Comprehensive analysis of the differential expression and prognostic value of COL1A2 in colon adenocarcinoma

Jian-Jiang Jin[1], Ting Zheng[1], Xiao-Xia Xu[1], Lei Zheng[1], Fang-Yuan Li[1], Xing-Xing Li[1], Li Zhou[1]

[1]Department of Medical Oncology, The First People's Hospital of Linping, Hangzhou 311103, Zhejiang, China

Correspondence to: Li Zhou; **email:** zl2002150302@163.com, https://orcid.org/0000-0002-2663-965X

ABSTRACT

Background: Colon adenocarcinoma (COAD) is a highly heterogeneous disease, which is the second most common cancer in females and third in males. Collagen type I alpha 2 (COL1A2) has been documented to be involved in the carcinogenesis of multiple tumors; however, the expression and prognostic significance of COL1A2 and its underlying mechanism in COAD remains unclarified.
Materials and Methods: The general profile of COL1A2, its expression pattern, and prognostic value were systematically assessed through various bioinformatics tools. The protein level of COL1A2 was verified in COAD patients using immunohistochemistry analysis. In addition, enrichment analyses were performed to explore the possible regulatory pathways of COL1A2 in COAD.
Results: The mRNA and protein levels of COL1A2 were significantly increased in COAD than that in normal tissues ($P < 0.05$). The COL1A2 expression tended to increase along with cancer stages and nodal metastasis status in COAD, while the promoter methylation levels of COL1A2 might negatively related to its mRNA expression. Survival analysis showed that COL1A2 was a reliable predictor for distinguishing the status of disease-specific survival (DSS), overall survival (OS), and progression-free survival (PFS), and might serve as a robust independent prognostic biomarker for DSS and OS in COAD patients ($P < 0.05$). The enrichment analysis showed focal adhesion as the most possible regulatory pathway by COL1A2.
Conclusion: Collectively, COL1A2 functioned as an independent prognostic biomarker and might be a potential therapeutic target in COAD.

INTRODUCTION

Colon adenocarcinoma (COAD) is the third most frequent cancer and the second main trigger of cancer-related death worldwide [1]. COAD is a heterogeneous digestive tract tumor characterized by dysregulated differentiation, proliferation, and apoptosis of intestinal epithelial cells [2]. According to statistics, 1.93 million new cases of COAD were identified, and 935,173 people died of this kind of cancer in 2020 [3]. It is the second most common cancer in females and third in males, which caused more than 500,000 mortalities

globally each year [4]. Most COAD cases are attributable to environmental factors such as dietary and chemicals, leading to genetic instabilities in intestinal cells, specific intestinal symbionts, and pathogens [5]. Over 50% of COAD-related deaths are closely associated with risk factors including alcohol consumption, smoking, and lack of exercise [6]. Currently, surgery, radiotherapy, chemotherapy, targeted therapy, or their combination are available therapeutic options administrated to COAD patients [7]. In spite of the continuous innovations and significant advances in treatment techniques, the 5-year overall

survival (OS) rate remains unsatisfactory at only 50%–65% and it is still incurable for advanced or metastatic COAD patients [8]. Therefore, it is imperative to understand the underlying mechanism responsible for COAD and develop a potential biomarker to prolong the survival time of COAD patients.

Collagen is the primary component of the extracellular matrix (ECM) and the most abundant type is collagen type I [9]. Collagen type I as a structural protein is observed in connective tissues such as tendon, bone, and skin [10]. It is a triple helix composed of two chains of collagen type I alpha 1 (COL1A1) and one chain of collagen type I alpha 2 (COL1A2) [10]. The structural integrity and coordinated biosynthesis of these chains are critical for tissue morphogenesis, growth, homeostasis, and repair [11]. Changes in collagen type I synthesis occur in embryogenesis, wound healing, and some pathological conditions, including fibrosis of the kidney, lung, and liver, scleroderma as well as cancers [12]. Moreover, the expression level of the fibrosis-related protein COL1A2 was obviously upregulated in the colonic tissue with intestinal fibrosis [13, 14]. Previous studies have demonstrated the involvement of COL1A2 in the cancer process, which can be both stimulatory and inhibitory. COL1A2 accelerated the development and angiogenesis of melanoma and medulloblastoma [15, 16]. On the other hand, COL1A2 suppressed the proliferation, migration, and invasion of bladder cancer cells [17]. Nevertheless, the prognostic role of COL1A2 and the actual significance across the various clinicopathological parameters such as age, gender, weight, histological subtypes, cancer stage, and nodal metastasis status in COAD have not been systemically studied yet.

Herein, multiple bioinformatics tools were adopted and the dataset of well-established cancer data from various demographic and clinicopathological patients was downloaded for comprehensive research of COAD. Firstly, we analyzed the expression and prognostic values of COL1A2 in COAD. Further, the underlying mechanisms of COAD based on COL1A2-related genes were explored by enrichment analysis.

MATERIALS AND METHODS

The general profile of COL1A2

The normalized RNA sequencing data based on TCGA Pan-Cancer were downloaded from the UCSC Xena database (https://xenabrowser.net/datapages/) to analyze the difference in COL1A2 mRNA expression levels between various tumors and normal tissues. The data were used to extract the mRNA expression values of COL1A2 and the tumors with less than three samples were deleted. Differential COL1A2 mRNA expression analyses in each tumor type were performed using the Student's *t*-test. $P < 0.05$ was considered to be statistically significant. Then, the Human Protein Atlas (HPA) (https://www.proteinatlas.org/) database was utilized to examine the localization of COL1A2 protein in human tumor cells in the "SUBCELL" column. The HPA database aims to map all the human proteins in cells, tissues, and organs using an integration of various omics technologies. As the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer, the Catalogue of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) was applied for analyzing the different mutation types related to the COL1A2 gene.

Expression analysis of COL1A2 in COAD

The expression data and corresponding clinical information in TCGA-COAD were obtained from the UCSC Xena database. The differential gene expression levels of COL1A2 in COAD and normal tissues were compared using unpaired and paired *t*-tests. A *P*-value less than 0.05 was regarded as the threshold of significance. Through the HPA database, we evaluated the protein levels of COL1A2 in COAD and normal tissues by immunohistochemical analysis, which was validated in an additional population with COAD.

Subsequently, the expression and promoter methylation level of the COL1A2 gene in COAD based on clinicopathological parameters such as age, gender, weight, histological subtypes, cancer stage, and nodal metastasis status was assessed using the UALCAN web server (http://ualcan.path.uab.edu/analysis.html), which is an interactive database for analyzing cancer OMICS data.

Prognosis analysis of COL1A2 in COAD

The relationship between COL1A2 expression and disease-free interval (DFI), disease-specific survival (DSS), overall survival (OS), and progression-free survival (PFS) in COAD was examined in the Gene Set Cancer Analysis (GSCA) (http://bioinfo.life.hust.edu.cn/GSCA/#/) which is an integrated database for genomic and immunogenomic gene set cancer analysis. $P < 0.05$ indicates statistical significance. Besides, the value of COL1A2 in distinguishing the survival status with regard to DFI, DSS, OS, and PFS was determined by the receiver operating characteristic (ROC) curve based on the TCGA-COAD data. The computed area under the curve (AUC) value ranging from 0.5 to 1.0 indicated the discrimination ability from 50–100%. Moreover, Cox regression analysis was performed

Table 1. The clinicopathological characteristics of patients in colon adenocarcinoma.

Clinical characteristics	Total number (%)
Age	
≤50	46 (15.7)
>50	247 (84.3)
Body mass index	
<25	75 (33.9)
≥25	146 (66.1)
Gender	
Female	133 (45.4)
Male	160 (54.6)
Nodal metastasis status	
N0	174 (59.4)
N1	71 (24.2)
N2	48 (16.4)
Cancer stage	
Stage I	48 (17.0)
Stage II	114 (40.3)
Stage III	82 (29.0)
Stage IV	39 (13.7)
Subtype	
Adenocarcinoma	250 (85.9)
Mucinous adenocarcinoma	41 (14.1)

to evaluate the independent prognostic value of COL1A2 by SPSS software (version 23.0), and $P < 0.05$ was considered to be statistically significant. Further, the R package "rms" was adopted to construct the nomogram and plot the calibration curves to predict 1-, 3-, and 5- year OS and DSS for COAD patients. The concordance index (C-index) was used to evaluate the predictive accuracy of the nomogram. C-index from 0.50 to 0.70 (lower accuracy), 0.71 to 0.90 (medium accuracy), above 0.90 (high accuracy) [18].

