Identification of a 5-Gene-Based Signature to Predict Prognosis and Correlate Immunomodulators for Rectal Cancer

Yi Lin
Capital Medical University
Qiang Ji
Capital Medical University
Yichen Peng
Capital Medical University
Chunna Yu
Capital Medical University
Yi Zheng
Capital Medical University
Xun Kang
Capital Medical University
Jianwei Zheng
Capital Medical University
Rixing Bai
Capital Medical University
Wenmao Yan
Capital Medical University
Xiaomin Wang
Capital Medical University
Wenbin Li (liwenbin@ccmu.edu.cn)
Capital Medical University

Research Article

Keywords: CLIC5,ENTPD8, PACSIN3, HGD, GNG7, Rectal Cancer, Prognosis

Posted Date: March 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1422386/v1

Abstract

Background

Specific tumour markers that are similar to the 21-gene assay in breast cancer, have yet to be identified in rectal cancer. The objective of this study is to identify a novel genetic signature in rectal cancer to help predict survival and provide clues for immunotherapy.

Methods

Differentially Expressed Genes were obtained from two GEO datasets of rectal cancer. By using data from TCGA and GSE133057, two cohorts of rectal cancer were applied to establish and evaluate the prognostic signature. Then, a nomogram was constructed to estimate survival predictability in both training and validation cohorts. Subsequently, we integrated the risk-score with clinicopathological features and assessed its interplay with immune cells and molecules. Finally, our study performed functional annotations, gene-targeted miRNAs, and single-cell analysis to reveal potential functions and relevant mechanisms.

Results

A total of 468 DEGs were identified between rectal cancer and normal tissues, and a signature consisting of 5 genes (CLIC5, ENTPD8, PACSIN3, HGD, and GNG7) was selected for further analyses in a prognostic model. According to the risk-score calculated by the model, results showed that overall survival was significantly worse in the high-risk group ($P = 0.0044$). The model exhibited high performance in time-dependent ROC as well as a nomogram. Notably, as an independent prognostic factor, the risk-score was associated with vascular invasion ($P = 0.038$). Furthermore, there was a dramatic difference in nonregulatory CD4 + and CD8 + T cells between the high and low-risk groups. Strikingly, the Heat map plot showed that genes in our model were correlated with immune inhibitors. Moreover, there was also a big difference in autophagy-, immune-, cell cycle-, infection-, and apoptosis-associated terms and pathways between high and low-risk groups by GO and KEGG enrichment. Then, hsa-miR-6887 was predicted to be a common microRNA of 5 genes, which was only correlated with GNG7 in expression. Markedly, the functional states of differentiation, apoptosis, and quiescence were highly related to the 5-gene signature in single-cell analysis.

Conclusion

Our results suggest that the 5-gene-based signature could serve as a novel prognostic biomarker in rectal cancer, which could be of benefit for decision-making in rectal cancer immunotherapy.

Introduction

According to the GLOBOCAN 2018 updates, colorectal cancer (CRC) is the fourth most frequently diagnosed cancer and the second leading cause of cancer death worldwide[1, 2]. In 2018, an estimated 1.85 million new cases were diagnosed, and 880,000 people died from this disease[1]. Rectal cancer accounts for more than 30% of CRCs in China and is associated with unfavourable clinical outcomes[3, 4]. To date, the standard strategy for locally advanced rectal cancer (LARC) is neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision (TME)[4, 5]. Nevertheless, assessments of the survival rates of rectal cancer are based largely on the TNM staging system, which has undergone changes resulting in different editions, and estimation of rates based on genetic markers has been limited due to a lack of compelling data.

Several studies have aimed to evaluate potential genetic prognostic factors as predictors of treatment response and disease outcome of locally advanced rectal cancer. In 2005, Ghadimi and his colleagues reported gene expression signatures from 30 LARC patients, in which a list of 54 differentially expressed genes (DEGs) between responders and nonresponders was identified and further used to establish expression profiling. This signature was able to successfully predict tumour response in 83% of patients ($P = 0.02$)[6]. In 2020, Ja Park et al. performed a gene expression study and reported a nine-gene signature (FGFR3, GNA11, H3F3A, IL12A, IL1R1, IL2RB, NKD1, SGK2, and SPRY2) for predicting the response using biopsy samples from 156 LARC patients[7]. In 2022, Kim S et al. reported that the levels of CXCL12 were elevated in LARC cells after nCRT and that CXCL12 expression in the plasma membrane of LARC cells after nCRT was correlated with a worse prognosis of LARC[8].

By collecting the increasing data regarding disease outcome, immune factors and microRNAs (miRNAs) had been identified as potential prognostic factors for rectal cancer[9–12]. Accumulating evidence suggested that appropriate molecular predictors may be more crucial for understanding prognosis and making treatment decisions than clinicopathological features[11, 13, 14]. However, unlike in breast cancer for which a series of prospective and retrospective studies of the 21-gene assay demonstrated an association with the prognosis, a specific tumour marker has not yet been identified for rectal cancer[15]. Thus, the lack of sufficiently informative biomarkers continues to hinder the molecular diagnosis and accurate prediction of prognosis for locally advanced rectal cancer.

The objective of this study was to determine whether differentially expressed genes profiles obtained from rectal cancer and normal mucosa could offer insight into the prognosis for locally advanced rectal cancer. In this study, relevant DEGs levels were analyzed from two transcriptional datasets for locally advanced rectal cancer, which were downloaded from the Gene Expression Omnibus (GEO) database. Survival analyses, including univariable Cox regression, were carried out on the training dataset to identify potential prognostic biomarkers based on data from The Cancer Genome Atlas (TCGA). A specific prognostic DEGs panel was established by multivariable Cox regression analysis, and a 5-gene-based risk-score signature model was built using the least absolute shrinkage and selection operator (LASSO) method. GSE133057 was applied as an independent dataset to validate the prognostic

performance of the constructed 5-gene-based signature. The low- and high-risk groups were defined based on risk-score calculated by the model and compared via Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), GSEA (Gene Set Enrichment Analysis), and CancerSEA. Finally, the low and high-risk groups were assessed for tumour mutation burden (TMB), microsatellite instability (MSI), immune cell infiltration, single-sample gene-set enrichment analysis (ssGSEA), and the presence of immune checkpoint molecules.

Materials And Methods

Pre-procession of Datasets and Database

Two independent datasets of locally advanced rectal cancer were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). GSE 15781, the platform of which was GPL2986, included 20 normal rectal tissues and 21 rectal cancer tissues. GSE 20842, which was performed on GPL4133, contained 65 normal rectal mucosa and 65 rectal cancer tissues. Additionally, the FPKM value of gene expression and related clinical phenotype of rectal cancer were downloaded from the TCGA database (UCSC website: http://xena.ucsc.edu/)[16]. Moreover, GSE 133057 (N = 33) data was applied for further prognosis validation, which was from the GEO database (https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE133057).

Identification of Differentially Expressed Genes

The R package “limma” was used to calculate statistical changes of expression levels between two datasets. The DEGs between normal tissues and rectal cancer tissues were extracted as previously described, and the Enhanced Volcano Plots were performed using the R package “EnhancedVolcano”.The gene dots with standards of adjusted *P-value* < 0.05 and |logFC|≥1 were considered as DEGs. In each plot, the dots of DEGs were marked in red to show the significance of expression changes, otherwise, which were defined as stable genes according to cut-off criteria. To evaluate the transcriptomic expression levels of two datasets, Heat Maps were constructed using the R package “pheatmap” across a comparison of normal and cancer tissues. In addition to heatmaps, the Venn diagram was approached to access the overlap of DEGs using the R package “Venn diagram”.

Construction and validation of the prognostic model

Variables and features were abstracted from the TCGA data for modelling, and patients were stratified into a high-expression group and a low-expression group according to the median cut-off values of single gene expression. Then, Kaplan-Meier survival and Cox proportional hazard analyses were used to determine overall survival by the “survival” R package. The “glmnet” R package was used for least absolute LASSO analyses, and the “survival” R package was used for multivariate Cox regression to obtain a risk classification score. The risk-score was calculated by using the “survival” R package, and the mathematical model is as follows: Risk score = $h0(t) * \exp(b1X1 + b2X2 + \ldots + bnXn)$ where n is the

representative number of modelling genes; b and X are the correlation coefficient and expression level of model gene prediction, respectively; and h0(t) is derived from the “predict” function.

Analysis of Clinicopathological Features and Survival

When it comes to a potential clinical significance, the interaction between 5 gene-based risk-score and characteristics of rectal cancer was investigated with the chi-squared test. Chi-squared analysis was performed using the “chisq.test” function in the R software to calculate the association of risk subgroups to clinical data. Then, Kaplan-Meier survival and Cox regression were performed to investigate the prognostic value of integrated risk subgroups and clinicopathological features by using the “survival” R package. This analysis helped in the identification of independent prognostic factors that were significantly associated with patients’ OS of rectal cancer.

Establishment and Validation of Time-ROC Curve and Predictive Nomogram

The Time-dependent Receiver Operating Characteristic (ROC) curve is a well-established analysis method to assess biomarkers with time-to-event outcomes. Patients were assigned to the high-risk and low-risk group according to the risk-score calculated by the model, and the GSE133057 dataset was used as a validation cohort to evaluate the prognostic value of the 5-gene signature. A nomogram was constructed to evaluate a prognostic scoring system for survival prediction in rectal cancer patients. The R package“rms” was used to build the nomogram and the calibration chart. The calibration chart was used to validate the performance of the nomogram. The R package “survivalROC” was performed to draw the receiver operating characteristic (ROC) curve to evaluate the accuracy of the nomogram. Decision curve analysis (DCA) was then employed to evaluate the clinical performance of the nomogram by using the R package “ggdca”.

Analysis of Functional annotation and Enrichment

By using the R package “clusterProfiler”, the low- and high-risk groups were analyzed to estimate the discrepancy of the potential biological processes and functions. $P < 0.05$ was set as the cut-off value for both Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. R-based package “GOplot” was used for visualization of the GO and KEGG analyses. Later, to estimate the significance of differences across the contrast of the HALLMARK gene set obtained from the MSigDB database, GSEA’s analysis was carried out via “GSEA (v4.0.3)”, and *P-value* < 0.05 was considered to be enriched significantly.

