

Construction of a Prognostic Model for Predicting Overall Survival of Patients with Colorectal Cancer

Shuwen Han

Huzhou Central Hospital

Kefeng Ding (✉ dingkefeng@zju.edu.cn)

Zhejiang University School of Medicine Second Affiliated Hospital <https://orcid.org/0000-0002-2380-3717>

Research article

Keywords: Colorectal cancer, colon adenocarcinoma, rectal adenocarcinoma, prognosis model, drug prediction

Posted Date: May 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-26225/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Colorectal cancer (CRC) is one of the most common malignancies. The purpose of this study is to construct a prognostic model for predicting the overall survival (OS) in patients with CRC.

Methods: The mRNA-seq and miRNA-seq data of colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) were downloaded from The Cancer Genome Atlas (TCGA) database. The differentially expressed RNAs (DE-RNAs) between tumor and normal tissues were screened. The Kaplan-Meier and univariate Cox regression analysis were used to screen the survival-related genes. Functional enrichment analysis of survival-related genes was conducted, followed by protein-protein interaction (PPI) analysis. Subsequently, the potential drugs targeting differentially expressed mRNAs (DE-mRNAs) were investigated. Multivariate Cox regression analysis was then conducted to screen the independent prognostic factors, and these genes were used to establish a prognostic model. A receiver operator characteristic (ROC) curve was constructed, and the area under the curve (AUC) value of ROC was calculated to evaluate the specificity and sensitivity of the model.

Results: A total of 855 survival-related genes were screened. These genes were mainly enriched in Gene Ontology (GO) terms, such as methylation, synapse organization, and methyltransferase activity; and pathway analysis showed that these genes were significantly involved in N-Glycan biosynthesis and the calcium signaling pathway. PPI analysis showed that aminolevulinic acid dehydratase (ALAD) and cholinergic receptor muscarinic 2 (CHRM2) served vital roles in the development of CRC. Aminolevulinic acid, levulinic acid, and loxapine might be potential drugs for CRC treatment. The prognostic models were built and the patients were divided into high-risk and low-risk groups based on the median of risk score (RS) as screening threshold. The OS for patients in the high-risk group was markedly shorter than that for patients in the low-risk group. Meanwhile, kazal type serine peptidase inhibitor domain 1 (KAZALD1), hippocalcin like 4 (HPCAL4), cadherin 8 (CDH8), synaptopodin 2 (SYNPO2), cyclin D3 (CCND3), and hsa_mir_26b may be independent prognostic factors that could be considered as therapeutic targets for CRC.

Conclusion: We established prognostic models that could predict the OS for CRC patients and may assist clinicians in providing personalized and precision treatment in this patient population.

Highlights:

1. *ALAD* served a vital role in the development of CRC.
2. *CHRM2* played a role in CRC development by affecting the calcium signaling pathway.
3. Aminolevulinic acid, levulinic acid, and loxapine might be potential drugs for treating CRC.
4. *KAZALD1* and *HPCAL4* were associated with the OS of CRC.
5. *CDH8*, *SYNPO2*, *CCND3*, and *hsa-mir-26b* were closely related to the prognostic of CRC staging.

Introduction:

Colorectal cancer (CRC) is one of the most common malignancies and leading causes of cancer death in the world [1]. The prognosis for CRC is widely variable, and approximately 20% of cases are metastatic at the time of emergence [2]. Meanwhile, there are significant differences in prognosis in CRC among different sites [3]. CRC includes colon cancer and rectal cancer, and adenocarcinoma is the most common pathological type of CRC. In addition, this adenocarcinoma comprised of colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) [4]. A previous study reported that the 5-year survival rate is approximately 10% in CRC patients with distant metastases [5]. In addition, COAD patients have a high rate of mortality due to late diagnosis, and lack of reliable biomarkers and therapeutic targets [6]. Therefore, it is necessary to clarify the survival events of CRC and screen novel prognostic factors and therapeutic targets.

Accumulating evidence has reported abnormal expressions of many genes that are involved in the development and carcinogenesis of CRC. Schmid et al. revealed that spondin 2 could induce CRC metastasis in a mice model, and served as a prognostic factor for CRC survival [7]. Kwak et al. observed that the mutation of *KRAS* genes seemed to be associated with the overall survival (OS) of CRC patients [8]. Meanwhile, BRCA1 interacting protein C-terminal helicase 1, ATM serine, homeobox B13, and partner and localizer of BRCA2 may be considered as CRC risk genes [9]. Despite extensive studies that have been carried out to screen prognostic markers, the mortality rate of CRC patients remains high. Additionally, due to the limits of specificity of individual molecules, the survival time cannot be accurately predicted. Thus, there is a strong need to establish a prognostic model to search multi-molecular biomarkers for predicting CRC prognosis.

With the breakthrough advancement of sequencing techniques and bioinformatics analysis, gene expression profiling has been confirmed to be a promising tool for classifying tumors and predicting cancer prognosis [10]. By integrating multidimensional genomic data from The Cancer Genome Atlas (TCGA), a prognostic model can be constructed and may provide personalized and precise treatment for cancer patients. Ying et al. established a prognostic model that generated a RS based on nine-gene expression for efficiently predicting endometrial carcinoma as the patient's prognosis [11]. In addition, Chen et al. developed a seven-gene signature that could be considered as an independent prognostic marker for OS prediction in CRC patients, providing new insights into identification of CRC patients with high-risk of mortality [12]. Since CRC has various common types, including COAD and READ, the developed prognostic signals are difficult to apply widely. Therefore, a prognostic model with high discriminant ability to effectively assist prognosis prediction for each type is needed in clinical practice.

Recently, CRC patients were staged according to the American Joint Committee on Cancer (AJCC) tumor, node, metastasis (TNM) system. A previous study reported that tumor staging is an important independent factor affecting the prognosis of CRC patients [13]. Additionally, OS is regarded as the ultimate measure of treatment benefit [14]. To comprehensively study the prognostic markers of CRC patients, a survival model for predicting prognosis in patients with CRC was constructed. The RNA-seq

and miRNA-seq datasets of COAD and READ based on the OS and tumor staging were respectively downloaded from TCGA database. The differentially expressed RNAs (DE-RNAs) between normal and tumor samples were screened, then Kaplan-Meier (K-M) analysis, univariate Cox regression analysis, and multivariate Cox regression analysis were performed to obtain the prognostic independent factors for CRC. Next, the receiver operator characteristic (ROC) curve was built, and the under the curve (AUC) values of the ROC curve were calculated to assess the specificity and sensitivity of all genes. Furthermore, potential drugs targeting DE-mRNAs were also predicted. This model represents a potentially useful tool for OS predicting, but it still needs to be further verified in clinical practice. The flow chart of the analysis procedure is displayed in Fig. 1.

Materials And Methods:

Raw data processing

The RNA-seq, miRNA-seq, and corresponding clinical information of TCGA-COAD and TCGA-READ patients up to July 19, 2019 were downloaded from the UCSC-Xena platform [15]. The RNA-seq and miRNA-seq were preprocessed by log₂ transformation. Patients were divided into three groups based on the OS: an OS less than five years was regarded as “S”; an OS more than five years and less than ten years was considered as “M”; and an OS more than ten years was defined as “L”. In addition, the tumor staging information of all patients was derived from TCGA database, including stage I, stage II, stage III, and stage IV. The number of each group is listed in Table 1.

Table 1
The number of each group in this analysis

	COAD	READ
Short OS	377	145
Medium OS	35	4
Long OS	5	1
Stage I	70	29
Stage II	157	43
Stage III	119	45
Stage IV	60	23

Re-annotation of mRNA, miRNA, lncRNA, and snoRNA

According to the annotation file gencode.v22.annotation.gene, which was obtained from the GENCODE database [16], the RNA-seq of TCGA-COAD and TCGA-READ datasets was re-annotated to obtain the mRNA, lncRNA, and snoRNA. Moreover, annotated miRNAs were obtained in the same way.

Screening of DE-RNAs

The differential expression of genes in COAD and READ datasets were respectively calculated according to survival time (S, M, and L) and tumor stage (I, II, III, and IV). DE-RNAs, including DEmRNAs, DEmiRNAs, DElncRNAs, and DESnoRNAs, were screened using `kruskal.test` from the R package `stats` [17], and the p value was corrected by Bonferroni-Holm method. DE-RNAs with p value < 0.05 were considered significant. Subsequently, hierarchical cluster analysis was performed to display the heat map using `pheatmap` package (version: 1.0.12) in R software [18].

Functional and pathways enrichment analysis

The univariate Cox regression analysis was used to screen the survival-related genes. To understanding the biological function of these genes, the Gene Ontology (GO) [19] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [20] enrichment analysis were conducted using `clusterProfiler` (version 3.12.0) [21]. The results of GO analysis contained biological process (BP), cellular component (CC), and molecular function (MF). A p value < 0.05 and count > 2 were defined as significant enrichment.

Construction of the protein-protein interaction (PPI) network

To explore the interactions between proteins, PPI analysis was performed using STRING (Version: 11.0, <http://www.string-db.org/>) database [22]. PPIs with confidence score > 0.4 were selected to constructed a PPI network, and this network was visualized using Cytoscape software [23].

Prediction of drug targeting survival-related DE-mRNAs

The drug gene interaction database (DGIdb) showed the drug-gene interaction and gene druggability information. In the present study, DGIdb 3.0 [24] was used to predict the potential drugs targeting the survival-related DE-mRNAs, and the drug-gene relationships were obtained. Subsequently, the drugs which were approved by the Food and Drug Administration agency and supported by the literature were screened. Finally, the drug-gene interaction network was structured by Cytoscape.

