Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A Novel Prognostic Score Based on *ZG16* for Predicting CRC Survival

This article was published in the following Dove Press journal:
Pharmacogenomics and Personalized Medicine

Wei Wang[1,2,*]
Jian-Fang Sun[3,*]
Xiao-Zhong Wang[1]
Hou-Qun Ying[4]
Xia-Hong You[1]
Fan Sun[1]

[1]Jiangxi Province Key Laboratory of Laboratory Medicine, Department of Clinical Laboratory, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi, People's Republic of China; [2]Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, People's Republic of China; [3]Neonatology Department, Dongguan Eighth People's Hospital, Dongguan Children's Hospital, Dongguan 523000, People's Republic of China; [4]Department of Nuclear Medicine, Jiangxi Province Key Laboratory of Laboratory Medicine, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, People's Republic of China

*These authors contributed equally to this work

Background: Colorectal cancer (CRC) is one of the lethal malignant tumors worldwide. However, the underlying mechanism of CRC and its biomarkers remain unclear. The aim of this study was to identify the key genes associated with CRC and to further explore their prognostic significance.

Methods: Four expression profile datasets (GSE41657, GSE74602, GSE113513, and GSE40967) downloaded from Gene Expression Omnibus (GEO) and one RNAseq dataset of CRC from The Cancer Genome Atlas (TCGA) database were included in our study. The Cox model was utilized for univariate or multivariate survival analysis. GEPIA and HAP database were adopted for verification of DEGs (*ZG16*). The decision curve analysis (DCA) and time-dependent ROC were chosen for evaluating the prognostic effectiveness of biomarkers.

Results: In total, 88 differentially expressed genes (DEGs) were identified, and the GO and KEGG enrichment analyses of DEGs were processed. After, the protein–protein interaction (PPI) network was constructed and 15 hub genes including *ZG16* were identified. The differential expression of *ZG16* between tumor and normal colorectal tissues were further verified in GEPIA and HAP database. Subsequent survival indicated that expression of *ZG16* is negatively correlated with overall survival of OS and is an independent prognostic factor for CRC patients. Furthermore, the construction of a prognostic score containing *ZG16*, TNM stage and age exhibited superior effectiveness for predicting long-term survival of CRC patients. Additionally, our results were verified using the GSE40967 dataset, which indicated an improved performance of combined risk score based on *ZG16* for predicting OS of CRC patients.

Conclusion: *ZG16* is a potential parameter for predicting prognosis in CRC. Furthermore, a combination of *ZG16*, TNM stage, and age allows improved prognosis of CRC.

Keywords: colorectal cancer, bioinformatics, *ZG16*, prognostic score, biomarker

Introduction

Colorectal cancer (CRC) is ranked as the third most common cancer diagnosed around the world by GLOBOCAN estimation.[1] Despite the fact that effective cancer screening measures and modern medicine have arisen, CRC remains the leading cause of cancer-related mortality worldwide.[2] During recent years, there has been a decline in mortality from CRC in the US (101,420 new cases of CRC in 2019, with an estimated 51,020 deaths).[3] Meanwhile, the incidence and mortality of CRC in China has increased, according to the Chinese Cancer Statistics.[4] Therefore, it is essential to identify novel prognostic biomarkers, as well as explore the underlying mechanism of CRC initiation and progression.

Correspondence: Fan Sun
Jiangxi Province Key Laboratory of Laboratory Medicine, Department of Clinical Laboratory, The Second Affiliated Hospital of Nanchang University, No. 1 Min De Road, Nanchang 330006, People's Republic of China
Tel/Fax +86 0791-86300410
Email 942114610@qq.com

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/PGPM.S275941

As expected, Tumor-node-metastasis (TNM) stage, defined by the American Joint Committee on Cancer, according to pathologic and clinical factors, has served as the fundamental diagnostic parameter and crucial index for CRC prognosis. However, an inaccurate prediction of CRC prognosis can lead to terrible consequences for patients (for example, the 5-year survival rate for stage IV was merely 10%).[5,6] Importantly, novel prognostic biomarkers have been identified on account of the limitation of traditional prediction.

The predisposition and incidence of CRC is due to multiple reasons, including genetic abnormality and environmental factors. Investigation of the molecular basis allows us to better understand the initiation and progression of CRC. Additionally, publicly accessible genome databases make bioinformatics analysis easy with increased development of gene sequencing technology. Numerous molecules and complex genomic alterations have been found in CRC from bioinformatics analysis. Nevertheless, significant variability among different studies has lowered the credibility of these results.

In this study, CRC-related differentially expressed genes (DEGs) were identified and validated using the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA). These DEGs were further subjected to gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) enrichment analyses. The protein–protein interaction (PPI) network of DEGs was constructed using the Cytoscape software, and the hub genes that were closely associated with CRC were captured and deciphered. Subsequently, we conducted univariate and multivariable Cox regression analyses for predicting CRC prognosis, which indicated *ZG16* expression is negatively correlated to the overall survival (OS) of CRC patients. Finally, we comprehensively established a risk score with better predictive effectiveness based on *ZG16*, age, and TNM stage.

Materials and Methods

Data Collection

The datasets enrolled in this study followed the criteria below: i) Human CRC tissue samples were utilized for the profiles; ii) Studies inculded normal groups as the control, and the sample size per dataset was >10; and iii) Studies with definite information about the technology and platform, and the expression data were pre-processing, which included background correcting, normalizing, and calculating expression. As a result, four gene expression datasets (GSE41657, GSE74602, GSE113513, GSE40967) were acquired from the GEO database, and were obtained using GPL6480 (Agilent-014850 Whole Human Genome Microarray), GPL6104 (Illumina humanRef-8 v2.0 expression beadchip), GPL15207 (Affymetrix Human Gene Expression Array), and GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array), respectively. The GSE41657 dataset included 12 normal mucosae, 51 adenomas, and 25 adenocarcinomas. The GSE74602 dataset included 30 paired normal and CRC samples. The GSE113513 dataset included 14 paired normal and CRC samples. The GSE40967 dataset included 562 samples from stage Ⅰ–Ⅳ CRC patients, with integrated molecular and survival characteristics. In addition, 383 CRC tissues and 51 adjacent normal tissues, including corresponding clinicopathological information, were obtained via the TCGA database. The study was granted approval by the institutional research ethics committee of the second affiliated hospital of Nanchang University.

Identification of DEGs

GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was applied to detect DEGs between normal and tumor tissues across different GEO datasets (GSE41657, GSE74602, and GSE113513). |log2FC|>2 and adjusted *P*-value<0.05 are considered statistically significant. The significant DEGs across three GEO datasets were validated using RNA sequencing data in the TCGA COADREAD dataset. Overlapping DEGs between the GEO and TCGA database were selected for the following studies. A Venn diagram containing these DEGs was outlined online (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

GO and KEGG Enrichment Analyses of DEGs

ClueGO, a Cytoscape 3.6.0 plug-in, integrates Gene Ontology (GO) terms and KEGG pathways of the overlapping DEGs.[7] Based on the binary gene–term matrix with selected genes, a term–term matrix is calculated using chance-corrected kappa statistics in order to determine the association strength between terms. After, the generated network represents the terms as nodes that are connected according to a predefined kappa score level (≥0.4 in the present study).