Enrichment analysis

To explore the underlying mechanism of COL1A2 in COAD, the TCGA-COAD samples were divided into high-COL1A2 and low-COL1A2 expression groups based on the median value of the COL1A2 gene. The DEGs between two expression groups were screened using the "limma" package. |Log2 Fold change (FC)| Cutoff >1 and $P < 0.05$ were set as the criterion of significant differences.

Following this, we conducted Gene Ontology (GO) annotations and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of the significant DEGs using the R package "cluster profiler". A P-value less than 0.05 was considered statistically significant. Gene Set Enrichment Analysis (GSEA) was conducted to further investigate the signaling pathways regulated by COL1A2-related genes in COAD. The gene set permutations were performed 1000 times. Clusters with

a false discovery rate (FDR) q-value < 0.05 and P-value < 0.05 were identified as significant.

Subsequently, we defined the enrichment level of a pathway in COAD samples as the single-sample gene-set enrichment analysis (ssGSEA) score [19]. The gene set represents the collection of all marker genes of a pathway. The association of the COL1A2 mRNA expression with the enrichment levels (ssGSEA scores) of the pathway was evaluated using Pearson's correlation test.

Data sharing statement

The datasets used and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

RESULTS

Patient characteristics

The TCGA-COAD dataset contains 293 samples with complete survival and gene expression data. As shown in Table 1, there were 46 patients (15.7%) diagnosed at the age of fewer than 50 years and 247 patients (84.3) over 50 years. Totally 75 patients (33.9%) had body mass index (BMI) < 25, and 133 patients (45.4%) were females. N0, N1, and N2 were found in 174 (59.4%), 71 (24.2%), and 48 (16.4%) patients, separately. Patients at stages I, II, III, and IV were 48 (17.0%), 114 (40.3%),

82 (29.0%), and 39 (13.7%), respectively. Besides, 250 patients were diagnosed with adenocarcinoma (85.9%). The TCGA-COAD dataset includes 42 paired samples.

The general profile of COL1A2

We firstly analyzed the expression levels of COL1A2 among various cancer types (Figure 1A). Subcellular location and immunofluorescence image of COL1A2 expression in human tumor cells were retrieved from the HPA (Figure 1B, 1C). The COL1A2 gene mutation in pan-cancer was evaluated through the COSMIC database. A total of 1246 samples were recorded for mutations, among which the missense substitution (52.77%) had the highest proportion, followed by synonymous substitution (15.81%), and other types (8.95%) (Figure 1D). Figure 1E presented the
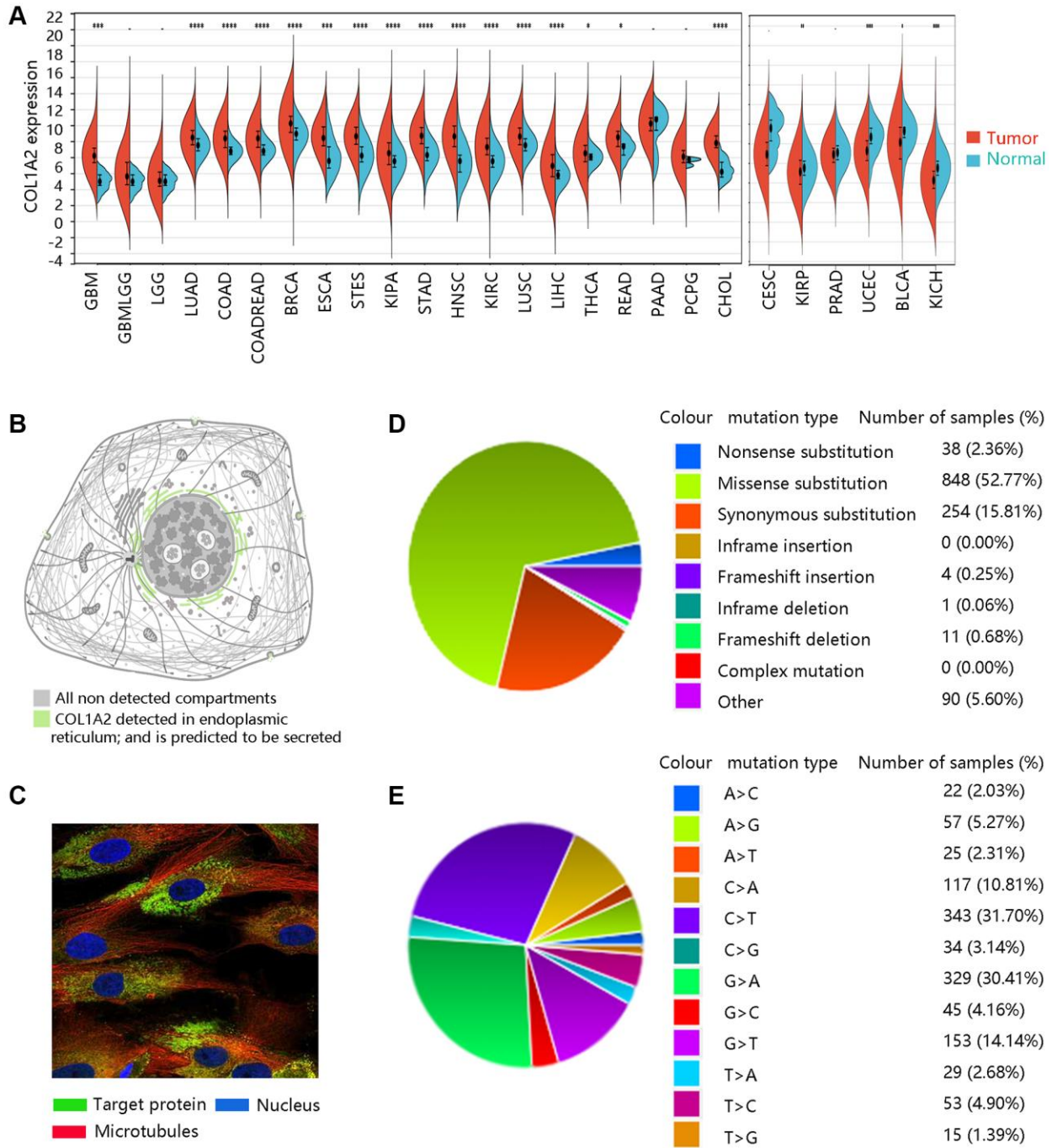

Figure 1. The basic characteristics of COL1A2. (**A**) The COL1A2 gene expression among multiple cancers. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. (**B**) Subcellular location and (**C**) immunofluorescence image of COL1A2 protein in human tumor cells. (**D**) Summary of different types of mutations related to the COL1A2 gene. (**E**) Various types of substitutional mutations.

breakdown of different substitution mutations, exhibiting the highest type of C > T (31.70%), and the lowest type of T > G (1.39%). Of note, no mutation of the COL1A2 gene occurred in COAD.