Survival analysis and prognostic survival models construction

Clinical and survival information were collected from COAD and READ datasets. Log-rank tests were conducted to evaluate the relationship between DE-RNAs and survival time, and p value < 0.05 was considered as statistically significant, then genes related to prognosis were screened and K-M survival curves were plotted. The `coxph` method [25] from the `survival` package was used for univariate Cox regression analysis to analyze the effects of differential RNA and clinical factors (tumor staging) on survival prognosis. The log-rank test was utilized to calculate the p value and the prognostic correlation coefficient β value, and genes with p value < 0.05 were selected for further study. Subsequently, a multivariate Cox regression analysis was employed to assess independent prognostic factors connected with patient survival, and the candidate DE-RNAs significantly related to the prognosis of patients were screened. Furthermore, the RS of each sample was calculated using the formula as follows: $RS = \beta_{gene}^T$

$\times \exp gene1 + \beta gene2 \times \exp gene2 + \dots + \beta geneN \times \exp geneN$ (where β represents β coefficient value and \exp represents the gene expression level). The COAD and READ patients were divided into high-risk and low-risk groups based on the median RS value, and K-M survival curve was used to assess the correlation between the risk model and the survival prognosis of the samples. A computational model for predicting clinical endpoints of samples from the expression of key prognostic genes was constructed by the support vector machine (SVM) method of e1071 package in R language [26]. To assess the specificity and sensitivity of genes, ROC curves were plotted and the AUC value of ROC was calculated. According to the DE-RNAs obtained in the prognosis model, the nomogram method in R package 'rms' [27] was applied to establish a nomogram model for predicting the three-year, five-year, and ten-year survival probability.

Results:

Screening of DE-RNAs in CRC patients

The DE-RNAs in the two datasets were screened based on the OS and tumor stage. A total of 3207 DE-RNAs were obtained in the COAD OS (Supplementary Table 1), including 2510 DEmRNAs, 375 DElncRNAs, 5 DESnoRNA, and 21 DEmiRNAs; a total of 2451 DE-RNAs were obtained in the COAD stage (Supplementary Table 2). Moreover, 1013 DE-RNAs and 1460 DE-RNAs were respectively screened in READ OS and stage (Supplementary Tables 3 and 4). The clustering heatmap of DE-RNAs in different groups is shown in Supplementary Fig. 1.

Results of functional and pathway enrichment analysis

A total of 143, 503, 61, and 148 survival-related genes were respectively obtained in COAD OS, COAD stage, READ OS, and READ stage groups. To explore the functional characteristics of the screened OS-related genes, functional enrichment analysis was performed (Fig. 2). The GO enrichment analysis showed that DE-RNAs in the COAD OS group were mainly associated with BP terms, such as methylation, positive regulation of synaptic plasticity, and response to cobalt ion; the genes in the COAD stage group were connected with BP terms including synapse organization, modulation of chemical synaptic transmission, and regulation of trans-synaptic signaling. Additionally, the genes in the READ OS group were primarily enriched in mRNA 5'-splice site recognition, regulation of the force of heart contraction, and muscle filament sliding; whereas the genes in the READ stage group were primarily correlated with muscle system process, muscle contraction, and cartilage development. Meanwhile, KEGG pathway enrichment analysis indicated that survival-related genes in the COAD groups were involved in N-Glycan biosynthesis, drug metabolism- other enzymes, and calcium signaling pathway; in addition, genes in the READ groups mostly participated in glycolysis/gluconeogenesis, carbon metabolism, PPAR signaling pathway, and calcium signaling pathway.

PPI network of survival-related RNAs

To clarify the relationship between proteins encoded by genes, STRING was utilized to establish PPI networks (Fig. 3). In the COAD OS group, there were 28 nodes and 19 edges. The degrees of surfeit 6

(*SURF6*), dolichyl-phosphate mannosyltransferase subunit 2 (*DPM2*), and aminolevulinic acid dehydratase (*ALAD*) were higher than that of others. In the PPI network of the COAD stage group, gastrin (*GAST*), dopamine receptor D4 (*DRD4*), and corticotropin releasing hormone receptor 1 (*CRHR1*) with higher degrees and might be recognized as hub genes. The PPI network of the READ OS group was composed of phosphoglycerate kinase 2 (*PGK2*) and fructose-bisphosphatase 2 (*FBP2*). Moreover, angiotensin II receptor type 1 (*AGTR1*), heart and neural crest derivatives expressed 2 (*HAND2*), and cholinergic receptor muscarinic 2 (*CHRM2*) had higher degrees than that of others in the PPI network of the READ stage group.

Drug prediction with regard to survival-related DE-mRNAs

The DE-mRNAs obtained from the univariate Cox regression analysis were further analyzed by DGIdb database. A total of 13 genes might be considered as druggable genes, such as *ALAD*, peptidylprolyl isomerase F (*PPIF*), melan-A (*MLANA*), N-Acetyltransferase 1 (*NAT1*), and *CHRM2* (Fig. 4). Furthermore, aminolevulinic acid was an inducer or inhibitor of *ALAD*, levulinic and chlorzoxazone could respectively regulate the expression level of *ALAD* and *MLANA*, and cyclosporine was a binder of *PPIF*.

Survival-related gene screening and prognostic model construction

Construction of a prognostic model based on genes in COAD OS

In the COAD OS group, the prognostic value of the 3207 DE-RNAs was assessed, and the results of log-rank test indicated that 143 genes were associated with survival of COAD patients. In addition, five genes were selected as independent predictors in COAD patients including hippocalcin like 4 (*HPCAL4*), *RP11_452L6.1*, *AC006042.6*, *RP11_465I4.3*, and kazal type serine peptidase inhibitor domain 1 (*KAZALD1*). These genes were used to build a prognostic model, and the RS of this model was calculated by using the formula: $RS = (0.58 * \text{Exp}HPCAL4) + (2.35 * \text{Exp}RP11_452L6.1) + (2.63 * \text{Exp}AC006042.6) + (0.53 * \text{Exp}RP11_465I4.3) + (0.49 * \text{Exp}KAZALD1)$. A total of 417 patients were divided into a high-risk group and low-risk group by applying the median of RS as the cut-off criteria, and subsequently assessed by K-M survival analysis. The results indicated a significant difference between the high-risk group and low-risk group (p value = $1.854e-2$), and the high-risk group was connected with poor prognosis (Fig. 5A). The AUC value of this prognostic model was 0.677 (Fig. 5B). The RS distribution, patient survival status, and heat map of five genes in the two groups are displayed in Fig. 5C-E. The results revealed that high-risk patients had high mortality and high expression levels of prognosis genes. Subsequently, a nomogram was plotted to predict the probability of the 3-, 5-, and 10-year OS (Fig. 5F). Five independent prognostic factors were obtained, including *KAZALD1*, *HPCAL4*, *AC006042.6*, *RP11_465I4.3*, and *RP11_452L6.1*.

Establishment of a prognostic model based on genes in the READ-OS group

In the READ-OS group, 61 survival-related genes were screened by univariate Cox analysis; among these, seven genes, namely *RNVU1_15*, *PGAM1P4*, *RP11_291L19.1*, *RN7SL535P*, *RP11_96D1.3*, *CTD_3220F14.2*, and *hsa_mir_4444_1* (all $P < 0.01$), were used to conduct a risk score model for READ patients. The results revealed that low expression level of these genes was associated with poor prognosis. The formula of RS was: $RS = 1.69 * \text{Exp}RNVU1_15 + 2.02 * \text{Exp}PGAM1P4 + 3.53 * \text{Exp}RP11_291L19.1 + 1.91 * \text{Exp}RN7SL535P + 5.62 * \text{Exp}RP11_96D1.3 + 0.79 * \text{Exp}CTD_3220F14.2 + 0.63 * \text{Exp}hsa_mir_4444_1$. The 150 READ patients were divided into a high-risk group and low-risk group based on the median value of RS. K-M survival analysis showed that patients with low RS had significantly longer OS than patients with high RS (p value = 0) (Fig. 6A). The AUC value of ROC curve was 0.917, indicating that the predictive model had high specificity (Fig. 6B). Meanwhile, the RS distribution, survival status of patients, and heat map of gene expression level are exhibited in Fig. 6C-E. Furthermore, seven independent prognostic factors including *RNVU1_15*, *PGAM1P4*, *RP11_291L19.1*, *RN7SL535P*, *RP11_96D1.3*, *CTD_3220F14.2*, and *hsa_mir_4444_1*, were revealed in the nomogram (Fig. 6). These genes may be potential therapeutic targets for READ patients.

Construction of the COAD stage prognostic model

In the COAD stage group, 267 prognosis-related genes were obtained using K-M survival analysis. A total of six genes significantly related to OS were detected using multi-Cox regression analysis. As a result, six genes, namely TUB like protein 1 (*TULP1*), cadherin 8 (*CDH8*), *AC002076.10*, *CTD_2561B21.10*, *RP11_266L9.8*, and *hsa_mir_26b*, were used to build the prognostic model. The RS value was calculated using the follow formula: $RS = (4.22 * \text{Exp}TULP1) + (3.95 * \text{Exp}CDH8) + (0.27 * \text{Exp}AC002076.10) + (0.42 * \text{Exp}CTD_2561B21.10) + (2.67 * \text{Exp}RP11_266L9.8) + (0.39 * \text{Exp}hsa_mir_26b)$. A total of 406 patients were divided into high-risk and low-risk groups according to the median RS. Survival analysis showed that patients in the high-risk group had a poorer prognosis than those in the low-risk group (p value = $2.842e-02$) (Fig. 7A). The AUC of prognostic model was 0.814 (Fig. 7B), suggesting that our model had a favorable efficiency in predicting prognosis of different stages of COAD patients. Notably, high-risk patients had more death and the expression level of survival-related genes was up-regulated in the high-risk group (Fig. 7C-E). In addition, six potential independent prognostic factors were predicted by nomograms, which were consistent with the genes obtained above (Fig. 7F).

