

Systematic Expression Analysis of the Diagnostic and Prognostic Value of *HEPACAM Family Member 2* in Colon Adenocarcinoma

Shuai Wang

Guangxi Medical University First Affiliated Hospital

Guo-Tian Ruan

Guangxi Medical University First Affiliated Hospital

Yi-Zhen Gong

Guilin Medical University Affiliated Hospital

Cun Liao

Guangxi Medical University First Affiliated Hospital

Lei Zhang

Guilin Medical University Affiliated Hospital

Feng Gao (✉ doctor0771@163.com)

Guangxi Medical University First Affiliated Hospital <https://orcid.org/0000-0002-1000-385X>

Research article

Keywords: HEPACAM2, Guangxi cohort, COAD, prognosis, diagnosis

Posted Date: January 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154487/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The diagnostic and prognostic value of HEPACAM family member 2 (*HEPACAM2*) gene in patients with colon adenocarcinoma (COAD) is rarely reported. Therefore, the purpose of this study is to explore the diagnostic and prognostic value of *HEPACAM2* gene in patients with COAD.

Methods: Firstly, we analyzed the differential expression levels of *HEPACAM2* gene and diagnostic value analysis from different databases. Secondly, univariate and multivariate survival analysis of the prognostic value of *HEPACAM2* gene in patients with COAD was performed. Finally, utilizing joint-effects analysis and comprehensive prognosis analysis to investigate the prognostic value of *HEPACAM2* and related genes.

Results: Differential analyses of multiple databases showed that the *HEPACAM2* expression level in COAD tumor tissue was significantly lower than that of adjacent normal tissues. The diagnostic ROC curve results indicated that *HEPACAM2* gene had a higher diagnostic value in COAD. The RT-qPCR verification results of COAD tissue in the Guangxi cohort showed that *HEPACAM2* expression level in COAD tumor tissue was significantly higher than that of adjacent normal tissues ($P < 0.001$), and the diagnostic value was high in COAD (AUC=0.892). The prognostic value analysis showed that low expression of *HEPACAM2* gene had a poorer prognosis of overall survival (OS) in patients with COAD when compared with those patients with high expression of *HEPACAM2* gene ($P = 0.038$, HR (95% CI) = 0.635(0.414-0.976)). Joint-effects analysis and comprehensive prognosis analysis showed that high expression of *HEPACAM2* combined with high expression of *CLCA1*, high expression of *REP15*, and high expression of *B3GNT6* were associated with a better prognosis of overall COAD OS.

Conclusion: The results of this study suggested that the *HEPACAM2* may be an independent diagnostic and prognostic biomarker for COAD.

Introduction

Colorectal cancer (CRC) is a malignant tumor that ranks second in cancer morbidity and third in mortality worldwide.(1) The morbidity and mortality of CRC are increasing globally, with approximately 1.4 million new CRC cases and 700,000 deaths in 2018.(2) Smoking, alcohol consumption, excess weight, consumption of red or processed meat, low dietary fiber or calcium intake are considered risk factors for increased CRC.(3) Colonoscopy and clinicopathological examinations are still the main methods of screening and diagnosis, but these operations are often invasive and could bring certain pain and economic burden to patients.(4) Serum carcinoembryonic antigen (CEA) has a certain prognostic value in CRC, but its low diagnostic value and poor early prognosis are still the main contradictions. (5, 6) At present, the main treatments for CRC are still surgery, radiotherapy, and chemotherapy. When diagnosed with advanced metastatic disease, CRC patients traditionally have a poor prognosis, with a 5-year survival rate of 5%-8%.(7) Therefore, we urgently need to find another novel, accurate, practical oncology biomarker to improve the early diagnosis and prognostic detection of patients with CRC. As an important regulatory molecule, mRNA expression is believed to play an important role in tumor cell survival and tumor progression, and the comparison of gene expression profiles between tumors and adjacent tissues plays an important role in the identification of biomarkers. (8, 9) Colon adenocarcinoma (COAD) is the most common type of CRC,(10) so this study aimed to find potential mRNA gene biomarkers with diagnostic and prognostic value in COAD.

Hepatic and glial cell adhesion molecule (*HEPACAM*) was first identified as a cell adhesion molecule in 2005, belonging to the immunoglobulin superfamily.(11) The *HEPACAM* gene is often down-regulated or lost in various human cancer cells or tissues.(12) It is reported that *HEPACAM* mediates the proliferation, differentiation, and migration of cancer cells.(13) However, as a member of the immunoglobulin superfamily, the HEPACAM family member 2 (*HEPACAM2*) gene has few reports on its role in cancer regulation and management. *HEPACAM2*, also called *MIKI* and *LOC253012*, is located in the chromosome region 7q21.(14) This gene encodes a protein related to the immunoglobulin superfamily and plays a role in mitosis. The poly (ADP-ribosylation) of the encoded protein can promote its translocation to the centrosome, thereby promoting the maturation of the centrosome.(15) It was reported that *HEPACAM2* participated in regulating the cell proliferation, and knockdown of this gene can lead to early and metaphase block of mitosis, abnormal nuclear morphology and apoptosis.(14, 16) It is noteworthy that the expression of *HEPACAM2* was downregulated in colorectal cancer(17–19). Nevertheless, the clinical importance of *HEPACAM2* in COAD remains unclear. This study is the first to investigate and explore the diagnostic and prognostic value of *HEPACAM2* gene in COAD.

Material And Methods

Public data collection

The *HEPACAM2* gene mRNA expression dataset and consistent clinicopathological parameter information were obtained from the global public cancer database of TCGA (<https://tcga-data.nci.nih.gov/>) (20). Additionally, we also obtained the co-expression gene expression dataset related to *HEPACAM2* gene in COAD from the TCGA database.

Differential expression analysis and diagnostic ROC Curve Analysis based on TCGA cohort and public networks

Firstly, we downloaded the related information of *HEPACAM2* gene in different cancers from UALCAN (<http://ualcan.path.uab.edu/index.html>)(21) and Tumor Immune Estimation Resource (TIMER, <https://cistrome.shinyapps.io/timer/>)(22) database. To further explore the differential expression of the *HEPACAM2* gene in COAD, we also obtained the different *HEPACAM2* gene expression data from the public databases of Gene Expression

Profiling Interactive Analysis (GEPIA; <http://gepia.cancer-pku.cn/index.html>)(23) and Metabolic gEne Rapid Visualizer (MERAV; <http://merav.wi.mit.edu/>)(24). Additionally, we also investigated and performed the differential expression of *HEPCAM2* gene in COAD cancer tissues and adjacent normal tissues, and analyzed the expression level of *HEPCAM2* gene in COAD tumor tissues under different tumor TNM stages based on the TCGA database. Finally, based on the TCGA database, we performed a receiver operating characteristic (ROC) curve to analyze the diagnostic value of the *HEPCAM2* gene in COAD tumor tissues and adjacent normal tissues.

Mining and validation of the differential expression and diagnostic value of *HEPCAM2*

Mining cohort

The mRNA expression of *HEPCAM2* in COAD or CRC was analyzed within the Oncomine dataset (<https://www.oncomine.org/>)(25), which based on the Gene Expression Omnibus (GEO). The differential expression scatter plot and diagnostic ROC curve of *HEPCAM2* gene were performed in two studies, including Skrzypczak COAD (26) and Hong CRC(27).

Validation cohort

We collected tumor tissue and adjacent normal tissue samples from 30 COAD patients from the Colorectal Surgery Department of the First Affiliated Hospital of Guangxi Medical University. Inclusion criteria: 1. The patient was older than 18 years; 2. Complete clinical parameter information; 3. The postoperative diagnosis was COAD; 4. Sign the consent form. Exclusion criteria: 1. The patient was under 18 years old; 2. Incomplete clinical parameter information; 3. Disagree to sign the consent form. The tissue samples were immersed in RNA protection solution immediately after the operation, and stored in a refrigerator at -80 degrees refrigerator according to the protection solution storage instructions. Tissue total RNA was extracted from the tissue, cDNA was synthesized through reverse transcription, and PCR reaction was performed afterwards. The reaction conditions were: 1 cycle at 95°C for 10 min; 40 cycles at 95°C for 15 s, 60°C for 1 min and 95°C for 30 s; 1 cycle at 95°C for 15 s, 60°C for 1 min, 95°C for 30 sec, 60°C for 15 s. The primer sequences were: *HEPCAM2* Forward 5'-TGCCACCCAATGCATCTCTGCT-3', *HEPCAM2* Reverse 5'-TCTGCACCACTGGCTTTGTGAC-3'; β -actin Forward 5'-CATGTACGTTGCTATCCAGGC-3', β -actin Reverse 5'-CTCCTTAATGTCACGCACGAT-3'. The final result was calculated using $2^{-\Delta\Delta Cq}$ (28) This study followed the Herki Declaration and was approved by the Ethics Association of the First Affiliated Hospital of Guangxi Medical University [No.: 2019(KY-E-001)], all patients signed an informed consent form.

Univariate and multivariate survival analysis of *HEPCAM2* Gene in COAD

The expression of *HEPCAM2* gene in COAD was divided into two groups according to the median cut-off value of gene expression level, namely, high- and low- expression groups. The Kaplan-Meier survival curve was performed to evaluate the survival prognosis of *HEPCAM2* gene in patients with COAD. We then constructed different prognostic COX risk models according to the different clinical parameters, the HR risk curve and multivariate survival analysis was investigated to explore the prognostic value of *HEPCAM2* gene in COAD. Finally, we constructed the *HEPCAM2* gene-related nomogram to estimate the prognosis risk in COAD.

