Prognostic Nomogram of Prognosis-Related Genes and Clinicopathological Characteristics to Predict the 5-Year Survival Rate of Colon Cancer Patients

Chao Huang, Jiefeng Zhao and Zhengming Zhu*

Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China

Background: The Cancer Genome Atlas (TCGA) has established a genome-wide gene expression profile, increasing our understanding of the impact of tumor heredity on clinical outcomes. The aim of this study was to construct a nomogram using data from the TCGA regarding prognosis-related genes and clinicopathological characteristics to predict the 5-years survival rate of colon cancer (CC) patients.

Methods: Kaplan–Meier and Cox regression analyses were used to identify genes associated with the 5-years survival rate of CC patients. Cox regression was used to analyze the relationship between the clinicopathological features and prognostic genes and overall survival rates in patients with CC and to identify independent risk factors for the prognosis of CC patients. A nomogram for predicting the 5-years survival rate of CC patients was constructed by R software.

Results: A total of eight genes (KCNJ14, CILP2, ATP6V1G2, GABRD, RIMKLB, SIX2, PLEKHA8P1, and MPP2) related to the 5-years survival of rate CC patients were identified. Age, stage, and PLEKHA8P1 were independent risk factors for the 5-years survival rate in patients with CC. The accuracy, sensitivity and specificity of the nomogram model constructed by age, TNM staging, and PLEKHA8P1 for predicting the 5-years survival of rate CC patients were 83.3, 83.97, and 85.79%, respectively.

Conclusion: The nomogram can correctly predict the 5-year survival rate of patients with CC, thus aiding the individualized decision-making process for patients with CC.

Keywords: colon cancer, prognostic genes, clinicopathological characteristics, overall survival, nomogram

INTRODUCTION

Colon cancer (CC) ranks third in incidence and second in mortality rates (1). Surgical treatment is the main method for managing CC to prolong survival time (2, 3). Adjuvant chemoradiotherapy can also significantly improve the prognosis of CC (4, 5). The 5-years survival rates of patients with stage I, II, and III CC are ∼93, 80, and 60%, respectively (6). The American Joint Committee on Cancer (AJCC) TNM staging system is widely used to assess the prognosis of patients with CC (5). However, the prognosis of patients with CC at the same stage varies widely, and the accuracy of

TMN staging as a predictive approach has certain limitations (7, 8). Therefore, another approach is needed to identify patients with poor prognosis to allow for the development of individualized treatment and monitoring approaches. Nomograms can provide an overall probability of a specific outcome for an individual patient and provide more accurate predictions than traditional staging systems, thereby improving personalized treatment decisions (9, 10). Previously developed microarray techniques can be used to predict the prognosis of many types of cancer (11–13). Previous studies have shown that gene expression profiles have certain application prospects in predicting patients' long-term prognosis (14). Meanwhile, prognostic gene expression profiles of colorectal cancer patients from tumor samples and adjacent normal mucosa have been described (15–17). Relevant studies have indicated that gene expression characteristics can improve the accuracy of prognosis prediction for stage II and III colorectal cancer (18, 19). However, few studies have combined prognostic genes with clinicopathological features to predict the long-term survival of patients with CC. In addition, The Cancer Genome Atlas (TCGA) has established a genome-wide gene expression profile, increasing our understanding of the impact of tumor heredity on clinical outcomes (20). Therefore, the aim of this study was to construct a nomogram using data from the TCGA regarding prognosis-related genes and clinicopathological characteristics to predict the 5-years survival rate of CC patients, thus providing an important basis for individualized decision-making for patients with CC.

MATERIALS AND METHODS

Data Download and Processing

RNA sequencing results from 437 tissues and 382 human colon adenoma and adenocarcinoma samples were obtained from the TCGA database (https://portal.gdc.cancer.gov). RNA sequencing results from 39 normal samples and 398 cancer samples were combined into a single matrix file using scripts in the Perl language (http://www.perl.org/). The Ensembl database (http://www.ensembl.org/index.html) was then used to convert the Ensembl ID in the matrix file to the gene name. Moreover, the clinical data of 385 cases were downloaded, and relevant clinical data were extracted.

Identification of Prognostic-Related Genes

First, Kaplan–Meier and Cox regression analyses were used to screen for genes associated with the 5-years survival of rate CC patients, and a $P < 0.05$ was used to define statistical significance. Next, the survivalROC package in R language was used to identify genes that were associated with 5-years survival and had an area under the curve (AUC) >0.6.

Survival Analysis

To determine the relationship between prognostic genes and CC survival, we used the survival package in R language for the survival analysis of the prognostic genes. The relationship between the clinicopathological characteristics and prognosis-related genes and the overall survival of patients with CC

was analyzed by a univariate analysis. The factors affecting CC survival in the univariate analysis were analyzed by multivariate Cox regression to identify independent risk factors for CC prognosis.

ROC Curve Analysis

To determine the accuracy of the combined factors to predict the 5-years survival rate of CC patients and the cutoff value of prognostic genes, we used the survivalROC package in R language for analysis. In addition, the sensitivity and specificity of the combined factors were calculated using the survivalROC package in R language.

Construction of Nomogram

The combined factors that predict the most accurate prognosis of CC were used to construct a nomogram model for predicting the 5-years survival rate of CC patients using the rms package in R language, and scores for various indicators were obtained. The scores corresponding to the indicators were added to obtain a total score; the higher the total score, the lower the 5-years survival rate of CC patients. Meanwhile, the survivalROC package was used to calculate the sensitivity and specificity of the model to evaluate its

TABLE 1 | Clinical characteristics of patients with colon cancer.