Hub Gene Analysis in PPI Network

The Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/) helped analyze the interactive relationships between overlapping DEGs. The results of this analysis were imported into Cytoscape 3.6.0 in order to establish a network model. Additionally, genes with top 15 node degree scores were captured using the Cytoscape plug-in cytoHubba.

Confirmation of *ZG16* Differential Expression

The web server for cancer and normal gene expression profiling (GEPIA) (Gene Expression Profiling Interactive Analysis, http://gepia.cancer-pku.cn/index.html) was used to further verify the differential expression of *ZG16*. Moreover, the protein levels of *ZG16* in colorectal tumor and normal tissues were investigated in the Human Protein Atlas (HPA, https://www.proteinatlas.org/), which contains immunohistochemistry-based expression data for cancer research.

Survival Analysis

In order to further identify the relationship between hub genes and CRC, univariate and multivariate Cox regression analysis was conducted using a survival package in R 3.6.2. The genes that correlated to overall survival (P<0.05) were retained for subsequent multivariate Cox

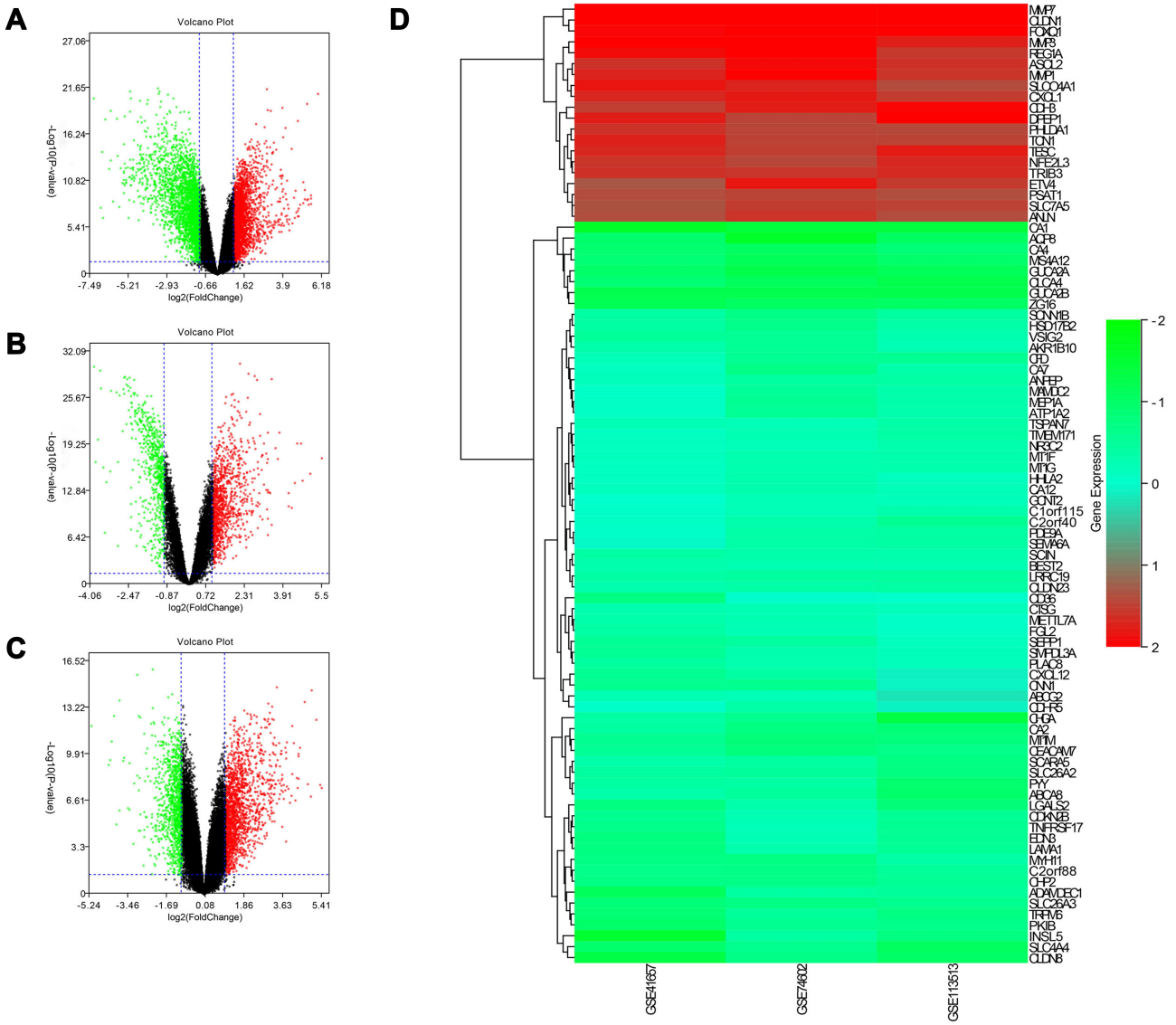

Figure 1 The volcano plots of GSE41657 (**A**), GSE74602 (**B**), and GSE113513 (**C**), and the heat map (**D**) describing the level of overlapping DEGs.

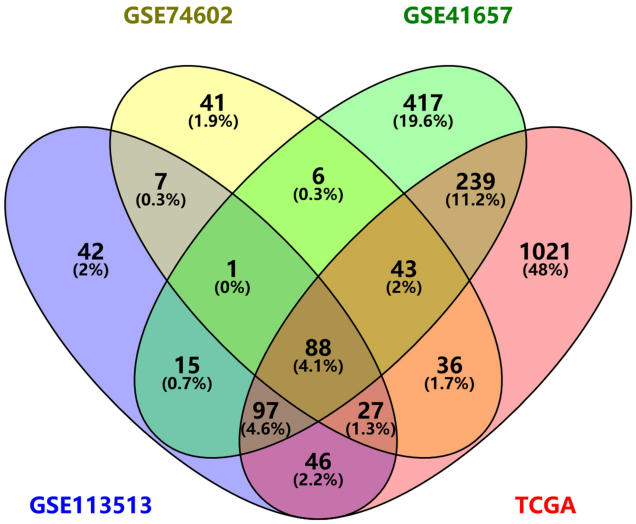

Figure 2 Venn plots of overlapping DEGs.

proportional hazards model. Then, a risk score was established based on the following formula:

$$Risk\ score = \sum xi * \beta i(xi : gene\ expression, \beta i : coefficient)$$

In this study, the risk score was as follows in the TCGA database: (−0.5664*ZG16)+(1.4050*TNM stage) +(0.7093*Age), in which patients with TNM stage I–III and Ⅳ were stratified as 1 or 2 and age ≤ or >60 years was considered as 1 or 2. The patients were classified into either low- or high-risk groups according to the median risk score, and the Kaplan–Meier curve was constructed using Graphpad Prism 6. Finally, the Decision curve analysis (DCA) was plotted to represent the clinical values for predicting CRC overall survival. In order to verify the efficacy of our result, we conducted an independent survival analysis of 562 CRC patients from the GSE40967 dataset. To classify the stage II–III CRC patients who can benefit from adjuvant chemotherapy, we proceeded with univariate or multivariate survival analysis in 203 stage II–III patients who had been treated with chemotherapy. Next, the risk score in GSE40967 was constructed according to the following formula: (−0.44792*ZG16)+(1.68657*TNM stage) +(0.79746*Age), in which patients with TNM stage I–III and Ⅳ were stratified as 1 or 2 and age ≤ or >70 years was considered as 1 or 2, respectively. In addition, time-dependent ROC was used to compare the prognostic effectiveness of different biomarkers across our study.