Analysis of Immune features with the 5-gene signature

To address the proportions of tumour-infiltrating immune cells, ssGSEA (single sample GSEA) analysis was performed by using the “gsva” package, to further obtain the immune cell types, functions, and pathways. Scores of each rectal cancer sample were calculated based on immune-related gene sets and estimated for relative proportions of tumour-infiltrating immune cells by CIBERSORT. To better understand the immune functions of these genes, the ssGSEA was performed to evaluate the enrichment degree of

547 immune-related genes, including immune cell types, functions, or pathways. A heat map plot was generated to show the correlation between the 5-gene in our model and immune inhibitors.

Prediction of miRNAs and Analysis of Functional State

Targertscan is an online tool that allows the user to identify candidate targeted MicroRNAs (miRNAs) by gene name[17]. The Targertscan database provides output screen ranks with which several features are included, such as the probability of conserved targeting (PCT), context + scores, and so forth. Then, the results of those predicted targets were intersected by the Venn diagram, and overlapping MicroRNAs indicated the common target of the input genes. Additionally, the co-expression of the 5 genes and MicroRNAs were analyzed by using the TCGA database. Kaplan-Meier analyses of TCGA datasets were performed to compare survival results of 5 genes and MicroRNAs. In addition, the CancerSEA tool provided analysis of genes' expression by single-cell and linked to functional states as previously described[18].

Statistical analysis

The chi-squared test (Fisher's exact test) was performed to analyze the correlations between genes' expression and clinicopathological variables. Overall survival was plotted and calculated using the Kaplan–Meier method, and differences between groups were compared by the log-rank test. The Cox proportional hazard model was used to estimate the relative risks of death correlated with genes' expression. Multivariate Cox analysis was carried out on the statistically significant variables in the univariate Cox regression. *P-value* < 0.05 (two-sided) were considered statistically significant. All analyses were performed by R (version 4.0.2) and visualized by R package "ggplot2".

Results

Differentially Expressed Genes between rectal cancer and normal tissues

The results of the analyses showed that unique DEGs were significantly different when comparing rectal cancer and normal tissues with thresholds of |log2FC| >1.0 and adjusted *P-value* < 0.05. In total, there were 1474 DEGs obtained from the GSE20842 dataset and 768 DEGs identified from the GSE15781 dataset, as displayed in the enhanced volcano map (Fig. 1A-B). To further elucidate the DEGs between cancer and normal tissues, heatmaps were grouped using hierarchical clustering. As illustrated in Fig. 1C-D, heatmap analyses revealed distinguishable trends of expression differences between rectal cancer and normal tissues. The rows of the heatmap represented the z score of the log2 fold change (log2 FC) from high to low expression in the two datasets, and the columns represented the contrasts of normal and cancer tissues. In the column labels, green indicated a low level of gene expression, while red indicated a high level of expression. In the row labels, blue represented normal tissue, and red represented tumours. In an effort to gain insight into the numbers of shared genes, a Venn diagram was constructed to assess the expression levels. Summary data showed that there were 468 genes in common between the two

datasets. Blue represented the significant DEGs from the GSE20842 dataset, while red represents the significant DEGs from the GSE15781 dataset. The DEGs found to be significant in both datasets were candidates for further research regarding prognostic assessment in rectal cancer.

Construction of prognostic gene signatures for rectal cancer

To investigate potential factors predictive of outcome in rectal cancer, survival data were obtained from the publicly available TCGA dataset to construct a training set. Univariable Cox regression was performed to assess the association between 468 DEGs and survival among rectal cancer patients. Then, 14 DEGs were identified by the survival analysis using the significance criterion of *P-value* < 0.05, and Figs. 2A visualized the results. The training set was used to construct a prognostic model for the survival of rectal cancer patients (Fig. 2B). Next, the 7 selected survival-related DEGs were further investigated using least absolute shrinkage and selection operator (LASSO) logistic regression, where the hyperparameter L1 was optimized by using 5-fold cross-validation with the minimal partial likelihood deviance (Fig. 2B-2C). Then, multivariate Cox regression was used to evaluate the risk-score of the 5 genes for the overall survival of rectal cancer patients (Fig. 2D). Cox regression was implemented under the proportional hazards (PH) assumption, and the risk-score was obtained for each patient using the product of the gene expression levels and estimated coefficients. As a result, a prognostic signature for rectal cancer was established, consisting of 5 genes (CLIC5, ENTPD8, PACSIN3, HGD, and GNG7). The formula used to calculate the risk-score of the 5-gene signature was as follows: CLIC5*0.99 + ENTPD8*1.11 + PACSIN3*2.61 + HGD*1.04 + GNG7*2.86.

Validating the risk-score in the time-ROC curves and survival analysis

To verify the prognostic value of the risk-score, the rectal cancer patients were divided into low-risk and high-risk groups according to the median risk-score across all rectal cancer patients (Fig. 3A-B). Thereafter, a time-dependent receiver operating characteristic (ROC) curve was generated to evaluate whether this risk-score had high predictive accuracy for prognosis, with an AUC of 0.926 at 3 years, an AUC of 0.881 at 4 years, and an AUC of 0.948 at 5 years for TCGA (Fig. 3C-D). The AUCs to predict survival for the GSE133057 dataset, as an independent validation shown in Fig. 3D, were 0.812 (3 years), 0.835 (4 years), and 0.828 (5 years). The model had a concordance index of 0.66 (95% CI 0.57 to 0.71). Finally, Kaplan–Meier survival curve analysis was used to assess the effectiveness of the 5 genes for the prognostic stratification of rectal cancer patients. In the TCGA data analysis, the high-risk group showed significantly unfavourable OS ($P = 0.0044$, log-rank = 6.328) compared with the low-risk group. A similar trend was observed in the GSE15781 dataset, where the high-risk group showed significantly worse OS (Fig. 3E-3F).

The prognostic signature is correlated with clinicopathological features and survival

To gain insight into the association between risk-score and clinical variables, clinicopathological parameters were compared between the high- and low-risk groups. Clinical data downloaded from TCGA are summarized and analyzed in Table I. Firstly, Kaplan–Meier analysis of TCGA datasets showed that rectal cancer tumours with high expression of the 5 genes showed worse outcomes (Fig. 4A-E). Then Kaplan–Meier curve was shown in Fig. 4F according to Distant Metastasis. A comparison suggested that prognostic values for CLIC5, ENTPD8, PACSIN3 HGD, and GNG7 were superior to Distant Metastasis. Secondly, the risk-score was associated with clinicopathological features, including vascular invasion (P = 0.038). Associations between the risk-score and other features, such as stage and lymphatic invasion, failed to reach statistical significance in the TCGA dataset (Table I). Thirdly, the results of Cox regression showed that after multivariable adjustments for clinicopathological factors, the risk-score remained significantly associated with patient OS (Fig. 4G). Our results also confirmed that the risk-score was an independent prognostic predictor of longer OS for rectal cancer patients (HR = 2.84; 95% CI, 1.4 to 5.77; P = 0.001, Fig. 4H). Stage, histologic grade, vascular/lymphatic invasion, and metastasis all had independent prognostic value in the multivariate analysis. Other clinicopathological parameters had no prognostic value in multivariate analyses.

Establishment of a predictive nomogram

To evaluate whether the 5-gene model was useful for survival prediction, a nomogram was developed to identify low- and high-risk patients based on the 3-year, 4-year, and 5-year OS rates. The nomogram demonstrated good discrimination of 5-year OS among patients with different clinical and pathologic parameters (Fig. 5A). Moreover, decision curve analysis (DCA) revealed that the model relative to the nomogram was associated with benefit gains (Fig. 5B). The calibration plot showed that the predicted power of both the training and validation set was near the ideal curve (Fig. 5C-D). Nonetheless, calculating the scores can be cumbersome, and the survival rate of patients at any time cannot be calculated conveniently. For this reason, the online predicted tool was made to provide access to the dynamic estimate of OS, and the final web-based nomogram is available online at https://xqccc.shinyapps.io/DynNoma/.

Functional enrichment analysis between the high and low-risk groups

To investigate the molecular function and signalling pathways between the high- and low-risk groups, further enrichment analyses were performed using GO terms and KEGG pathway annotations. According to GO analysis, the top 30 terms were visualized in Fig. 6A, and the most significant enrichments were autophagy (GO:0006914), vacuolar membrane (GO:0005774), and ubiquitin-like protein transferase activity (GO:0019787), which contained 227, 187, and 176 genes, respectively(Supplemental Table I). The

top 30 enriched KEGG pathways are visualized in Fig. 6B, while the results of KEGG analysis indicated that the main enrichments were cell cycle (hsa04110), measles (hsa05162), apoptosis (hsa04210), p53 signalling pathway (hsa04115), and nucleotide excision repair (hsa03420). These results also indicated that infection (by human cytomegalovirus, Escherichia coli, or Epstein–Barr virus) likely played an important role in rectal cancer, which contained 95, 90, and 89 genes, respectively(Supplemental Table II). As shown in Fig. 6C-D, the Ridge plot and Enrichment plot showed the top 15 enrichments of GSEA, and these results confirmed that genes are predominantly involved in the cell cycle, autophagy, apoptosis and immunity pathways between the high- and low-risk groups in rectal cancer.

Efficacy of the model with signature immunotherapeutic relevant genes

To assess the appropriateness of the 5 genes as clinically accepted biomarkers for immunotherapy, correlation analyses were carried out for microsatellite instability (MSI), tumour mutation burden (TMB), and the tumour microenvironment (TME) between the high- and low-risk groups. As shown, no significant differences were found in terms of microsatellite instability between the high- and low-risk groups (Fig. 7A). Nevertheless, the low-risk group had a higher proportion of MSI scores. In terms of the TMB, there were no significant differences between the two groups (Fig. 7B). Correlation analysis was evaluated between the 5-gene signature and immune cell infiltration. The tumour immune microenvironment consists of massive immune cell subsets surrounding cancer cells, including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells. Notably, the high- and low-risk groups showed significant differences in nonregulatory $CD4^+$ T cells and $CD8^+$ T cells (Fig. 7C-D). Myeloid dendritic cells were found to be almost significantly different between the two groups (Fig. 7C-D). More importantly, the Heat map plot showed that genes in our model were correlated with immune inhibitors, especially CLIC5 and GNG7 (Fig. 7E).