Clinical characteristics	n (%)
Age	
≤65 years	148 (40.66)
>65 years	216 (59.34)
Sex	
Male	192 (52.75)
Female	172 (47.25)
Stage	
I	63 (17.31)
II	146 (40.11)
III	102 (28.02)
IV	53 (14.56)
T classification	
T1	7 (1.92)
T2	64 (17.58)
T3	251 (68.96)
T4	42 (11.54)
M classification	
M0	278 (76.37)
M1	53 (14.56)
Mx	33 (9.07)
N classification	
N0	216 (59.34)
N1	85 (23.35)
N2	63 (17.31)
Survival status	
Death	65 (17.86)
Survival	299 (82.14)

clinical value. Moreover, the concordance index (C-index) was calculated to evaluate the performance of the model prediction results, and the calibration curve was plotted to observe the relationship between the predicted probability and the actual incidence (21, 22).

RESULTS

Clinical Characteristics of Patients

From the clinical data of 385 patients, the patients' age, sex, stage, TNM staging, survival time and survival status were extracted. After deleting samples with incomplete clinical data, a total of 364 cases were retained for further analysis (**Table 1**).

Prognosis and Survival Analysis

A total of eight genes (KCNJ14, CILP2, ATP6V1G2, GABRD, RIMKLB, SIX2, PLEKHA8P1, and MPP2) associated with the

5-years survival rate of CC patients were identified (**Table 2**). The Kaplan–Meier analysis showed that all eight genes were correlated with the prognosis of CC (all $P < 0.05$) (**Figure 1**). The univariate analysis showed that the factors related to the overall survival rate of patients with CC were as follows: age ($P = 0.010$), stage ($P < 0.001$), T classification ($P < 0.001$), M classification ($P < 0.001$), N classification ($P < 0.001$), KCNJ14 ($P < 0.001$), CILP2 ($P = 0.014$), ATP6V1G2 ($P < 0.001$), GABRD ($P = 0.005$),

TABLE 2 | Prognostic-related gene in patients with colon cancer.

Gene	Log-rank *p*-value	Hazard ratio (95%CI)	Cox *p*-value	AUC
KCNJ14	0.021	2.04 (1.19–3.51)	0.009	0.611
CILP2	0.025	1.25 (1.05–1.49)	0.014	0.613
ATP6V1G2	0.020	6.18 (2.48–15.36)	0.000	0.627
GABRD	0.017	1.21 (1.06–1.38)	0.005	0.634
RIMKLB	0.008	1.65 (1.14–2.39)	0.008	0.635
SIX2	0.013	1.17 (1.09–1.26)	0.000	0.638
PLEKHA8P1	0.000	1.54 (1.12–2.12)	0.009	0.656
MPP2	0.000	4.11 (2.13–7.92)	0.000	0.710

TABLE 3 | Univariate analysis and multivariate analysis of the correlation of prognostic-related genes with overall survival among patients with colon cancer.

Clinicopathologic variable	Univariate analysis		Multivariate analysis	
	HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Age	1.03 (1.01–1.05)	0.010	1.05 (1.02–1.07)	0.000
Sex	1.33 (0.81–2.18)	0.267		
Stage	2.26 (1.70–3.00)	0.000	2.13 (1.32–3.43)	0.002
T classification	2.68 (1.64–4.37)	0.000	1.27 (0.68–2.35)	0.455
M classification	1.85 (1.37–2.51)	0.000	1.45 (0.97–2.16)	0.072
N classification	1.99 (1.50–2.65)	0.000	0.95 (0.60–1.49)	0.815
KCNJ14	2.05 (1.19–3.52)	0.009	1.02 (0.55–1.91)	0.941
CILP2	1.25 (1.05–1.49)	0.012	1.16 (0.90–1.49)	0.243
ATP6V1G2	5.61 (2.20–14.30)	0.000	0.67 (0.08–5.41)	0.705
GABRD	1.21 (1.06–1.38)	0.005	0.97 (0.76–1.24)	0.819
RIMKLB	1.31 (1.01–1.71)	0.043	1.08 (0.59–1.97)	0.803
SIX2	1.17 (1.09–1.26)	0.000	1.08 (0.97–1.21)	0.164
PLEKHA8P1	1.54 (1.12–2.13)	0.008	1.48 (1.03–2.13)	0.035
MPP2	4.08 (2.12–7.83)	0.000	2.94 (0.90–9.65)	0.075
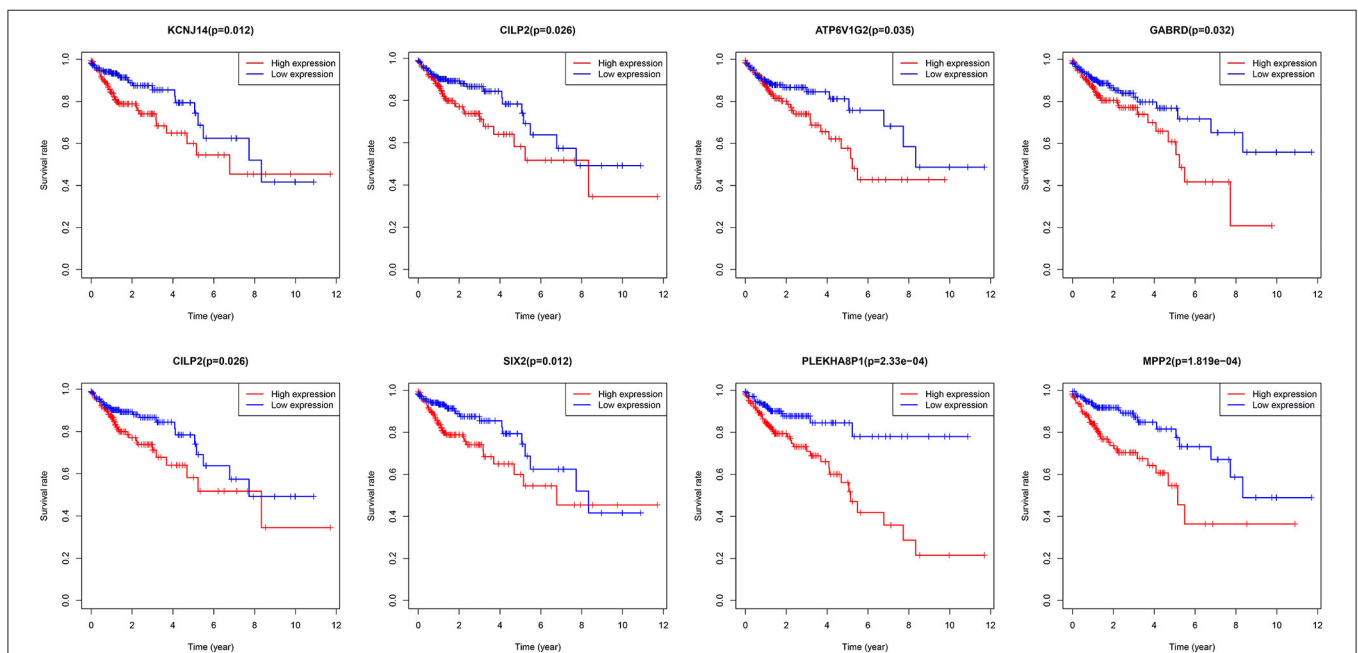
FIGURE 1 | Survival curves of eight genes, KCNJ14, CILP2, ATP6V1G2, GABRD, RIMKLB, SIX2, PLEKHA8P1, and MPP2.