Collection of *HEPCAM2* gene mutation and immune infiltration information of COAD

The mutation status and mutation frequency of *HEPCAM2* related to COAD were obtained from the cBio Cancer Genomics Portal (cbioportal, <https://www.cbioportal.org/>) database.(29) Additionally, we obtained the potential connection between *HEPCAM2* gene and different immune infiltrating cells through the TIMER database.

Correlation analysis and correlation-genes survival analyses in COAD patients

We collected COAD gene sets associated with *HEPCAM2* gene from different datasets, including GEPIA, ULCAN, and LinkedOmics (<http://www.linkedomics.org/>)(30) dataset, and drew a Venn diagram to select the intersection genes of the three datasets. Then, we also performed a correlation analysis of intersection genes based on the TCGA cohort. The intersection genes were also separated into two compartments based on median cut-off value, the univariate and multivariate survival analysis to evaluate the prognostic value of these genes in COAD.

Joint-effects analysis and comprehensive prognosis analysis

The joint-effects survival analysis of correlation genes was investigated the combination effects in COAD patients' survival, the survival curves and multivariate survival analyses were also utilized to evaluate the prognostic value in COAD. Meanwhile, we also conducted different prognostic nomograms based on joint-effects survival results. Finally, we constructed different prognostic risk scored models based on joint-effects survival analysis. The formula for calculating the prognostic risk score was: Risk score = gene₁ expression × β_1 gene₁ + gene₂ expression × β_2 gene₂ + ... Gene_n expression × β_n Gene_n. The result of β came from the coefficients of the multivariate COX regression risk model. (31) COAD patients were divided into different risk models according to the level of median risk score values. The prognostic time-dependent ROC curves were performed by R software to estimate the predictive accuracy in COAD patients.

Statistical analysis

The comparison of *HEPACAM2* gene expression in COAD tumor tissues and adjacent normal tissues in the TCGA cohort was performed by unpaired Student's t-test, and the validation cohort of RT-PCR result was validated by paired t-test, and the mean \pm Standard deviation represented gene expression level. The multivariate Cox risk model was adjusted by different models, namely, model 1: Unadjusted model; model 2: Adjusted by TNM stage; model 3: Adjusted by age, sex, and TNM stage. The results were presented with hazard ratio (HR), 95% confidence interval (CI) and *P*-value. All results in this study were considered to be statistically different at *P*<0.05. All statistical calculations were performed by SPSS 25.0 (IBM, New York, NY, USA) and R software, version 4.0.3 (<http://www.R-project.org/>).

Results

Data resource processing

We downloaded a total of 461 COAD patients' clinical data and 456 gene expression levels (including 480 cancer tissue expression levels and 41 adjacent normal tissue expression levels) from the TCGA database. We integrated the two data to obtain 438 cases of COAD tumor tissue and 41 cases of adjacent normal tissue (removal of no prognostic information, mismatched information, and repeated expression of cancer tissue). The univariate survival analysis of clinical parameters was shown in Table 1. The results showed that the TNM stage was correlated with the OS of COAD patients (Log-rank *P*< 0.001).

Table 1
Baseline patient characteristics in TCGA cohort

Variables	Patients	OS		HR (95%CI)	Log-rank <i>P</i>
	(n = 438)	No. of events	MST (days)		
Age(years)					
≤ 65	268	57	3042	1	0.196
> 65	168	41	1910	1.305(0.871–1.965)	
Missing*	2				
Sex					
Male	234	54	2475	1	0.545
Female	204	44	NA	0.884(0.593–1.318)	
TNM Stage					
I	73	4	NA	1	< 0.001
II	168	28	2821	2.308(0.807–6.602)	
III	126	31	NA	4.101(1.446–11.634)	
IV	61	31	858	11.355(4.003–32.208)	
Missing#	10				
Notes: Missing*, information of age was unknown in 2 patients; Missing#, information of TNM stage was not reported in 10 patients; TCGA, The Cancer Genome Atlas; OS, overall survival; MST, median survival time; 95 % CI, 95 % confidence interval; HR, hazards ratio; NA, not available.					

Differential expression analysis and diagnostic ROC Curve Analysis based on TCGA cohort and public networks

We downloaded the *HEPACAM2* gene expression level in various cancer tissues and normal tissues from the UALCAN and TIMER databases. The results showed that the expression level of *HEPACAM2* gene in COAD tumor tissue was lower than that in normal colon tissue. (Fig. 1) We further downloaded the expression box diagrams of *HEPACAM2* gene in COAD tissue and normal colon tissue from the GEPIA and MERAV databases, and the results were consistent with the previous description. (Figure S1A-1B) Meanwhile, the expression level of *HEPACAM2* gene in COAD patients of different TNM stages did not reach statistical differences (*P*> 0.05). (Figure S1C)

Based on the TCGA database, we also investigated the differential expression of *HEPACAM2* gene between COAD tumor tissues and adjacent normal tissues, the result showed that the *HEPACAM2* expression level was higher in COAD adjacent normal tissues than that in tumor tissues. We also found that the *HEPACAM2* expression level didn't show differential expression in different TNM stages. (Fig. 2A) The diagnostic ROC curve showed that *HEPACAM2* had a higher diagnostic value in patients with COAD (*P*< 0.001, Area Under Curve (AUC) = 0.940, 95%CI = 0.805–0.979). (Fig. 2B)

Finally, we also investigated the differential expression and diagnostic value of *HEPACAM2* gene in COAD or CRC using an Oncomine dataset based on GEO cohort, the result showed that the *HEPACAM2* gene was highly expressed in Skrzypczak COAD or Hong CRC normal tissue than that in tumor tissues, and the *HEPACAM2* gene had a high diagnostic value in Skrzypczak COAD ($P < 0.001$, AUC = 0.896, 95%CI = 0.812–0.980) and Hong CRC ($P < 0.001$, AUC = 0.976, 95%CI = 0.944–1.000). (Fig. 3)

Validation and analysis of *HEPACAM2* in the diagnostic value of COAD based on the Guangxi cohort

We collected 30 pairs of COAD patients' tumor tissues and adjacent normal tissues. After RT-qPCR detection, we found that the expression level of *HEPACAM2* gene in COAD tumor tissue (0.036942 ± 0.062463) was significantly lower than that in adjacent normal colon tissue (0.167750 ± 0.179779). ($P < 0.001$) Meanwhile, the diagnostic ROC curve suggested that the *HEPACAM2* gene had a higher diagnostic value in patients with COAD ($P < 0.001$, AUC = 0.892, 95%CI = 0.805–0.979). (Fig. 2C-2E)

Univariate and multivariate survival analysis of *HEPACAM2* gene in COAD

We performed a survival analysis of *HEPACAM2* gene in patients with COAD in accordance with the median cutoff value of *HEPACAM2* gene expression, the patients with high expression of *HEPACAM2* gene had better survival than those with lowly expressed *HEPACAM2* gene. (Fig. 4A) In terms of the survival results of univariate clinical parameters, we constructed two different adjusted models and found that the *HEPACAM2* gene was related to the OS of patients with COAD, namely model 2: Adjusted by TNM stage ($P = 0.044$, HR (95% CI) = 0.643(0.419–0.988)) and model 3: Adjusted by age, sex, and TNM stage ($P = 0.038$, HR (95% CI) = 0.635(0.414–0.976)). (Table 2) Finally, the association between *HEPACAM2* gene and death risk of COAD patients was presented in Fig. 5A-5C. In short, the death risk of COAD decreased with the increased expression of *HEPACAM2* gene. The *HEPACAM2* gene-related nomogram showed that *HEPACAM2* gene made a certain contribution to COAD OS. (Figure S2 A)

Table 2
Multivariate survival analysis of *HEPACAM2* gene and *HEPACAM2* related genes expression in COAD of TCGA cohort

Gene	Patients (n = 438)	OS (model 0)		OS (model 1)		OS (model 2)	
		HR (95%CI)	Crude P^*	HR (95%CI)	Adjusted $P^\#$	HR (95%CI)	Adjusted $P^\&$
<i>HEPACAM2</i>			0.006		0.044		0.038
Low	219	1		1		1	
High	219	0.560(0.370–0.846)		0.643(0.419–0.988)		0.635(0.414–0.976)	
<i>CLCA1</i>			0.001		0.008		0.006
Low	219	1		1		1	
High	219	0.499(0.330–0.755)		0.561 (0.366–0.860)		0.550 (0.358–0.844)	
<i>REP15</i>			0.023		0.046		0.042
Low	219	1		1		1	
High	219	0.623(0.414–0.938)		0.651(0.427–0.992)		0.654(0.423–0.984)	
<i>B3GNT6</i>			0.003		0.005		0.004
Low	219	1		1		1	
High	219	0.528(0.348–0.801)		0.540(0.352–0.830)		0.535(0.348–0.823)	

Notes: Crude P^* , Univariate survival analysis; Adjusted $P^\#$, Adjustment for TNM stage; Adjusted $P^\&$, Adjustment for Age, Sex, and TNM stage; TCGA, The Cancer Genome Atlas; OS, Overall survival; COAD, Colon adenocarcinoma; HR, hazard ratio; CI, confidence interval; TNM, Tumor Node Metastasis; NA, not available.