Expression analysis of COL1A2

Based on the TCGA-COAD data, the mRNA expression levels of COL1A2 in COAD and normal tissues were compared. Both unpaired and paired *t*-tests demonstrated that COL1A2 gene expression was remarkably upregulated in the COAD compared with that in normal tissues (all *P* < 0.001) (Figure 2A, 2B). The HPA database showed the different COL1A2 protein expressions in COAD and normal colon samples (Figure 2C, 2D). The immunohistochemistry analysis validated the higher protein level of COL1A2 in COAD tissue (Figure 2E, 2F).
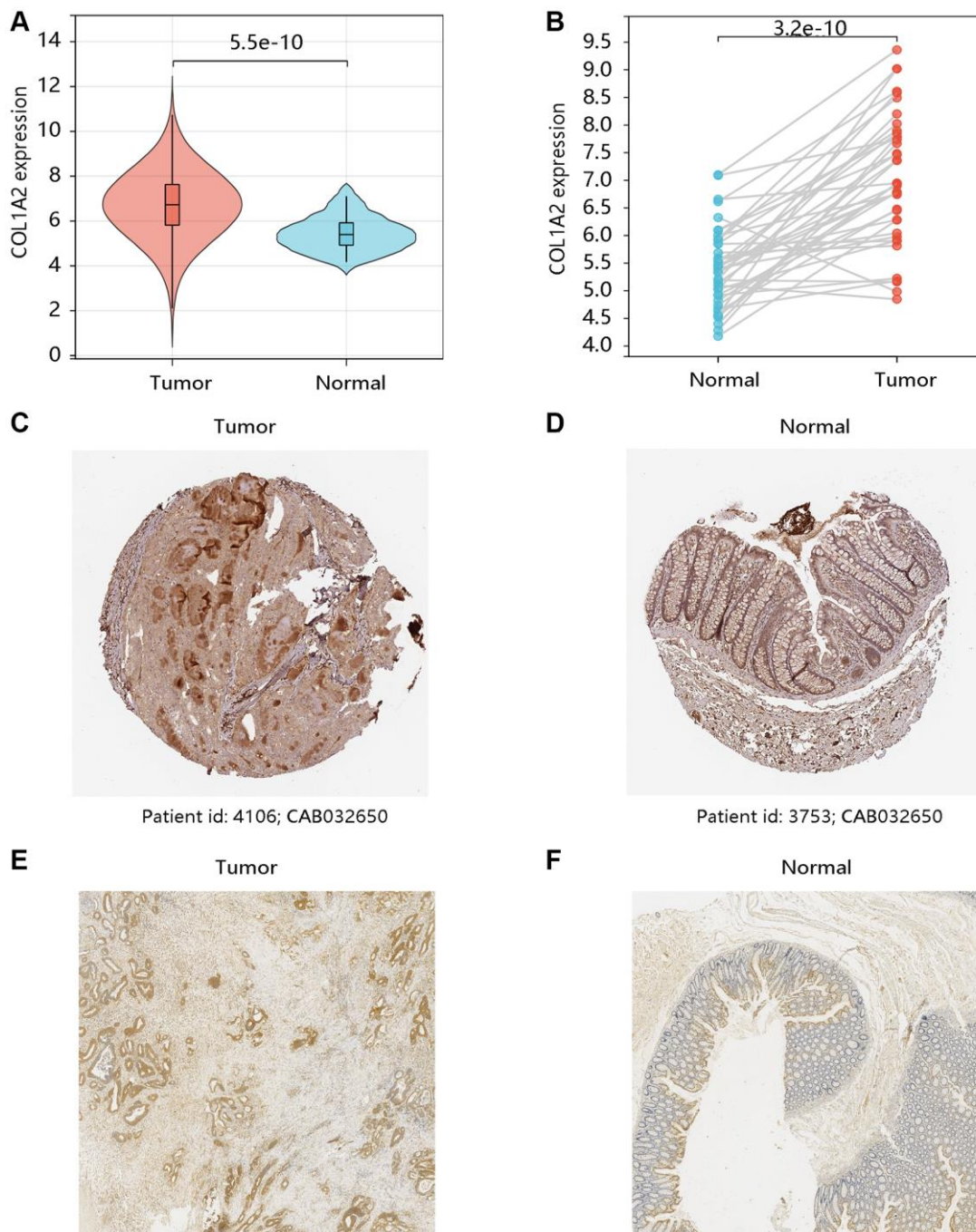

Figure 2. The mRNA and protein levels of COL1A2 in colon adenocarcinoma (COAD). The differential gene expression of COL1A2 in COAD and normal tissues using (**A**) unpaired and (**B**) paired *t*-tests. The COL1A2 protein levels in (**C**) COAD and (**D**) normal colon tissues using the HPA database. The COL1A2 proteins in (**E**) COAD and (**F**) normal colon tissues by the immunohistochemistry analysis.

Transcriptional expression and epigenetic regulation of COL1A2 across different clinicopathological factors

We have observed the elevated mRNA level of the COL1A2 in COAD in the earlier section, and then we evaluated the COL1A2 expression in COAD based on various clinicopathological characteristics like age, gender, weight, histological subtypes, cancer stage, and nodal metastasis status (Figure 3A–3F). Patients with mucinous adenocarcinoma had remarkably higher COL1A2 mRNA expression than those with adenocarcinoma ($P < 0.05$) (Figure 3D). It tended to increase the COL1A2 expression at advanced cancer stages (stage 1 < stage 2 < stage 3) and COL1A2 expression increased along with the nodal metastasis status (N0 < N1 < N2) (Figure 3E, 3F).
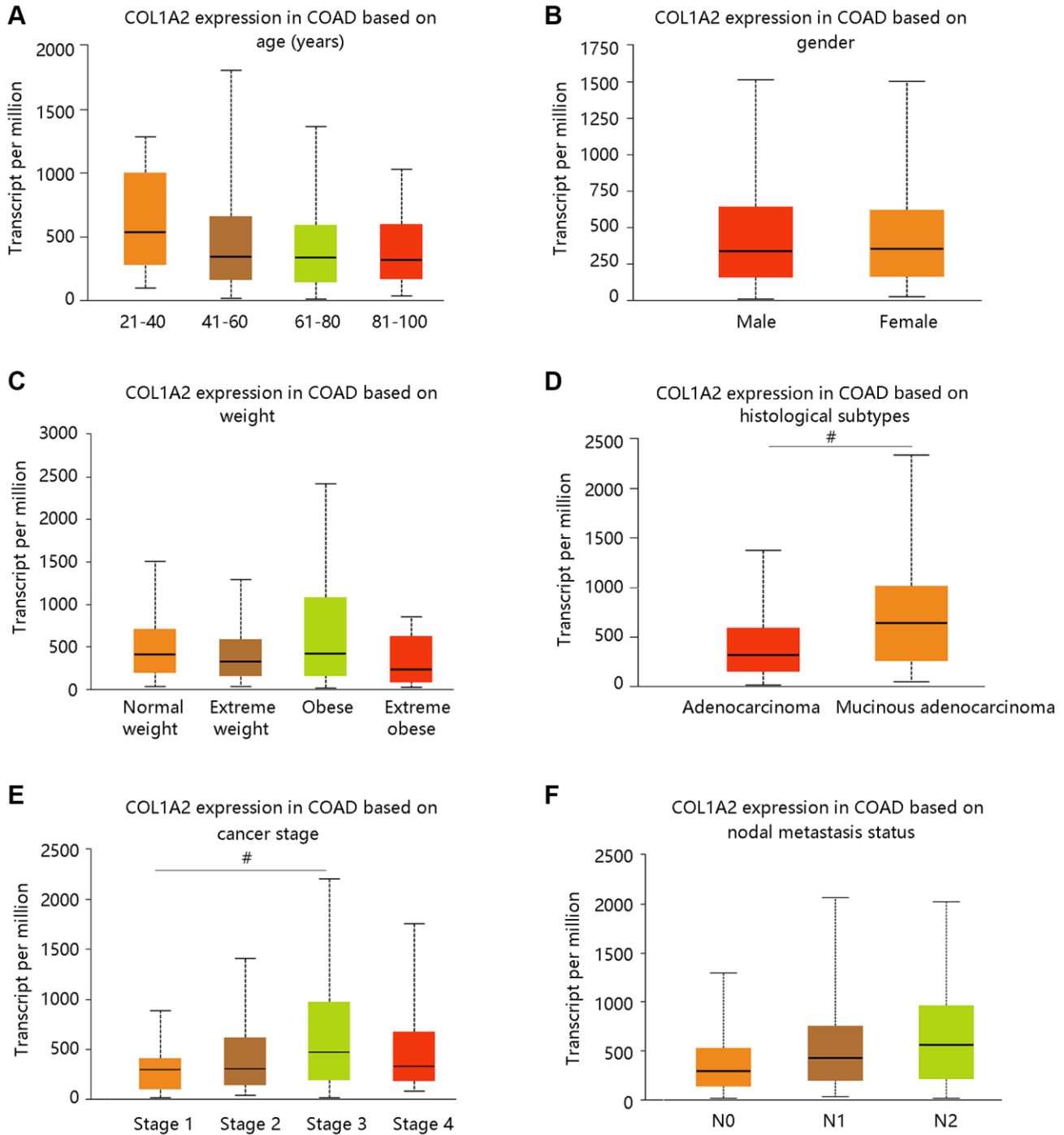

Figure 3. The COL1A2 expression in colon adenocarcinoma based on various clinicopathological parameters. (**A**) Age (years). (**B**) Gender. (**C**) Weight. Normal weight: 18.5 ≤BMI < 25; Overweight: 25 ≤ BMI < 30; Obese: 30 ≤ BMI < 40; Extreme obese: BMI > 40. Abbreviation: BMI: body-mass-index. (**D**) Histological subtypes. (**E**) Cancer stage. (**F**) Nodal metastasis status. #$P < 0.05$.