FIGURE 2 | Forest plot of the correlation of prognostic-related genes and clinicopathological features with overall survival among patients with colon cancer. *P < 0.05, **P < 0.01, ***P < 0.001.

FIGURE 3 | Expression level of PLEKHA8P1 in colon tumors and normal tissues.
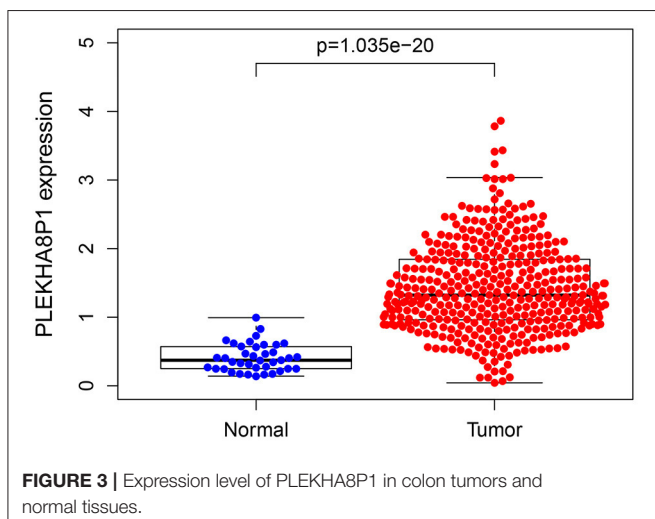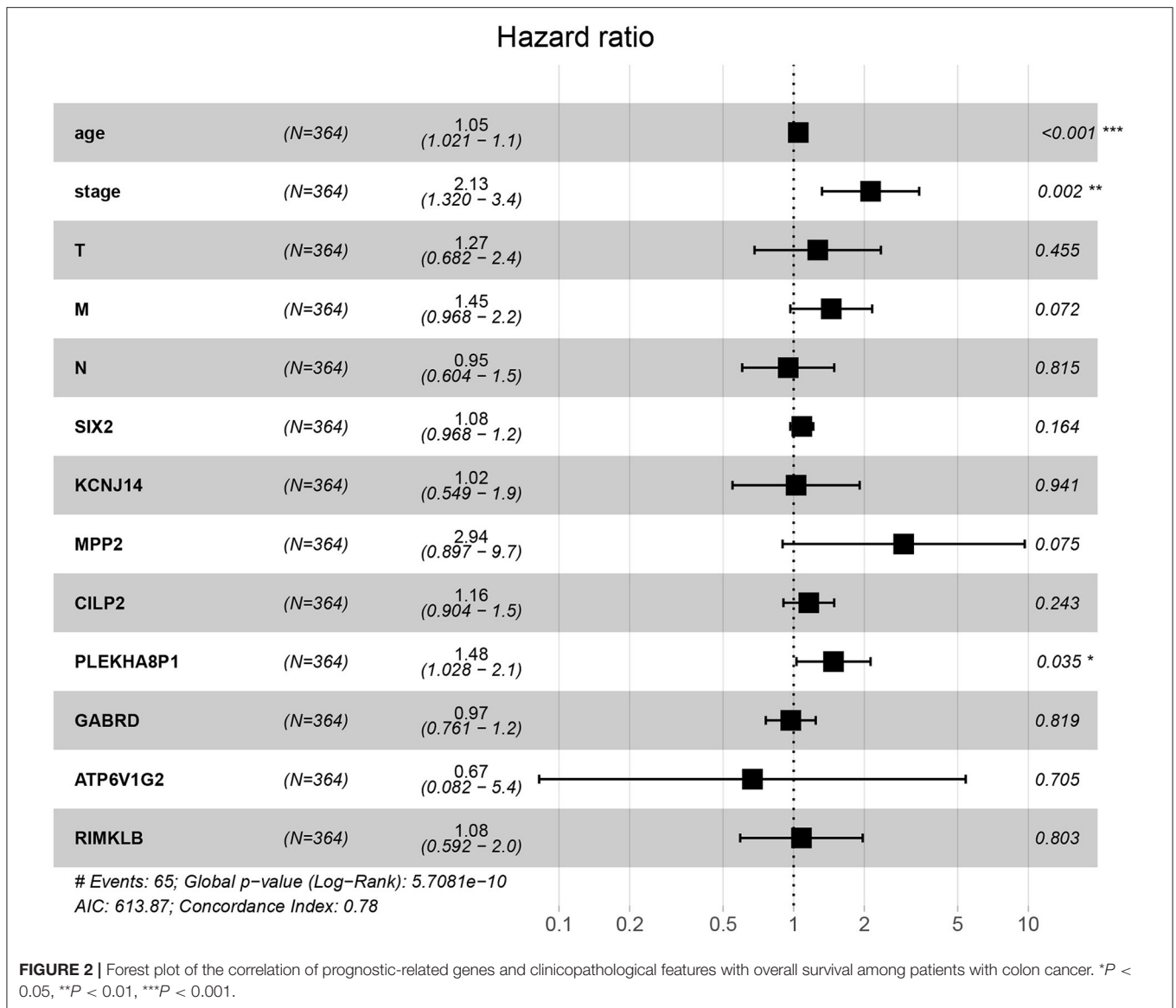
RIMKLB ($P = 0.008$), SIX2 ($P < 0.001$), PLEKHA8P1 ($P = 0.009$), and MPP2 ($P < 0.001$). The multivariate Cox regression analysis showed that age ($P = 0.010$), stage ($P < 0.001$), and PLEKHA8P1 ($P = 0.009$) were independent risk factors for the prognosis of patients with CC (**Table 3** and **Figure 2**). Moreover, PLEKHA8P1 was highly expressed in colon tumors ($P < 0.001$) (**Figure 3**).