Results

Identification of DEGs in CRC

According to different conditions, 906 DEGs in GSE41657, 249 DEGs in GSE74602, 323 DEGs in GSE113513, and 1597 DEGs in TCGA were identified. They were shown on a volcano map based on the value of |logFC| (Figure 1A–C). The overlap among datasets contained 88 genes, as shown in the heat map (Figure 1D) and the Venn diagram (Figure 2), and consisted of 68 downregulated and 20 upregulated genes between CRC and normal tissues.

Enrichment of DEGs

In order to analyze the biological classification of DEGs, GO and KEGG enrichment analyses were performed using Cytoscape. The enriched GO terms of DEGs were divided into three parts, including biological process (BP), cell composition (CC), and molecular function (MF). Nitrogen metabolism was the most highly correlated GO/pathway term (Supplemental Figure 1). Furthermore, the overview chart indicated that bicarbonate transport, mineral absorption, metallopeptidase activity, brush border membrane, and monovalent inorganic cation homeostasis were the functional groups of these DEGs (Figure 3A). In addition, this functionally grouped network with terms as nodes were linked according to the kappa score level ≥0.4 (Figure 3B). Changes in BP of DEGs were significantly enriched in metallopeptidase activity, metalloendopeptidase activity, monovalent inorganic cation homeostasis, and bicarbonate transport. The CC were mainly enriched in brush border and brush border membrane. Within the MF category, carbon-oxygen lyase activity, hydro-lyase activity, and carbonate dehydratase activity were predominant. KEGG pathway analysis revealed that DEGs were primarily associated with mineral absorption, nitrogen metabolism, and pancreatic and bile secretion (Table 1).

PPI Analysis of DEGs

The PPI network of 88 DEGs was constructed and displayed using STRING and Cytoscape (Figure 4A). The PPI network contained 49 nodes and 71 edges. Furthermore, the top 15 genes (*ZG16*, *GUCA2B*, *CXCL12*, *AQP8*, *CXCL1*, *GUCA2A*, *MMP1*, *MMP7*, *SLC26A3*, *MS4A12*, *CLCA4*, *PYY*, *MMP3*, *SLC4A4*, *ABCG2*) were extracted for subsequent study according to the "Degree" algorithm (Figure 4B).

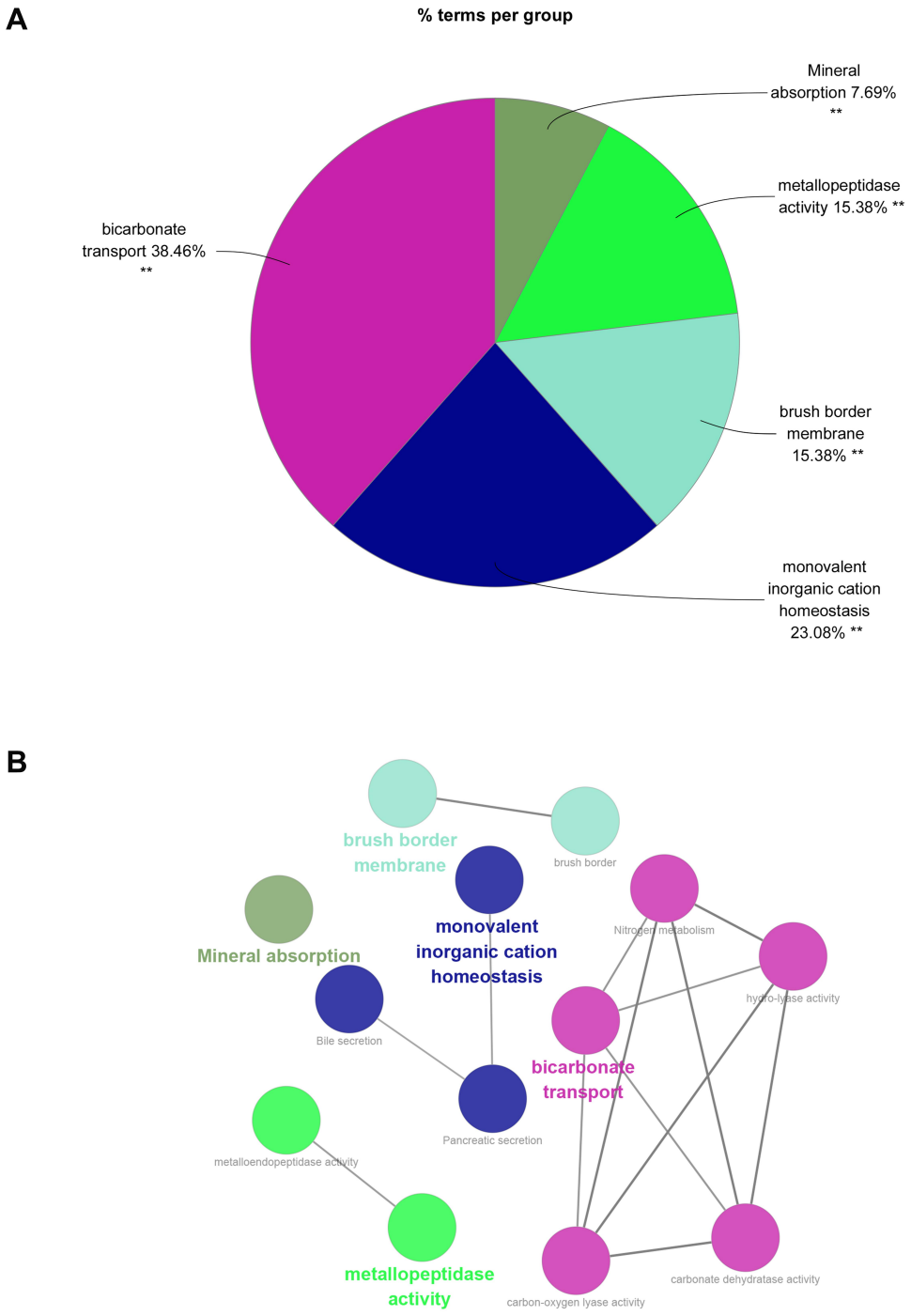

Figure 3 Overview chart with functional groups including specific terms for these genes (**A**). Functionally grouped network (**B**) with terms based on their kappa score level (≥0.4).
Note: **The representative Terms and Pathways.