DNA methylation is closely related to the development of cancer in the human body [20]. From our data, the promoter methylation of COL1A2 was significantly related to age, cancer stage, and nodal metastasis status, but had no notable relation with gender, weight, and histological subtypes (Figure 4A–4F). Intriguingly, the promoter methylation level of COL1A2 decreased with the development of cancer stages (stage 4 < stage 3 < stage 2 < stage 1) and nodal metastasis status (N2 < N1 < N0) (Figure 4E, 4F). The above findings suggested that the promoter methylation of COL1A2 might have a negative relation with its mRNA expression in COAD.
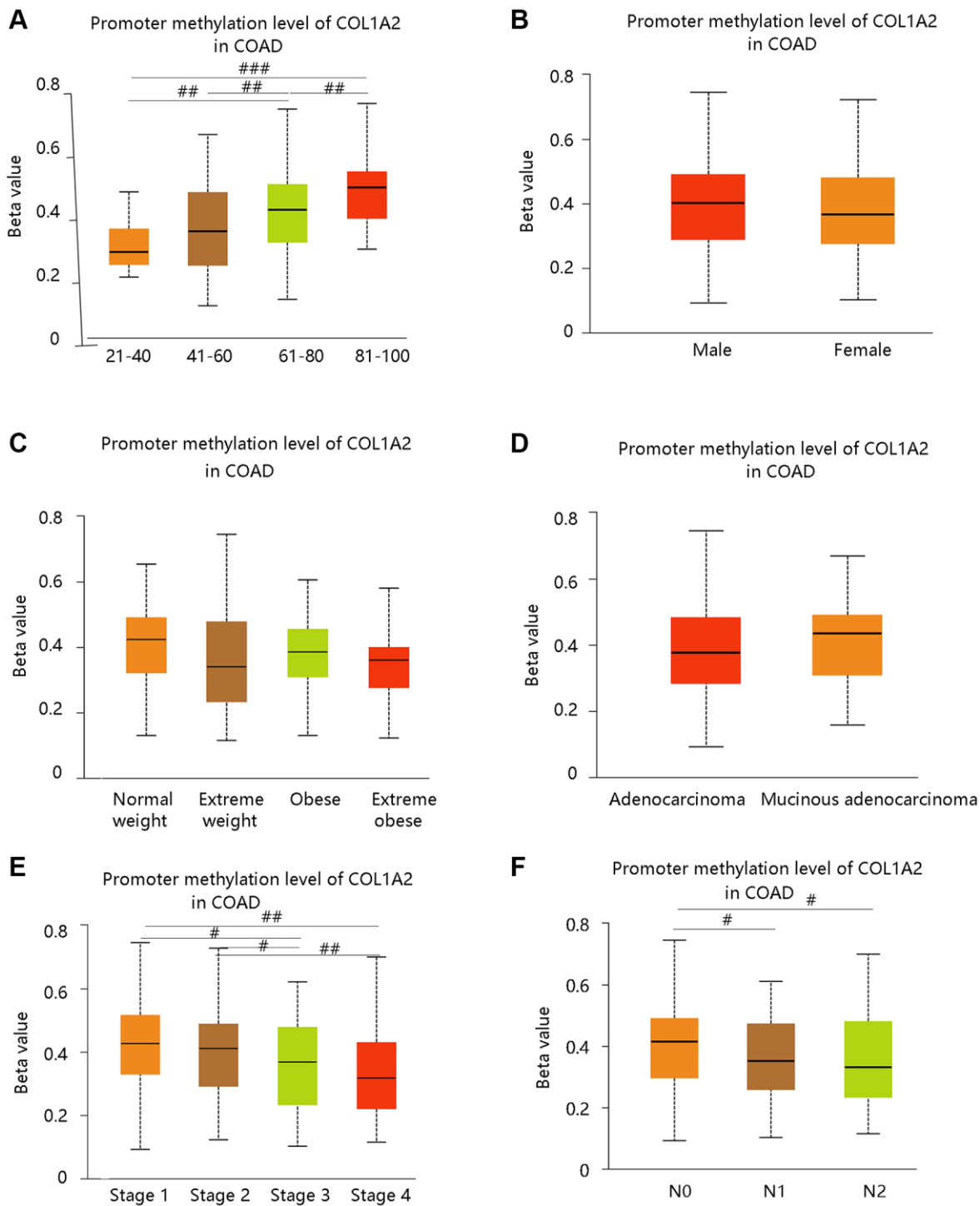

Figure 4. The COL1A2 promoter methylation level in colon adenocarcinoma based on various clinicopathological parameters. (**A**) Age. (**B**) Gender. (**C**) Weight. Normal weight: $18.5 \leq BMI < 25$; Overweight: $25 \leq BMI < 30$; Obese: $30 \leq BMI<40$; Extreme obese: $BMI > 40$. Abbreviation: BMI: body-mass-index. (**D**) Histological subtypes. (**E**) Cancer stage. (**F**) Nodal metastasis status. #$P < 0.05$; ##$P < 0.01$; ###$P < 0.001$.

COL1A2 expression predicted survival status and served as an independent prognostic biomarker in COAD