The prognostic signature correlated with differentiation, apoptosis, and quiescence

To understand how the potential molecular function of the 5-gene signature would impact the survival of rectal cancer patients, TargetScan was used to predict the targets of the 5 genes. As expected, it provided 5 sets of miRNAs targeted to 5 genes, and these were further used to evaluate the consensus prediction. As shown in the Venn diagram, the overlap of the miRNAs revealed one common microRNA (Fig. 8A). The TCGA rectal cancer dataset was stratified into high versus low hsa-miR-6887-expressing groups, and then, upon Kaplan–Meier survival analysis, high hsa-miR-6887-expressing tumours showed a tendency toward an improved outcome for patients, but this did not reach statistical significance (Fig. 8B). Thus, Analyses of TCGA cohorts showed that the expression of GNG7 was correlated with hsa-miR-6887 (Fig. 8C). The CancerSEA tool was used to identify genes correlated with functional state, in addition, differentiation, apoptosis, and quiescence were significantly related to the 5-gene signature in the single-cell dataset GSE81861 (Fig. 8D-F).

Discussion

In this study, we identified a 5-gene-based prognostic model by comparing normal tissues with rectal cancer and further validated the predictive power of the model in two independent rectal cancer datasets. Our findings indicated that the OS rates increased with the risk-score in the TCGA and GSE133057 rectal cancer datasets. Strikingly, both the high- and low-risk groups showed a significant association with the TME, especially non-regulatory $CD4^+$ T cells and $CD8^+$T cells. Moreover, the data showed that the risk-score was significantly associated with vascular invasion (Table I). Based on these results, we propose the use of the 5-gene-based classification model as a novel molecular-based prognostic tool to evaluate the survival of rectal cancer patients. The model provides a starting place for further research regarding the role of genes in the development of rectal cancer.

Rectal cancer is a tumour with a relatively high prevalence but without convincing prognostic and predictive molecular markers. It has been established that biomarkers can classify rectal cancer patients with certain subtypes, some of which may benefit from tailored therapy. Several important factors have been linked to rectal cancer survival, and these molecules participate in key processes involved in cancer development, including cell migration and invasion[19–23]. Although these factors were considered to be significant in colorectal cancer, the findings were not verified in rectal cancer via validation in an independent dataset. Our results were validated in a separate cohort with rectal cancer patients. These data supported the findings that the 5-gene-based signature well served as a predictor for survival in rectal cancer patients.

Our study showed a strong correlation between the 5-gene-based tumour markers and clinical outcomes of rectal cancer. The molecular function of CLIC5, ENTPD8, PACSIN3, HGD, and GNG7 remains unclear, partly due to the lack of original research in rectal cancer. Chloride intracellular channel 5 (CLIC5) belongs to the family of chloride (Cl^-) channels which are responsible for encoding chloride intracellular channel (CLIC) proteins. To date, there are six known members in the CLIC family (CLIC1-6), and accumulating evidence supports their role in tumour biology, especially gastrointestinal cancer[24]. A previous study identified upregulated changes in CLIC1 in colorectal cancer (CRC), which was shown to be associated with poor prognosis in CRC patients[25]. CLIC4 was also found to be overexpressed in CRC, and its upregulation was correlated with unfavourable 5-year clinical outcomes[26]. To date, there have been no reports of a relationship between CLIC5 and rectal cancer. Ectonucleoside triphosphate diphosphohydrolase 8 (ENTPD8) is a member of the ectonucleoside triphosphate diphosphohydrolases (E-NTPDase) family, plays an essential role in ATP metabolism and is mainly expressed in the intestine[27]. Although ENTPD8 is still poorly understood, there is evidence that it plays a crucial role in pancreatic cancer and exhibits metabolic activity toward gene-metabolite networks[28]. PACSIN3 was identified as an intracellular adapter protein that regulates endocytosis, vesicle transport, membrane internalization, and actin reorganization. As reported, PACSIN3 is one of the mobility-related genes that is downregulated in ING5-overexpressing SGC-7901 gastric cancer cells. Another study also found that PACSIN3 was decreased in prostate cancer. To our knowledge, there is no report of a study about PACSIN3 in rectal cancer. The HGD gene encodes one of the enzymes called homogentisate 1,2

dioxygenase, which is required for the catabolism of the amino acids tyrosine and phenylalanine and is generally active in the kidneys and liver to catalyze oxidation-reduction reactions. The current study showed that high HGD mRNA expression (≥ 3-fold) was associated with poorer survival of, histological grade, advanced stage, and metastasis of cholangiocarcinoma patients[29]. Previous results have shown that HGD is a potential key factor in the regulatory mechanism of BRAFV600E-mediated PTC. However, there was no significant discrepancy in overall survival[30]. G protein γ subunit 7 (GNG7), which is a component of the large G γ family, was first identified to be a downregulated differentially expressed gene in pancreatic cancer[31] and was then found in gastrointestinal tract cancer (including oesophagal, gastric, and colorectal cancer)[32]. In a previous study, it was shown that GNG7 acts as a potential tumour suppressor both in vitro and in vivo[33]. A similar study also demonstrated that it was a tumour suppressor gene in clear cell renal cell carcinoma and lung adenocarcinoma[34]. Taken together, our research is the first work to evaluate the survival value of a 5-gene-based tumour marker in rectal cancer and might aid in the improvement in molecular prognosis[11].

Collectively, several studies have shown the clinical importance of immune infiltrates in colorectal cancer. Galon and his colleagues examined tumour-infiltrating lymphocytes (TILs) in approximately 400 colorectal cancer samples and found that $CD8^+$ and $CD45RO^+$ T cells in the tumour were superior predictors to the histopathological staging methods[12]. Previously, studies have also shown that $CD4^+$ and $CD8^+$ T cells were promised survival predictors in colorectal cancer patients. A significant correlation was observed between the density of $CD8^+$ T cells in the peritumoral region and a longer disease-free interval ($P = 0.009$), and Kaplan–Meier analysis later suggested that the percentage of $CD8^+$ T cells might have clinical application in the stratification of the relapse risk of patients ($P = 0.006$)[35]. Yasuda K et al. reported that tumour-infiltrating lymphocytes (TILs), especially density $CD4^+$ T cells and $CD8^+$ TILs, were strongly associated with the tumour treatment response of rectal cancer after neoadjuvant chemoradiotherapy (nCRT)[12]. As mentioned earlier, our study found that the 5-gene signature was strongly correlated with tumour-specific $CD4^+$ and $CD8^+$ type T cells. Obviously, these data were consistent with studies of TILs in colorectal cancer.

Another interesting finding in our study was that the 5-gene signature was predicted to be intersected in hsa-miR-6887-3p, while only GNG7 correlated with hsa-miR-6887-3p. MicroRNAs (miRNAs) are critical tumorigenesis mediators in many human cancers. The role of miRNAs as clinical biomarkers in colorectal cancer research is promising [10]. Many studies have documented aberrant miRNA levels as biomarkers for colorectal cancer and have reported the evaluation of their potential roles as diagnostic and prognostic indicators. Li H demonstrated that hsa-miR-6887-3p inhibited the tumorigenesis of colorectal cancer by downregulating Mex3a expression and functioned as an important regulator in the hsa-miR-6887-3p/Mex3a/RAP1GAP signalling axis[36]. However, the clinical application of miRNAs as predictive biomarkers in rectal cancer remains to be seen[9]. Our data also suggested the need to further studies to investigate the biological role of hsa-miR-6887-3p and GNG7 in rectal cancer.

It is noteworthy that the 5-gene signature was correlated with the vascular invasion ($P = 0.038$), but not the lymphatic invasion. The vascular invasion has been associated with an increased risk of regional and distant metastasis in colorectal cancer patients[37]. Some authors have reported that vascular invasion was a strong prognosticator for rectal cancer that influenced disease progression and survival[37–39]. Thus, a few analyses have failed to indicate the prognostic value of vascular invasion in colorectal, colon, and rectal cancer survival[40, 41]. It is therefore important to identify vascular invasion-related biomarkers to better diagnose and classify rectal cancer patients. However, until now, no validated tumour markers have been identified[41–43]. Our findings suggested that the vascular invasion is critical for the progression of colorectal cancer which warrants further investigation. The lymphatic invasion has also been reported to be linked with an adverse prognosis of colorectal cancer[38]. However, our data did not show any statistically significant association between lymphatic invasion and prognosis. Based on our results, further research into the function of the 5-gene signature in vascular invasion is warranted.

One potential limitation of our research was that it was predominantly based on bioinformatics results. Our study was designed to decrease bias and increase the repeatability of the analytic results based on the use of two independent large-sample datasets. Nevertheless, our findings needed to be validated in a more prospective study, and the correlation between gene markers and other factors requires further investigation. Another key question is whether the 5 genes indeed affect the progression of rectal cancer. Consequently, future research will be essential to understand the biological association between the expression of these 5 genes and rectal cancer.

To summarize, as highlighted in this paper, the 5-gene-based signature was a robust prognosticator for rectal cancer patients. Moreover, the risk-score generated by the signature was shown to have the ability to further classify the patients into low- and high-risk groups. In addition, the prognosis of the two groups was significantly different, and the high-risk group had a more unfavourable OS. Our findings were confirmed in an independent validation set and subsequently evaluated by time-ROC curves and nomograms. Further analyses demonstrated that the two groups of patients differed significantly with respect to nonregulatory $CD4^+$ T cells and $CD8^+$ T cells. Our results also suggested that there were clear differences in vascular invasion ($P = 0.038$) between the high- and low-risk groups, and the high-risk group exhibiting a higher risk of vascular invasion.