Establishment of the READ stage prognostic model

In the READ stage group, Cox regression analysis, K-M survival analysis, and ROC curve analysis were performed to assess the prognostic value of the 1460 DE-RNAs. A total of four survival-related genes were selected by a series of analyses: adducin 2 (*ADD2*), cyclin D3 (*CCND3*), synaptopodin 2 (*SYNPO2*), and homeobox A4 (*HOXA4*). These four genes were used to construct a predictive model. RSs were calculated using the formula: $RS = (1.6e + 03 * \text{Exp}ADD2) + (8.8e + 00 * \text{Exp}CCND3) + (4.1e - 04 * \text{Exp}SYNPO2) + (1.5e - 02 * \text{Exp}HOXA4)$. A total of 140 patients were divided into high- and low-risk groups based on the median of the RS. There was a significant difference between the high-risk and low-risk groups (p value = $2.8021e-01$) (Fig. 8A), and the patients in the low-risk group had markedly longer OS

than those in the high-risk group. Moreover, the ROC curve analysis was conducted and the AUC of the prognostic model was 0.859 (Fig. 8B). The RS, survival status of patients, and heat map of gene expression levels are shown in Fig. 8C-E. Subsequently, a nomogram was plotted to predict the READ patients' outcome (Fig. 8F). *ADD2*, *CCND3*, *SYNPO2*, and *HOXA4* were considered as independent prognostic factors.

Discussion:

CRC is one of the most common malignancies, with multiple common types. Despite surgical resection, chemotherapy, and radiotherapy used in CRC treatment, the prognosis of CRC remains poor and CRC has high morbidity and mortality rates [28]. In the present study, we established a prognostic model of COAD and READ according to the OS and tumor staging. The clinical implementation of this developed risk score model is likely to provide more relevant information for clinical diagnoses and may improve the prognosis of CRC patients. *ALAD* and *CHRM2* were considered as hub genes in the PPT network. Furthermore, aminolevulinic acid, levulinic, and loxapine may be potential drugs for CRC treatment. Meanwhile, *KAZALD1*, *HPCAL4*, *CDH8*, *SYNPO2*, *CCND3*, and *hsa_mir_26b* were significantly associated with the OS of CRC patients, and served as potential independent prognostic factors.

PPI analysis showed that *ALAD* and *CHRM2* were connected with the development of CRC, and were also related to the prognosis of CRC patients. *ALAD* catalyzed the second step in heme biosynthesis and was an inhibitor of the 26S proteasome [29]. Proteasome has been found to be associated with cancer and is a therapeutic target for cancer [30]. Meanwhile, the genetic variants in *ALAD* are connected with the risk of genitourinary cancers [31]. Neslund et al. indicated the potential interactions between *ALAD* genetic variants and occupational Pb exposure, suggesting that *ALAD* may be a current target of prostate cancer [32]. Additionally, a polymorphism in *ALAD* genes may affect lead toxicokinetics; Van Bommel et al. revealed that *ALAD* may alter the risk of renal cell carcinoma related to lead exposure [33]. *ALAD* has also been connected with prognosis of cancers, such as breast cancer [34]. Drug prediction analysis showed that *ALAD* was targeted by aminolevulinic acid (ALA) and levulinic acid. ALA is an endogenous metabolite formed by succinyl-CoA and glycine in mitochondria. ALA, as an already approved therapeutic strategy, is considered as a successful prodrug used in cancer treatment. The photosensitizer excited by ALA could contribute to the generation of reactive oxygen species, which induces oxidative damage and consequently leads to cell death [35]. Ferreira et al. indicated that melanoma cancer could be treated by photodynamic therapy using ALA [36]. Prior studies observed that 5-ALA was a promising candidate for CRC diagnosis, and 5-ALA mediated photodynamic therapy could be a therapeutic strategy for CRC [37, 38], which is consistent with our finding that ALA is an effective drug for CRC treatment. Another drug, levulinic acid, was detected in this study. Evidence demonstrated that levulinic acid has antimicrobial activity, and could reduce *Salmonella* contamination on the surface of various materials. *Salmonella* infection is the common cause of foodborne diseases [39]; meanwhile, food-borne mutagens and pollution are known risk factors for CRC [40]. Thus, we speculated that levulinic acid may be a potential drug for CRC therapy. *CHRM2* belongs to a larger family of G protein-coupled receptors with adenylate cyclase inhibitory effect [41]. Furthermore, the *CHRM2* gene is prone to alcohol dependence and drug

dependence [42] and may affect the internal nerves of gastrointestinal function [43]. However, there have been no reports on the association of *CHRM2* and CRC; thus, further experiments should be performed to explore the effect of *CHRM2* on CRC. We also found that *CHRM2* was involved in the calcium signaling pathway. Accumulating evidence suggested that intracellular Ca²⁺ homeostasis is changed in cancer cells, and this alteration participates in tumor initiation, progression, and metastasis. Therefore, targeting calcium signaling is a novel approach for treating malignant tumors [44]. Bush et al. showed that PPAR γ -regulated calcium signaling was correlated with CRC [45]. Taken together, *CHRM2* serves an important role in the development of CRC by affecting the calcium signaling pathway. In addition, *CHRM2* has been targeted by loxapine. Previous studies reported that loxapine was used in the treatment of breast cancer and mental disorder [46]. However, the therapeutic effect of loxapine on CRC remains unclear, and further experiments in preclinical models must be conducted to validate this result.

The results of the prognostic model showed that *KAZALD1* and *HPCAL4* may be considered as significant independent prognostic factors for CRC. *KAZALD1* was over-expressed in cancer samples and was associated with shorter survival of patients with high-grade glioma [47]. Moreover, Alvi et al. indicated that *KAZALD1* hypomethylation was related to poor survival of small bowel adenocarcinoma, and could be used as a prognostic biomarker [48]. The proteins encoded by *HPCAL4* may regulate intracellular calcium homeostasis, which is connected to tumorigenesis and cancer progression [49]. Bavarva et al. observed that *HPCAL4* was relevant to various cancer types, such as breast cancer, lung cancer, and CRC [50]. Preoperative staging of CRC is the basis of surgical strategy, and incomplete staging can lead to poor prognosis [51]. Meanwhile, tumor staging can provide valuable prognostic information and contribute to the design of treatment strategies [52]. In the present study, we observed that *CDH8*, *SYNPO2*, *CCND3*, and *hsa_mir_26b* may be independent prognostic factors of CRC staging. The proteins encoded by *CDH8* were type II classical cadherin from the cadherin superfamily [53]. A study comparing the gene expression profiles of cancer and health samples showed that the abnormalities of genes such as *CDH8* appeared only in cancer tissues [54]. Lee et al. revealed that *CDH8* was a prognosis factor of non-small cell lung cancer (NSCLC) [55]. Furthermore, Lu et al. demonstrated that *CDH8* may be used for predicting the survival of patients with stage I NSCLC [56]. *SYNPO2* is a member of the podin family and is a repressor of tumor cell invasion [57]. Kai et al. indicated that *SYNPO2* could induce the formation of the complex stress fiber network in the cell body, and promote the migration of PC3 prostate cancer cells under serum stimulation [58]. Additionally, Kandimalla et al. revealed that *SYNPO2* methylation was significantly correlated with the tumor stage of bladder cancer [59]. The proteins encoded by *CCND3* belong to the highly conserved cyclin family, and these proteins are involved in the phosphorylation of tumor suppressor protein Rb [60]. Several studies reported that *miR-592* could inhibit cell proliferation and tumor growth of CRC cells by suppressing *CCND3* expression [61]. Edelman et al. found that the expression level of *CCND3* was significantly altered in patients with stage I squamous cell lung cancer [62]. Another gene, *hsa_mir_26b*, was significantly related to the OS of CRC. Lin et al. found that *miR-26b* could induce pig granulosa cell apoptosis [63]. Meanwhile, Ma et al. observed that *miR-26b* was markedly connected with the invasiveness and metastasis of CRC cells, and it might serve as a prognostic factor for CRC [64]. However, the effect of *hsa-mir-26b* on the prognosis of tumor staging had

not been reported. Thus, the relationship between *hsa-mir-26b* and cancer staging needs further study. Notably, the AUC values of the ROC curve for the COAD OS, COAD stage, READ OS, and READ stage prognostic models were 0.677, 0.917, 0.814, and 0.859, respectively, indicating that these genes had a good performance in survival prediction. Taken together, *KAZALD1*, *HPCAL4*, *CDH8*, *SYNPO2*, *CCND3*, and *hsa_mir_26b* were considered as potential prognostic markers for CRC. However, the specific action mechanism of these genes needs to be elucidated in further studies.

Overall, we systematically investigated the molecular mechanism underlying the development and prognosis of CRC. First, the DE-RNAs related to the prognosis of CRC were screened. Even though the action mechanisms of these genes are still not fully understood, our study has provided a theoretical basis for further research on CRC prognostic. Second, a number of drugs were predicted by drug-target analysis. Although these drugs have not been clinically proven, our work has provided guidance for clinical drug use. Some anticancer drugs may improve disease-free survival or relapse-free survival of CRC [65], but there is currently no effective drugs for improving OS. In the present study, OS was regarded as the endpoint for developing molecular markers; thus, these drugs may be beneficial for improving the OS of CRC. Third, the prognostic models may assist clinicians in providing individualized treatment in CRC patients.

Conclusion:

In conclusion, the results in this study suggest that the prognostic models developed are reliable tools for predicting the OS of CRC patients, and these findings could assist clinicians in providing personalized and precise treatment for CRC patients. *KAZALD1*, *HPCAL4*, *CDH8*, *SYNPO2*, *CCND3*, and *hsa_mir_26b* might be therapeutic targets for CRC. Additionally, aminolevulinic acid, levulinic acid, and loxapine may be potential drugs for CRC treatment.

Abbreviations

CRC: colorectal cancer

COAD: colon adenocarcinoma

READ: rectal adenocarcinoma

ALAD: aminolevulinic acid dehydratase

CHRM2: cholinergic receptor muscarinic 2

KAZALD1: kazal type serine peptidase inhibitor domain 1

CDH8: cadherin 8

ADD2: adducin 2

CCND3: cyclin D3

OS: overall survival

AJCC: American Joint Committee on Cancer

TCGA: The Cancer Genome Atlas

DE-RNAs: differentially expressed RNAs

ROC: receiver operator characteristic

AUC: area under the curve

RS: risk score

K-M: Kaplan-Meier

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

BP: biological process

CC: cellular component

MF: molecular function

PPI: protein-protein interaction

DGIdb: drug gene interaction database

HPCAL4: hippocalcin like 4

TULP1: TUB like protein 1

SYNPO2: synaptopodin 2

NSCLC: non-small cell lung cancer

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing Interests

The authors declare that no conflicts of interest exist.