To investigate the effect of COL1A2 mRNA expression on the clinical outcomes of COAD patients, the GSCA database was employed to analyze the association of COL1A2 with DFI, DSS, OS, and PFS. High COL1A2 expression led to a worse prognosis except for DFI (Figure 5A). Then, the ROC curves of COL1A2 expression were drawn to examine the value of COL1A2 in predicting the survival status. Similarly, COL1A2 could not predict DFI status (AUC = 0.487), but exhibited the satisfactory performance in predicting survival status of DSS (AUC = 0.711), OS (AUC = 0.622), and PFS (AUC = 0.554) (Figure 5B–5E).
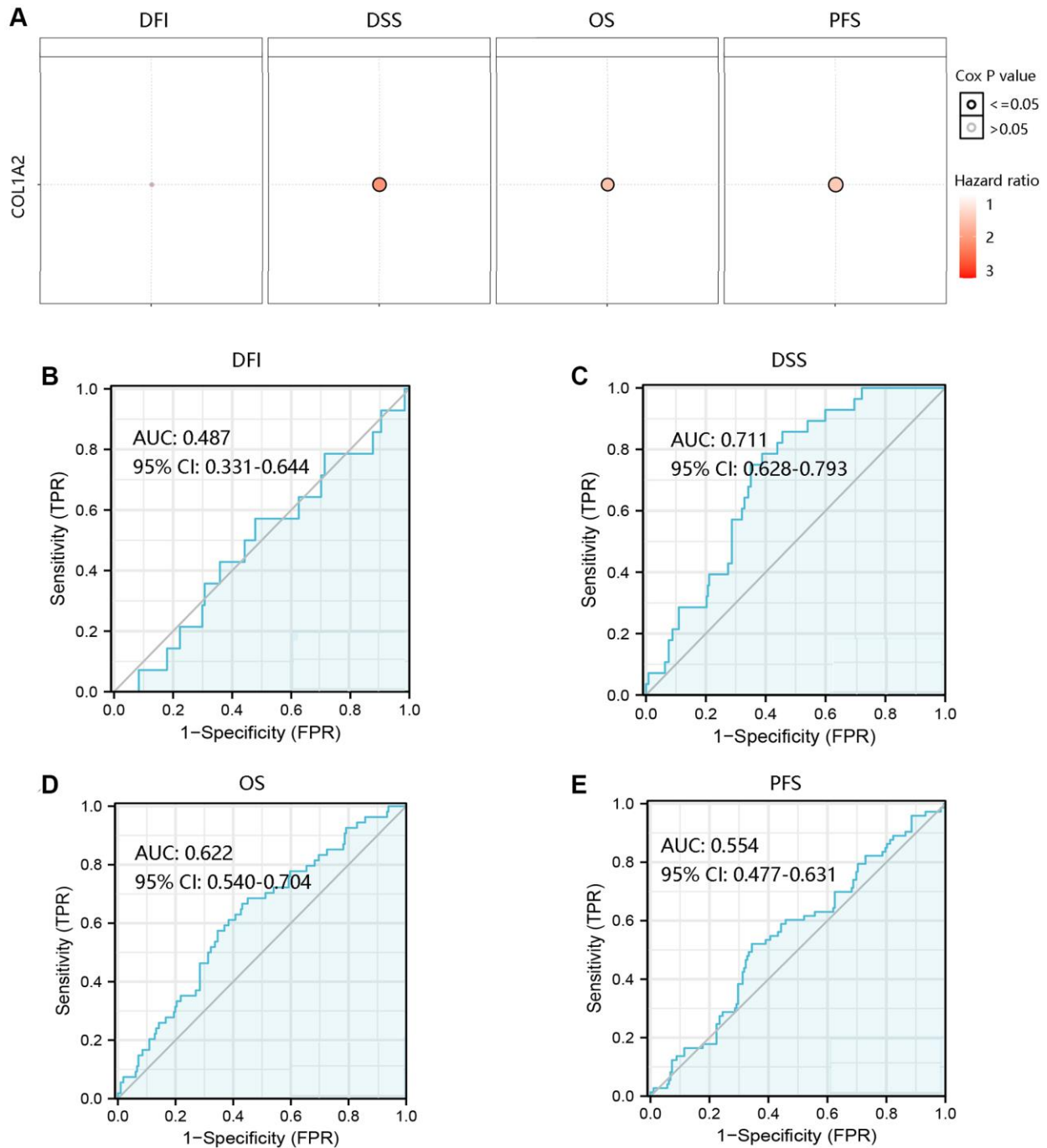

Figure 5. Prognostic and predictive analysis of COL1A2 in colon adenocarcinoma (COAD). (**A**) The effect of COL1A2 on DFI, DSS, OS, and PFS in COAD. The value of COL1A2 in predicting survival status of (**B**) DFI, (**C**) DSS, (**D**) OS, and (**E**) PFS. Abbreviations: DFI: disease-free interval; DSS: disease-specific survival; OS: overall survival; PFS: progression-free survival; AUC: area under the curve; 95% CI: 95% confidence interval; FPR: false-positive rate; TPR: true-positive rate.

Table 2. Cox regression analyses of disease-specific survival in colon adenocarcinoma patients.

Variable	Univariate analysis HR (95% CI)	*P*-value	Multivariate analysis HR (95% CI)	*P*-value
Age	0.970 (0.944–0.997)	0.031	0.954 (0.912–0.999)	0.043
BMI	0.991 (0.940–1.043)	0.721	0.985 (0.894–1.086)	0.763
Gender				
Male vs. female	1.552 (0.714–3.371)	0.267	0.925 (0.292–2.927)	0.894
N status				
N1 vs. N0	3.840 (1.540–9.577)	0.004	0.333 (0.024–4.629)	0.413
N2 vs. N0	5.302 (2.040–13.780)	0.001	0.630 (0.049–8.079)	0.722
Cancer stage				
II vs. I	1.399 (0.151–12.914)	0.767	9931.693 (0.000-Inf)	0.943
III vs. I	4.754 (0.593–38.099)	0.142	48372.047 (0.000-Inf)	0.934
IV vs. I	23.294 (3.003–180.674)	0.003	244290.187 (0.000-Inf)	0.924
Subtype				
Mucinous adenocarcinoma vs. adenocarcinoma	0.759 (0.228–2.529)	0.653	1.006 (0.186–5.432)	0.994
COL1A2	1.954 (1.481–2.579)	<0.001	1.674 (1.040–2.696)	0.034

Abbreviations: HR: hazard ratio; 95% CI: 95% confidence interval; BMI: body mass index; N status: nodal metastasis status.

Table 3. Cox regression analyses of overall survival in colon adenocarcinoma patients.

Variable	Univariate analysis HR (95% CI)	*P*-value	Multivariate analysis HR (95% CI)	*P*-value
Age	1.011 (0.990–1.033)	0.296	1.008 (0.981–1.035)	0.572
BMI	0.943 (0.886–1.005)	0.070	0.933 (0.870–1.002)	0.056
Gender				
Male vs. female	1.471 (0.846–2.560)	0.172	0.918 (0.439–1.922)	0.821
N status				
N1 vs. N0	1.711 (0.883–3.317)	0.112	0.317 (0.033–3.078)	0.322
N2 vs. N0	3.430 (1.815–6.481)	<0.001	1.200 (0.127–11.364)	0.874
Cancer stage				
II vs. I	6.078 (0.807–45.787)	0.080	13156.010 (0.000-Inf)	0.888
III vs. I	10.318 (1.372–77.596)	0.023	36880.177 (0.000-Inf)	0.876
IV vs. I	25.016 (3.248–192.678)	0.002	108029.274 (0.000-Inf)	0.863
Subtype				
Mucinous adenocarcinoma vs. adenocarcinoma	1.388 (0.694–2.775)	0.354	1.927 (0.799–4.645)	0.144
COL1A2	1.585 (1.297–1.937)	<0.001	1.298 (1.006–1.674)	0.045

Abbreviations: HR: hazard ratio; 95% CI: 95% confidence interval; BMI: body mass index; N status: nodal metastasis status.

To further probe into the independent prognostic value of COL1A2 in predicting DSS, OS, and PFS of COAD patients, Cox regression analysis was conducted using the TCGA-COAD data. The results showed that age, N1, N2, stage IV, and COL1A2 had a significant relationship with DSS in COAD patients (all P < 0.05), but other clinical factors were not associated with DSS in the univariate analysis. When these variables were integrated into multivariate Cox regression analysis, age and COL1A2 were independent prognostic factors (all P < 0.05) (Table 2). In addition, N2, stage III, stage IV, and COL1A2 were closely connected with OS (all P < 0.05), while only COL1A2 (P < 0.05) was still significantly associated with OS in the multivariate analysis (Table 3). However, COL1A2 could not

Table 4. Cox regression analyses of progression-free survival in colon adenocarcinoma patients.

Variable	Univariate analysis HR (95% CI)	P-value	Multivariate analysis HR (95% CI)	P-value
Age	0.985 (0.968–1.003)	0.102	0.992 (0.970–1.015)	0.506
BMI	0.991 (0.963–1.019)	0.515	0.989 (0.956–1.024)	0.547
Gender				
Male vs. female	1.371 (0.854–2.203)	0.192	1.023 (0.569–1.839)	0.938
N status				
N1 vs. N0	1.640 (0.933–2.881)	0.086	0.698 (0.138–3.536)	0.664
N2 vs. N0	3.859 (2.201–6.766)	<0.001	2.053 (0.412–10.224)	0.380
Cancer stage				
II vs. I	2.059 (0.785–5.396)	0.142	1.788 (0.597–5.351)	0.299
III vs. I	3.072 (1.157–8.161)	0.024	1.777 (0.263–12.005)	0.555
IV vs. I	9.563 (3.498–26.141)	<0.001	6.725 (1.156–39.130)	0.034
Subtype				
Mucinous adenocarcinoma vs. adenocarcinoma	0.684 (0.313–1.498)	0.342	0.941 (0.395–2.240)	0.891
COL1A2	1.307 (1.106–1.544)	0.002	1.146 (0.924–1.422)	0.215

Abbreviations: HR: hazard ratio; 95% CI: 95% confidence interval; BMI: body mass index; N status: nodal metastasis status.

independently predict the PFS of COAD patients ($P <$ 0.05) (Table 4). These results indicated that high COL1A2 expression was an independent prognostic factor for predicting worse DSS and OS in COAD.