Confirmation of *ZG16* Differential Expression

Using the GEPIA, a website tool to provide customizable functions such as tumor/normal differential expression analysis based on TCGA and the Genotype-Tissue Expression (GTEx) data, we compared the ZG16 expression of Colon adenocarcinoma (COAD) and Rectal adenocarcinoma (ROAD) with that of normal tissues. In total, not only 275 COAD samples and 349 normal tissues but also 92 ROAD samples and 318 normal rectal tissues were

Table 1 Significantly Enriched GO Terms and KEGG Pathways of DEGs

GOID	GOTerm	*P*-value	FDR	-Log10(FDR)	% Associated Genes	Nr. Genes
KEGG:04978	Mineral absorption	2.000E-07	6.500E-07	6.187	10.34	6.00
GO:0008237	Metallopeptidase activity	6.024E-07	1.566E-06	5.805	4.02	9.00
GO:0004222	Metalloendopeptidase activity	2.067E-04	2.067E-04	3.685	4.07	5.00
GO:0005903	Brush border	4.896E-06	9.093E-06	5.041	4.58	7.00
GO:0031526	Brush border membrane	1.480E-06	3.206E-06	5.494	7.41	6.00
KEGG:04972	Pancreatic secretion	8.563E-05	9.276E-05	4.033	4.90	5.00
KEGG:04976	Bile secretion	1.594E-05	2.073E-05	4.683	6.94	5.00
GO:0055067	Monovalent inorganic cation homeostasis	9.778E-06	1.412E-05	4.850	4.12	7.00
KEGG:00910	Nitrogen metabolism	8.520E-09	5.538E-08	7.257	29.41	5.00
GO:0016835	Carbon-oxygen lyase activity	4.462E-05	5.273E-05	4.278	5.62	5.00
GO:0016836	Hydro-lyase activity	8.930E-06	1.451E-05	4.838	7.81	5.00
GO:0004089	Carbonate dehydratase activity	1.176E-08	5.095E-08	7.293	27.78	5.00
GO:0015701	Bicarbonate transport	1.006E-10	1.307E-09	8.884	14.55	8.00

Abbreviation: FDR, false discovery rate.

analyzed. On the basis of |log2FC|>2 and adjusted *P*-value<0.05, we found obviously lower *ZG16* expression in tumor patients compared with the normal ones (Figure 5). According to the immunostaining data from the HPA database, we verified that expression levels of *ZG16* decreased in colon or rectal cancer tissues compared with normal tissues (Figure 6).

Survival Analysis of CRC Patients

In total, 373 CRC patients from TCGA were included in our survival analysis (Table 2). The prognostic role of hub genes was analyzed using univariate and multivariate Cox regression. The results indicated that high expression of *ZG16*, *AQP8*, and *SLC26A3* were correlated with OS of CRC patients. In addition, the prognostic signature involved in TNM stage, age, and the above-mentioned genes was developed using the multivariate Cox proportional hazards regression model. Significantly, TNM stage Ⅳ, age >60 years, and low *ZG16* expression were closely associated with worse OS (Table 3). Thus, we constructed a risk score including TNM stage, age, and ZG16 expression. The risk score was calculated based on three values and relevant coefficients (0.5664 for ZG16, 1.4050 for TNM stage, and 0.7093 for Age). Then, all patients were divided into either high- or low-risk groups, according to the median score (Figure 7A). In addition, the distribution of dead or alive patients reveals that the risk score was closely correlated with mortality risk (Figure 7B). Furthermore, the Kaplan–Meier curve indicated that patients in the high-risk group had worse OS compared to the low-risk group (Figure 8A). In order to explore the accuracy of the prediction value, a decision curve analysis (DCA) was employed. From the curves

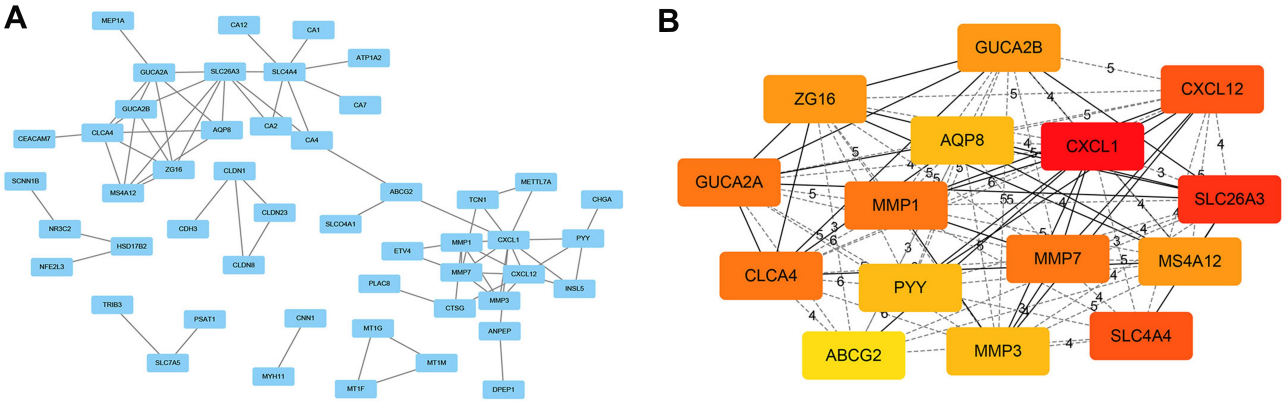

Figure 4 The PPI network (**A**) of DEGs and top 15 hub genes (**B**).

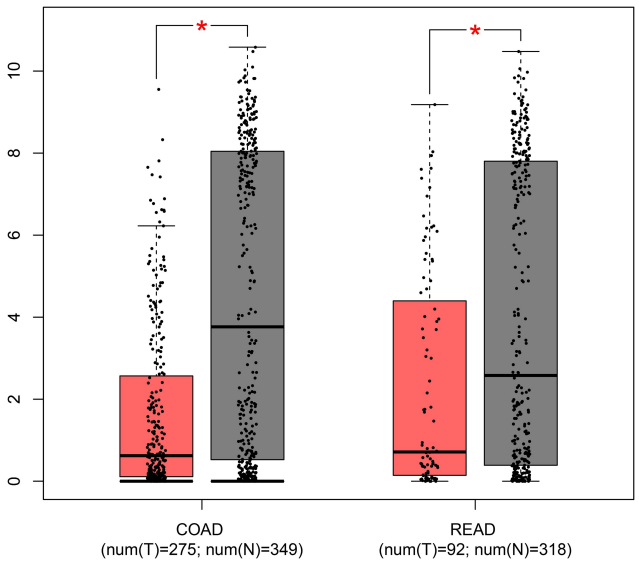

Figure 5 The *ZG16* expression of Colon adenocarcinoma (COAD) and Rectal adenocarcinoma (ROAD) with that of normal tissues.
Note: *P<0.05.

depicted in Figure 8B, we were able to discern that the risk score owned the best predictive ability of CRC OS.