Funding

This work was supported by the Zhejiang Science and Technology Projects (No. Q20H160001).

Authors' Contributions

All authors participated in the conception and design of the study;

Wrote the paper: Han Shuwen, and Ding Kefeng;

Processed the data: Han Shuwen;

Drew figures: Ding Kefeng;

All authors read and approved the paper.

Acknowledgements

The authors gratefully acknowledge the database available to us for this study.

Availability of data and materials

The datasets generated during the current study are not publicly available but obtained from corresponding authors on reasonable request.

References

1. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RG, et al. (2017) Colorectal cancer statistics, 2017. *CA: a cancer journal for clinicians* 67: 177-193.
2. Sarshekeh AM, Advani S, Overman MJ, Manyam G, Kee BK, et al. (2017) Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. *PLoS One* 12.
3. Xi Y, Yuefen P, Wei W, Quan Q, Jing Z, et al. (2019) Analysis of prognosis, genome, microbiome, and microbial metabolome in different sites of colorectal cancer. *Journal of translational medicine* 17:

353.

4. Hajmanoochehri F, Asefzadeh S, Kazemifar AM, Ebtehaj M (2014) Clinicopathological features of colon adenocarcinoma in Qazvin, Iran: a 16 year study. *Asian Pacific Journal of Cancer Prevention* 15: 951-955.
5. Sharkas GF, Arqoub KH, Khader YS, Tarawneh MR, Nimri OF, et al. (2017) Colorectal cancer in Jordan: survival rate and its related factors. *Journal of oncology* 2017.
6. McLaughlin C, Kim N-K, Bandyopadhyay D, Deng X, Kaplan B, et al. (2019) Adjuvant radiation therapy for T4 non-rectal colon adenocarcinoma provides a cause-specific survival advantage: A SEER database analysis. *Radiotherapy and Oncology* 133: 50-53.
7. Schmid F, Wang Q, Huska M, Andrade-Navarro M, Lemm M, et al. (2016) SPON2, a newly identified target gene of MACC1, drives colorectal cancer metastasis in mice and is prognostic for colorectal cancer patient survival. *Oncogene* 35: 5942-5952.
8. Kwak MS, Cha JM, Yoon JY, Jeon JW, Shin HP, et al. (2017) Prognostic value of KRAS codon 13 gene mutation for overall survival in colorectal cancer: direct and indirect comparison meta-analysis. *Medicine* 96.
9. Cercek A, Kemel Y, Mandelker D, Stewart C, Arnold AG, et al. (2018) Prevalence of germline genetic alterations in colorectal cancer patients. *American Society of Clinical Oncology*.
10. Quackenbush J (2006) Microarray analysis and tumor classification. *New England Journal of Medicine* 354: 2463-2472.
11. Ying J, Wang Q, Xu T, Lyu J (2018) Establishment of a nine-gene prognostic model for predicting overall survival of patients with endometrial carcinoma. *Cancer medicine* 7: 2601-2611.
12. Chen H, Sun X, Ge W, Qian Y, Bai R, et al. (2017) A seven-gene signature predicts overall survival of patients with colorectal cancer. *Oncotarget* 8: 95054.
13. Batista WR, Santos G, Vital FMR, Matos D (2019) Immunoexpression of TS, p53, COX2, EGFR, MSH6 and MLH1 biomarkers and its correlation with degree of differentiation, tumor staging and prognostic factors in colorectal adenocarcinoma: a retrospective longitudinal study. *Sao Paulo Medical Journal* 137: 33-38.
14. Verma S, McLeod D, Batist G, Robidoux A, Martins IR, et al. (2011) In the end what matters most? A review of clinical endpoints in advanced breast cancer. *The oncologist* 16: 25.
15. Tomczak K, Czerwińska P, Wiznerowicz M (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology* 19: A68.
16. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* 22: 1760-1774.
17. Breslow N (1970) A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* 57: 579-594.
18. Kolde R, Kolde MR (2015) Package 'pheatmap'. R Package 1.

19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. *Nature genetics* 25: 25-29.
20. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28: 27-30.
21. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* 16: 284-287.
22. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, et al. (2010) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research* 39: D561-D568.
23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13: 2498-2504.
24. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, et al. (2018) DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic acids research* 46: D1068-D1073.
25. Cox DR (1972) Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34: 187-202.
26. Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A (2008) Misc functions of the Department of Statistics (e1071), TU Wien. R package 1: 5-24.
27. Harrell Jr FE, Harrell Jr MFE, Hmisc D (2019) Package ‘rms’. Vanderbilt University 229.
28. Liang R, Lin Y, Yuan CL, Liu ZH, Li YQ, et al. (2018) High expression of estrogen-related receptor α is significantly associated with poor prognosis in patients with colorectal cancer. *Oncology letters* 15: 5933-5939.
29. Schmitt SM, Neslund-Dudas C, Shen M, Cui C, Mitra B, et al. (2016) Involvement of ALAD-20S Proteasome Complexes in Ubiquitination and Acetylation of Proteasomal α 2 Subunits. *Journal of cellular biochemistry* 117: 144-151.
30. Voorhees PM, Dees EC, O’Neil B, Orlowski RZ (2003) The proteasome as a target for cancer therapy. *Clinical cancer research* 9: 6316-6325.
31. da Rocha Filho HN, da Silva EC, Silva FR, Courrol LC, de Mesquita CH, et al. (2015) Expression of genes involved in porphyrin biosynthesis pathway in the human renal cell carcinoma. *Journal of fluorescence* 25: 1363-1369.
32. Neslund-Dudas C, Levin AM, Rundle A, Beebe-Dimmer J, Bock CH, et al. (2014) Case-only gene–environment interaction between ALAD tagSNPs and occupational lead exposure in prostate cancer. *The Prostate* 74: 637-646.
33. Van Bommel DM, Boffetta P, Liao LM, Berndt SI, Menashe I, et al. (2011) Comprehensive analysis of 5-aminolevulinic acid dehydrogenase (ALAD) variants and renal cell carcinoma risk among individuals exposed to lead. *PloS one* 6.
34. Ge J, Yu Y, Xin F, Yang ZJ, Zhao HM, et al. (2017) Downregulation of delta-aminolevulinic acid dehydratase is associated with poor prognosis in patients with breast cancer. *Cancer science* 108:

604-611.

35. Wachowska M, Muchowicz A, Firczuk M, Gabrysiak M, Winiarska M, et al. (2011) Aminolevulinic acid (ALA) as a prodrug in photodynamic therapy of cancer. *Molecules* 16: 4140-4164.
36. M Ferreira D, Y Saga Y, Aluicio-Sarduy E, Tedesco A (2013) Chitosan nanoparticles for melanoma cancer treatment by photodynamic therapy and electrochemotherapy using aminolevulinic acid derivatives. *Current medicinal chemistry* 20: 1904-1911.
37. Kondo Y, Murayama Y, Konishi H, Morimura R, Komatsu S, et al. (2014) Fluorescent detection of peritoneal metastasis in human colorectal cancer using 5-aminolevulinic acid. *International journal of oncology* 45: 41-46.
38. Hatakeyama T, Murayama Y, Komatsu S, Shiozaki A, Kuriu Y, et al. (2013) Efficacy of 5-aminolevulinic acid-mediated photodynamic therapy using light-emitting diodes in human colon cancer cells. *Oncology reports* 29: 911-916.
39. Zhao T, Zhao P, Cannon JL, Doyle MP (2011) Inactivation of Salmonella in biofilms and on chicken cages and preharvest poultry by levulinic acid and sodium dodecyl sulfate. *Journal of food protection* 74: 2024-2030.
40. Yesudhas D, Gosu V, Anwar MA, Choi S (2014) Multiple roles of toll-like receptor 4 in colorectal cancer. *Frontiers in immunology* 5: 334.
41. Hendershot CS, Bryan AD, Ewing SWF, Claus ED, Hutchison KE (2011) Preliminary evidence for associations of CHRM2 with substance use and disinhibition in adolescence. *Journal of abnormal child psychology* 39: 671-681.
42. Luo X, Kranzler HR, Zuo L, Wang S, Blumberg HP, et al. (2005) CHRM2 gene predisposes to alcohol dependence, drug dependence and affective disorders: results from an extended case-control structured association study. *Human Molecular Genetics* 14: 2421-2434.
43. Davis EA, Zhou W, Dailey MJ (2018) Evidence for a direct effect of the autonomic nervous system on intestinal epithelial stem cell proliferation. *Physiological reports* 6.
44. Cui C, Merritt R, Fu L, Pan Z (2017) Targeting calcium signaling in cancer therapy. *Acta pharmaceutica sinica B* 7: 3-17.
45. Bush CR, Havens JM, Necela BM, Su W, Chen L, et al. (2007) Functional genomic analysis reveals cross-talk between peroxisome proliferator-activated receptor γ and calcium signaling in human colorectal cancer cells. *Journal of Biological Chemistry* 282: 23387-23401.
46. Allen MH, Feifel D, Lesem MD, Zimbhoff DL, Ross R, et al. (2011) Efficacy and safety of loxapine for inhalation in the treatment of agitation in patients with schizophrenia: a randomized, double-blind, placebo-controlled trial. *The Journal of clinical psychiatry* 72: 1313-1321.
47. Wang H, Feng Y, Bao Z, Jiang C, Yan W, et al. (2013) Epigenetic silencing of KAZALD1 confers a better prognosis and is associated with malignant transformation/progression in glioma. *Oncology reports* 30: 2089-2096.
48. Alvi MA, McArt DG, Kelly P, Fuchs M-A, Alderdice M, et al. (2015) Comprehensive molecular pathology analysis of small bowel adenocarcinoma reveals novel targets with potential for clinical utility.