To better predict the prognosis of COAD patients, two nomogram models on basis of the multivariate Cox regression analysis results were constructed and calibration curves were drawn to evaluate their efficiency. Two statistically significant prognostic factors, age, and COL1A2 were included in the nomogram model based on DSS, which had a C-index of 0.765 (Figure 6A). N status and COL1A2 were enrolled in the construction of the nomogram model based on OS, which had a C-index of 0.734 (Figure 6B). The calibration curves for 1-, 3-, and 5-year DSS and OS of COAD patients were presented in Figure 6C, and Figure 6D, respectively. Importantly, the calibration curve showed good agreement between prediction and observation in the 5-year survival probability. These findings revealed that the nomogram models had high efficiency in anticipating DSS and OS of COAD patients.

Enrichment analysis

To elucidate the pathological function of COL1A2, the DEGs between the high- and low- COL1A2 expression groups were firstly identified as shown in the volcano plot (Figure 7A) and heat map (Figure 7B). According to the selection criterion, the significant DEGs were used for GO and KEGG enrichment analyses. GO annotations predicted the functional effect of the target genes in three aspects: biological process (BP), cellular component (CC), and molecular function (MF). We found that cellular response to an organic substance, extracellular matrix organization, extracellular matrix, vesicle, signaling receptor binding, and growth factor binding were significantly regulated in COAD (Figure 7C). The KEGG enrichment analysis showed that the pathways related to the functions of COL1A2 interactive genes were oxidative phosphorylation, phagosome, focal adhesion, fatty acid metabolism (Figure 7D).

To further reveal the underlying mechanism of COL1A2 in COAD, GSEA was performed on gene expression microarray data of TCGA-COAD. The results showed that the top six positive pathways were cytokine-cytokine receptor interaction (NES = 2.192, $P <$ 0.001), ECM receptor interaction (NES = 2.260, $P <$ 0.001), focal adhesion (NES = 2.334, $P <$ 0.001), Hedgehog signaling pathway (NES = 2.237, $P <$ 0.001), JAK-STAT signaling pathway (NES = 2.228, $P <$ 0.001), and pathways in cancer (NES = 2.277, $P <$ 0.001) (Figure 8) (Table 5). Combined with the previous enrichment analysis results, focal adhesion was the consistent pathway.

Following this, we analyzed the enrichment levels of the focal adhesion pathway using ssGSEA. The COL1A2 gene expression level had a significant positive correlation with the enrichment levels (ssGSEA scores) of the focal adhesion pathway ($P <$ 0.001)

(Figure 9). Therefore, the COL1A2 might affect the development of COAD via positive regulation of the focal adhesion pathway.