Validation in GSE40967

The baseline characteristics of 562 CRC patients from GSE40967 are summarized in Table 4. Using the X-tile software, we sought the optimal cut-off point of *ZG16* based on OS (Supplemental Figure 2). Furthermore, we found that males, >70 years, low *ZG16*, or mutated *KRAS* patients were all correlated to poor OS from survival analysis. As shown in Table 5, age >70 years (crude HR=1.967, 95% CI=1.476–2.622; adjusted HR=2.220, 95% CI=1.548–3.185), and stage Ⅳ (crude HR=4.738, 95% CI=3.312–6.777; adjusted HR=5.401, 95% CI=3.732–7.817) were significantly associated with increased risk of death due to CRC. On the other hand, high *ZG16* expression (crude HR=0.644, 95% CI=0.463–0.896; adjusted HR=0.639, 95% CI=0.458–0.892) was significantly correlated to a decreased risk of disease from the disease. The Cox-model survival analysis in 203 stage II–III CRC patients who received chemotherapy revealed that low *ZG16* (P<0.05, adjusted HR=2.804, 95% CI=1.446–5.438) and mutated *KRAS* (P<0.05, adjusted HR=2.120, 95% CI=1.215–3.697) were closely related with worse OS (Figure 9). Time-dependent ROC helped compare the prognostic efficacy of each biomarker in this study. Comparatively, the areas under the ROC (AUCs) of combined risk score stayed at the relatively high level than TNM stage, age, and *ZG16* in the 120 months (Figure 10).

Discussion

Colorectal cancer (CRC) is triggered by an accumulation of genetic mutations and epigenetic changes. Notably, CRC is characterized by extensive tumor heterogeneity and significant genomic instability, which influences greatly the treatment and prognosis of the patient.[8,9] Based on microarray technology and next-generation sequencing technology, aberrantly expressed genes and pathways have been identified in CRC. Furthermore, these genes are considered to be potential candidates for diagnosis or predicting survival of CRC patients. However, inconsistent conclusions have been drawn due to many uncertain factors.

As early as 1998, Zymogen Granule Protein 16 (ZG16) was found to be a linker molecule that was binding to the luminal surface of pancreatic cells in rats.[10] Subsequently, it was identified in humans and detected in the intestinal goblet cells, pancreatic acinar cells, serosanguineous acinar cells of the parotid gland, as well as in serum.[11] A previous study showed high *ZG16* expression in the colon,[12] and protein sequencing analysis revealed that ZG16 contains a signal peptide, which suggests that it might monitor colon condition by direct secretion. Several researchers have demonstrated an absence of the mucosal barrier consisting of ZG16, which may allow bacteria to closely affect the epithelium and cause an enhanced bacterial uptake and local inflammation.[13] Besides, the crystal structure analysis uncovered that ZG16p has a β-prism fold structure that resembles the mannose-binding Jacalin-related lectins.[14] A study showed that Jacalin can exert an anti-tumor effect by interacting with the Thomsen–Friedenreich (TF) antigen or by binding to the oncoprotein – MUC1.[15,16] The remarkable similarity of ZG16 to Jacalin implied that ZG16 may play an important role in CRC immunity. Interestingly, ZG16 was proven to suppress growth and sphere formation of stem-like CRC cells, which suggests that loss of ZG16 was likely to make a great difference in CRC initiation and stemness.[17] Additionally, bioinformatics analysis suggested that miR-196a promoted CRC cell growth and migration by down-regulating *ZG16* expression.

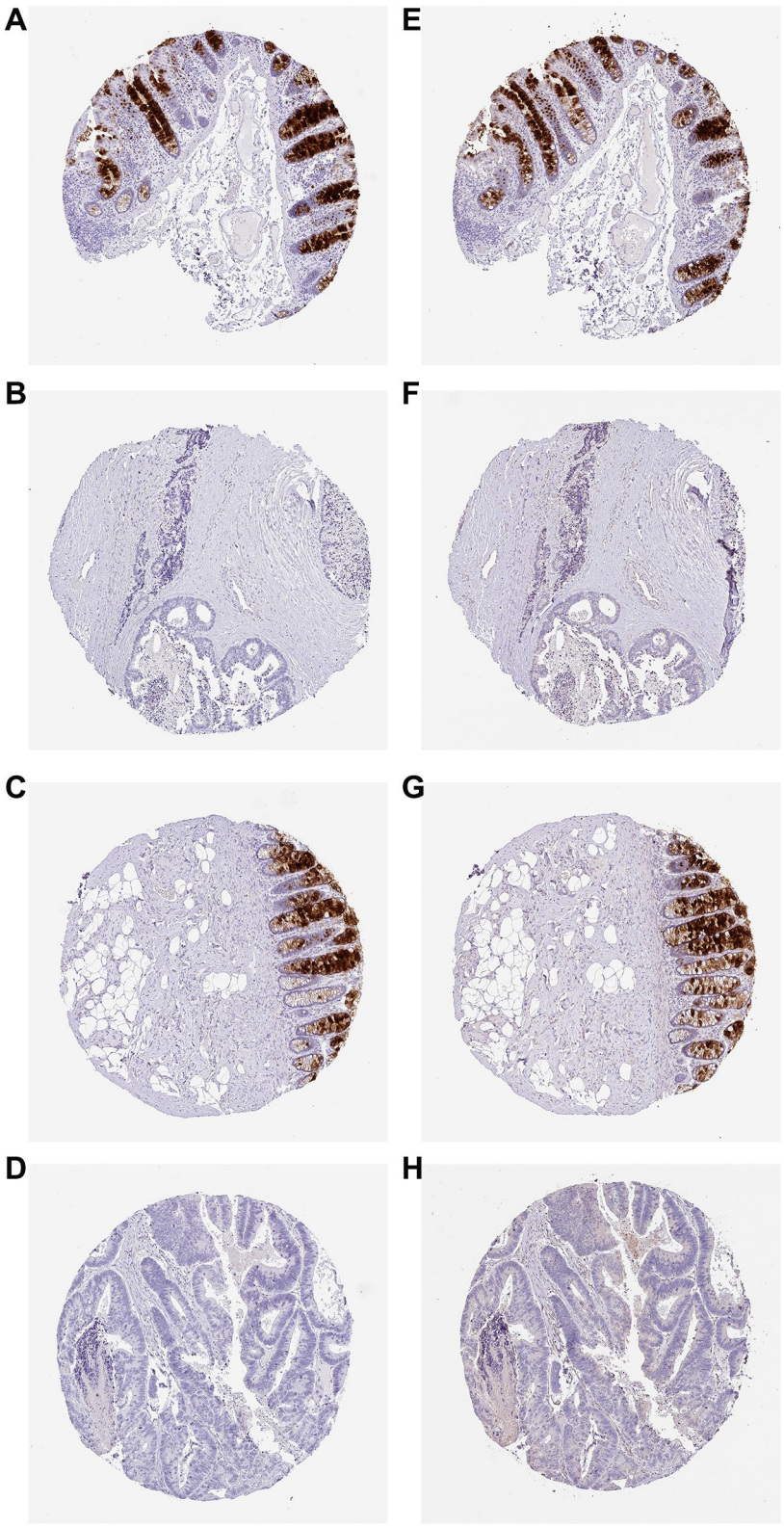

Figure 6 The ZG16 protein expression in colorectal normal tissue and tumor tissue. ((**A** and **E**) Slices of colonic normal tissue staining by Antibody HPA052066 and HPA052512; (**B** and **F**) slices of colonic tumor tissue staining by Antibody HPA052066 and HPA052512; (**C** and **G**) slices of rectal normal tissue staining by Antibody HPA052066 and HPA052512; and (**D** and **H**) slices of rectal tumor tissue staining by Antibody HPA052066 and HPA052512.)