ROC Curve Analysis

The cutoff values for age and PLEKHA8P1 to evaluate 5-years survival in patients with CC were 65 years and 1.545, respectively. The AUC, sensitivity and specificity of age combined stage and PLEKHA8P1 for assessing the 5-year survival rate of CC patients were 0.761, 74.50, 67.40%, respectively. The AUC, sensitivity and specificity of TNM staging for assessing the 5-years survival rate of CC patients were 0.718, 78.25, and 63.57%, respectively. The AUC, sensitivity and specificity of age combined with TNM staging and PLEKHA8P1 for assessing the 5-years survival rate
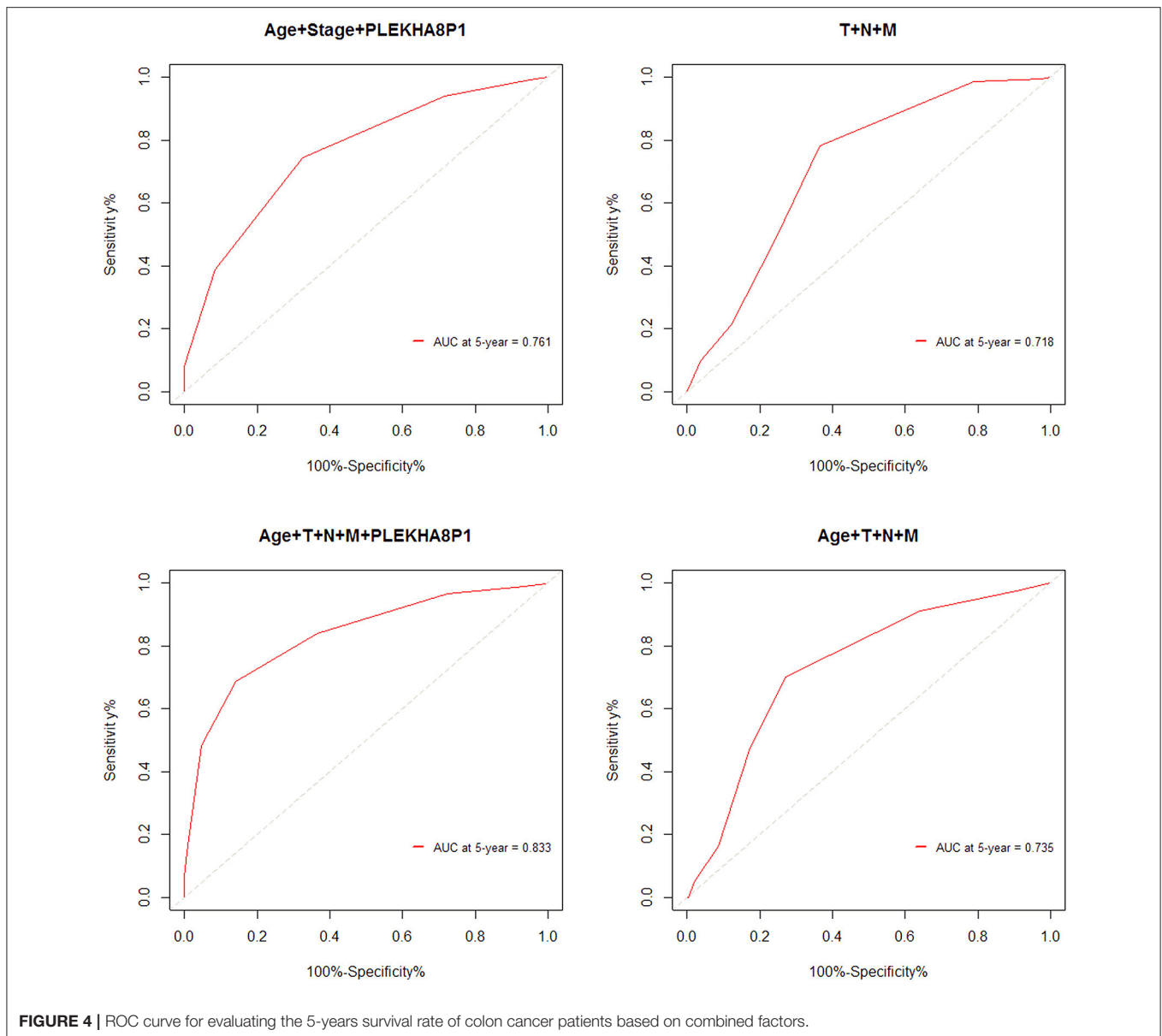
FIGURE 4 | ROC curve for evaluating the 5-years survival rate of colon cancer patients based on combined factors.