Oncotarget 6: 20863.

49. Vastagh C, Solymosi N, Farkas I, Liposits Z (2019) Proestrus differentially regulates expression of ion channel and calcium homeostasis genes in GnRH neurons of mice. *Frontiers in Molecular Neuroscience* 12.
50. Bavarva JH, Tae H, Settlage RE, Garner HR (2013) Characterizing the genetic basis for nicotine induced cancer development: a transcriptome sequencing study. *PloS one* 8.
51. Levy M, Visokai V, Lipska L, Topolcan O (2008) Tumor markers in staging and prognosis of colorectal carcinoma. *Neoplasma* 55: 138.
52. Tsoi H, Wong K-F, Luk JM, Staunton D (2019) Clinical utility of CDH17 biomarker in tumor tissues and liquid biopsies for detection and prognostic staging of colorectal cancer (CRC). *American Society of Clinical Oncology*.
53. Hulpiau P, Van Roy F (2009) Molecular evolution of the cadherin superfamily. *The international journal of biochemistry & cell biology* 41: 349-369.
54. Papatotiriou I, Ntanovasilis D, Apostolou P (2017) Comparison of genomic and gene expression profile between cancer and healthy samples. *Journal of clinical oncology*: e13001.
55. Lee Y, Yoon K-A, Joo J, Lee D, Bae K, et al. (2013) Prognostic implications of genetic variants in advanced non-small cell lung cancer: a genome-wide association study. *Carcinogenesis* 34: 307-313.
56. Lu Y, Lemon W, Liu P-Y, Yi Y, Morrison C, et al. (2006) A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS medicine* 3.
57. Weins A, Schwarz K, Faul C, Barisoni L, Linke WA, et al. (2001) Differentiation-and stress-dependent nuclear cytoplasmic redistribution of myopodin, a novel actin-bundling protein. *The Journal of cell biology* 155: 393-404.
58. Xia E, Zhou X, Bhandari A, Zhang X, Wang O (2018) Synaptopodin-2 plays an important role in the metastasis of breast cancer via PI3K/Akt/mTOR pathway. *Cancer management and research* 10: 1575.
59. Kandimalla R, Van Tilborg AA, Zwarthoff EC (2013) DNA methylation-based biomarkers in bladder cancer. *Nature Reviews Urology* 10: 327.
60. Liu Q, Fu H, Sun F, Zhang H, Tie Y, et al. (2008) miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic acids research* 36: 5391.
61. Liu Z, Wu R, Li G, Sun P, Xu Q, et al. (2015) MiR-592 inhibited cell proliferation of human colorectal cancer cells by suppressing of CCND3 expression. *International journal of clinical and experimental medicine* 8: 3490.
62. Edelman MJ, Redman MW, Albain KS, McGary EC, Rafique NM, et al. (2019) SWOG S1400C (NCT02154490)—A Phase II Study of Palbociclib for Previously Treated Cell Cycle Gene Alteration—Positive Patients with Stage IV Squamous Cell Lung Cancer (Lung-MAP Substudy). *Journal of Thoracic Oncology* 14: 1853-1859.

63. Lin F, Li R, xiang Pan Z, Zhou B, bing Yu D, et al. (2012) miR-26b promotes granulosa cell apoptosis by targeting ATM during follicular atresia in porcine ovary. PLoS One 7.
64. Ma YL, Zhang P, Wang F, Moyer MP, Yang JJ, et al. (2011) Human embryonic stem cells and metastatic colorectal cancer cells shared the common endogenous human microRNA-26b. Journal of cellular and molecular medicine 15: 1941-1954.
65. Tejpar S, Saridaki Z, Delorenzi M, Bosman F, Roth AD (2011) Microsatellite instability, prognosis and drug sensitivity of stage II and III colorectal cancer: more complexity to the puzzle. Oxford University Press. pp. 841-844.

Figures

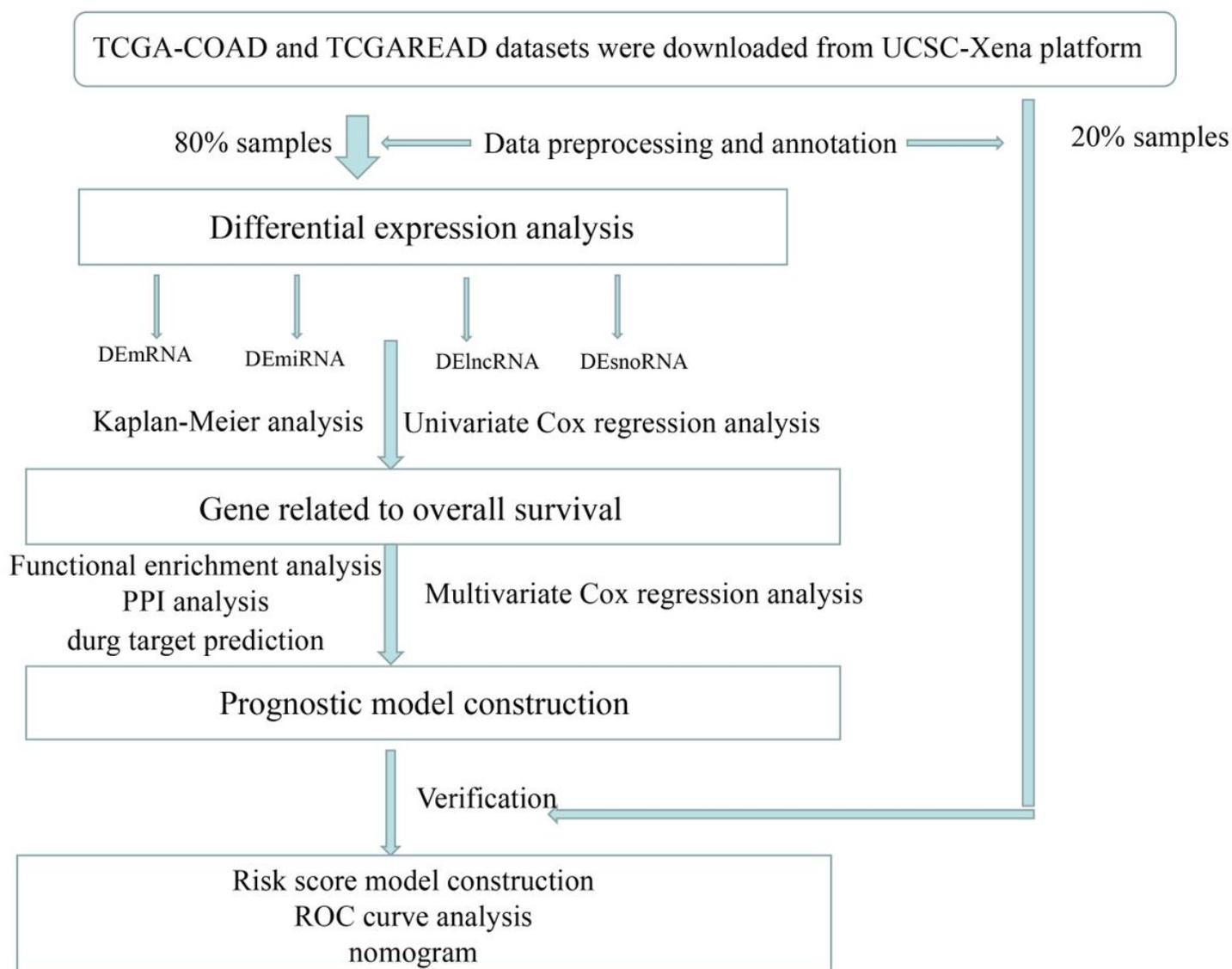


Figure 1

Flow diagram of the analysis procedure COAD: colon adenocarcinoma, READ: rectal adenocarcinoma, OS: overall survival, ROC: receiver operator characteristic The RNA-seq and miRNA-seq datasets of COAD and READ were downloaded from TCGA database, then the DE-RNAsGs between normal and tumor samples were screened. Next, K-M analysis, univariate Cox regression analysis, and multivariate Cox regression analysis were performed to obtain the prognostic independent factors related to OS. Then, the ROC curve was built and the AUC values of the ROC curve were calculated. Furthermore, potential drugs targeting key DE-RNAsGs were also predicted.