Table 2 Clinical and Pathological Parameters of 373 Patients with CRC from the TCGA Cohort

Variables	Categories	No. of Patients	MST (Days)	Log Rank Test, *P*-value
Gender	Male Female	204 169	2,475 -	0.373
Age	≤60years >60years	144 229	- 2,134	0.006*
STAGE	I II III IV	58 138 118 59	- 3,042 1,741 1,566	<0.001*
Chemotherapy	Yes No NA	39 31 303	1,910 1,711	0.348
MSI Status	MSI-L MSS MSI-H NA	63 248 51 11	2,003 - 2,532	0.566
KRAS	Mutation Wild Type NA	28 30 315	1,162 1,741	0.351
BRAF	Mutation Wild Type NA	3 32 338	- -	0.219
SLC26A3	High Low	186 187	- 2,475	0.018*
AQP8	High Low	186 187	- 2,047	0.011*
ZG16	High Low	186 187	- 1,741	0.002*

Note: P-values were calculated by the Kaplan–Meier method with a Log rank test (**P*-value<0.05).

Abbreviations: MST, median survival time; MSI, microsatellite instability (MSI-L, microsatellite instability-low; MSS, microsatellite stability; MSI-H, microsatellite instability-high); NA, not available.

During the last decade, it has become clear that genomic differences are highly relevant to the progression and prognosis of CRC. In this study, we discovered 88 DEGs by comprehensively analyzing three microarray datasets from GEO and RNA sequencing data from TCGA. GO enrichment and KEGG pathway analysis revealed that the DEGs were primarily involved in metallopeptidase activity, brush border and brush border membrane, carbon-oxygen lyase activity, mineral absorption, and nitrogen metabolism. We then conducted PPI network analysis to identify the interactions between DEGs. Using the CytoHubba, 15 hub genes, including *ZG16*, were detected in the PPI network. Next, the differential expression of *ZG16* between tumor and normal colorectal tissues were further verified in the GEPIA and HAP database. Subsequent univariate and multivariate Cox regression analysis indicated that *ZG16* expression was inversely correlated with OS of CRC patients. In general, the results are in line with multiple previous studies, which indicated that loss of *ZG16* expression may make a great difference in CRC development and worse OS.[18–21]

In fact, the TNM system based on the depth of tumor invasion (T), the number of affected lymph nodes (N), and the presence of metastasis (M) remains the most common prognostic indicator for stratifying patients. However, survival rates are highly heterogeneous for even patients who are within the same TNM stage.[22] Furthermore, many risk factors (eg, age, family history, and inflammatory bowel disease) can influence the prognosis of CRC patients. In the present study, we developed a prediction model based on TNM status, age, and *ZG16* for the long-term survival of CRC patients. Finally, Kaplan–Meier curve and DCA curve indicated the significant prognostic value of the

Table 3 The Univariate and Multivariate Cox Regression Analysis of 373 CRC Patients' OS from the TCGA Cohort

Factor	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Gender (Female)	0.821 (0.532–1.268)	0.374	–	–
Age (>60 years)	1.988 (1.203–3.286)	0.007*	2.032 (1.213–3.404)	0.007*
TNM stage (Ⅳ)	2.780 (1.695–4.558)	<0.001*	4.076 (2.388–6.970)	<0.001*
ZG16 (High/Low)	0.489 (0.309–0.774)	0.002*	0.567 (0.347–0.929)	0.024*
AQP8 (High/Low)	0.566 (0.363–0.882)	0.012*	0.803 (0.468–1.378)	0.426
SLC26A3 (High/Low)	0.583 (0.371–0.916)	0.019*	0.665 (0.376–1.179)	0.163

Note: **P*-value<0.05.

Abbreviations: HR, hazard ratio; CI, confidence interval.

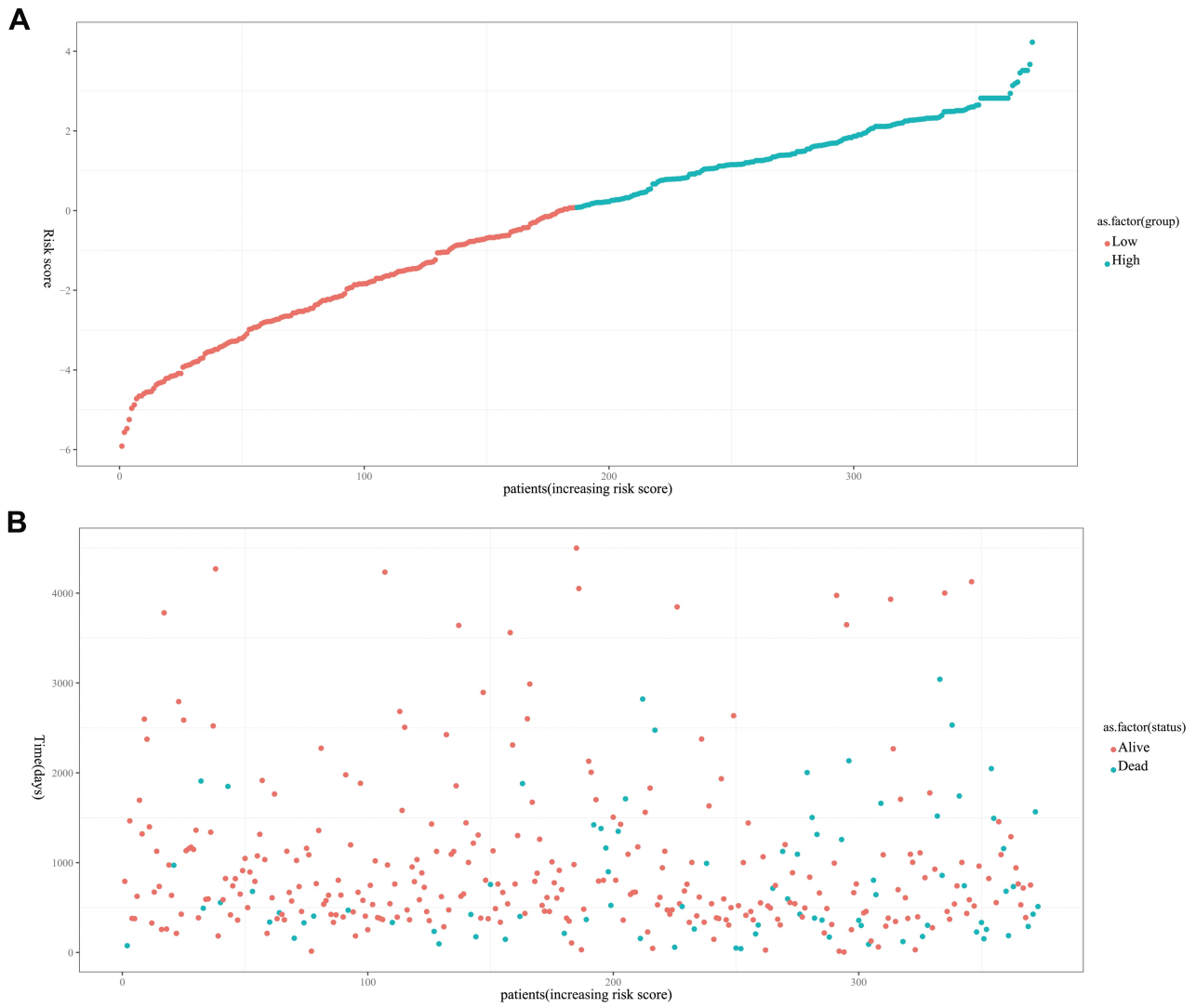

Figure 7 The risk score distribution (**A**) and survival time statistic (**B**) for CRC patients.