of CC patients were 0.833, 83.97, and 85.79%, respectively. The AUC, sensitivity and specificity of age combined with TNM staging for assessing the 5-years survival rate of CC patients were 0.735, 70.21, and 73.06%, respectively (**Figure 4**). These results indicated that age combined with TNM staging and PLEKHA8P1 were most accurate for evaluating the 5-years survival rate of CC patients.

Construction of Nomogram

The rms package in R language was used to construct a logistic regression model constructed by age, TNM staging and PLEKHA8P1, and the C-index for evaluation was 0.74, indicating that the prediction model was accurate. Then, the plotting function was constructed, and the nomogram was plotted (**Figure 5**). A score of age ≤65 years was 0 points, while a

score of age >65 years was 60 points; a score of T1 was 0 points; a score of T2 was 33 points; a score of T3 was 67 points; a score of T4 was 100 points; a score of N0 was 0 points; a score of N1 was 36 points; a score of N2 was 72 points; a score of M0 was 0 points; a score of M1 was 39 points; a score of Mx was 78 points; and a score of PLEKHA8P1 ≤ 1.545 U/ml was 0 points, while a score of PLEKHA8P1 > 1.545 U/ml was 48 points. The highest score was 358 points, suggesting that the 5-years survival probability of patients with CC was <10%. The probability of 5-years survival of CC can be predicted based on the total points (**Table 4**). The accuracy, sensitivity and specificity of this prediction model were 83.30, 83.97, and 85.79%, respectively, indicating the validity of the model. The calibration curve was closer to the ideal curve, which indicated that the prediction was in good agreement with the actual results (**Figure 6**).
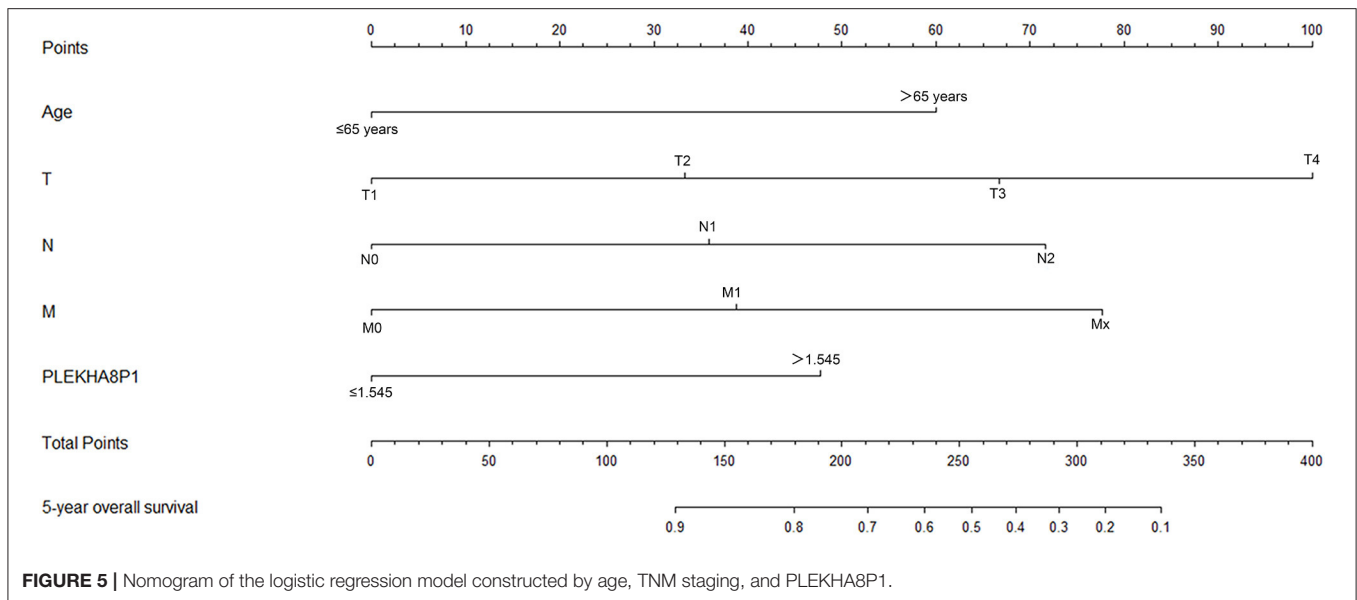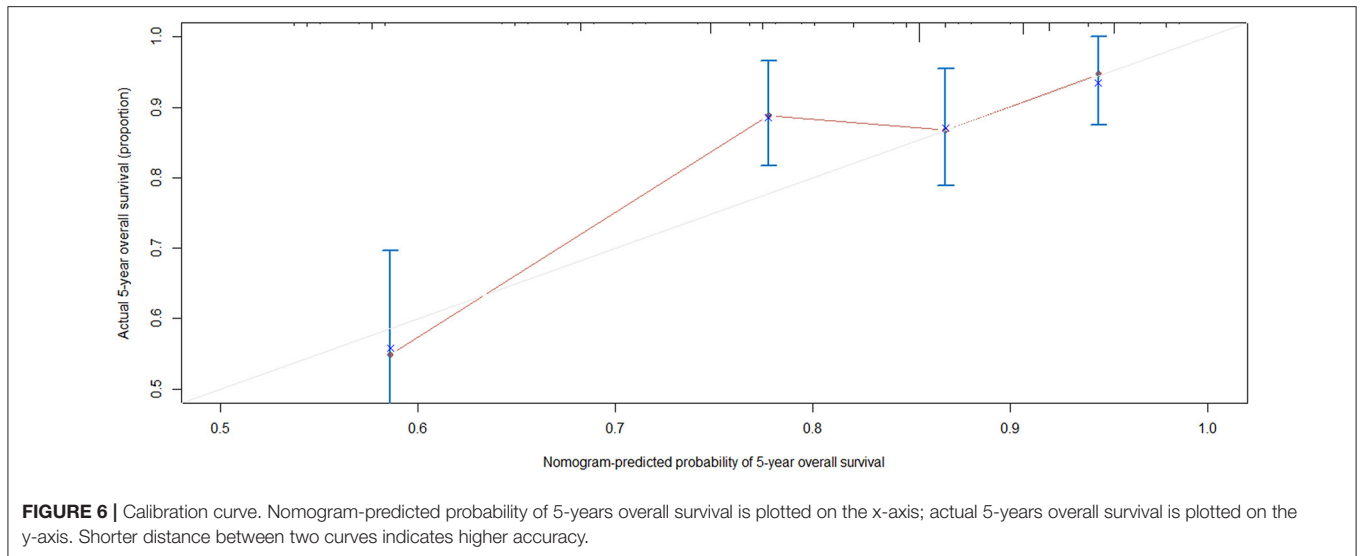
FIGURE 5 | Nomogram of the logistic regression model constructed by age, TNM staging, and PLEKHA8P1.