More importantly, this was an independent prognostic factor based on the Cox regression model. Our research aimed to uncover the complex interaction between 5 genes and clinical outcomes in rectal cancer through the use of microRNA and single-cell analysis tools. In summary, our results indicated that the 5-gene-based signature could contribute to the prognostic evaluation of rectal cancer and might pave the way for new therapeutic strategies in the foreseeable future.

Abbreviations

DEGs: Differentially expressed genes; GEO: Gene Expression Omnibus; OS:overall survival; TCGA:The Cancer Genome Atlas; LASSO:least absolute shrinkage and selection operator; nCRT: neoadjuvant

chemoradiotherapy; TME: total mesorectal excision; LARC: locally advanced rectal cancer; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene Set Enrichment Analysis; TMB: tumour mutation burden; MSI: microsatellite instability; ssGSEA: single-sample gene-set enrichment analysis; ROC: Receiver Operating Characteristic; DCA: Decision curve analysis.

Declarations

Acknowledgements

The authors appreciate countless individuals who have contributed to The Cancer Genome Atlas (TCGA) Program.

Funding

Our study was supported by funds from the National Natural Science Foundation of China (Li Wenbin, Grant No. 81972338); the National Natural Science Foundation of China (Wang Xiaomin, Grant No. 82070169); and the Natural Foundation of Capital Medical University (Lin Yi, Grant No. PYZ2017160).

Availability of data and materials

The dataset including transcriptome profiling oflocally advanced rectal cancerwasretrieved from the Gene Expression Omnibus (GEO) repository, access numberGSE 15781(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15781).Transcriptome profiling of 65 normal rectal mucosaand 65 rectal cancer tissuesis availablein GEO (accession number GSE 20842, HTTP:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20842).

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Chen W, Zheng R, Zuo T, Zeng H, Zhang S, He J: **National cancer incidence and mortality in China, 2012**. *Chin J Cancer Res* 2016, **28**(1):1-11.
3. Cao W, Chen HD, Yu YW, Li N, Chen WQ: **Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020**. *Chin Med J (Engl)* 2021, **134**(7):783-791.

4. Wu AW, Cai Y, Li YH, Wang L, Li ZW, Sun YS, Ji JF: **Pattern and Management of Recurrence of Mid-Low Rectal Cancer After Neoadjuvant Intensity-Modulated Radiotherapy: Single-Center Results of 687 Cases**. *Clin Colorectal Cancer* 2018, **17**(2):e307-e313.
5. Dong C, Ding Y, Weng S, Li G, Huang Y, Hu H, Zhang Z, Zhang S, Yuan Y: **Update in version 2021 of CSCO guidelines for colorectal cancer from version 2020**. *Chin J Cancer Res* 2021, **33**(3):302-307.
6. Ghadimi BM, Grade M, Difilippantonio MJ, Varma S, Simon R, Montagna C, Fuzesi L, Langer C, Becker H, Liersch T *et al*: **Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy**. *J Clin Oncol* 2005, **23**(9):1826-1838.
7. Park IJ, Yu YS, Mustafa B, Park JY, Seo YB, Kim GD, Kim J, Kim CM, Noh HD, Hong SM *et al*: **A Nine-Gene Signature for Predicting the Response to Preoperative Chemoradiotherapy in Patients with Locally Advanced Rectal Cancer**. *Cancers (Basel)* 2020, **12**(4).
8. Kim S, Yeo MK, Kim JS, Kim JY, Kim KH: **Elevated CXCL12 in the plasma membrane of locally advanced rectal cancer after neoadjuvant chemoradiotherapy: a potential prognostic marker**. *J Cancer* 2022, **13**(1):162-173.
9. Waldron RM, Moloney BM, Gilligan K, Lowery AJ, Joyce MR, Holian E, Kerin MJ, Miller N: **MicroRNAs as biomarkers of multimodal treatment for rectal cancer**. *Br J Surg* 2021, **108**(8):e260-e261.
10. Ghafouri-Fard S, Hussen BM, Badrlou E, Abak A, Taheri M: **MicroRNAs as important contributors in the pathogenesis of colorectal cancer**. *Biomed Pharmacother* 2021, **140**:111759.
11. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P *et al*: **Type, density, and location of immune cells within human colorectal tumors predict clinical outcome**. *Science* 2006, **313**(5795):1960-1964.
12. Yasuda K, Nirei T, Sunami E, Nagawa H, Kitayama J: **Density of CD4(+) and CD8(+) T lymphocytes in biopsy samples can be a predictor of pathological response to chemoradiotherapy (CRT) for rectal cancer**. *Radiat Oncol* 2011, **6**:49.
13. Hamid HKS, Davis GN, Trejo-Avila M, Igwe PO, Garcia-Marin A: **Prognostic and predictive value of neutrophil-to-lymphocyte ratio after curative rectal cancer resection: A systematic review and meta-analysis**. *Surg Oncol* 2021, **37**:101556.
14. Zhao K, Wang M, Kang H, Wu A: **A prognostic five long-noncoding RNA signature for patients with rectal cancer**. *J Cell Biochem* 2019.
15. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein LJ, Chia SKL *et al*: **21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer**. *N Engl J Med* 2021, **385**(25):2336-2347.
16. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN *et al*: **Visualizing and interpreting cancer genomics data via the Xena platform**. *Nat Biotechnol* 2020, **38**(6):675-678.
17. Agarwal V, Bell GW, Nam JW, Bartel DP: **Predicting effective microRNA target sites in mammalian mRNAs**. *Elife* 2015, **4**.

18. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z *et al*: **CancerSEA: a cancer single-cell state atlas**. *Nucleic Acids Res* 2019, **47**(D1):D900-D908.
19. Pages F, Berger A, Camus M, Sanchez-Cabo F, Costes A, Molidor R, Mlecnik B, Kirilovsky A, Nilsson M, Damotte D *et al*: **Effector memory T cells, early metastasis, and survival in colorectal cancer**. *N Engl J Med* 2005, **353**(25):2654-2666.
20. Jeong D, Park S, Kim H, Kim CJ, Ahn TS, Bae SB, Kim HJ, Kim TH, Im J, Lee MS *et al*: **RhoA is associated with invasion and poor prognosis in colorectal cancer**. *Int J Oncol* 2016, **48**(2):714-722.
21. Aykut B, Ochs M, Radhakrishnan P, Brill A, Hocker H, Schwarz S, Weissinger D, Kehm R, Kulu Y, Ulrich A *et al*: **EMX2 gene expression predicts liver metastasis and survival in colorectal cancer**. *BMC Cancer* 2017, **17**(1):555.
22. Paauwe M, Schoonderwoerd MJA, Helderman R, Harryvan TJ, Groenewoud A, van Pelt GW, Bor R, Hemmer DM, Versteeg HH, Snaar-Jagalska BE *et al*: **Endoglin Expression on Cancer-Associated Fibroblasts Regulates Invasion and Stimulates Colorectal Cancer Metastasis**. *Clin Cancer Res* 2018, **24**(24):6331-6344.
23. Tian Y, Sun F, Zhong Y, Huang W, Wang G, Liu C, Xiao Y, Wu J, Mu L: **Expression and Clinical Significance of POLR1D in Colorectal Cancer**. *Oncology* 2020, **98**(3):138-145.
24. Anderson KJ, Cormier RT, Scott PM: **Role of ion channels in gastrointestinal cancer**. *World J Gastroenterol* 2019, **25**(38):5732-5772.
25. Petrova DT, Asif AR, Armstrong VW, Dimova I, Toshev S, Yaramov N, Oellerich M, Toncheva D: **Expression of chloride intracellular channel protein 1 (CLIC1) and tumor protein D52 (TPD52) as potential biomarkers for colorectal cancer**. *Clin Biochem* 2008, **41**(14-15):1224-1236.
26. Deng YJ, Tang N, Liu C, Zhang JY, An SL, Peng YL, Ma LL, Li GQ, Jiang Q, Hu CT *et al*: **CLIC4, ERp29, and Smac/DIABLO derived from metastatic cancer stem-like cells stratify prognostic risks of colorectal cancer**. *Clin Cancer Res* 2014, **20**(14):3809-3817.
27. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A *et al*: **Proteomics. Tissue-based map of the human proteome**. *Science* 2015, **347**(6220):1260419.
28. An Y, Cai H, Yang Y, Zhang Y, Liu S, Wu X, Duan Y, Sun D, Chen X: **Identification of ENTPD8 and cytidine in pancreatic cancer by metabolomic and transcriptomic conjoint analysis**. *Cancer Sci* 2018, **109**(9):2811-2821.
29. Aukkanimart R, Boonmars T, Juasook A, Sriraj P, Boonjaraspinyo S, Wu Z, Laummuanwai P, Pairojkul C, Khuntikeo N, Rattanasuwan P: **Altered Expression of Oxidative Metabolism Related Genes in Cholangiocarcinomas**. *Asian Pac J Cancer Prev* 2015, **16**(14):5875-5881.
30. Yu X, Zhong P, Han Y, Huang Q, Wang J, Jia C, Lv Z: **Key candidate genes associated with BRAF(V600E) in papillary thyroid carcinoma on microarray analysis**. *J Cell Physiol* 2019, **234**(12):23369-23378.
31. Shibata K, Mori M, Tanaka S, Kitano S, Akiyoshi T: **Identification and cloning of human G-protein gamma 7, down-regulated in pancreatic cancer**. *Biochem Biophys Res Commun* 1998, **246**(1):205-

209.

