) 267, <https://doi.org/10.1038/s41419-020-2434-z>.
- [38] R.J. DeBerardinis, N.S. Chandel, Fundamentals of cancer metabolism, *Sci. Adv.* 2 (5) (2016), e1600200, <https://doi.org/10.1126/sciadv.1600200>.
- [39] N.N. Pavlova, C.B. Thompson, The emerging hallmarks of cancer metabolism, *Cell Metabol.* 23 (1) (2016) 27–47, [10.1016/j.cmet.2015.12.006](https://doi.org/10.1016/j.cmet.2015.12.006).
- [40] A. Andres-Hernando, C. Cicerchi, M. Kuwabara, et al., Umami-induced obesity and metabolic syndrome is mediated by nucleotide degradation and uric acid generation, *Nat. Metab.* 3 (9) (2021) 1189–1201, <https://doi.org/10.1038/s42255-021-00454-z>.
- [41] J. Wang, K. Liu, T. Xiao, P. Liu, R.A. Prinz, X. Xu, Uric acid accumulation in DNA-damaged tumor cells induces NKG2D ligand expression and antitumor immunity by activating TGF- β -activated kinase 1, *Oncol Immunology* 11 (1) (2022), 2016159, <https://doi.org/10.1080/2162402x.2021.2016159>.
- [42] M.A. Üstüner, L. Dogan, Relationship of preoperative serum uric acid level with survival in colorectal cancer, *Journal of the College of Physicians and Surgeons-Pakistan : JCPSP.* 30 (7) (2020) 717–721, [10.29271/jcpsp.2020.07.717](https://doi.org/10.29271/jcpsp.2020.07.717).
- [43] V.T. Lin, F.T. Lin, TRIP6: an adaptor protein that regulates cell motility, antiapoptotic signaling and transcriptional activity, *Cell. Signal.* 23 (11) (2011) 1691–1697, [10.1016/j.cellsig.2011.06.004](https://doi.org/10.1016/j.cellsig.2011.06.004).

- [44] T.G. Grunewald, S. Willier, D. Janik, et al., The Zyxin-related protein thyroid receptor interacting protein 6 (TRIP6) is overexpressed in Ewing's sarcoma and promotes migration, invasion and cell growth, *Biol. Cell.* 105 (11) (2013) 535–547, <https://doi.org/10.1111/boc.201300041>.
- [45] Y. Yang, X.M. Li, J.R. Wang, et al., TRIP6 promotes inflammatory damage via the activation of TRAF6 signaling in a murine model of DSS-induced colitis, *J. Inflamm.* 19 (1) (2022) 1, <https://doi.org/10.1186/s12950-021-00298-0>.
- [46] K.A. Katsura, J.A. Horst, D. Chandra, et al., WDR72 models of structure and function: a stage-specific regulator of enamel mineralization, *Matrix Biol. : journal of the International Society for Matrix Biology* 38 (2014) 48–58, [10.1016/j.matbio.2014.06.005](https://doi.org/10.1016/j.matbio.2014.06.005).
- [47] W. El-Sayed, R.C. Shore, D.A. Parry, C.F. Inglehearn, A.J. Mighell, Hypomaturation amelogenesis imperfecta due to WDR72 mutations: a novel mutation and ultrastructural analyses of deciduous teeth, *Cells Tissues Organs* 194 (1) (2011) 60–66, <https://doi.org/10.1159/000322036>.
- [48] J. Mahmoudian, R. Ghods, M. Nazari, et al., PLAC1: biology and potential application in cancer immunotherapy, *Cancer immunology, immunotherapy : CII* 68 (7) (2019) 1039–1058, <https://doi.org/10.1007/s00262-019-02350-8>.
- [49] M. Dratwa, B. Wysoczańska, P. Łacina, T. Kubik, K. Bogunia-Kubik, TERT-regulation and roles in cancer formation, *Front. Immunol.* 11 (2020), 589929, <https://doi.org/10.3389/fimmu.2020.589929>.
- [50] B. Schoeps, C. Eckfeld, O. Prokopchuk, et al., TIMP1 triggers neutrophil extracellular trap formation in pancreatic cancer, *Cancer Res.* 81 (13) (2021) 3568–3579, <https://doi.org/10.1158/0008-5472.Can-20-4125>.
- [51] G. Song, S. Xu, H. Zhang, et al., TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway, *Journal of experimental & clinical cancer research : CR* 35 (1) (2016) 148, <https://doi.org/10.1186/s13046-016-0427-7>.
- [52] B. Farhood, M. Najafi, K. Mortezaee, CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: a review, *J. Cell. Physiol.* 234 (6) (2019) 8509–8521, [10.1002/jcp.27782](https://doi.org/10.1002/jcp.27782).
- [53] T.F. Gajewski, H. Schreiber, Y.X. Fu, Innate and adaptive immune cells in the tumor microenvironment, *Nat. Immunol.* 14 (10) (2013) 1014–1022, <https://doi.org/10.1038/ni.2703>.
- [54] T. Yamamoto, K. Kawada, K. Obama, Inflammation-related biomarkers for the prediction of prognosis in colorectal cancer patients, *Int. J. Mol. Sci.* 22 (15) (2021), <https://doi.org/10.3390/ijms22158002>.