TABLE 4 | Relationship between total points and 5-years overall survival in colon cancer.

Total points	5-years overall survival
>336	<10%
312–336	10–20%
292–311	21–30%
274–291	31–40%
255–273	41–50%
235–254	51–60%
211–234	61–70%
179–210	71–80%
129–178	81–90%
<129	>90%

DISCUSSION

This study is the first to combine prognostic genes and clinicopathological characteristics of CC to predict the 5-years survival rate of CC patients. We first performed a series of analyses to identify genes that significantly affected 5-years survival in CC. Then, the relationship between these genes and clinicopathological characteristics and the overall survival rate of CC was analyzed, and independent risk factors for CC survival were identified. Finally, a logistic regression model was constructed based on the AUC of the combined factors, and a nomogram was drawn to predict the 5-years overall survival rate of CC patients. This study found that age, stage, and PLEKHA8P1 were independent risk factors for the 5-years survival rate in patients with CC. PLEKHA8P1 belongs to the pseudogene family. Only ∼2% of the genes in the human genome encode proteins. Non-coding RNAs include microRNAs, long non-coding RNAs, and pseudogenes (23–25). Currently, the functions and mechanisms of lncRNAs and pseudogenes have not been fully elucidated (24–26). However, an increasing number of studies have shown that pseudogenes have important biological functions (27, 28). In the process of homologous recombination, pseudogenes may result in the loss of some bases, thus affecting the transcription level of genes (29). Pseudogenes can also induce endogenous small interfering RNAs to inhibit the expression of functional genes (25). Pseudogene RNAs can play a regulatory role as competing endogenous RNAs (26, 30). On the other hand, the results of an increasing number of studies have indicated that pseudogenes play a crucial role in cancer. Chen et al. (31) found that the pseudogene CTNNAP1 promotes the growth of human tumors by regulating the expression of its homologous gene, CTNNA1. Lin et al. (32) showed that the pseudogene OCT4-pg could inhibit the growth and differentiation of mesenchymal stem cells. Rutnam et al. indicated that the pseudogene TUSC2p1 protects the expression of the tumor suppressor gene TUSC2 by competitively binding with miRNA, thereby inhibiting the proliferation of breast cancer cells (33). Poliseno et al. (25) demonstrated that the pseudogene PTENP1 had the ability to produce the corresponding mRNA and can interact with the transcription products of the parent gene PTEN, thus playing a role in inhibiting cell growth. Poliseno et al. (34) also found deletion of the pseudogene PTENP1 in some CC, gastric cancer and malignant melanoma. Moreover, the expression of some pseudogenes is related to the staging and grading of cancer and can be a molecular marker for the prognosis of cancer. The increased expression level of the pseudogene OCT4-pgq1 was closely associated with poor prognosis in gastric cancer and could lead to worse overall patient survival rates (35). PLEKHA8P1 expression was significantly correlated with the monthly survival rate and monthly disease-free survival rate of renal cell carcinoma patients, suggesting that its expression changes play a key role in predicting the prognosis of renal cell carcinoma (36). This study also showed that PLEKHA8P1 was significantly associated with the 5-years survival rate of CC patients and was highly expressed in colon tumors.

FIGURE 6 | Calibration curve. Nomogram-predicted probability of 5-years overall survival is plotted on the x-axis; actual 5-years overall survival is plotted on the y-axis. Shorter distance between two curves indicates higher accuracy.