32. Shibata K, Tanaka S, Shiraishi T, Kitano S, Mori M: **G-protein gamma 7 is down-regulated in cancers and associated with p 27kip1-induced growth arrest**. *Cancer Res* 1999, **59**(5):1096-1101.
33. Ohta M, Mimori K, Fukuyoshi Y, Kita Y, Motoyama K, Yamashita K, Ishii H, Inoue H, Mori M: **Clinical significance of the reduced expression of G protein gamma 7 (GNG7) in oesophageal cancer**. *Br J Cancer* 2008, **98**(2):410-417.
34. Zheng H, Tian H, Yu X, Ren P, Yang Q: **G protein gamma 7 suppresses progression of lung adenocarcinoma by inhibiting E2F transcription factor 1**. *Int J Biol Macromol* 2021, **182**:858-865.
35. Makkai-Popa ST, Lunca S, Dimofte G, Vranceanu A, Franciug D, Ivanov I, Zugun F, Tarcoveanu E, Carasevici E: **Corelation of lymphocytic infiltrates with the prognosis of recurrent colo-rectal cancer**. *Chirurgia (Bucur)* 2013, **108**(6):859-865.
36. Li H, Liang J, Wang J, Han J, Li S, Huang K, Liu C: **Mex3a promotes oncogenesis through the RAP1/MAPK signaling pathway in colorectal cancer and is inhibited by hsa-miR-6887-3p**. *Cancer Commun (Lond)* 2021, **41**(6):472-491.
37. Krasna MJ, Flancbaum L, Cody RP, Shneibaum S, Ben Ari G: **Vascular and neural invasion in colorectal carcinoma. Incidence and prognostic significance**. *Cancer* 1988, **61**(5):1018-1023.
38. de Ridder JA, Knijn N, Wiering B, de Wilt JH, Nagtegaal ID: **Lymphatic Invasion is an Independent Adverse Prognostic Factor in Patients with Colorectal Liver Metastasis**. *Ann Surg Oncol* 2015, **22 Suppl 3**:S638-645.
39. Sternberg A, Amar M, Alfici R, Groisman G: **Conclusions from a study of venous invasion in stage IV colorectal adenocarcinoma**. *J Clin Pathol* 2002, **55**(1):17-21.
40. Bianchi G, Annicchiarico A, Morini A, Pagliai L, Crafa P, Leonardi F, Dell'Abate P, Costi R: **Three distinct outcomes in patients with colorectal adenocarcinoma and lymphovascular invasion: the good, the bad, and the ugly**. *Int J Colorectal Dis* 2021.
41. Campanati RG, Sancio JB, Sucena LMA, Sanches MD, Resende V: **Primary Tumor Lymphovascular Invasion Negatively Affects Survival after Colorectal Liver Metastasis Resection?***Arq Bras Cir Dig* 2021, **34**(1):e1578.
42. Tao H, Li J, Liu J, Yuan T, Zhang E, Liang H, Huang Z: **Construction of a ceRNA Network and a Prognostic lncRNA Signature associated with Vascular Invasion in Hepatocellular Carcinoma based on Weighted Gene Co-Expression Network Analysis**. *J Cancer* 2021, **12**(13):3754-3768.
43. Guner OS, Tumay LV: **Persistent extramural vascular invasion positivity on magnetic resonance imaging after neoadjuvant chemoradiotherapy predicts poor outcome in rectal cancer**. *Asian J Surg* 2021, **44**(6):841-847.

Tables

Table I Association between the patient's clinicopathological features and risk- score in TCGA rectal cancer patients

Clinicopathological Features		Risk-score			
	No.	High (%)	Low (%)	χ^2	*P*-value
Gender				0.10717	0.743
Male	45	24 (61.5)	21 (55.3)		
Female	32	15 (38.5)	17 (44.7)		
Age				0.32765	0.567
≤60	23	10 (25.6)	13 (34.2)		
>60	54	29 (74.4)	25 (65.8)		
History of Polyps				0.010969	0.917
Absent	34	18 (54.5)	16 (59.3)		
Present	26	15 (45.5)	11 (40.7)		
No record*	17				
Tumour Depth				1.199	0.2735†
T1+T2	13	10	3		
T3+T4	61	29	33		
No record*	3				
Lymph-Node Metastasis				0	1
Absent	42	21 (53.8)	21 (55.3)		
Present	35	18 (46.2)	17 (44.7)		
Distant Metastasis				2.536	0.324†
M0	58	27 (69.2)	31 (81.6)		
M1	11	8 (20.5)	3 (7.9)		
Mx	8	4 (10.3)	4 (10.5)		
Stage				2.4232	0.489†
I	13	8 (21.6)	5 (13.9)		
II	28	11 (29.7)	16 (44.4)		
III	22	11 (29.7)	11 (30.6)		
IV	11	7 (18.9)	4 (11.1)		
No record*	3				

Lymphatic Invasion				2.4368	0.119
Absent	38	15 (44.1)	23 (65.7)		
Present	31	19 (55.9)	12 (34.3)		
No record*	8				
Vascular Invasion				4.2871	0.038
Absent	50	20 (60.6)	30 (85.7)		
Present	18	13 (39.4)	5 (14.3)		
No record*	9				
Neoadjuvant Treatment				0.40822	0.483
Responders	57	26 (74.3)	31 (83.8)		
Nonresponders	15	9 (25.7)	6 (16.2)		
No record*	5				
Microsatellite Test				0.88399	0.347
Microsatellite stable	63	34 (87.2)	29 (76.3)		
Microsatellite instability	13	5 (12.8)	9 (23.7)		
No record*	1				

Abbreviations: *Data incomplete; †, Fisher's exact test,

Figures

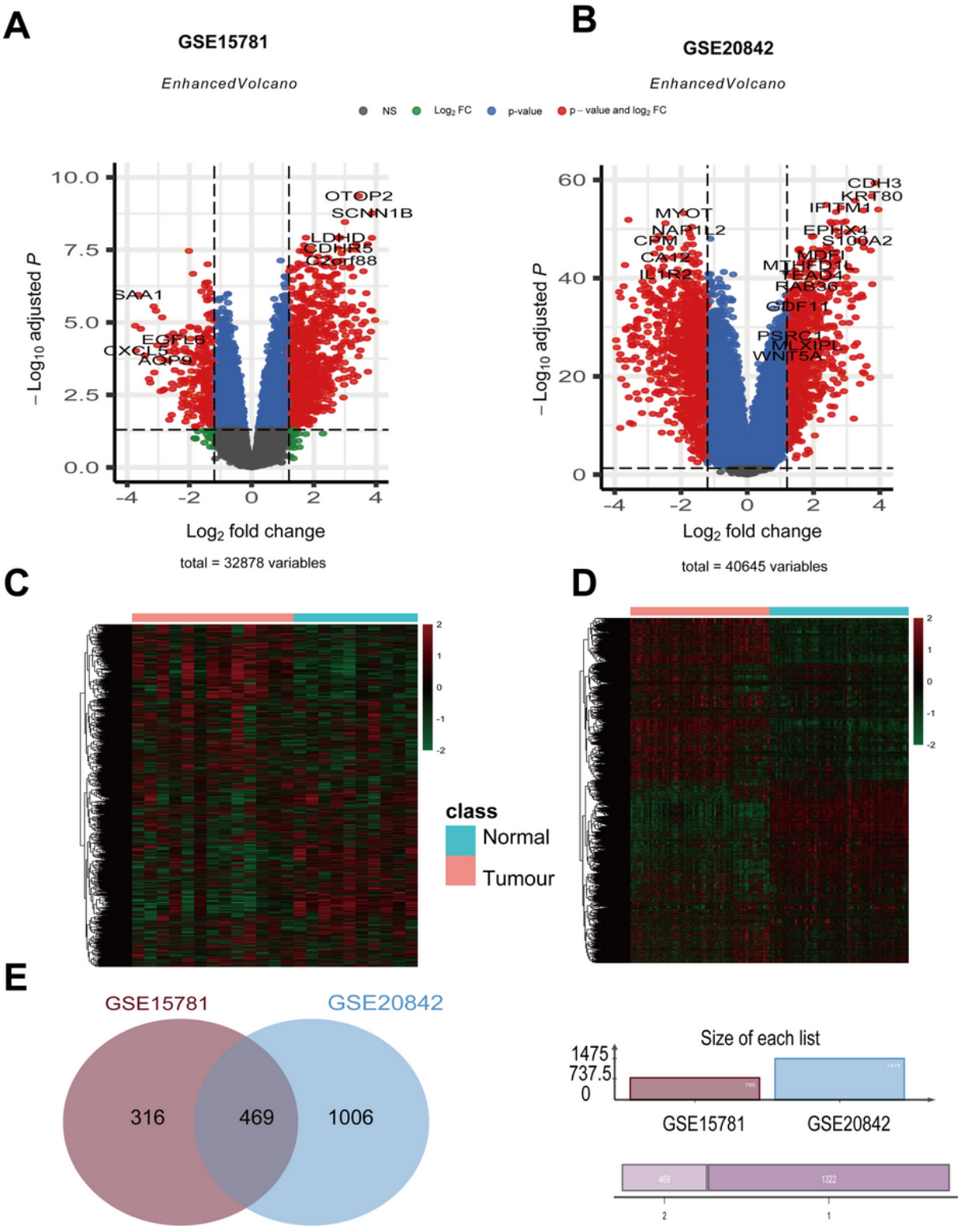

Figure 1

Identification of differential expression genes (DEGs) between normal tissue and cancer tissue in two rectal cancer datasets.

(A) Enhanced Volcano plot of DEGs for GSE15781 dataset. (B) Enhanced Volcano plot of DEGsfor the GSE20842 dataset. (C) The heatmap plot of DEGs forthe GSE15781 dataset. (D) The heatmap plot of

DEGs in the GSE20842 dataset. (E)The Venn diagram showed thenumber of DEGs and common DEGs identified bytwo profiling datasets.(DEGs according to the value of *P*< 0.05 and |logFC| < 1; Red, high expression; green, low expression.)