Table 3
Joint effects analysis of *HEPACAM2* and related genes expression in COAD patients OS

Group	<i>HEPACAM2</i>	Related genes	Patients (n = 438)	OS (model 0)		OS (model 1)		OS (model 2)	
				HR (95% CI)	Crude <i>P</i> *	HR (95% CI)	Adjusted <i>P</i> #	HR (95% CI)	Adjusted <i>P</i> &
<i>HEPACAM2&CLCA1</i>									
A	Low	Low	117	1		1		1	
B	Low	High	45	0.654(0.333–1.286)	0.218	0.694(0.351–1.372)	0.294	0.724(0.365–1.439)	0.357
C	High	Low	42	0.752(0.358–1.581)	0.452	0.832(0.392–1.765)	0.631	0.905(0.426–1.924)	0.796
D	High	High	174	0.433(0.270–0.695)	0.001	0.499(0.304–0.819)	0.006	0.489(0.298–0.803)	0.005
<i>HEPACAM2&REP15</i>									
A	Low	Low	183	1		1		1	
B	Low	High	39	0.927(0.493–1.743)	0.814	0.856(0.453–1.619)	0.633	0.864(0.448–1.666)	0.663
C	High	Low	36	0.685(0.311–1.510)	0.348	0.815(0.367–1.809)	0.615	0.823(0.365–1.856)	0.639
D	High	High	180	0.515(0.323–0.820)	0.005	0.568(0.350–0.923)	0.022	0.560(0.344–0.910)	0.019
<i>HEPACAM2&B3GNT6</i>									
A	Low	Low	184	1		1		1	
B	Low	High	35	0.724(0.365–1.439)	0.357	0.413(0.173–0.984)	0.046	0.397(0.166–0.947)	0.037
C	High	Low	35	0.905(0.426–1.924)	0.796	0.698(0.314–1.551)	0.377	0.653(0.291–1.462)	0.300
D	High	High	184	0.489(0.298–0.803)	0.005	0.544(0.340–0.870)	0.011	0.534(0.333–0.856)	0.009
Notes: Crude <i>P</i> *, Univariate survival analysis; Adjusted <i>P</i> #, Adjustment for TNM stage; Adjusted <i>P</i> &, Adjustment for Age, Sex, and TNM stage; TCGA, The Cancer Genome Atlas; OS, Overall survival; COAD, Colon adenocarcinoma; HR, hazard ratio; CI, confidence interval; TNM, Tumor Node Metastasis; NA, not available.									

HEPACAM2 gene mutation and immune infiltration information of COAD

We investigated the mutation status of *HEPACAM2* gene in patients with COAD and found that mutation frequency was low and genomic alterations occurred in COAD patients. (Fig. 6A-6B) Additionally, we utilized the TIMER dataset to analyze possible correlations between *HEPACAM2* gene and immune infiltration of different Immune cells in COAD. The result showed that there was no significant and positive association between *HEPACAM2* gene and different Immune cells. (Fig. 6C)

Correlation analysis and correlation-genes survival analyses in COAD patients

We selected correlation genes that were related to the *HEPACMA2* gene in COAD. The correlation coefficients of the genes were greater than 0.7 in the three datasets, including GEPIA, ULCAN, and LinkedOmics databases. The correlation analysis by Venn diagram showed that *HEPACAM2* gene was strongly associated with chloride channel accessory 1 (*CLCA1*) gene, RAB15 effector protein (*REP15*) gene, and UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 6 (*B3GNT6*) gene. (Figure 7 and Figure 8) The survival curves showed that the low expression of *CLCA1* gene, *REP15* gene, and *B3GNT6* gene had worse survival in COAD. (Figure 4B-4D) Finally, the multivariable survival analyses indicated that the expression of *CLCA1* gene (Model 1: *P* = 0.008, HR (95% CI) = 0.561 (0.366-0.860); Model 2: *P* = 0.006, HR (95% CI) = 0.550 (0.358-0.844)), *REP15* gene (Model 1: *P* = 0.046, HR (95% CI) = 0.651(0.427-0.992); Model 2: *P* = 0.044, HR (95% CI) = 0.654(0.423-0.984)), and *B3GNT6* gene (Model 1: *P* = 0.005, HR (95% CI) = 0.540(0.352-0.830); Model 2: *P* = 0.004, HR (95% CI) = 0.535(0.348-0.823)) were significantly associated with the OS of patients with COAD.

Joint-effects analysis and comprehensive prognosis analysis

We carried out the survival analysis of each combination with *HEPACAM2* gene. (Figure 5D-5F) Highly expressed *HEPACAM2* gene combined with highly expressed *CLCA1* gene (Model 1: $P = 0.006$, HR (95% CI) = 0.499 (0.304-0.819); Model 2: $P = 0.005$, HR (95% CI) = 0.489 (0.298-0.803)) or *REP15* gene (Model 1: $P = 0.022$, HR (95% CI) = 0.568 (0.350-0.923); Model 2: $P = 0.019$, HR (95% CI) = 0.560 (0.344-0.910)) or *B3GNT6* gene (Model 1: $P = 0.011$, HR (95% CI) = 0.544 (0.340-0.870); Model 2: $P = 0.009$, HR (95% CI) = 0.534 (0.333-0.856)) predicted longer OS of COAD patients. In other words, the combination of these highly expressed genes was associated with a reduced risk of death in COAD. However, we also observed that the combination of low expression of *HEPACAM2* gene and high expression of *B3GNT6* associated with a better prognosis of COAD OS (Model 1: $P = 0.046$, HR (95% CI) = 0.413 (0.173-0.984); Model 2: $P = 0.037$, HR (95% CI) = 0.397 (0.166-0.947)).

The nomograms of *HEPACAM2* and *CLCA1*, *HEPACAM2* and *REP15*, *HEPACAM2* and *B3GNT6* showed that these different combinations displayed a higher prognostic contribution to COAD OS than the only *HEPACAM2* gene-related nomogram. (Figure S2B- S2D)

Finally, the risk scored model of *HEPACAM2* and *CLCA1* was constructed by the following formula: $(-0.453 \times \text{HEPACAM2 expression}) + (-0.598 \times \text{CLCA1 expression})$. The risk scored model of *HEPACAM2* and *REP15* was as follows: $(-0.453 \times \text{HEPACAM2 expression}) + (-0.615 \times \text{REP15 expression})$. The risk scored model of *HEPACAM2* and *B3GNT6* was as follows: $(-0.453 \times \text{HEPACAM2 expression}) + (-0.429 \times \text{B3GNT6 expression})$. All nomograms we constructed showed that low risk scored group had better survival and fewer deaths of COAD OS. The prognosis 1-, 3-, and 5-year AUC of *HEPACAM2* and *CLCA1* was 0.547, 0.561 and 0.635, respectively. The prognosis 1-, 3-, and 5-year AUC of *HEPACAM2* and *REP15* was 0.559, 0.582 and 0.650, respectively. The prognosis 1-, 3-, and 5-year AUC of *HEPACAM2* and *B3GNT6* was 0.550, 0.569 and 0.619, respectively. (Figure 9)

Discussion

Although with the advancement of medical technology, surgical techniques, and chemotherapy, the treatment of COAD has made great progress, the mortality rate of COAD remains at a high level,(32) which may be due to insufficient understanding of its underlying mechanism and lack of effectiveness of biomarkers. Previous studies have investigated and found that dysregulated genes may play an important role in cancer and have great potential as biomarkers.(33) However, little is known about the potential clinical value of *HEPACAM2* in patients with COAD. Thus, our study aimed to investigate and explore the value of *HEPACAM2* gene expression in the diagnosis and prognosis of COAD.

Data in multiple databases showed that the expression of *HEPACAM2* gene in tumor tissues of patients with COAD was significantly lower than that of adjacent normal COAD tissues ($P < 0.05$). We also investigated the diagnostic value of *HEPACAM2* gene expression in COAD on the TCGA data cohort (AUC = 0.940, 95%CI = 0.805–0.979) and GEO data cohort (Skrzypczak COAD (26), AUC = 0.896, 95%CI = 0.812–0.980) showed the high diagnostic value of *HEPACAM2* gene in COAD, and the data of Hong CRC (27) (AUC = 0.976, 95%CI = 0.944-1.000) indirectly reflected the high diagnostic value of *HEPACAM2* gene in COAD. Then, we validated the diagnostic value of *HEPACAM2* gene in COAD based on the Guangxi cohort and found that the expression of *HEPACAM2* gene in adjacent normal COAD tissues was significantly higher than that in COAD tumor tissues, and it also had a higher diagnostic value in COAD (AUC = 0.892, 95 %CI = 0.805–0.979). Finally, when compared with the low expression of *HEPACAM2*, the survival prognosis analysis found that the highly expressed *HEPACAM2* gene had a better prognosis and the reduced risk of death in patients with COAD based on different adjusted models. Therefore, we hypothesized that the *HEPACAM2* gene might be an independent diagnostic and prognostic biological indicator in patients with COAD.