The American Joint Committee on Cancer TNM staging system is widely used for the prognostic evaluation of CC patients (5). However, Liu et al. (8) indicated that the MSKCC nomogram was better than the AJCC staging system for predicting the 5-years survival rate, and the C-index of the MSKCC nomogram in the studied Chinese cohort was 0.71. Weiser et al. (37) demonstrated that a prognostic model including prognostic factors was superior to the current AJCC system, and its C-index increased from 0.60 to 0.68. The applicability of gene expression profiles for predicting the prognosis of colorectal cancer patients has been demonstrated in several studies (19, 38, 39). Barrier et al. (38) showed that microarray gene expression profile analysis can predict the prognosis of patients with stage II CC. Lee et al. (40) found that a nomogram model including TNM staging and genetic risk score obtained from the TCGA database could successfully predict the overall survival rate of colorectal cancer patients, and its C-index was higher than that of TNM staging alone (0.75 vs. 0.69). The prognostic prediction model constructed by pathologic M combined with pathologic T had a prognostic prediction efficiency with a 5-years AUC of 0.712 and C-index of 0.680 for patients with colon adenocarcinoma (41). Another prognostic model composed of six significant prognostic factors (age, first-degree relative cancer history, differentiation grade, vessel/nerve invasion, TNM stage, and HALP) had a 5-years AUC of 0.73 for patients with locally advanced colorectal cancer (42). The prognostic nomogram constructed by age, sex, histological grade, T stage, number of lymph nodes retrieved, tumor size and N stage had a 5-years AUC of 0.729 for patients with non-metastatic CC (43). In this study, the accuracy, sensitivity and specificity of age combined with TNM staging and PLEKHA8P1 for predicting the 5-years survival rate of CC patients were higher than those of the TNM staging system. In addition, the C-index of the model constructed by age, TNM staging, and PLEKHA8P1 for predicting the 5-years survival rate was 0.74, and its accuracy, sensitivity, and specificity were 83.3,

83.97, and 85.79%, respectively, indicating that the model has high validity. There are some limitations in this study. First, the mRNA gene expression value is difficult to obtain due to the high cost in clinical practice. However, when the cost is reduced, this approach could be widely used in clinical practice. Second, other prognostic factors, such as tumor markers and inflammatory markers, were not included.

In conclusion, age, PLEKHA8P1 and stage were risk factors for poor patient prognosis in CC. The nomogram model constructed by age, TNM staging, and PLEKHA8P1 can correctly predict the 5-years survival rate of patients with CC, thus aiding individualized decision-making for patients with CC. Moreover, the results of this study also provide some direction for future fundamental research. However, the biological function and molecular mechanism of PLEKHA8P1 need further study.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found at https://portal.gdc.cancer.gov.

AUTHOR CONTRIBUTIONS

CH designed the study, analyzed the data, and wrote the manuscript with contributions from all authors. CH and JZ collected data. ZZ provided critical comments for this paper. All authors read and approved the final version of the paper.

FUNDING

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) 68:394–424. doi: 10.3322/caac.21492

2. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet.* (2019) 394:1467–80. doi: 10.1016/s0140-6736(19)32319-0

3. Schmoll HJ, Van Cutsem E, Stein A, Valentini V, Glimelius B, Haustermans K, et al. ESMO consensus guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. *Ann Oncol.* (2012) 23:2479–516. doi: 10.1093/annonc/mds236

4. Ng SC, Wong SH. Colorectal cancer screening in asia. *Br Med Bull.* (2013) 105:29–42. doi: 10.1093/bmb/lds040

5. Edge SB, Compton CC. The american joint committee on cancer: the 7th edition of the aJCC cancer staging manual and the future of tNM. *Ann Surg Oncol.* (2010) 17:1471–4. doi: 10.1245/s10434-010-0985-4

6. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new american joint committee on cancer sixth edition staging. *J Natl Cancer Instit.* (2004) 96:1420–5. doi: 10.1093/jnci/djh275

7. Weiser MR, Landmann RG, Kattan MW, Gonen M, Shia J, Chou J, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol.* (2008) 26:380–5. doi: 10.1200/JCO.2007.14.1291

8. Liu M, Qu H, Bu Z, Chen D, Jiang B, Cui M, et al. Validation of the memorial sloan-Kettering cancer center nomogram to predict overall survival after curative colectomy in a chinese colon cancer population. *Ann Surg Oncol.* (2015) 22:3881–7. doi: 10.1245/s10434-015-4495-2

9. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J Clin Oncol.* (2013) 31:1188–95. doi: 10.1200/JCO.2012.41.5984

10. Han D-S, Suh Y-S, Kong S-H, Lee H-J, Choi Y, Sano SAT, et al. Nomogram predicting long-term survival after d2 gastrectomy for gastric cancer. *J Clin Oncol.* (2012) 30:3834–40. doi: 10.1200/JCO.2012.4.8343

11. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, et al. Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med.* (2002) 8:68. doi: 10.1038/nm0102-68

12. Van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* (2002) 415:530. doi: 10.1038/415530a

13. Van De Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* (2002) 347:1999–2009. doi: 10.1056/NEJMoa021967

14. Quackenbush J. Microarray analysis and tumor classification. *N Engl J Med.* (2006) 354:2463–72. doi: 10.1056/NEJMra042342

15. Kerr D, Gray R, Quirke P, Watson D, Yothers G, Lavery I, et al. A quantitative multigene rT-PCR assay for prediction of recurrence in stage iI colon cancer: selection of the genes in four large studies and results of the independent, prospectively designed qUASAR validation study. *J Clin Oncol.* (2009) 27:4000. doi: 10.1200/jco.2009.27.1_suppl.4000

16. Wang Y, Jatkoe T, Zhang Y, Mutch MG, Talantov D, Jiang J, et al. Gene expression profiles and molecular markers to predict recurrence of dukes' b colon cancer. *J Clin Oncol.* (2004) 22:1564–71. doi: 10.1200/JCO.2004.08.186