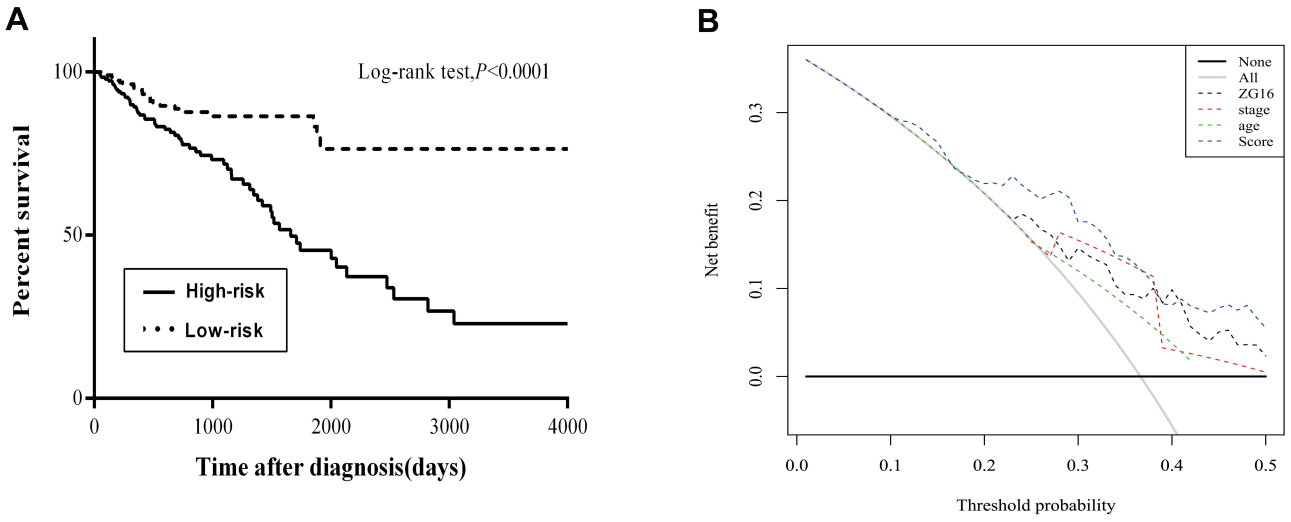

Figure 8 The overall survival curves of high-risk group and low-risk group divided by prognostic score based on *ZG16* (**A**). The DCA curves of multiple predictive models (**B**).

Table 4 Clinical and Baseline Parameters of 562 CRC Patients from the GSE40967 Cohort

Variables	Categories	No. of Patients	MST (Months)	Log Rank Test, *P*-value
Gender	Male Female	307 255	106 -	0.029*
Age	≤70 years >70 years NA	321 240 1	- 83	<0.001*
STAGE	I II III IV	33 264 205 60	- 183 112 27	<0.001*
Chemotherapy	Yes No NA	233 312 17	145 106	0.641
MMR status	dMMR pMMR NA	74 441 47	- 112	0.308
TP53	Mutation Wild Type NA	190 160 212	105 -	0.314
KRAS	Mutation Wild Type NA	213 328 21	132 145	0.040*
BRAF	Mutation Wild Type NA	51 457 54	- 145	0.677
ZG16	High Low	464 98	145 76	0.009*

Note: *P*-values were calculated by the Kaplan–Meier method with a Log rank test (**P*-value<0.05).
Abbreviations: MST, median survival time; MMR, mismatch repair (pMMR, proficient mismatch repair; dMMR, deficient mismatch repair); NA, not available.

prognostic score incorporating *ZG16*. To validate these results, we carried out comprehensive survival analysis of 562 CRC patients from the GSE40967, which revealed that low *ZG16*, older age, and stage Ⅳ was associated with worse OS of CRC patients. And we found that *ZG16* and *KRAS* might contribute to classify the stage II–III patients who could benefit from the adjuvant chemotherapy. Additionally, the combined use of *ZG16*, TNM stage, and age greatly increased the prognostic efficacy for CRC patients after treatment. Although numerous biomarkers arose over the past few years, only the *KRAS* gene has entered routine clinical practice. Some research delivered that mutations in KRAS exon were associated with poor prognosis as well as lower survival,[23] which were coincident with ours result of the stage II–III CRC patients receiving chemotherapy. However, *KRAS* was used to predictthe response to EGFR-targeted therapies in stage Ⅳ CRC.[24] As for the low-grade CRC patients and those who did not treat with target therapy, *ZG16* or the risk score might provide us new choices for prediction patients' outcome.

In conclusion, the expression of *ZG16* was negatively correlated with the OS of CRC patients, and is, therefore, considered as a promising biomarker for predicting the prognosis of CRC. In addition, our study constructed a combined score with superior performance as a potential predictor for CRC patients. These findings can contribute to the development of novel strategies for diagnosis and prognostic prediction of CRC patients. However, there were some limitations in our study: 1) our results were not be validated by experiment using local specimens; and 2) there were no definite treatment lines in the data enrolled in our research.

Table 5 The Univariate and Multivariate Cox Regression Analysis of 562 CRC Patients' OS from the GSE40967 Cohort

Factor	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Gender (Female)	0.685 (0.486–0.965)	0.030*	0.978 (0.640–1.495)	0.920
Age (>70 years)	1.967 (1.476–2.622)	<0.001*	2.220 (1.548–3.185)	<0.001*
TNM stage (Ⅳ)	4.738 (3.312–6.777)	<0.001*	5.401 (3.732–7.817)	<0.001*
ZG16 (High/Low)	0.644 (0.463–0.896)	0.008*	0.639 (0.458–0.892)	0.008*
Chemotherapy (Yes)	1.073 (0.692–1.255)	0.642	–	–
MMR (dMMR/pMMR)	0.776 (0.476–1.265)	0.310	–	–
TP53 (M/WT)	1.196 (0.844–1.696)	0.315	–	–
KRAS (M/WT)	1.355 (1.013–1.812)	0.041*	1.188 (0.883–1.599)	0.255
BRAF (M/WT)	1.115 (0.666–1.868)	0.679	–	–

Note: **P*-value<0.05.
Abbreviations: HR, hazard ratio; CI, confidence interval.