In the study of Huang et al., they obtained a total of 116 normal colorectal mucosal tissues and 930 tumor tissues of patients with CRC from the TCGA and GEO (GSE71187, GSE21815, and GSE21510) databases. The Cox proportional hazards stepwise regression analysis was used to search out the prognostic genes and found that high expression of *HEPACAM2* decreased the death risk of CRC (HR = 0.937, $P = 0.017$). (18) Another study by Wu et al., five datasets (GSE39582, GSE4183, GSE10714, GSE8671, and GSE14333) from GEO were selected to identify differentially expressed genes (DEGs) associated with the colorectal normal mucosa–adenoma-carcinoma sequence, the result showed that the expression of *HEPACAM2* gene showed a downward trend with tumor progression. The highly expressed *HEPACAM2* showed a better prognostic value in patients with CRC ($P = 0.00029$). They also found that *HEPACAM2* participated in the biology process of mitotic nuclear division and cell division.(19) A similar result was also found in Klopffleisch et al.'s study, the expression level of *HEPACAM2* gene was up-regulated in the benign stage of tumor and adenoma, such as the small intestinal mucosa and colon. Additionally, *HEPACAM2* gene might be involved in cell-cell adhesion and tumor metastasis. (34) It seems that *HEPACAM2* gene may play as a potential tumor suppressor candidate.

Our gene mutation analysis found that the mutation rate of *HEPACAM2* gene in COAD was low. Additionally, as we have known that immune cells infiltrating in the tumor microenvironment played an important role in the initiation and progression of the tumor through directly contacting with tumor cells to promote or suppress tumor cell growth.(35) But the *HEPACAM2* gene didn't show a significantly positive relationship with immune cells infiltrating. We obtained the genes associated with the *HEPACAM2* gene through different data websites and found that *CLCA1*, *REP15* and *B3GNT6* genes were strongly related to *HEPACAM2* gene in COAD. Based on the TCGA database, multivariate survival analysis found that *CLCA1*, *REP15* and *B3GNT6* genes had statistical significance in the survival of COAD patients. After the combined survival analysis and comprehensive survival prognostic analysis, it was found that *HEPACAM2* gene combined with *CLCA1* gene, *HEPACAM2* gene combined with *REP15* gene, and *HEPACAM2* gene combined with *B3GNT6* gene could improve the survival prediction of patients with COAD.

CLCA1 gene is expressed in the intestinal epithelium.(36) This gene may act as a tumor suppressor and play an important function in regulating the differentiation and inhibiting the proliferation in colon cell lines.(37) Previous studies reported that low expression of *CLCA1* was associated with a poorer OS in patients with CRC and related to tumorigenesis, metastasis, and high chromosomal instability. (38, 39) The protein encoded by the *REP15* gene interacts with Rab15 bound to GTP and participates in the circulation of the transferrin receptor from the endocytic circulation chamber to the cell surface.(40) A study by Xu et al. identified that *REP15* could serve as a novel prognostic gene for CRC and low expression of *REP15* was associated with unfavorable prognosis of CRC, besides, they also found that the high expression of *REP15* was positively correlated with the p53 pathway, and negatively correlated with glycerophospholipid metabolism, hedgehog and insulin pathways.(41) The gene *B3GNT6* encodes a glycosyltransferase, which is capable of adding progressive carbohydrates to form a core 3 O-glycan structure and only appears in specific tissues, such as the colon. It has been reported that the core 3 structure in colon cancer tissue decreases with the decrease of core 3 synthase activity.(42) Besides, B3GNT6 was down-expressed in the grade of CRC progress and in tumor metastasis when compared with normal tissues.(43, 44) These prognostic genes seem to have changes in their expression during the occurrence and development of tumors. When these genes are combined with the *HEPACAM2* gene, they can improve the survival prediction ability of patients with COAD. Of course, these assumptions still need to be further verified in the future.

Of course, this research still has some shortcomings. Firstly, the clinical parameter information we obtained from public databases is still not perfect. Secondly, the diagnostic and prognostic value of *HEPACAM2* still needs to be further explored in vivo and in vitro. Finally, multi-center and larger samples are still needed to verify our findings.

Conclusions

This study explored the diagnostic and prognostic value of *HEPACAM2* gene in patients with COAD for the first time. The results of multiple data cohorts and the Guangxi cohort found that *HEPACAM2* gene had a higher diagnostic value in COAD, and the results of multivariate COX survival analysis showed that low expression of *HEPACAM2* gene had a poorer prognosis in patients with COAD when compared with those patients with high expression of *HEPACAM2* gene. *HEPACAM2* gene may be an independent biomarker for diagnosis and prognosis in patients with COAD. However, these findings need to be verified in the future.

Declarations

Ethics approval and consent to participate

All patients signed an informed consent form, and the experimental protocol was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University. (No. 2019(KY-E-001))

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was sponsored in part by The Project to Improve the Basic Research Ability of Young and Middle-aged Teachers in Universities (2020KY12026) and Innovation Project of Guangxi Graduate Education (YCBZ2020048). 2020KY12026 and YCBZ2020048 play roles in the analysis, interpretation of data, and writing of the manuscript.

Acknowledgements

The authors would like to acknowledge the support of the National Key Clinical Specialty Programs (General Surgery and Oncology) and the Key Laboratory of Early Prevention and Treatment for Regional High-Incidence-Tumor (Guangxi Medical University), Ministry of Education, China. The authors also thank the contributors of The Cancer Genome Atlas (<https://cancergenome.nih.gov/>).

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. Kalmar A, Nagy ZB, Galamb O, Csabai I, Bodor A, Wichmann B, et al. Genome-wide expression profiling in colorectal cancer focusing on lncRNAs in the adenoma-carcinoma transition. *BMC Cancer*. 2019;19(1):1059.
3. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin*. 2018;68(1):31-54.
4. Zhao QQ, Jiang C, Gao Q, Zhang YY, Wang G, Chen XP, et al. Gene expression and methylation profiles identified CXCL3 and CXCL8 as key genes for diagnosis and prognosis of colon adenocarcinoma. *J Cell Physiol*. 2020;235(5):4902-12.
5. Huang EY, Chang JC, Chen HH, Hsu CY, Hsu HC, Wu KL. Carcinoembryonic antigen as a marker of radioresistance in colorectal cancer: a potential role of macrophages. *BMC Cancer*. 2018;18(1):321.
6. Shinkins B, Nicholson BD, Primrose J, Perera R, James T, Pugh S, et al. The diagnostic accuracy of a single CEA blood test in detecting colorectal cancer recurrence: Results from the FACS trial. *PLoS One*. 2017;12(3):e0171810.
7. Zhang M, Miao F, Huang R, Liu W, Zhao Y, Jiao T, et al. RHBDD1 promotes colorectal cancer metastasis through the Wnt signaling pathway and its downstream target ZEB1. *J Exp Clin Cancer Res*. 2018;37(1):22.
8. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nature medicine*. 2015;21(5):449-56.
9. Vishnubalaji R, Hamam R, Abdulla MH, Mohammed MA, Kassem M, Al-Obeed O, et al. Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. *Cell Death Dis*. 2015;6:e1614.
10. Mutch MG. Molecular profiling and risk stratification of adenocarcinoma of the colon. *J Surg Oncol*. 2007;96(8):693-703.
11. Chung Moh M, Hoon Lee L, Shen S. Cloning and characterization of hepaCAM, a novel Ig-like cell adhesion molecule suppressed in human hepatocellular carcinoma. *J Hepatol*. 2005;42(6):833-41.
12. Moh MC, Zhang T, Lee LH, Shen S. Expression of hepaCAM is downregulated in cancers and induces senescence-like growth arrest via a p53/p21-dependent pathway in human breast cancer cells. *Carcinogenesis*. 2008;29(12):2298-305.
13. Jiang XL, Zhang Y, Tan B, Luo CL, Wu XH. Renal tumor-derived exosomes inhibit hepaCAM expression of renal carcinoma cells in a p-AKT-dependent manner. *Neoplasma*. 2014;61(4):416-23.
14. Saugier-veber P, Doummar D, Barthez MA, Czernecki V, Drouot N, Apartis E, et al. Myoclonus dystonia plus syndrome due to a novel 7q21 microdeletion. *Am J Med Genet A*. 2010;152A(5):1244-9.
15. Ozaki Y, Matsui H, Asou H, Nagamachi A, Aki D, Honda H, et al. Poly-ADP ribosylation of Miki by tankyrase-1 promotes centrosome maturation. *Mol Cell*. 2012;47(5):694-706.
16. Asou H, Matsui H, Ozaki Y, Nagamachi A, Nakamura M, Aki D, et al. Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. *Biochem Biophys Res Commun*. 2009;383(2):245-51.
17. Zhang GL, Pan LL, Huang T, Wang JH. The transcriptome difference between colorectal tumor and normal tissues revealed by single-cell sequencing. *J Cancer*. 2019;10(23):5883-90.
18. Huang Z, Yang Q, Huang Z. Identification of Critical Genes and Five Prognostic Biomarkers Associated with Colorectal Cancer. *Med Sci Monit*. 2018;24:4625-33.
19. Wu Z, Liu Z, Ge W, Shou J, You L, Pan H, et al. Analysis of potential genes and pathways associated with the colorectal normal mucosa-adenoma-carcinoma sequence. *Cancer Med*. 2018;7(6):2555-66.
20. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell*. 2018;173(2):283-5.
21. Chandrashekar DS, Bachel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. 2017;19(8):649-58.
22. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017;77(21):e108-e10.
23. Ruan GT, Gong YZ, Liao XW, Wang S, Huang W, Wang XK, et al. Diagnostic and prognostic values of CXC motif chemokine ligand 3 in patients with colon cancer. *Oncol Rep*. 2019;42(5):1996-2008.
24. Zhu J, Sanborn JZ, Benz S, Szeto C, Hsu F, Kuhn RM, et al. The UCSC Cancer Genomics Browser. *Nat Methods*. 2009;6(4):239-40.
25. Fahrman JF, Vykoukal J, Fleury A, Tripathi S, Dennison JB, Murage E, et al. Association Between Plasma Diacetylspermine and Tumor Spermine Synthase With Outcome in Triple-Negative Breast Cancer. *J Natl Cancer Inst*. 2020;112(6):607-16.
26. Skrzypczak M, Goryca K, Rubel T, Paziewska A, Mikula M, Jarosz D, et al. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One*. 2010;5(10).
27. Hong Y, Downey T, Eu KW, Koh PK, Cheah PY. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin Exp Metastasis*. 2010;27(2):83-90.
28. Ruan GT, Gong YZ, Zhu LC, Gao F, Liao XW, Wang XK, et al. The Perspective of Diagnostic and Prognostic Values of Lipoygenases mRNA Expression in Colon Adenocarcinoma. *Onco Targets Ther*. 2020;13:9389-405.

29. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-4.
30. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* 2018;46(D1):D956-D63.
31. Liao X, Zhu G, Huang R, Yang C, Wang X, Huang K, et al. Identification of potential prognostic microRNA biomarkers for predicting survival in patients with hepatocellular carcinoma. *Cancer Manag Res.* 2018;10:787-803.
32. Issa IA, Nouredine M. Colorectal cancer screening: An updated review of the available options. *World J Gastroenterol.* 2017;23(28):5086-96.
33. Eldai H, Periyasamy S, Al Qarni S, Al Rodayyan M, Muhammed Mustafa S, Deeb A, et al. Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner. *PLoS One.* 2013;8(10):e76251.
34. Klopffleisch R, Klose P, da Costa A, Brunnberg L, Gruber AD. HEPACAM1 and 2 are differentially regulated in canine mammary adenomas and carcinomas and its lymph node metastases. *BMC Vet Res.* 2010;6:15.
35. Zhang X, Quan F, Xu J, Xiao Y, Li X, Li Y. Combination of multiple tumor-infiltrating immune cells predicts clinical outcome in colon cancer. *Clin Immunol.* 2020;215:108412.
36. Elble RC, Widom J, Gruber AD, Abdel-Ghany M, Levine R, Goodwin A, et al. Cloning and characterization of lung-endothelial cell adhesion molecule-1 suggest it is an endothelial chloride channel. *J Biol Chem.* 1997;272(44):27853-61.
37. Yang B, Cao L, Liu B, McCaig CD, Pu J. The transition from proliferation to differentiation in colorectal cancer is regulated by the calcium activated chloride channel A1. *PLoS One.* 2013;8(4):e60861.
38. Yang B, Cao L, Liu J, Xu Y, Milne G, Chan W, et al. Low expression of chloride channel accessory 1 predicts a poor prognosis in colorectal cancer. *Cancer.* 2015;121(10):1570-80.
39. Li X, Hu W, Zhou J, Huang Y, Peng J, Yuan Y, et al. CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway. *Cell Commun Signal.* 2017;15(1):38.
40. Elferink LA, Strick DJ. Functional properties of rab15 effector protein in endocytic recycling. *Methods Enzymol.* 2005;403:732-43.
41. Xu X, Gong C, Wang Y, Hu Y, Liu H, Fang Z. Multi-omics analysis to identify driving factors in colorectal cancer. *Epigenomics.* 2020;12(18):1633-50.
42. Kim YS. Mucin glycoproteins in colonic neoplasia. *Keio J Med.* 1998;47(1):10-8.
43. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, et al. An integrative framework identifies alternative splicing events in colorectal cancer development. *Mol Oncol.* 2014;8(1):129-41.
44. Iwai T, Kudo T, Kawamoto R, Kubota T, Togayachi A, Hiruma T, et al. Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc Natl Acad Sci U S A.* 2005;102(12):4572-7.

Figures

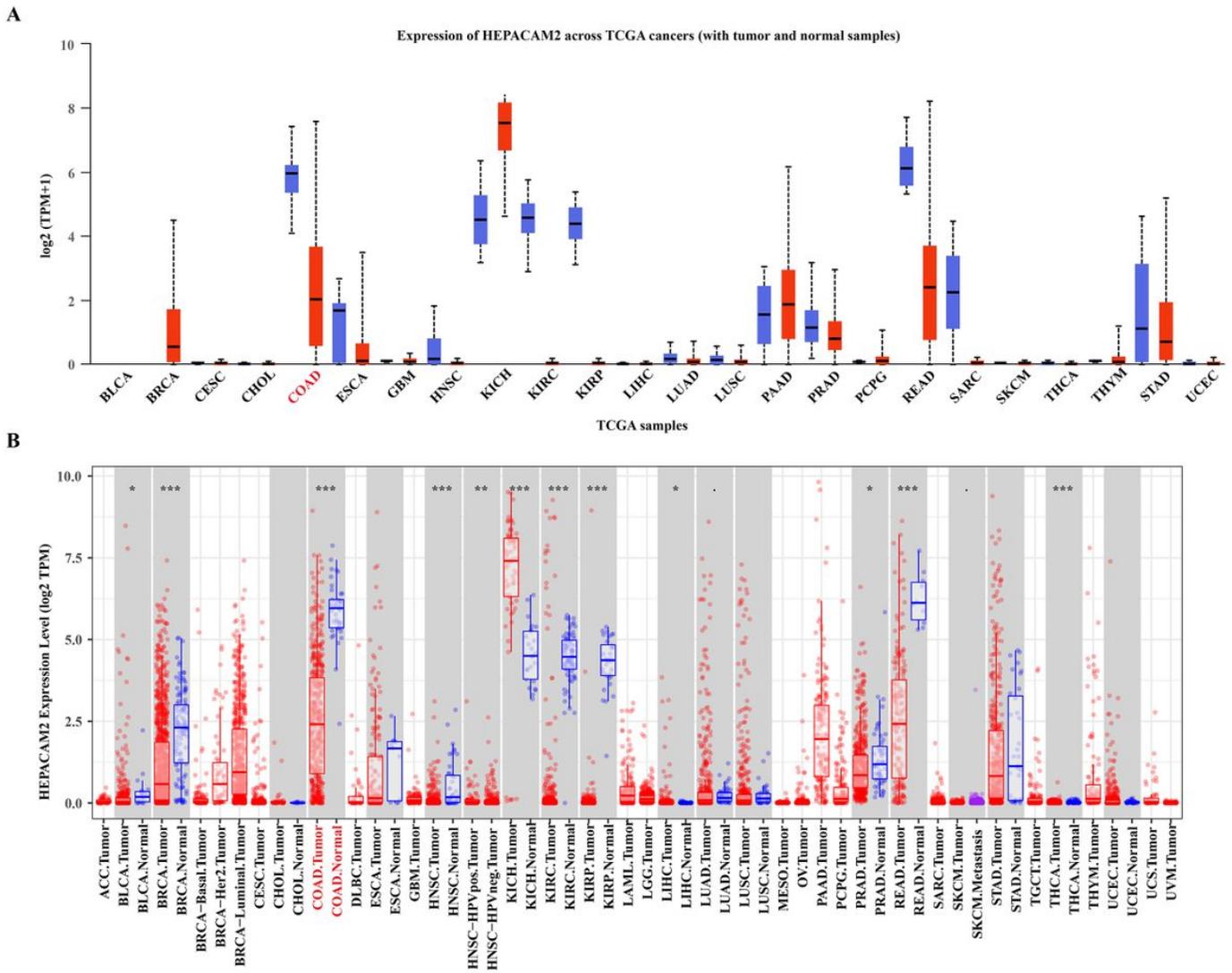


Figure 1

The differential expression of HEPACAM2 in different types of cancer based on UALCAN and TIMER databases. (A) UALCAN (B) TIMER. Notes: HEPACAM2: HEPACAM family member 2; TCGA: The Cancer Genome Atlas; COAD: colon adenocarcinoma.