17. Barrier A, Roser F, Boëlle P-Y, Franc B, Tse C, Brault D, et al. Prognosis of stage iI colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene.* (2007) 26:2642. doi: 10.1038/sj.onc.1210060

18. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene expression signature to improve prognosis prediction of stage iI and iII colorectal cancer. *J Clin Oncol.* (2011) 29:17–24. doi: 10.1200/JCO.2010.30.1077

19. Lin Y-H, Friederichs J, Black MA, Mages J, Rosenberg R, et al. Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer. *Clin Cancer Res.* (2007) 13:498–507. doi: 10.1158/1078-0432.CCR-05-2734

20. Network CGAR. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* (2008) 455:1061. doi: 10.1038/nature07385

21. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, Survival Analysis.* Springer (2015). doi: 10.1007/978-3-319-19425-7

22. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *Jama.* (2017) 318:1377–84. doi: 10.1001/jama.2017.12126

23. Amiel J, Henrion-Caude A. miRNA, development and disease. *Adva Genet.* (2012) 80:1–36. doi: 10.1016/B978-0-12-404742-6.00001-6

24. Kung JT, Colognori D, Lee JT. Long noncoding rNAs: past, present, and future. *Genetics.* (2013) 193:651–69. doi: 10.1534/genetics.112.146704

25. Poliseno L, Marranci A, Pandolfi PP. Pseudogenes in human cancer. *Front Med.* (2015) 2:68. doi: 10.3389/fmed.2015.00068

26. Yu G, Yao W, Gumireddy K, Li A, Wang J, Xiao W, et al. Pseudogene pTENP1 functions as a competing endogenous rNA to suppress clear-cell renal cell carcinoma progression. *Mol Cancer Therap.* (2014) 13:3086–97. doi: 10.1158/1535-7163.MCT-14-0245

27. Guo X, Tang Y. OCT4 pseudogenes present in human leukemia cells. *Clin Exp Med.* (2012) 12:207–16. doi: 10.1007/s10238-011-0163-4

28. Zhao S, Yuan Q, Hao H, Guo Y, Liu S, Zhang Y, et al. Expression of oCT4 pseudogenes in human tumours: lessons from glioma and breast carcinoma. *J Pathol.* (2011) 223:672–82. doi: 10.1002/path.2827

29. Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: real functions and promising signature. *J Med Genet.* (2015) 52:17–24. doi: 10.1136/jmedgenet-2014-102785

30. Korneev SA, Park J-H O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense rNA transcribed from an nOS pseudogene. *J Neurosci.* (1999) 19:7711–20. doi: 10.1523/JNEUROSCI.19-18-07711.1999

31. Chen X, Zhu H, Wu X, Xie X, Huang G, Xu X, et al. Downregulated pseudogene cTNNAP1 promote tumor growth in human cancer by downregulating its cognate gene cTNNA1 expression. *Oncotarget.* (2016) 7:55518. doi: 10.18632/oncotarget.10833

32. Lin H, Shabbir A, Molnar M, Lee T. Stem cell regulatory function mediated by expression of a novel mouse oct4 pseudogene. *Biochem Biophys Res Commun.* (2007) 355:111–6. doi: 10.1016/j.bbrc.2007.01.106

33. Rutnam ZJ, Du WW, Yang W, Yang X, Yang BB. The pseudogene tUSC2P promotes tUSC2 function by binding multiple microRNAs. *Nat Commun.* (2014) 5:2914. doi: 10.1038/ncomms3914

34. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* (2010) 465:1033. doi: 10.1038/nature09144

35. Hayashi H, Arao T, Togashi Y, Kato H, Fujita Y, De Velasco M, et al. The oCT4 pseudogene pOU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene.* (2015) 34:199. doi: 10.1038/onc.2013.547

36. Chen B, Wang C, Zhang J, Zhou Y, Hu W, Guo T. New insights into long noncoding rNAs and pseudogenes in prognosis of renal cell carcinoma. *Cancer Cell Int.* (2018) 18:157. doi: 10.1186/s12935-018-0652-6

37. Weiser MR, Gönen M, Chou JF, Kattan MW, Schrag D. Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. *J Clin Oncol.* (2011) 29:4796. doi: 10.1200/JCO.2011.36.5080

38. Barrier A, Boelle P-Y, Roser F, Gregg J, Tse C, et al. Stage iI colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol.* (2006) 24:4685–91. doi: 10.1200/JCO.2005.05.0229

39. Eschrich S, Yang I, Bloom G, Kwong KY, Boulware D, Cantor A, et al. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol.* (2005) 23:3526–35. doi: 10.1200/JCO.2005.00.695

40. Lee J-H, Jung S, Park WS, Choe EK, Kim E, et al. Prognostic nomogram of hypoxia-related genes predicting overall survival of colorectal cancer–Analysis of tCGA database. *Sci Rep.* (2019) 9:1803. doi: 10.1038/s41598-018-38116-y

41. Jin LP, Liu T, Meng FQ, Tai JD. Prognosis prediction model based on competing endogenous rNAs for recurrence of colon adenocarcinoma. *BMC cancer.* (2020) 20:968. doi: 10.1186/s12885-020-07163-y

42. Jiang H, Li H, Li A, Tang E, Xu D, Chen Y, et al. Preoperative combined hemoglobin, albumin, lymphocyte and platelet levels predict survival in patients with locally advanced colorectal cancer. *Oncotarget.* (2016) 7:72076–83. doi: 10.18632/oncotarget.12271

43. Pei JP, Zhang CD, Liang Y, Zhang C, Wu KZ, Zhao ZM, et al. Novel nomograms individually predicting overall survival of non-metastatic colon cancer patients. *Front Oncol.* (2020) 10:733. doi: 10.3389/fonc.2020.00733

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.