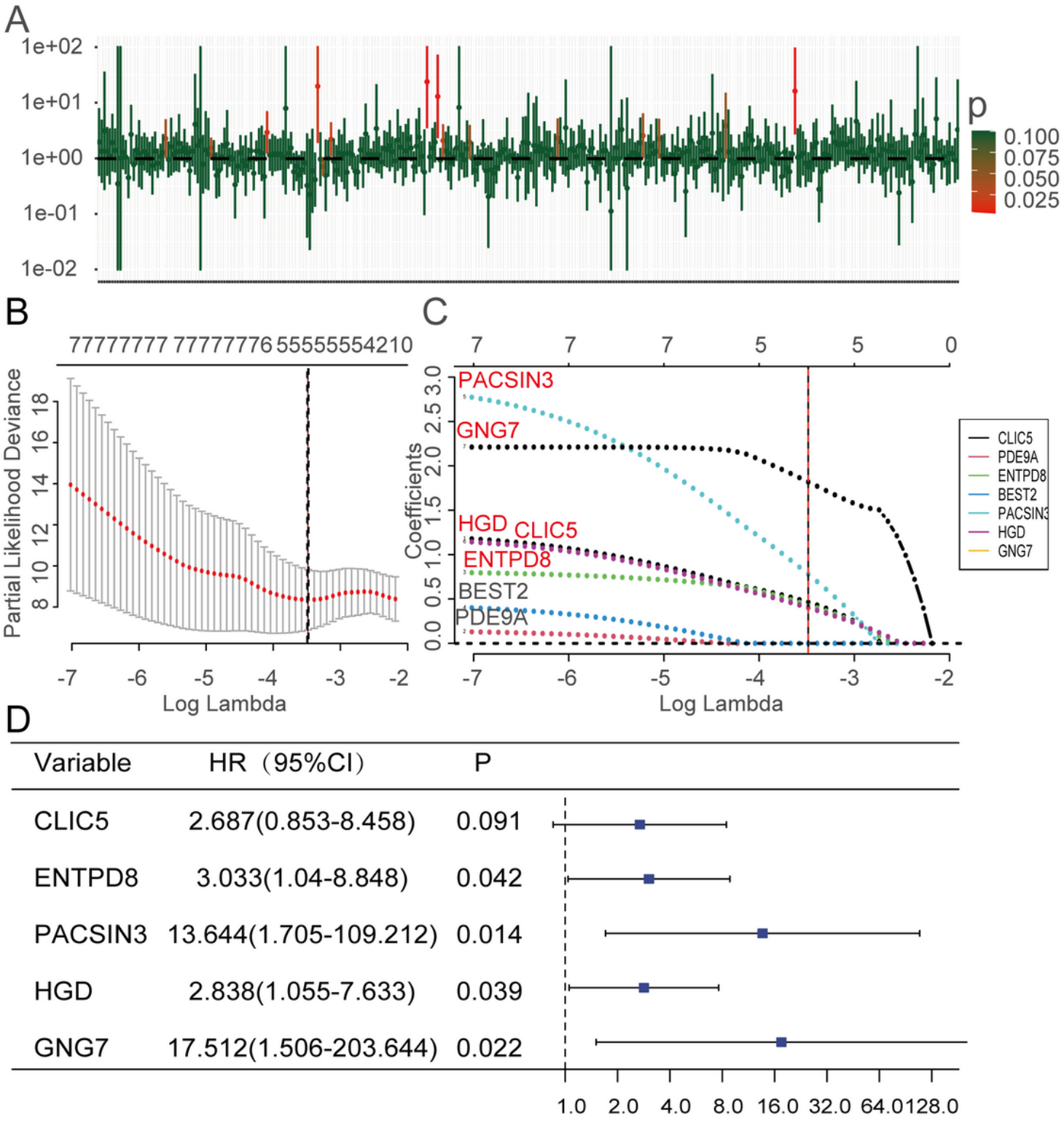

Figure 2

Construction of prognostic genes signaturefor rectal cancer from DEGs.

(A)Univariable Cox regressionto evaluate theassociation between 468DEGs and survival of rectal cancer patients. The x-axis indicates the value of *P* while the y-axis indicates the 95% Confidence Interval (CI). (B)Least absolute shrinkage and selection operator (LASSO) coefficient profiles (y-axis) of the genes. (C)The optimal penalization coefficient (λ) via 5-fold cross-validation is based on partial likelihood deviance. (D) Multivariate Cox regressionto establish the best prognostic genes signature for rectal cancer patients.(Green,value of $P \geq 0.05$; Red, value of $P < 0.05$.)

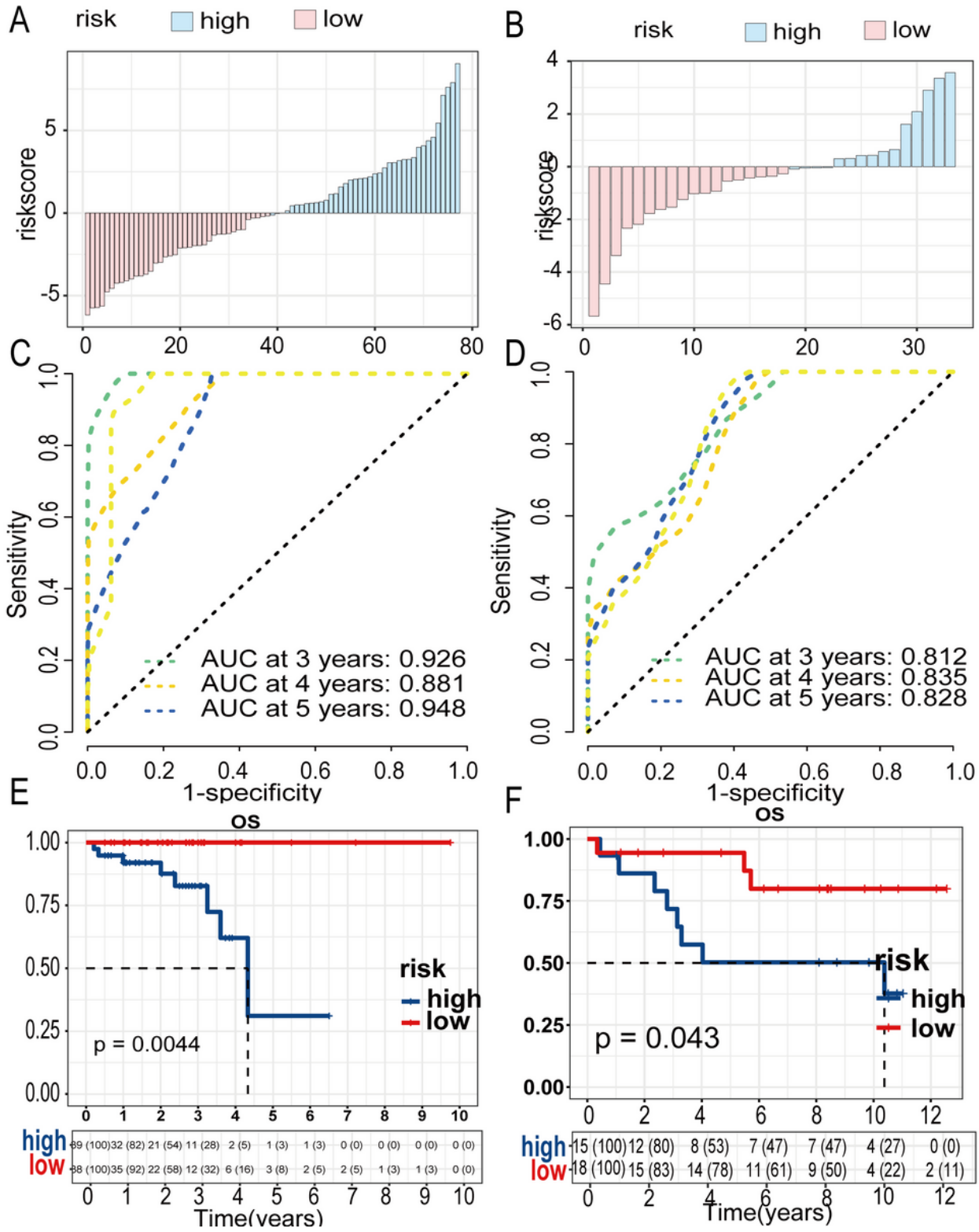

Figure 3

Prognosticperformance of the5-gene signaturemodel in TCGA and GSE133057 cohorts

(A-B)The distribution and median value of the risk-score in the TCGA and GSE133057 cohorts. (C-D)The ROC curves derived from models indicate the prognostic accuracy of the risk-score for the overall survival in TCGA and GSE133057 rectal cancer cohorts. (E-F) TheKaplan-Meier curves showed the overallsurvival in patients with rectal cancer according to the low and high-risk groups in TCGAand GSE133057 cohorts.