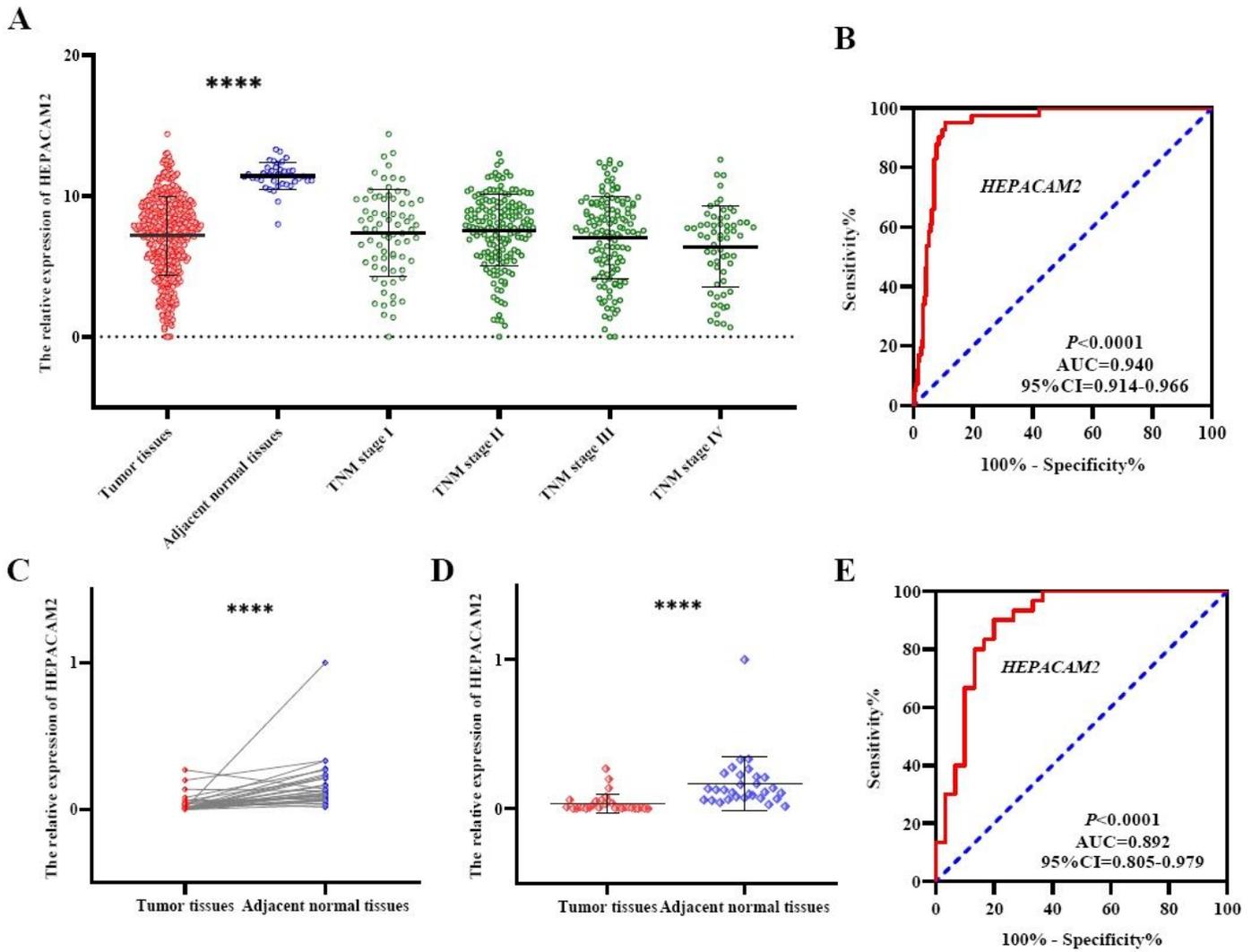


Figure 2

The differential expression and diagnostic ROC analysis of HEPACAM2 in COAD tumor tissues and adjacent normal tissues based on TCGA cohort and Guangxi cohort. (A)-(B) TCGA cohort; (C)-(E) Guangxi cohort. Notes: HEPACAM2: HEPACAM family member 2; TCGA: The Cancer Genome Atlas; COAD: colon adenocarcinoma. ROC: receiver operating characteristic; AUC: area under the curve; CI: confidence interval.

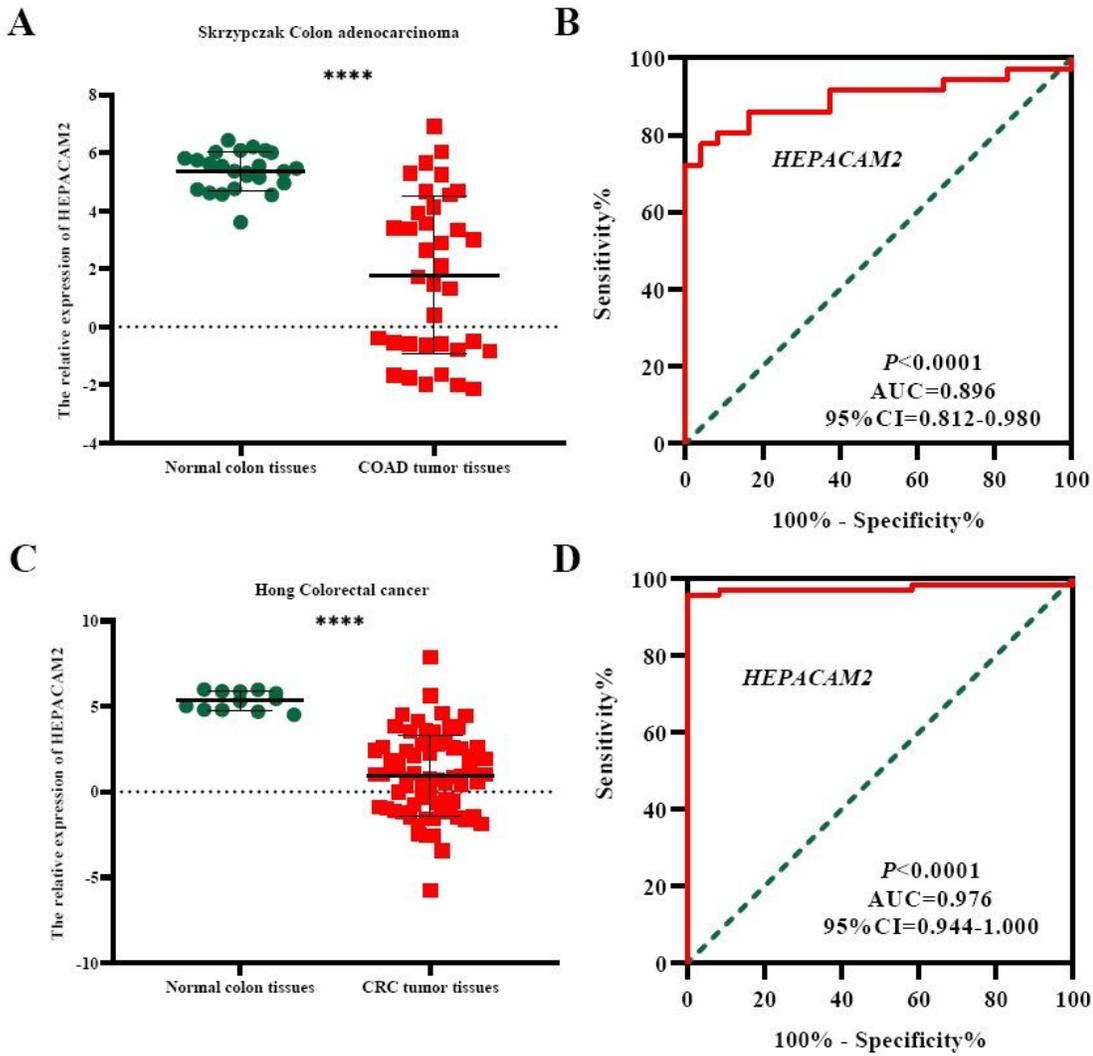


Figure 3

The differential expression and diagnostic ROC analysis of HEPACAM2 in COAD and CRC based on the GEO database. (A)-(B) Skrzypczak COAD; (C)-(D) Hong CRC. Notes: HEPACAM2: HEPACAM family member 2; GEO: Gene Expression Omnibus; COAD: colon adenocarcinoma; CRC: Colorectal cancer; ROC: receiver operating characteristic; AUC: area under the curve; CI: confidence interval.