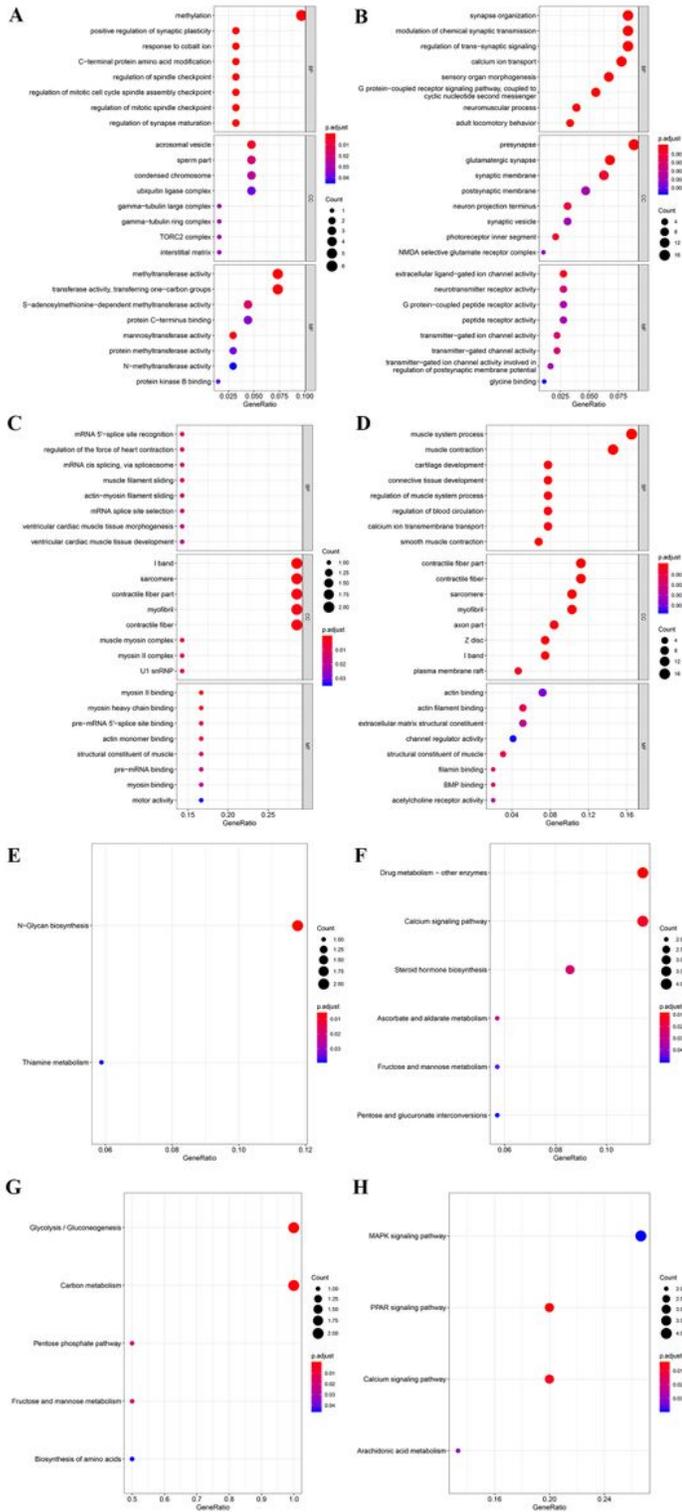


Figure 2

Functional enrichment analysis of survival-related genes. A: GO enrichment analysis of genes in the COAD OS group. B: GO enrichment analysis of genes in the COAD stage group. C: GO enrichment analysis of genes in the READ OS group. D: GO enrichment analysis of genes in the READ stage group. E: KEGG pathway analysis of genes in the COAD OS group. F: KEGG pathway analysis of genes in the COAD stage group. G: KEGG pathway analysis of genes in the READ OS group. H: KEGG pathway analysis of genes in the READ stage group. The Y-axis represents the GO terms and KEGG pathways, the size of each dot represents the gene count, and the color represents the adj-P. The functional enrichment analysis of survival-related genes was conducted using clusterProfiler. GO enrichment analysis showed that survival-related genes were mainly associated with BP terms, such as methylation, synapse organization, mRNA 5'-splice site recognition, and cartilage development. KEGG pathway enrichment analysis indicated that survival-related genes in the COAD groups were involved in N-Glycan biosynthesis, drug metabolism-other enzymes, and calcium signaling pathway; genes in the READ groups mostly participated in glycolysis/gluconeogenesis, carbon metabolism, PPAR signaling pathway, and calcium signaling pathway.

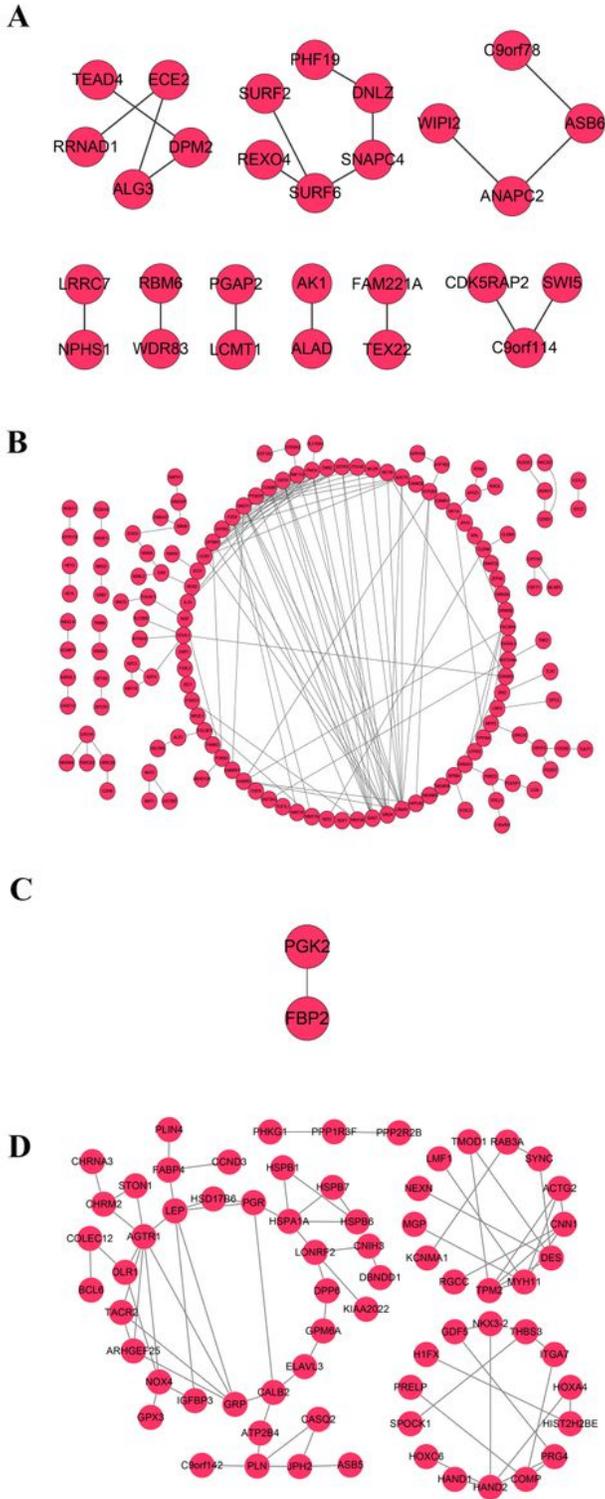


Figure 3

The PPI network of survival-related genes. A: COAD OS group. B: COAD stage group. C: READ OS group. D: READ stage group. PPI analysis was performed using the STRING database. In the COAD OS group, the degrees of SURF6, DPM2, and ALAD were higher than others. In the COAD stage group, GAST, DRD4, and CRHR1 were considered as hub genes. The PPI network of the READ OS group was composed of PGK2 and FBP2. In the READ stage group, AGTR1, HAND2, and CHRM2 had higher degrees than others.

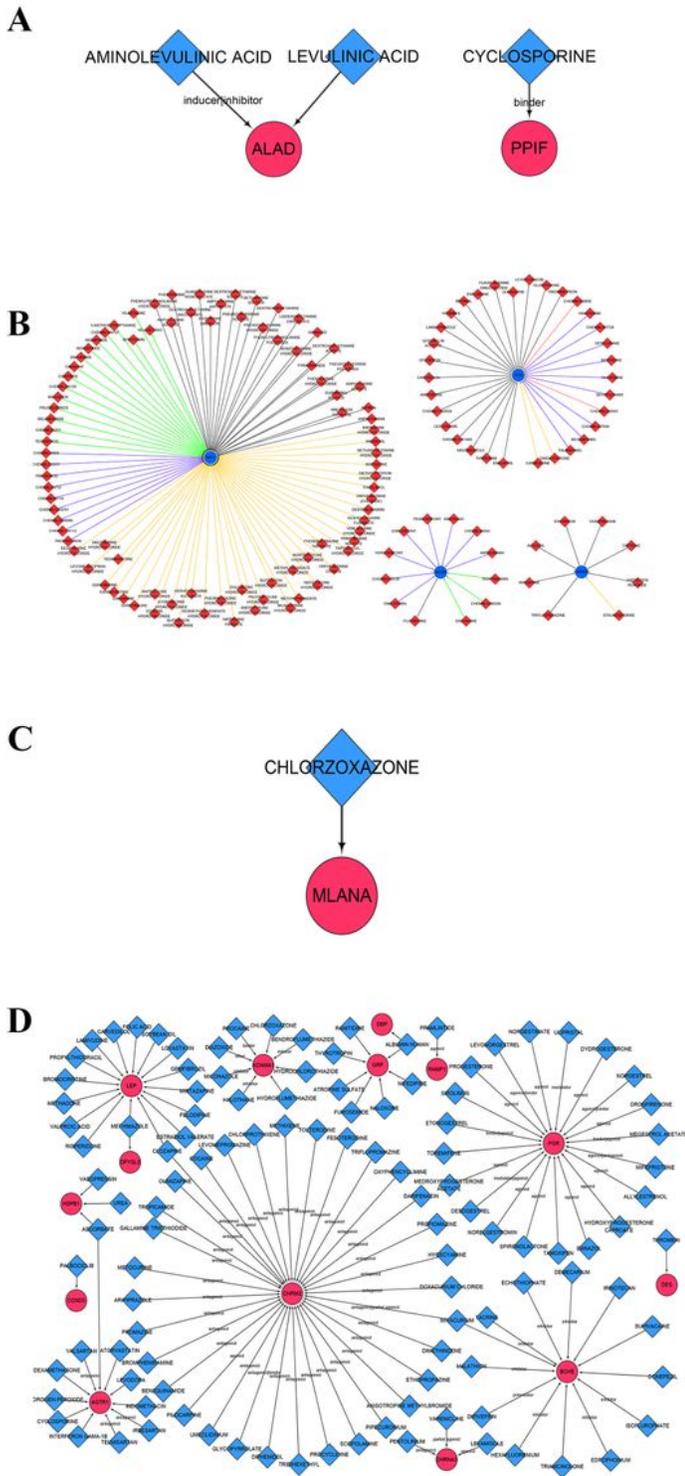


Figure 4

The constructed drug target gene interaction network. A: COAD OS group. B: COAD stage group. C: READ OS group. D: READ stage group. Circle nodes represent genes and diamond nodes represent drugs. The genes obtained in the PPI network were further analyzed by the DGIdb database. ALAD, PPIF, MLANA, NAT1, and CHRM2 may be considered as druggable genes. Aminolevulinic acid, levulinic, chlorzoxazone, and cyclosporine could be potential drugs for CRC treatment.