DISCUSSION

COAD is a heterogeneous disease always occurring in the elderly [21]. Besides, COAD is highly invasive and is easily spread to other organs with the liver as the most common site of metastasis. Although significant achievements have been made in the treatment strategies, the 5-year survival rate is still low and the 5-year OS for metastatic patients is approximately 13% [22]. Hence, searching for novel molecular biomarkers for carcinogenesis and pathogenesis of COAD was essential for the COAD diagnosis and treatment. COL1A2 encodes the alpha 2 chain of collagen type I which is the main ECM component of bone and skin [23]. It has been demonstrated that the abnormal COL1A2 mRNA expression led to a worse prognosis of
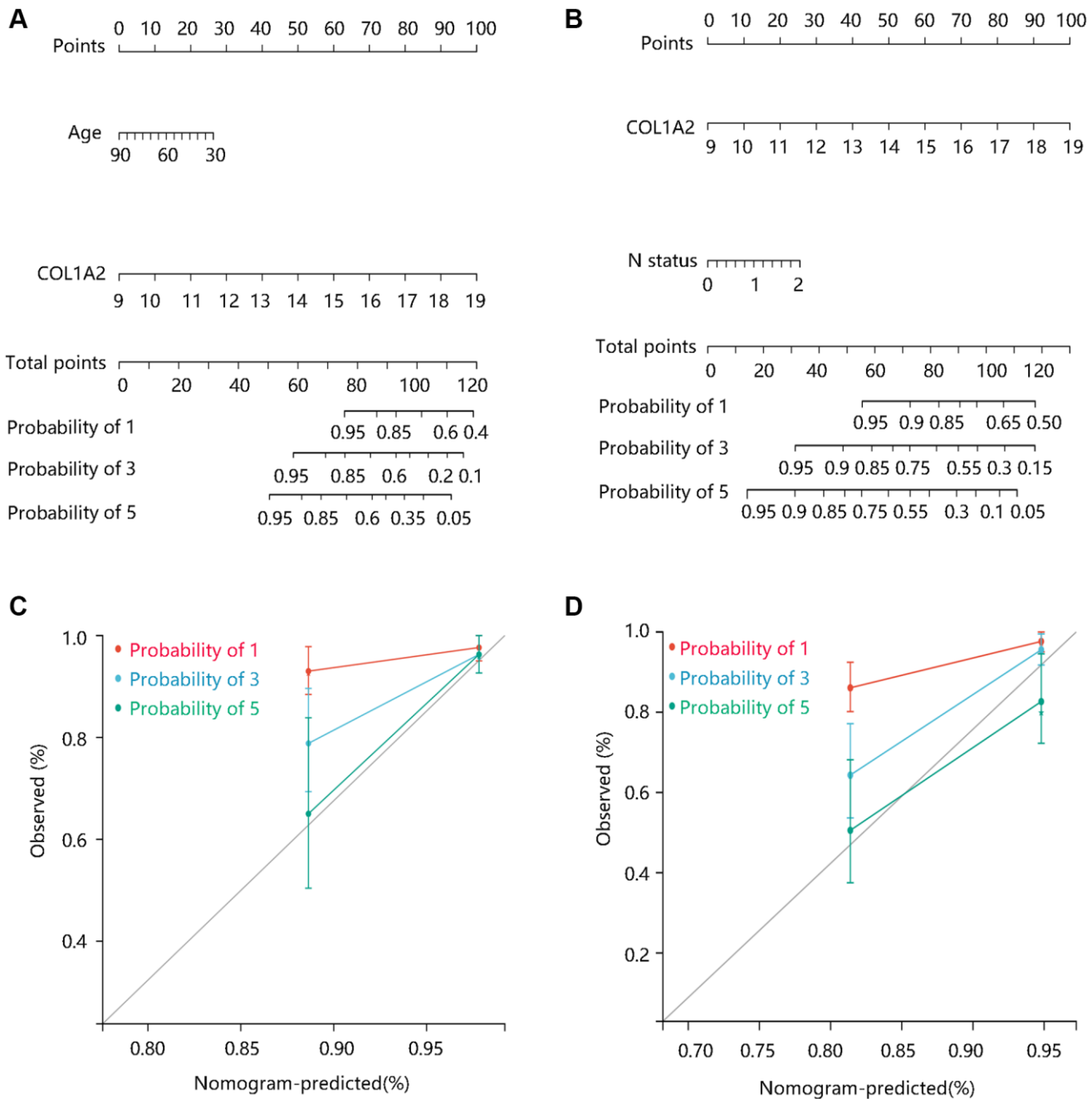
Figure 6. The construction and evaluation of nomogram models. Nomogram models based on Cox regression analysis results in terms of (**A**) DSS, and (**B**) OS. The calibration curves for the (**C**) DSS nomogram model and (**D**) OS nomogram model. Abbreviations: DSS: disease-specific survival; OS: overall survival.
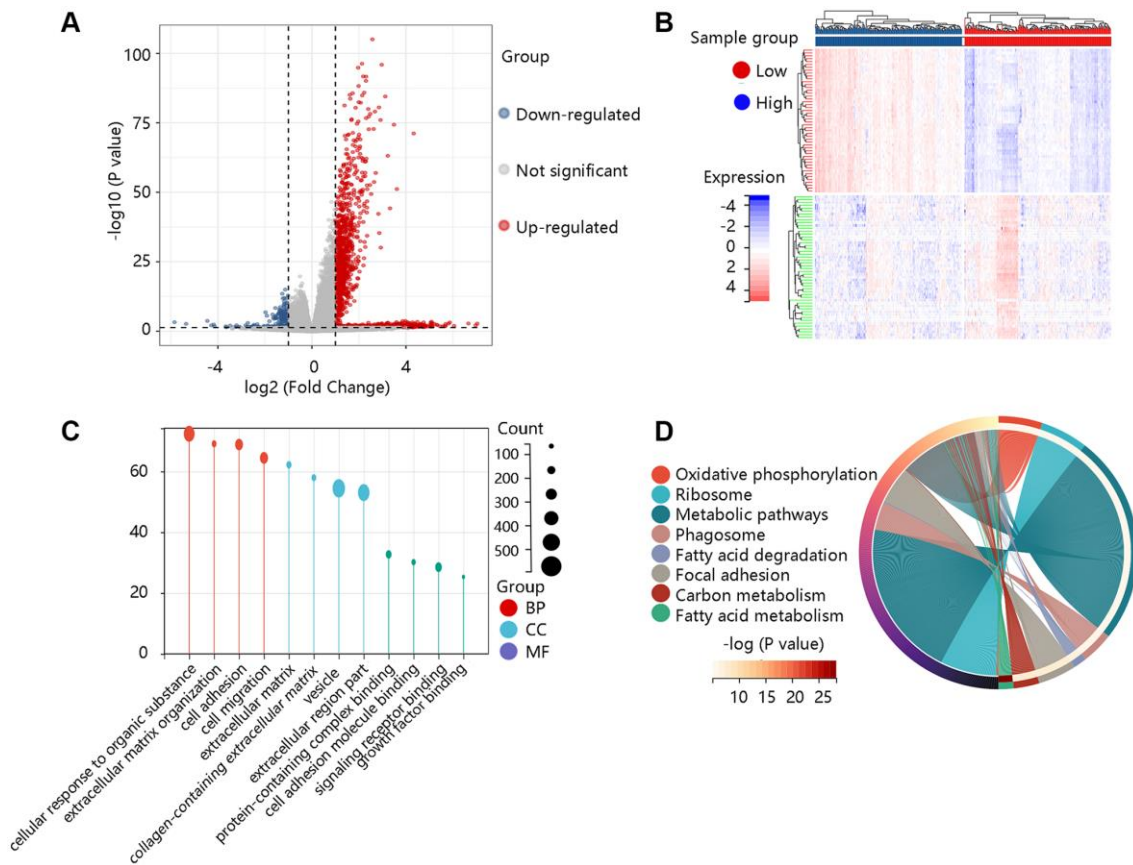
Figure 7. Functional enrichment analysis of the differentially expressed genes (DEGs). (**A**) The volcano plot exhibited the DEGs between high- and low- COL1A2 expression groups. (**B**) The heat map showed the top 50 significant DEGs. (**C**) The gene ontology annotations of the DEGs. (**D**) The KEGG pathway enrichment of the DEGs.
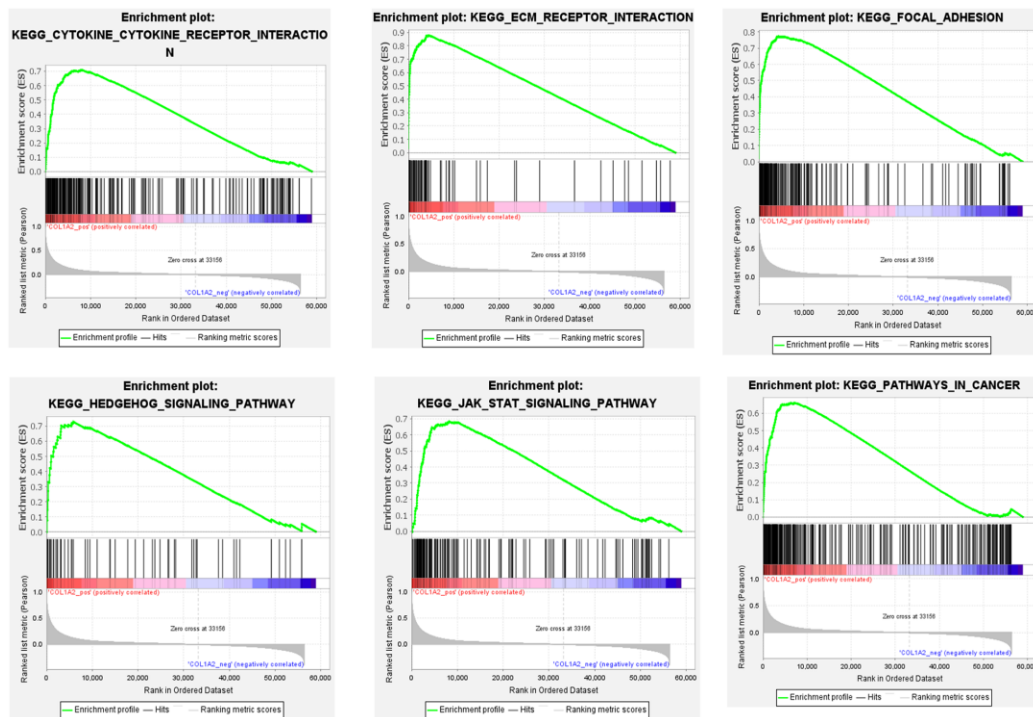

Figure 8. Gene set enrichment analysis of COL1A2 in colon adenocarcinoma.

Table 5. Gene set enrichment analysis of COL1A2 in colon adenocarcinoma.

KEGG name	ES	NES	NOM *p*-value	FDR *q*-value
Cytokine-cytokine receptor interaction	0.709	2.192	<0.001	0.002
ECM receptor interaction	0.877	2.260	<0.001	0.002
Focal adhesion	0.773	2.334	<0.001	0.001
Hedgehog signaling pathway	0.729	2.237	<0.001	0.002
JAK-STAT signaling pathway	0.684	2.228	<0.001	0.002
Pathways in cancer	0.660	2.277	<0.001	0.002

Abbreviations: ES: enrichment score; NES: normalized enrichment score; NOM: nominal; FDR: false discovery rate.

gastric cancer patients and enhances the proliferation of prostate cancer cell [24, 25]. In addition, COL1A2 overexpression was a significant risk factor for intracranial aneurysm susceptibility [26]. Whereas, COL1A2 decreased the proliferative activity of liver epithelial cells [27]. These results suggested that COL1A2 was a double-edged sword in the biological process. This study evaluated the utility of COL1A2 as a potential biomarker in COAD and we explored the underlying mechanisms by which COL1A2 affected COAD through enrichment analysis.