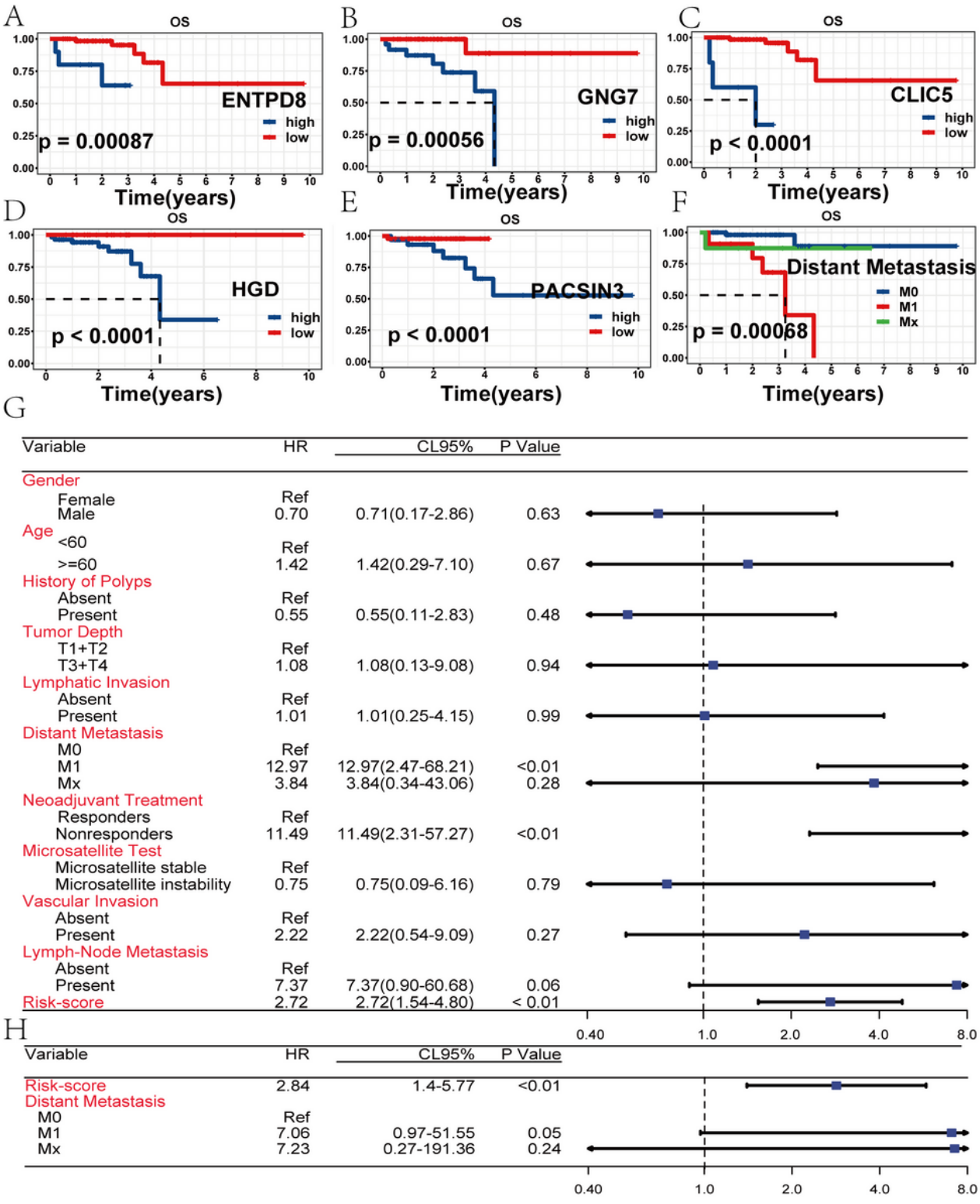

Figure 4

Theprognostic estimations of clinicopathological features and the risk-score among TCGA rectal cancer cohort

(A-E)The Kaplan–Meier analyses of single gene expression from the 5-Gene-Based Signature in TCGA rectal cancer cohort. (F)The prognostic value ofa representative clinical variable (Distant Metastasis) in

TCGA rectal cancer cohort. (B) The multivariate Cox analyses according to overall survival in the TCGA rectal cancer cohort. (H)The independent prognostic predictor of overall survival for TCGA rectal cancer patients.

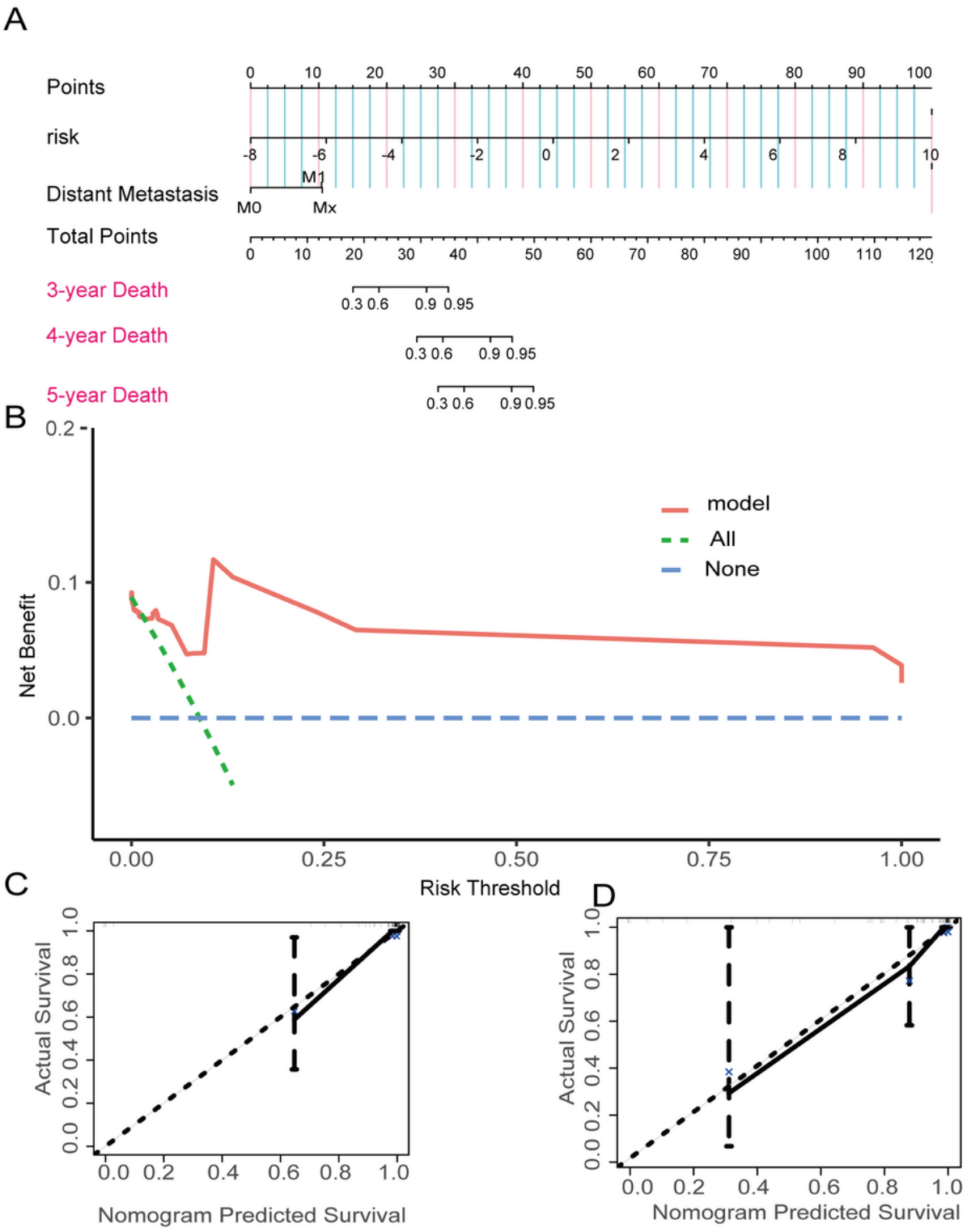

Figure 5

Nomogram, calibration plots, and decision curves for the prediction of survival for patients with rectal cancer.

(A) Nomogram for the prediction of survival possibility at 3, 4 and 5 years.(b) DCA for assessing the clinical utilityof the nomogram. The x-axisindicated the percentage of threshold probability while the y-axis indicates the net benefit. (C, D) Calibration plots for predicting survival possibility at 3, 4 and 5 years, Diagonal line: ideal model, vertical bars: 95% confidence interval. (AUC: areaunder the receiver operating characteristic curve) .

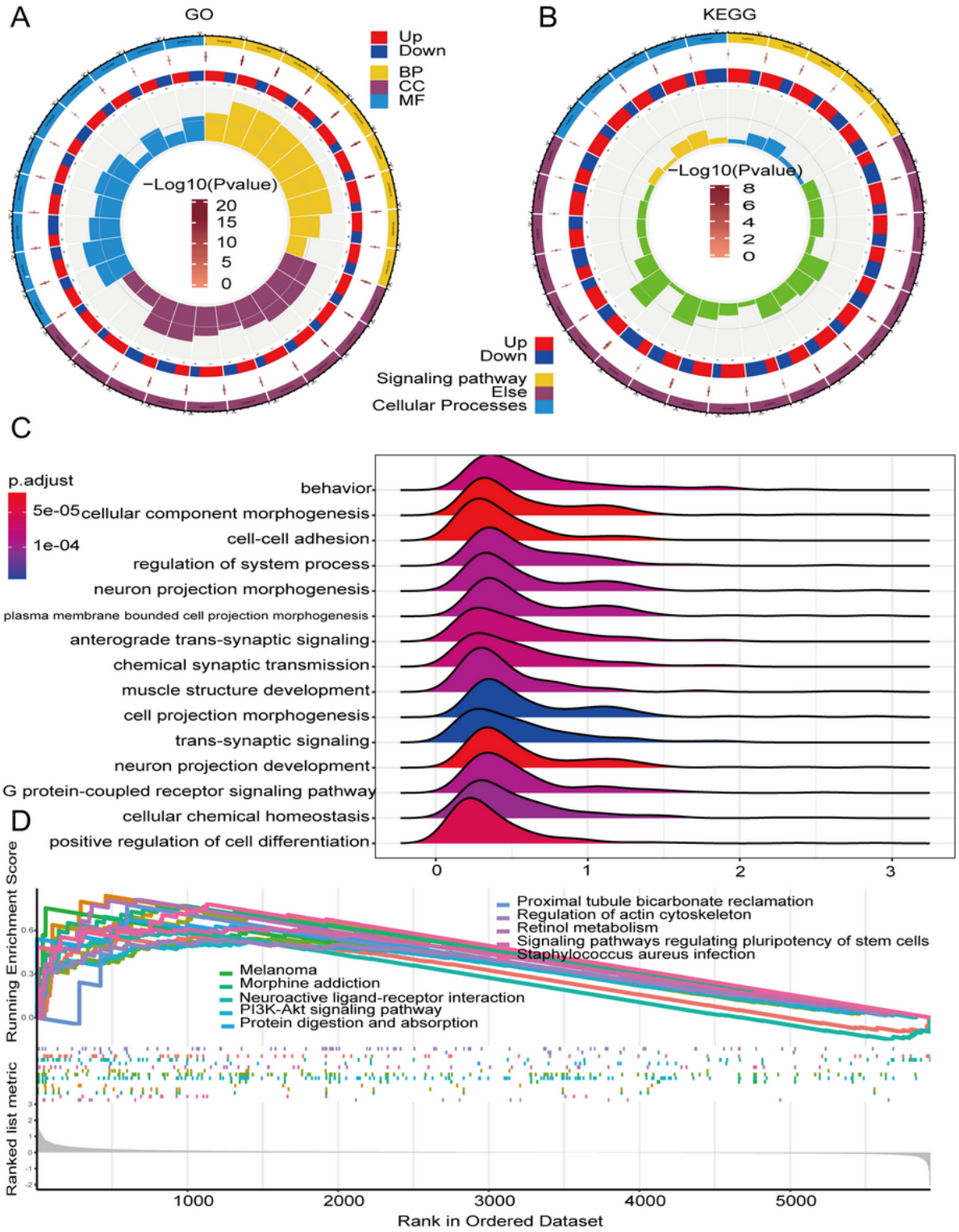

Figure 6

Results of the enrichment analyses in the TCGA cohort.