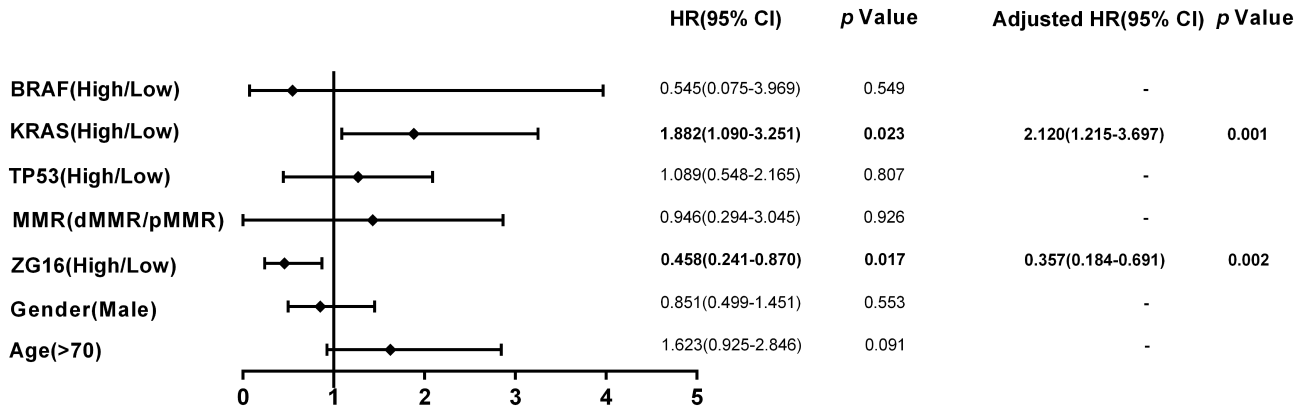

Figure 9 The Forest plot based on univariate and multivariate Cox regression analysis of 203 stage II–III CRC patients who received chemotherapy from GSE40967.

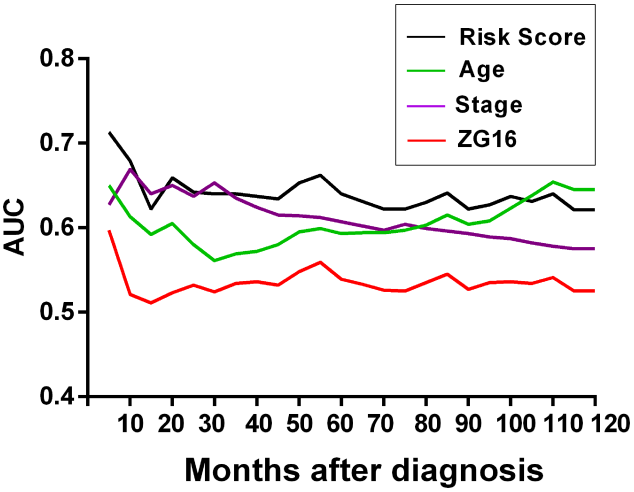

Figure 10 Time-dependent ROC of biomarkers for predicting overall survival of CRC patients after treatment.

Data Sharing Statement

The datasets analyzed for this study can be found in the GEO and TCGA database.

Funding

This work was supported by the Health and Wellness Committee of Jiangxi province, China [grant numbers 20204251].

Disclosure

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

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*. 2018.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683–691. doi:10.1136/gutjnl-2015-310912
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34. doi:10.3322/caac.21551
4. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: good or bad news from the 2018 global cancer statistics? *Cancer Commun*. 2019;39(1):22. doi:10.1186/s40880-019-0368-6
5. Simon K. Colorectal cancer development and advances in screening. *Clin Interv Aging*. 2016;11:967–976. doi:10.2147/CIA.S109285
6. Nitsche U, Maak M, Schuster T, et al. Prediction of prognosis is not improved by the seventh and latest edition of the TNM classification for colorectal cancer in a single-center collective. *Ann Surg*. 2011;254(5):-793–800; discussion 800–791. doi:10.1097/SLA.0b013e3182369101
7. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091–1093. doi:10.1093/bioinformatics/btp101
8. Markovic S, Dimitrijevic I, Zogovic B, Markovic V, Barisic G, Krivokapic Z. Current trends in clinical genetics of colorectal cancer. *J buon*. 2016;21(5):1042–1049.
9. Kolenčík D, Shishido SN, Pitule P, Mason J, Hicks J, Kuhn P. Liquid biopsy in colorectal carcinoma: clinical applications and challenges. *Cancers*. 2020;12:6.
10. Mito A, Nakano Y, Saitoh T, et al. Lectin ZG16p inhibits proliferation of human colorectal cancer cells via its carbohydrate-binding sites. *Glycobiology*. 2018;28(1):21–31. doi:10.1093/glycob/cwx088
11. Tateno H, Yabe R, Sato T, et al. Human ZG16p recognizes pathogenic fungi through non-self polyvalent mannose in the digestive system. *Glycobiology*. 2012;22(2):210–220. doi:10.1093/glycob/cwr130
12. Rodríguez-Piñeiro AM, Bergström JH, Ermund A, et al. Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. *Am J Physiol Gastrointest Liver Physiol*. 2013;305(5):G348–356. doi:10.1152/ajpgi.00047.2013
13. Johansson ME, Gustafsson JK, Holmén-Larsson J, et al. Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut*. 2014;63(2):281–291. doi:10.1136/gutjnl-2012-303207
14. Kanagawa M, Satoh T, Ikeda A, et al. Crystal structures of human secretory proteins ZG16p and ZG16b reveal a Jacalin-related β-prism fold. *Biochem Biophys Res Commun*. 2011;404(1):201–205. doi:10.1016/j.bbrc.2010.11.093
15. Meng H, Li W, Boardman LA, Wang L. Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer. *BMC Cancer*. 2018;18(1):433. doi:10.1186/s12885-018-4337-2

16. Pelaseyed T, Bergström JH, Gustafsson JK, et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev*. 2014;260(1):8–20. doi:10.1111/imr.12182
17. Chen X, Du P, She J, Cao L, Li Y, Xia H. Loss of ZG16 is regulated by miR-196a and contributes to stemness and progression of colorectal cancer. *Oncotarget*. 2016;7(52):86695–86703. doi:10.18632/oncotarget.13435
18. Escudero-Paniagua B, Bartolomé RA, Rodríguez S, et al. PAUF/ZG16B promotes colorectal cancer progression through alterations of the mitotic functions and the Wnt/β-catenin pathway. *Carcinogenesis*. 2020;41(2):203–213.
19. Gao X, Yang J. Identification of genes related to clinicopathological characteristics and prognosis of patients with colorectal cancer. *DNA Cell Biol*. 2020;39(4):690–699. doi:10.1089/dna.2019.5088
20. Xu H, Ma Y, Zhang J, et al. Identification and verification of core genes in colorectal cancer. *Biomed Res Int*. 2020;2020:8082697.
21. Zhao ZW, Fan XX, Yang LL, et al. The identification of a common different gene expression signature in patients with colorectal cancer. *Math Biosci Eng*. 2019;16(4):2942–2958. doi:10.3934/mbe.2019145
22. Sagaert X, Vanstapel A, Verbeek S. Tumor heterogeneity in colorectal cancer: what do we know so Far? *Pathobiology*. 2018;85(1–2):72–84. doi:10.1159/000486721
23. Li W, Qiu T, Zhi W, et al. Colorectal carcinomas with KRAS codon 12 mutation are associated with more advanced tumor stages. *BMC Cancer*. 2015;15:340.
24. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int J Mol Sci*. 2017;18:1. doi:10.3390/ijms18010197