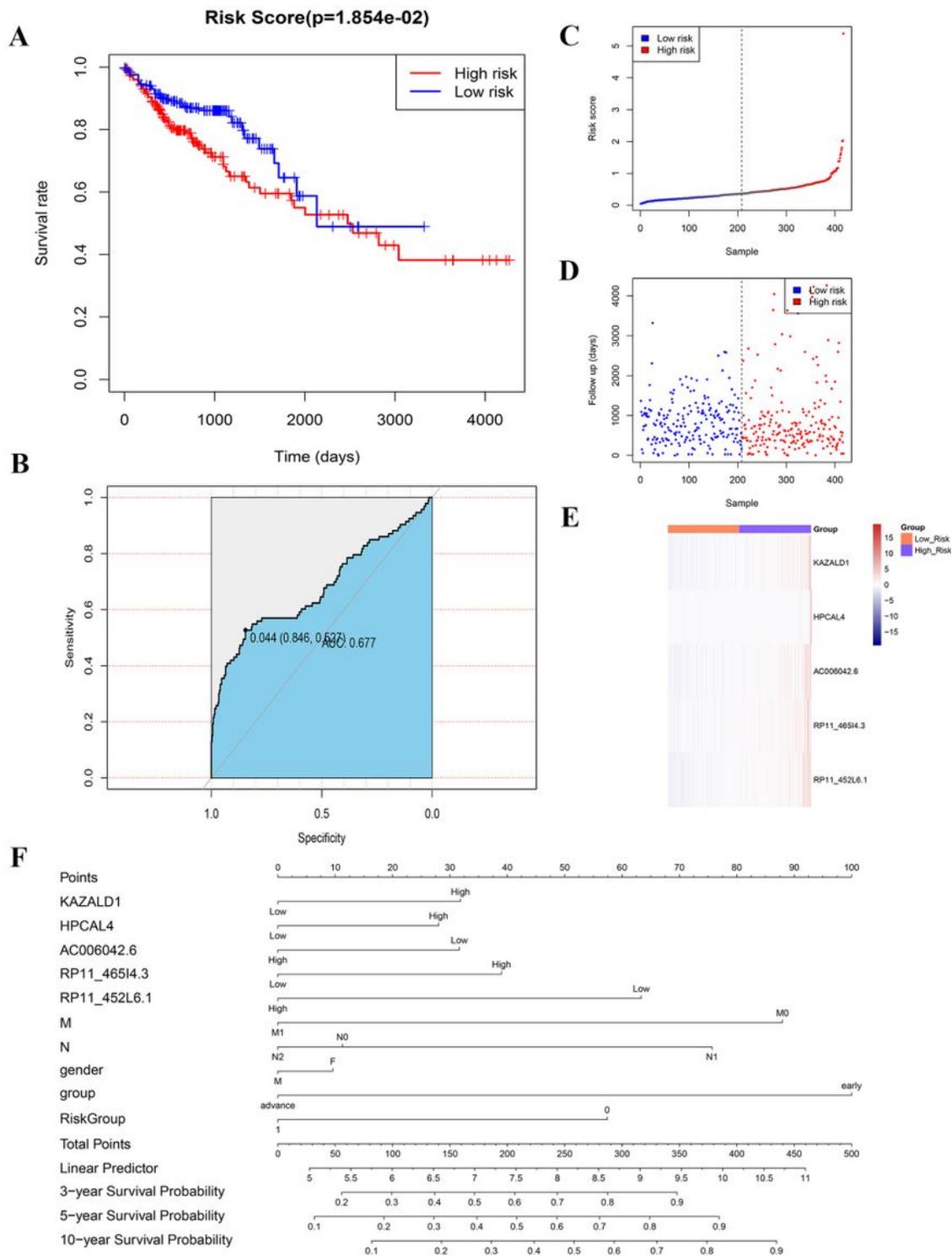


Figure 5

Prognostic risk score model analysis of five prognostic genes in the COAD-OS group. A: The K-M curve for the low-risk group and high-risk group. The abscissa axis indicates time and the vertical axis indicates survival rate. B: The ROC curve of five prognosis genes. The X-axis indicates specificity of the model and the Y-axis indicates sensitivity of the model. C: Risk score distribution. The X-axis represents the number of patients and Y-axis represents the risk score. D: Patient survival status. The X-axis indicates the

number of samples and the Y-axis indicates the follow-up time. E: The expression levels of the five genes. Red represents the low-risk group and blue represents the high-risk group. F: Nomogram was used to predict overall survival (OS) at 3, 5, and 10 years in the COAD-OS group. The total points and the predicted 3, 5, 10 years OS were obtained by adding up the points of each multivariate regression variable. Five genes (HPCAL4, RP11_452L6.1, AC006042.6, RP11_465I4.3, and KAZALD1) were used to build a prognostic model. A total of 417 patients were divided into a high-risk group and low-risk group using the median of the RS, and the high-risk group was connected with poor prognosis. The AUC value of this prognostic model was 0.677. The results revealed that high-risk patients had high mortality and high expression levels of these prognosis genes. These genes may be independent prognostic factors.

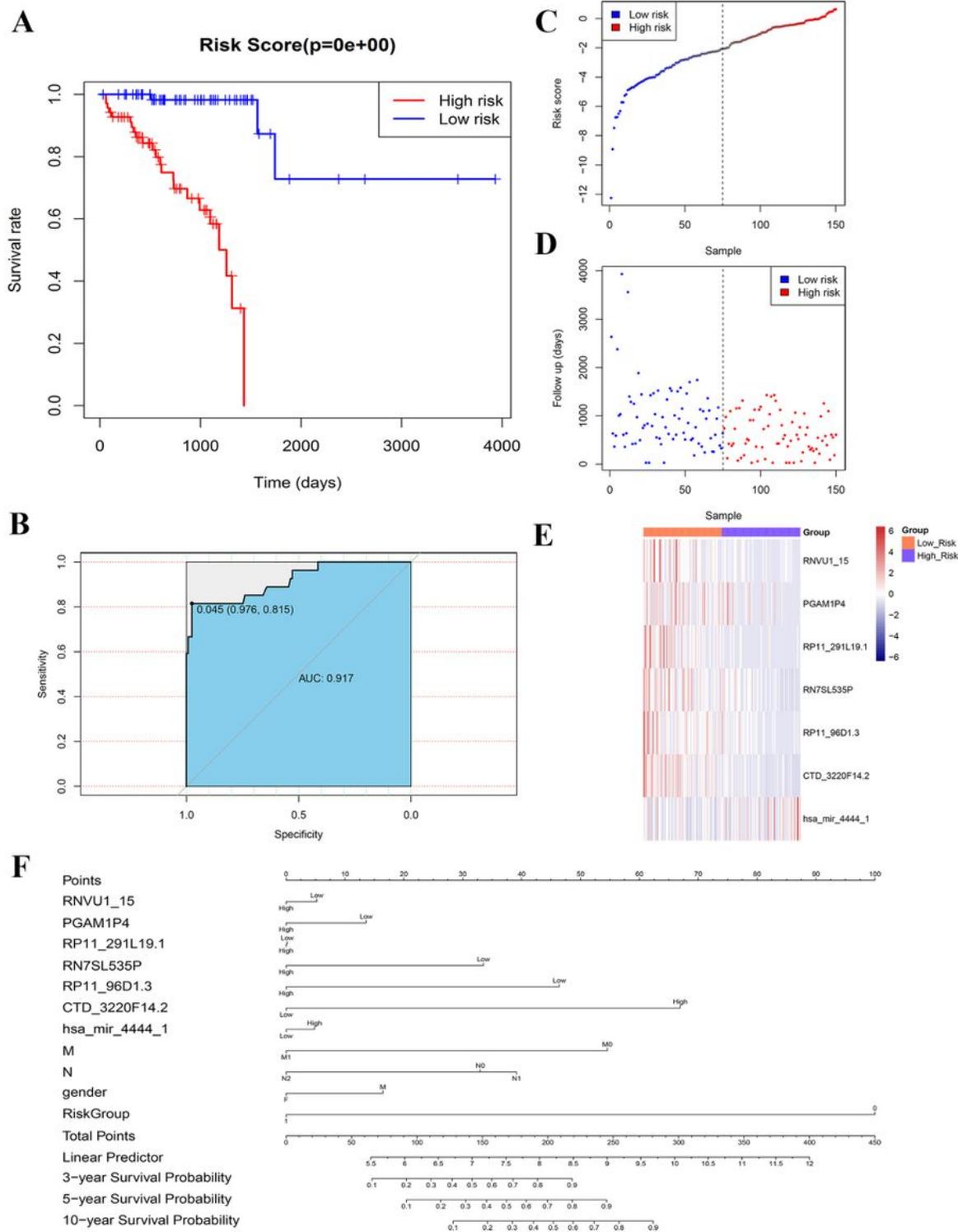


Figure 6

Prognostic model analysis of seven survival-related genes in the READ-OS group. A: K-M curve for the low-risk and high-risk groups. B: ROC analysis of seven prognosis genes. C: Risk score distribution. The X-axis represents the number of patients and the Y-axis represents the risk score. D: Survival status of patients. The abscissa axis represents the number of patients and the vertical axis represents the follow-up time. E: Heat map of expression of the seven genes. F: Nomogram was used to predict overall survival

(OS) at 3, 5, and 10 years in the READ-OS group. RNVU1_15, PGAM1P4, RP11_291L19.1, RN7SL535P, RP11_96D1.3, CTD_3220F14.2, and hsa_mir_4444_1 were used to conduct a risk score model. The 150 READ patients were divided into high-risk and low-risk groups, and patients with low risk score had significantly longer OS than patients with high risk score. The AUC of the prognostic model was 0.814. Additionally, these seven survival-related genes were considered as independent prognostic factors.