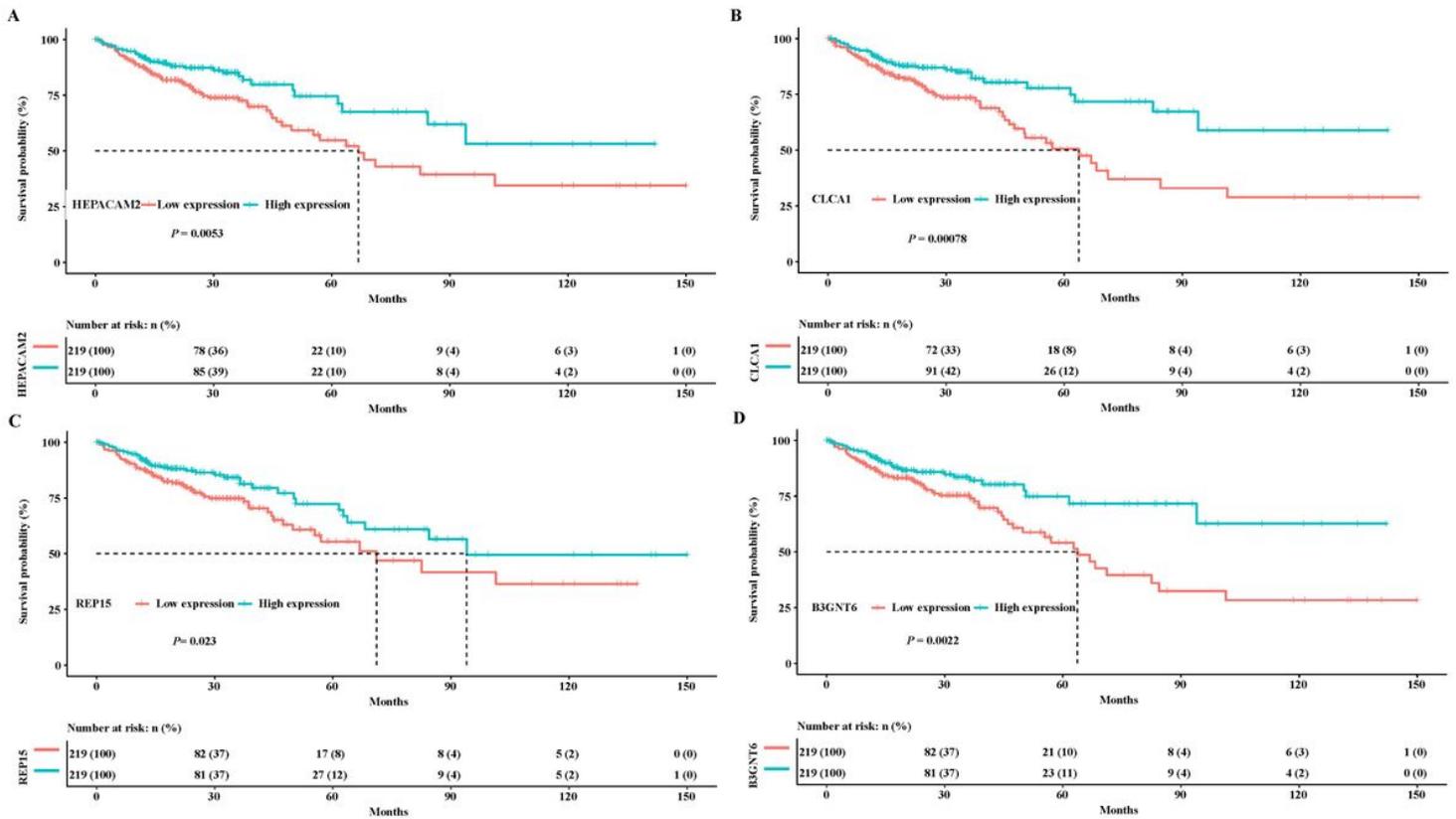


Figure 4 Survival curve of OS in COAD based on TCGA cohort. (A) HEPACAM2; (B) CLCA1; (C) REP15; (D) B3GNT6. Notes: OS, overall survival; COAD, colon adenocarcinoma; TCGA: The Cancer Genome Atlas; HEPACAM2: HEPACAM family member 2; CLCA1: chloride channel accessory 1; REP15: RAB15 effector protein; B3GNT6: UDP-GlcNAc: betaGal beta-1,3-N-acetylglucosaminyltransferase 6.

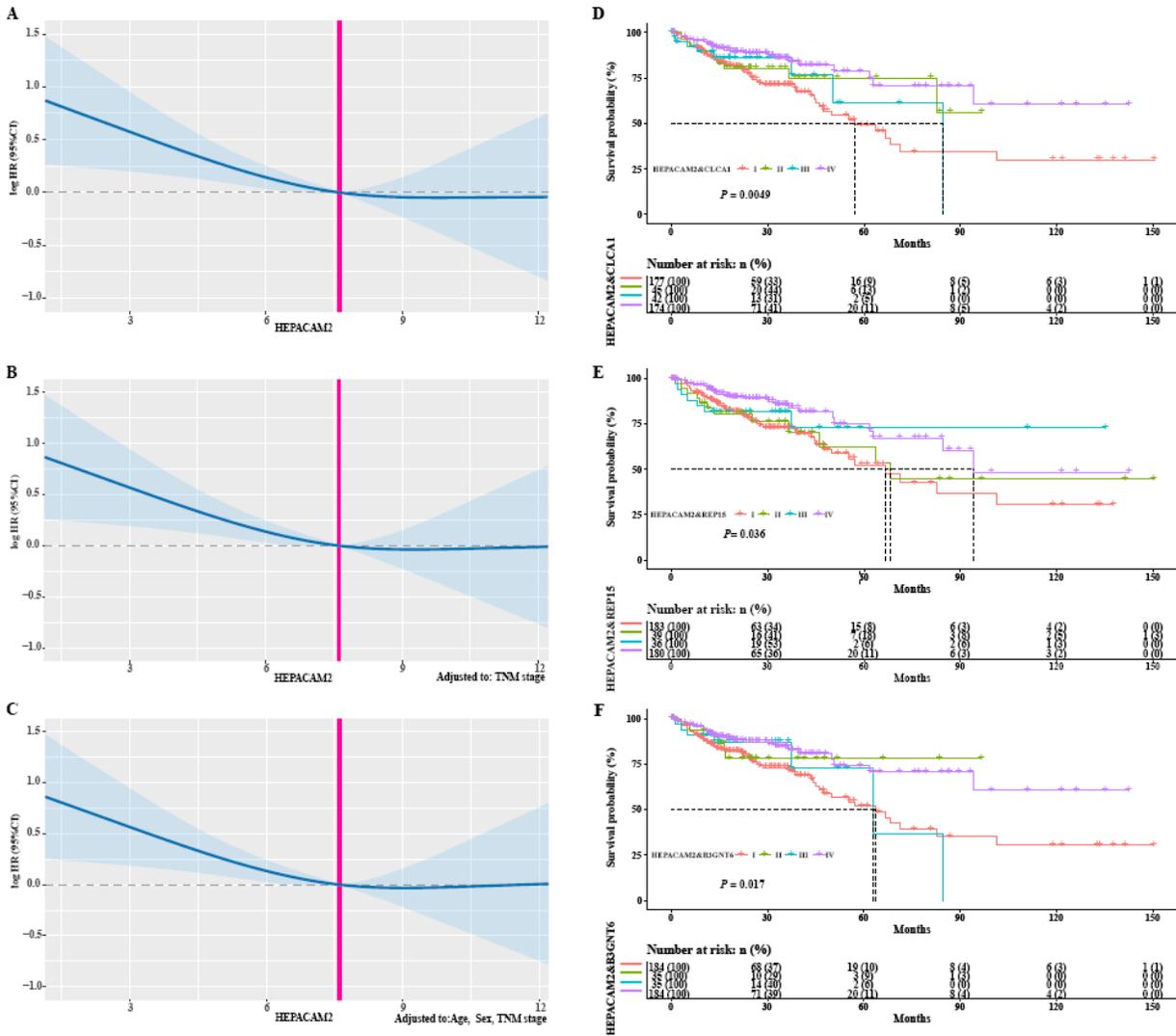


Figure 5

The relationship between HEPACAM2 gene and COAD death risk and joint-effect survival curve based on TCGA cohort. (A)-(C) The relationship between HEPACAM2 gene and COAD death risk. (A) Univariate; (B) Adjusted by model 1: TNM stage; (C) Adjusted by model 2: Age, Sex, and TNM stage; (D)-(F) Joint-effect survival curve. (D) HEPACAM2 combined with CLCA1; (E) HEPACAM2 combined with REP15; (F) HEPACAM2 combined with B3GNT6. Notes: COAD, colon adenocarcinoma; TCGA: The Cancer Genome Atlas; HEPACAM2: HEPACAM family member 2; CLCA1: chloride channel accessory 1; REP15: RAB15 effector protein; B3GNT6: UDP-GlcNAc: betaGal beta-1,3-N-acetylglucosaminyltransferase 6.