(A)The significant GO term analyses of differentially expressed genes between the high- and low-risk groups. (Red indicated up-regulation; Blue indicated down-regulation; Yellow indicated biological process class; blue indicated molecular function class; green indicated cellular component class.) (B) The

significant KEGG term analyses of differentially expressed genes between the high- and low-risk groups. (Red indicated up-regulation; Blue indicated down-regulation; Yellow indicated Signaling Pathway class; Blue label belonged to cellular process class. Purple label belonged to else class).

(C) The Ridge plot by GSEA between the high- and low-risk groups showed the top 15significant GOterms

(D) The Enrichment plotby GSEA between the high- and low-risk groups showed the top 15 significant KEGG terms (Permutation tests $P < 0.05$, FDR < 0.25).

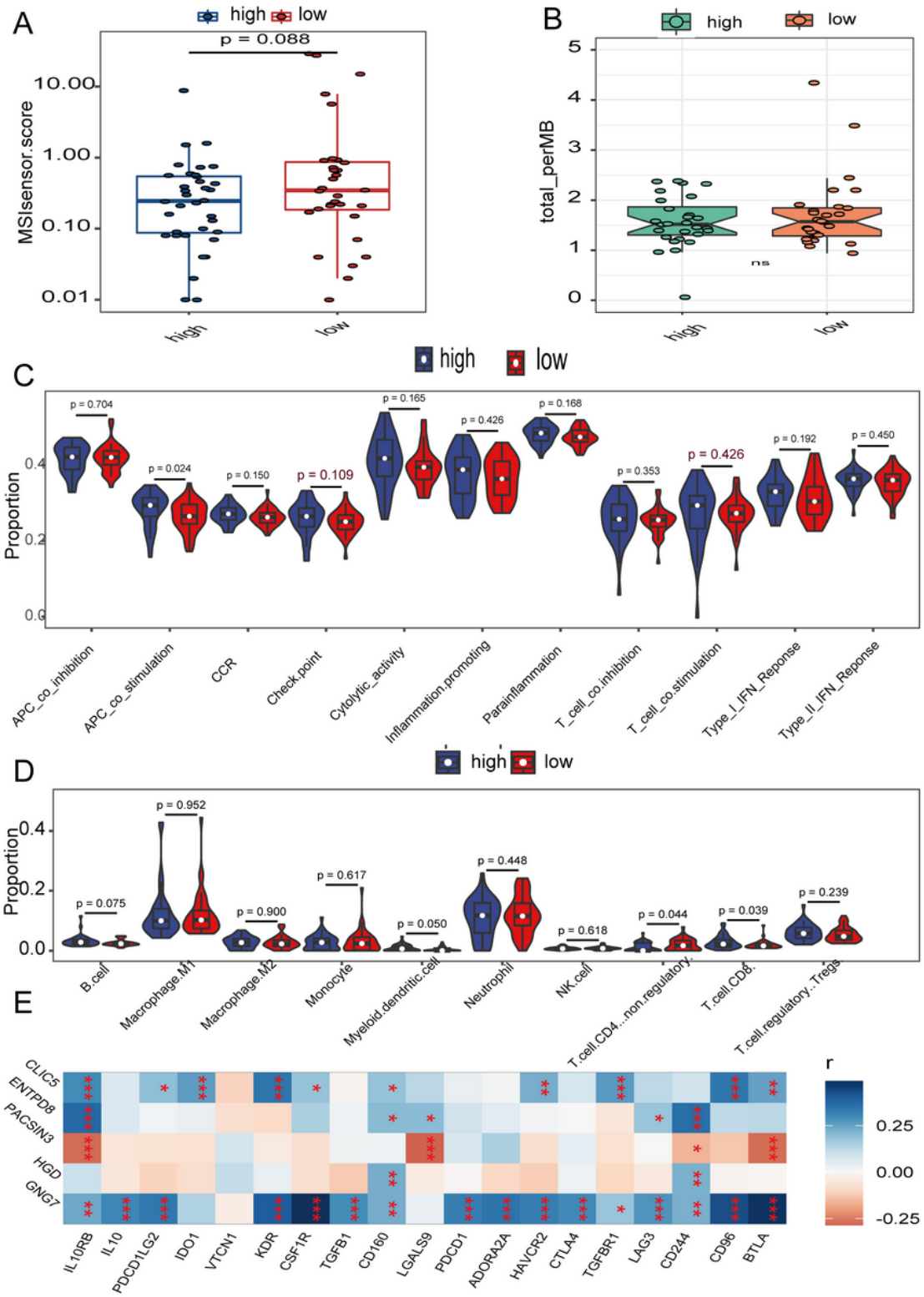

Figure 7

Efficacy ofthe model with signature immunotherapeutic relevantgenes

The MIS scores in the boxplot between different risk groups in the TCGA rectal cancer cohort. (B)Differential TMB levels in the boxplot between high- and low-risk groups among TCGA rectal cancer cohort. (C)The ssGSEA scores of the 10 immune cells in the Violin Plot between high- and low-risk groups

among TCGA rectal cancer cohort. (D)Comparison of the 11 immune-related functions in the Violin Plot between high- and low-risk groups among TCGA rectal cancer cohort.(E) Correlation of the 5-gene expression with immune checkpoint molecules among TCGA rectal cancer cohort. (ns, not significant; ***P*< 0.01; ****P*< 0.001.

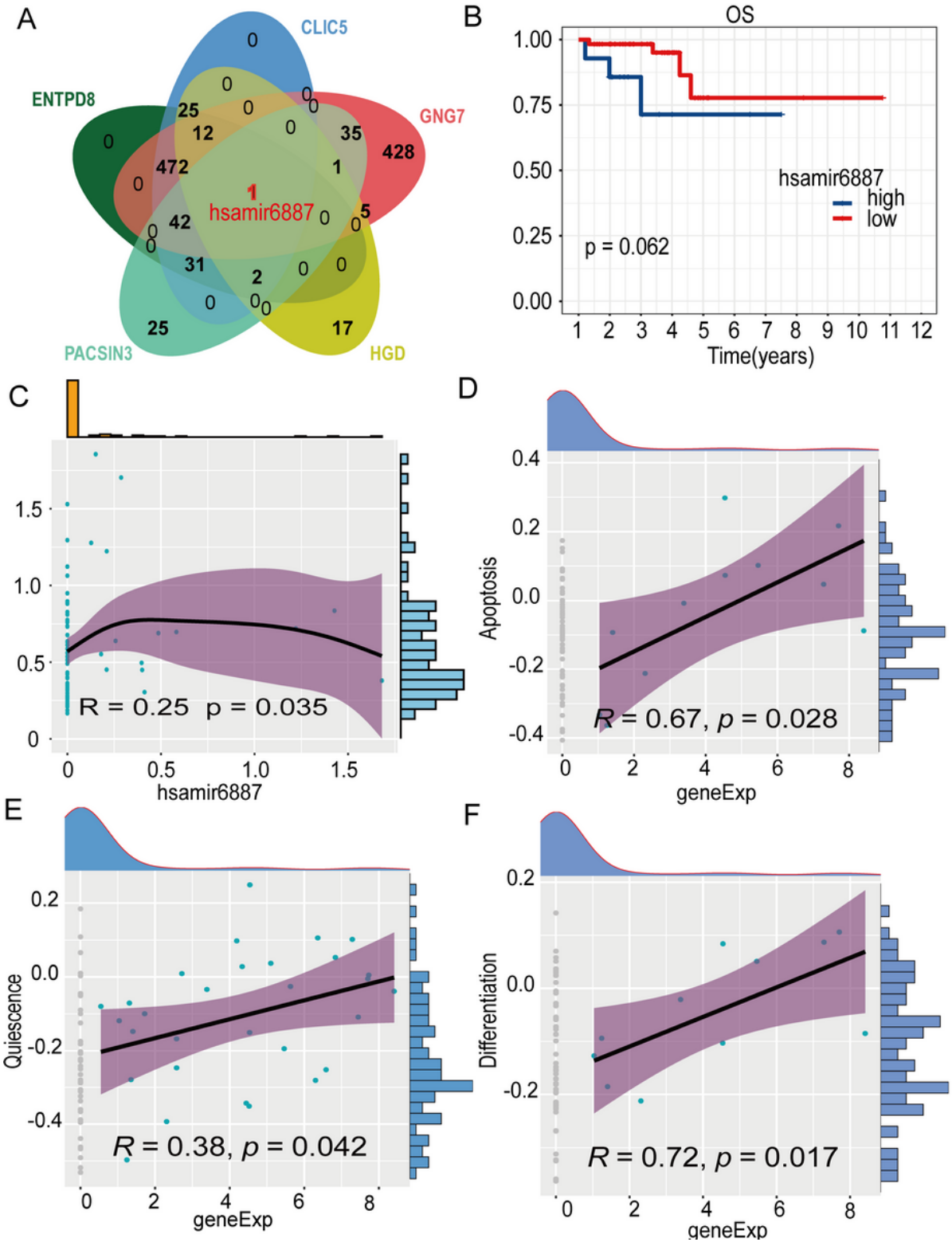

Figure 8

Analyses of the function of the 5-gene signature

(A)Venn diagram to show the overlap of the miRNAs targeted to 5 genes predicted by TargetScan. (B) Kaplan-Meier survival analyses of high and low hsa-miR-6887-expressing groups. (C)The expression of GNG7 correlated with hsa-miR-6887 in TCGA rectal cancer cohort. (D-F)Visualization of correlationsbetween the 5-gene signature and functional states (differentiation, apoptosis, and quiescence) originated fromCancerSEA

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalI.zip
- SupplementalII.zip