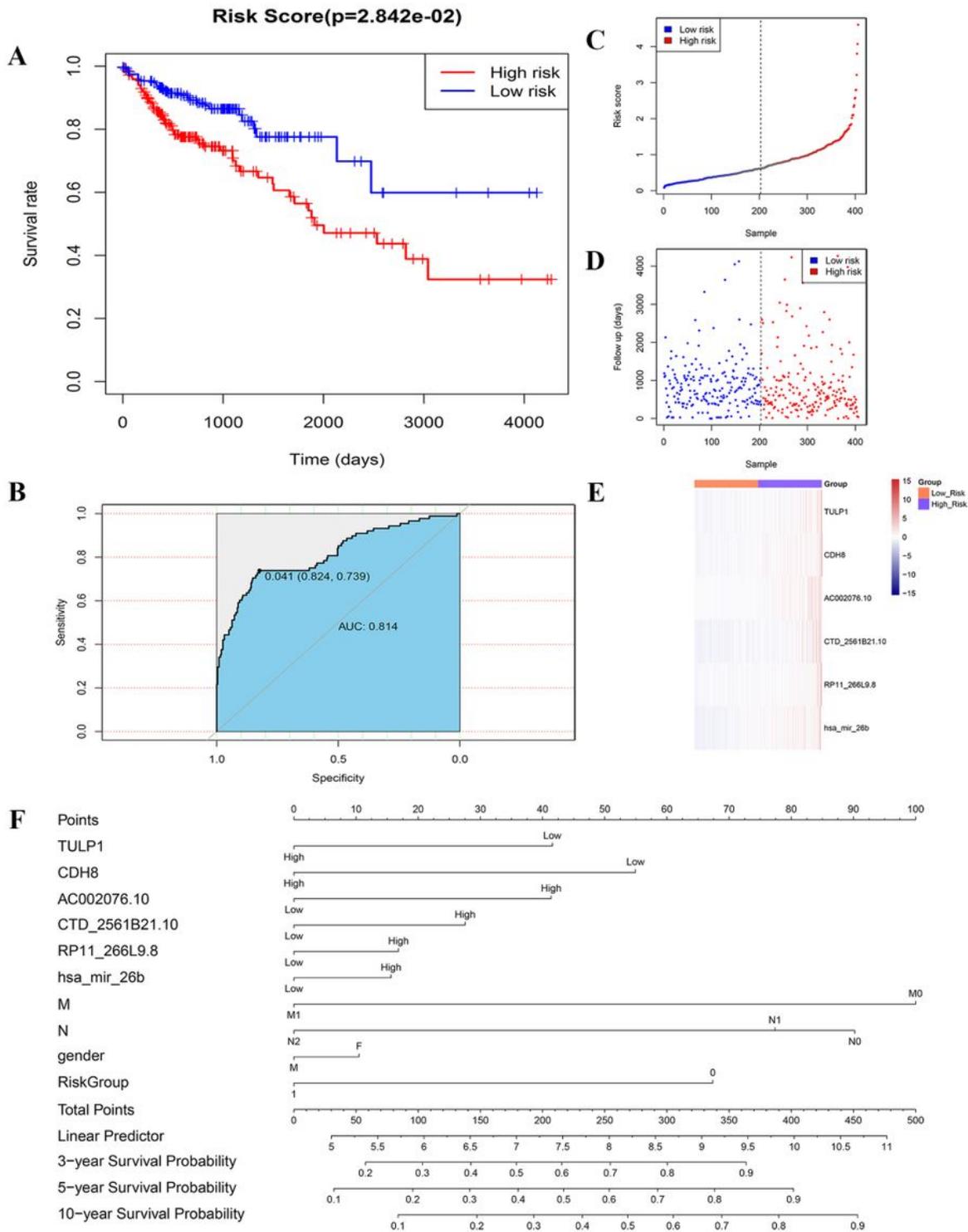


Figure 7

Prognostic model analysis of seven survival-related genes in the COAD-stage group. A: K-M curve for the low-risk and high-risk groups. B: ROC analysis of six prognosis genes. C: Risk score distribution. The X-axis represents the number of patients and the Y-axis represents the risk score. D: Survival status of patients. The abscissa axis represents the number of patients and the Y-axis represents the follow-up time. E: Heat map of expression of the six genes. F: Nomogram for predicting overall survival (OS) at 3, 5, and 10 years in COAD-stage group. TULP1, CDH8, AC002076.10, CTD_2561B21.10, RP11_266L9.8, and hsa_mir_26b, were used to build the prognostic model. A total of 406 patients were divided into high-risk and low-risk groups, and patients in the high-risk group had a poorer prognosis than those in the low-risk group. The AUC of the prognostic model was 0.814. These genes were up-regulated in the high-risk group. Patients in the high-risk group had a higher mortality rate and the expression level of survival-related genes was up-regulated in the high-risk group. These six potential independent prognostic factors were predicted using nomograms.

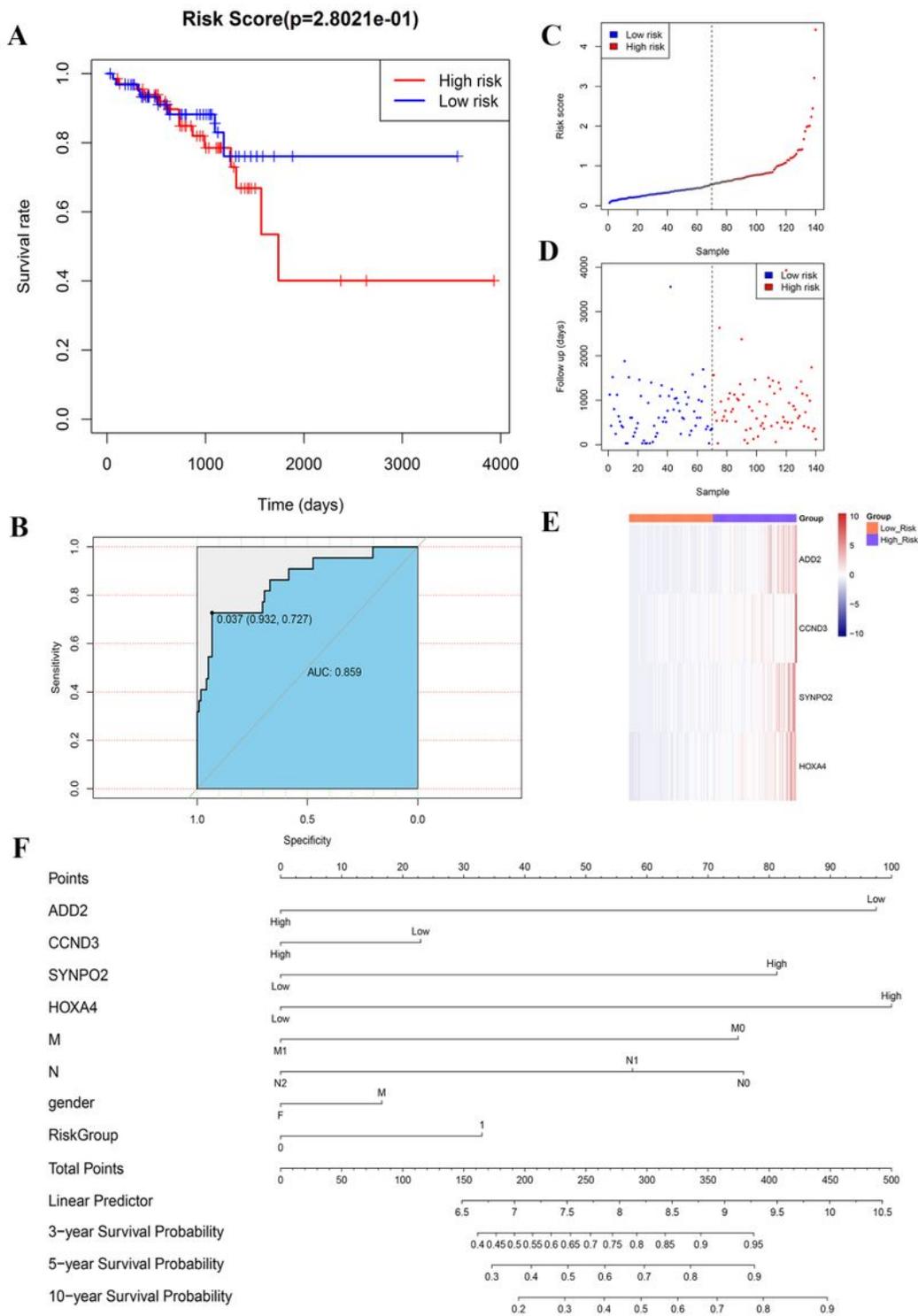


Figure 8

Prognostic model analysis of four survival-related genes in the READ-stage group. A: K-M curve for the low-risk and high-risk groups. B: ROC analysis of prognosis genes. C: Risk score distribution. The X-axis indicates the number of patients and the Y-axis indicates the risk score. D: Survival status of patients. The abscissa axis represents the number of patients and the Y-axis represents the follow-up time. E: Expression heat map of the four genes. F: Nomogram for predicting overall survival (OS) at 3, 5, and 10

years in the READ-stage group. ADD2, CCND3, SYNPO2, and HOXA4 were used to construct a predictive model. A total of 140 patients were divided into high- and low-risk groups, and the low-risk patients had markedly longer OS than those in the high-risk group. The AUC of the prognostic model was 0.859. The high expression level of these genes was observed in the high-risk group. ADD2, CCND3, SYNPO2, and HOXA4 were considered as independent prognostic factors.

Supplementary Files

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

- [SupplementaryTable1.xls](#)
- [SupplementaryFigure1.tif](#)
- [CERTIFICATEOFENGLISHEDITING.pdf](#)
- [SupplementaryTable2.xls](#)
- [SupplementaryTable4.xls](#)
- [SupplementaryTable1.xls](#)
- [SupplementaryTable3.xls](#)
- [SupplementaryFigure1.tif](#)
- [CERTIFICATEOFENGLISHEDITING.pdf](#)
- [SupplementaryTable2.xls](#)
- [SupplementaryTable4.xls](#)
- [SupplementaryTable3.xls](#)