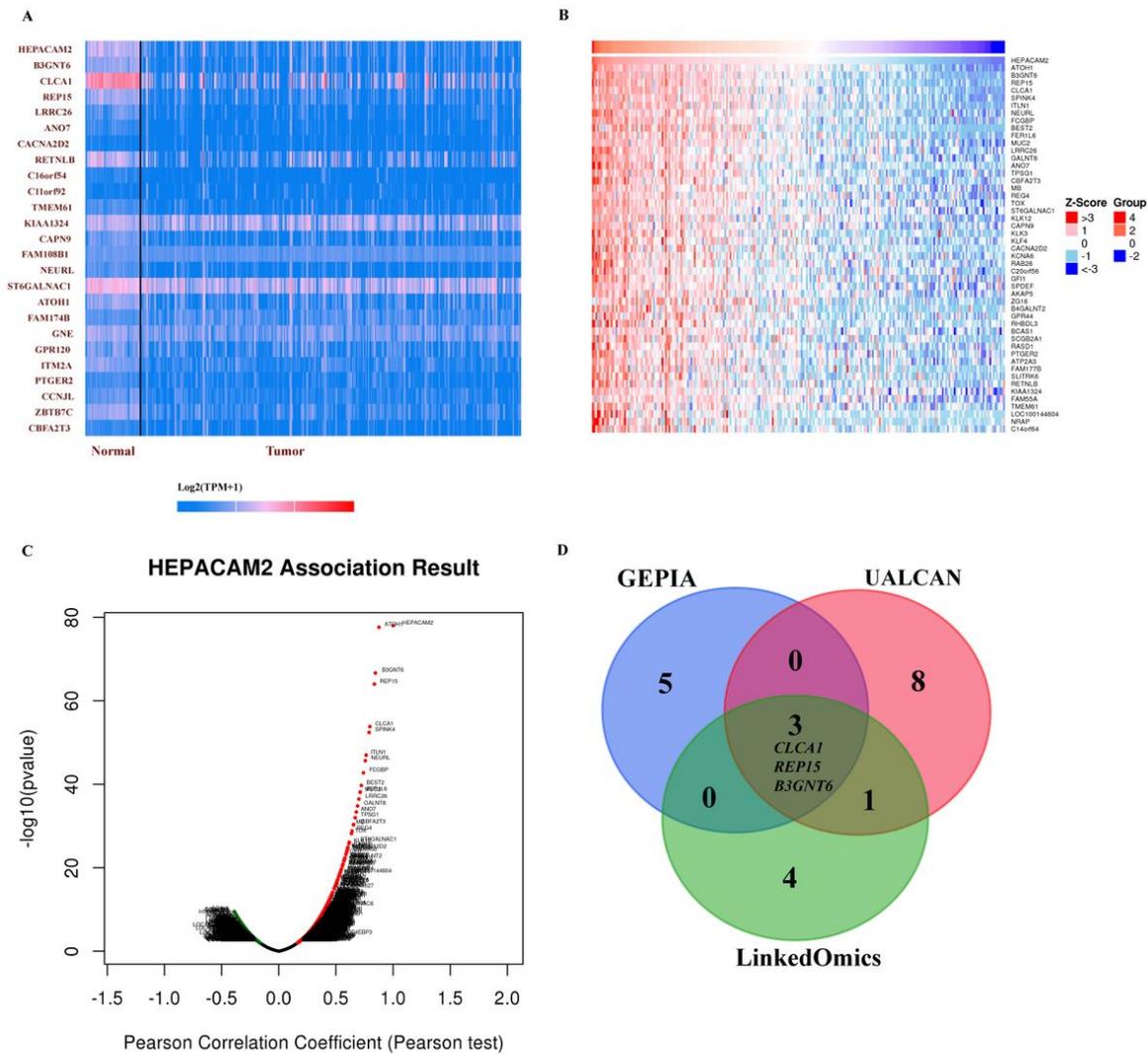


Figure 7

The heat map and the volcano map of LinkedOmics database and the Venn diagrams of among GEPIA, UALCAN, and LinkedOmics databases.

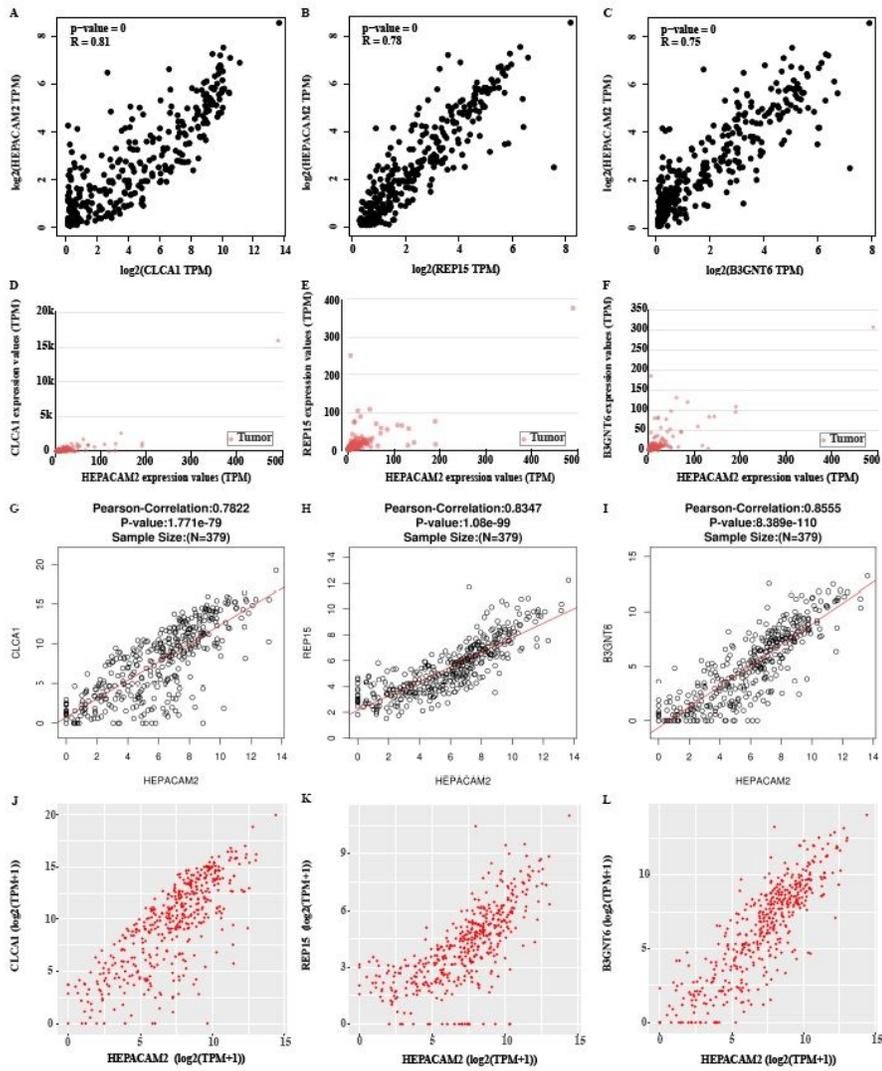


Figure 8

The correlation analyses among different datasets. (A)-(C) GEPIA; (D)-(F) UALCAN; (G)-(I) LinkedOmics; (J)-(L) TCGA cohort.

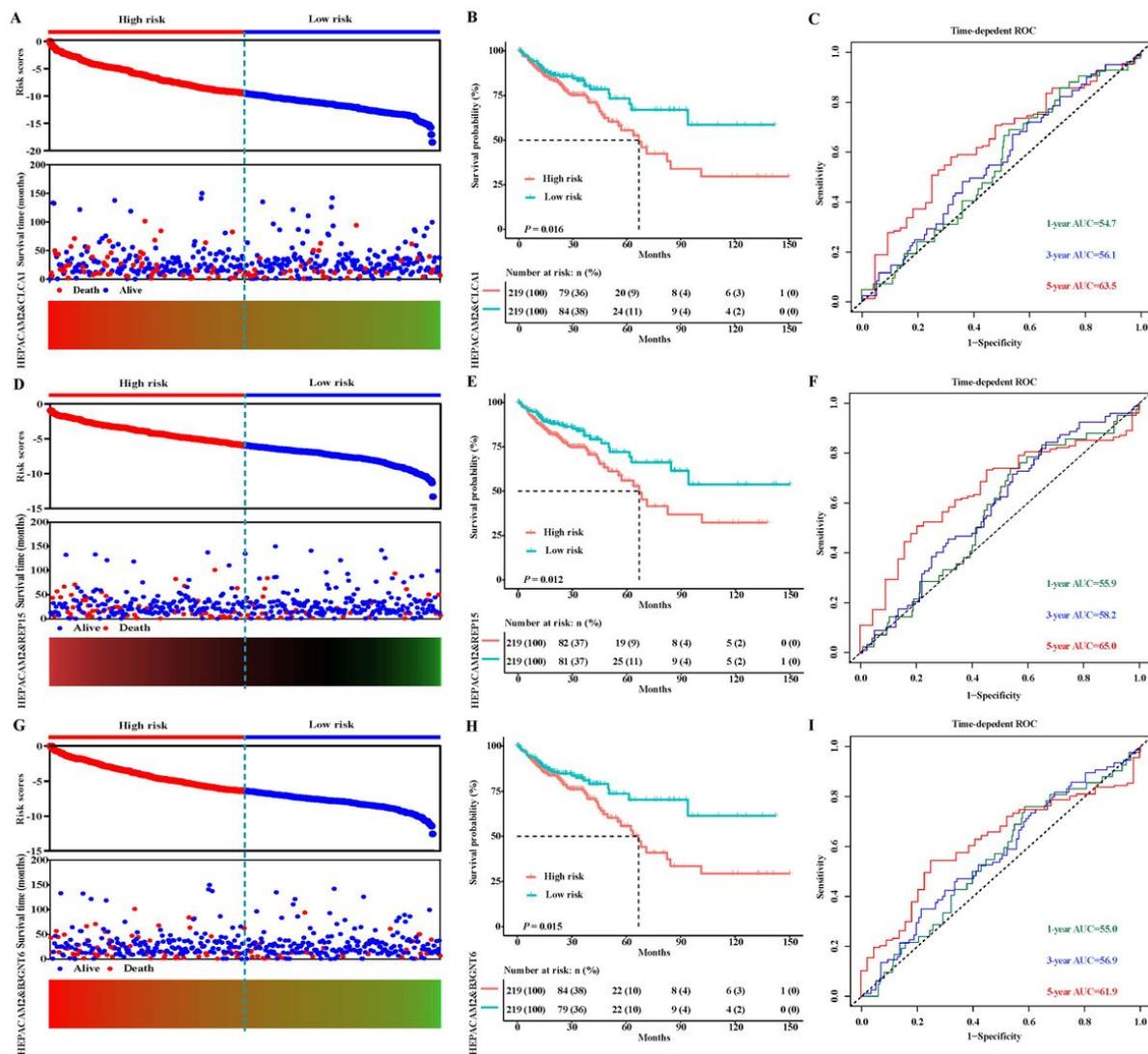


Figure 9

The risk score model and ROC curve of HEPACAM2 and correlation genes for COAD OS based on TCGA cohort. (A)-(C) Risk score model of HEPACAM2 combined with CLCA1; (D)-(F) Risk score model of HEPACAM2 combined with REP15; (G)-(I) Risk score model of HEPACAM2 combined with B3GNT6. Notes: OS, overall survival; COAD, colon adenocarcinoma; TCGA: The Cancer Genome Atlas; HEPACAM2: HEPACAM family member 2; CLCA1: chloride channel accessory 1; REP15: RAB15 effector protein; B3GNT6: UDP-GlcNAc: betaGal beta-1,3-N-acetylglucosaminyltransferase 6.

Supplementary Files

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

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)