Firstly, we had a general overview of COL1A2 from the following aspects: the differential mRNA expression of COL1A2 in various tumors, the location of COL1A2 protein in human tumor cells, and the different mutation types related to the COL1A2 gene. We found that COL1A2 was differentially expressed in most of the tumors, and the COL1A2 protein was located in the endoplasmic reticulum. The missense substitution of COL1A2 had the highest proportion of the mutation type. Then, we explored the COL1A2 expression in COAD and observed significant upregulation of COL1A2 transcriptional levels in COAD, and tended to increase along with stages and the nodal metastasis status. Epigenetic changes in genes are thought to be the leading cause of neoplastic transformation [28], and hence we analyzed the promoter methylation level of
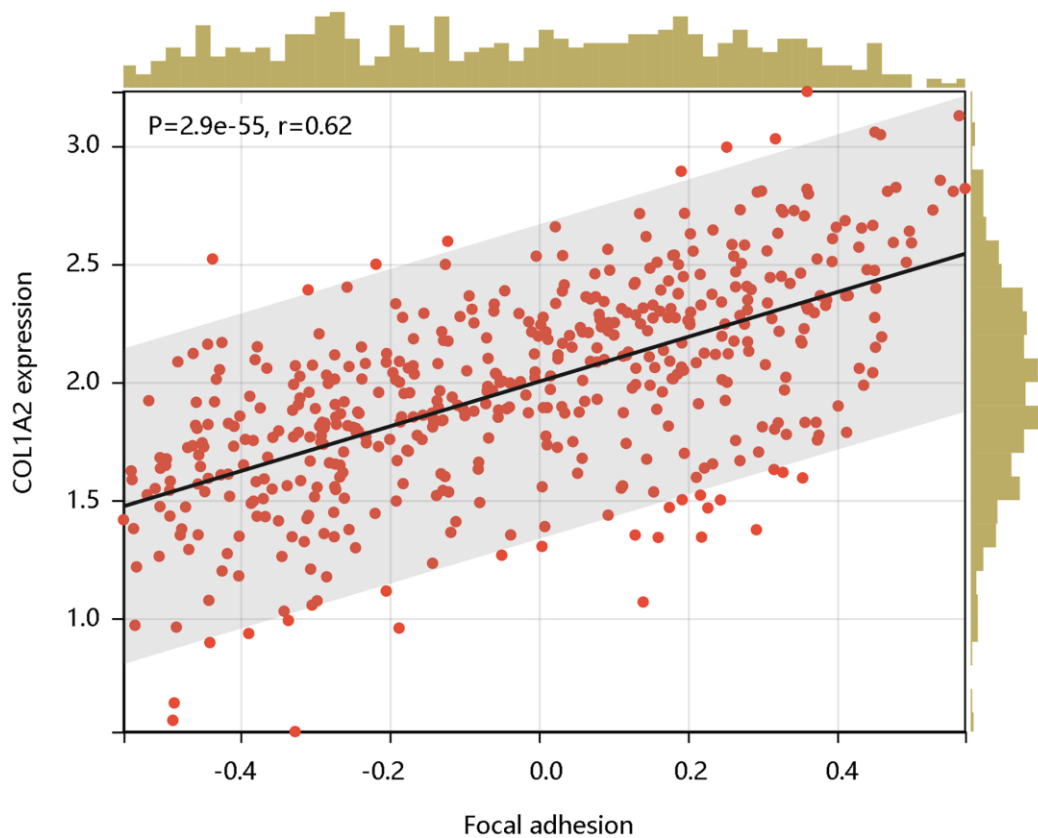

Figure 9. The significant positive correlation of COL1A2 mRNA expression with the enrichment levels of the focal adhesion pathway.

COL1A2 across various clinicopathological characteristics and found the possible negative correlation with the expression profile, particularly in terms of cancer stage and nodal metastasis status.

Next, we explored the relationship between COL1A2 mRNA expression and the clinical outcomes of COAD patients. Upon the survival analyses, COL1A2 overexpression contributed to a lower survival probability of DSS, OS, and PFS. Meanwhile, COL1A2 exhibited satisfactory performance in predicting the DSS, OS, and PFS status of patients with COAD. Further, Cox regression analyses showed that the elevated mRNA level of COL1A2 was an independent predictor for worse DSS and OS, revealing COL1A2 as a potential prognostic biomarker for COAD. Changes in COL1A2 in the ECM microenvironment are often accompanied by stromal invasion and the occurrence of epithelial neoplasms [29]. It is also involved in the induction of epithelial-mesenchymal transformation (EMT) in breast and lung cancer cells [30, 31]. Importantly, COAD arises from the epithelial cells of the colon or rectum, whose functions such as cellular differentiation, migration, and invasion are regulated by the physical interactions with ECM [29, 32]. Thus, it is reasonable to speculate that COL1A2 might promote tumorigenesis, invasion, and metastasis of COAD, which might be an explanation for the high expression of COL1A2 leading to a worse prognosis for COAD patients. Additionally, COL1A2 was secreted by stromal fibroblasts, while tumor-associated fibroblasts stimulate the occurrence of COAD in part by inducing inflammation in the tumor microenvironment [33–35]. The cancer cells and stromal cells in the COAD microenvironment generate high levels of pro-inflammatory eicosanoids, which are robust lipid mediators implicated in cancer cell angiogenesis, proliferation, and metastasis by the following mechanisms: direct activation of receptors on cancer cells; enforcing the epithelial cells to release angiogenic factors, pro-inflammatory mediators, and tumor growth factors [36]. This might be another explanation for the worse clinical outcomes of COAD patients with COL1A2 overexpression.
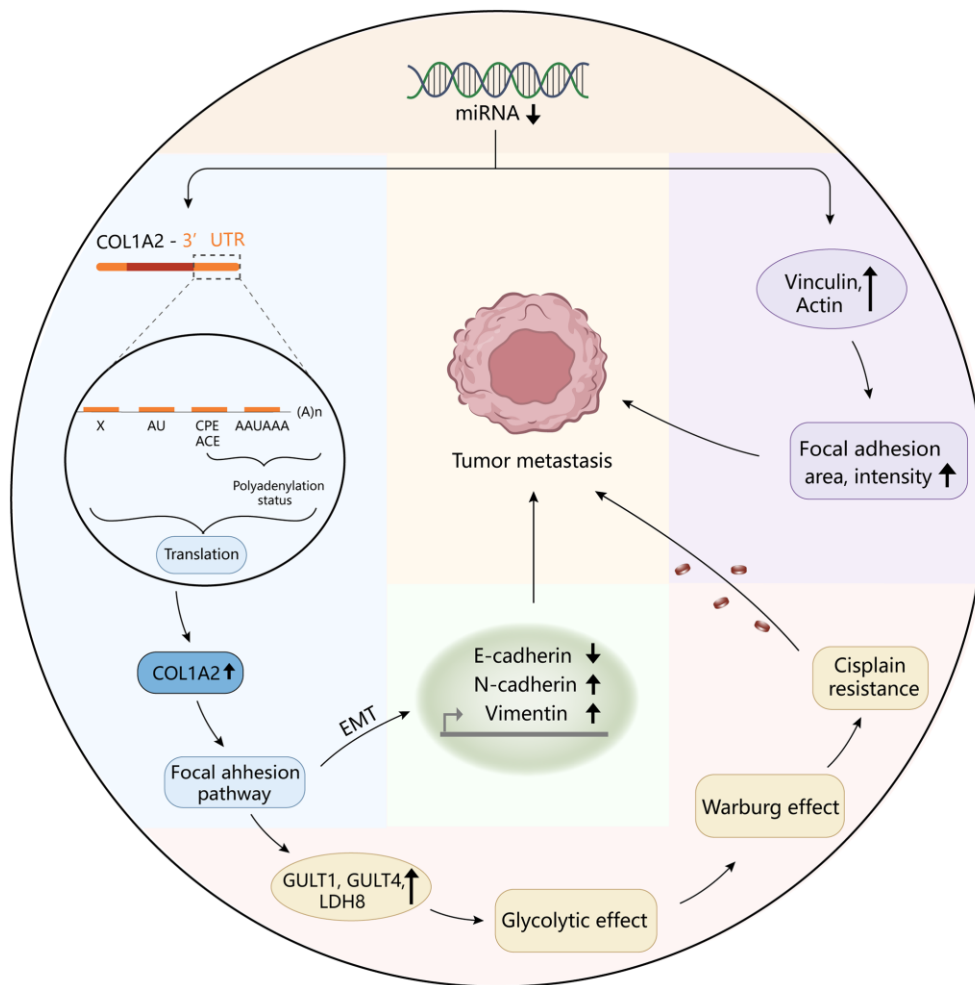

Figure 10. The diagram revealed the possible mechanism of COL1A2 involvement in colon adenocarcinoma.

To reveal the pathological function of COL1A2 in COAD, enrichment analyses were performed. GO enrichment of the COL1A-related DEGs showed the cellular response to an organic substance, extracellular matrix organization, extracellular matrix, vesicle, signaling receptor binding, and growth factor binding. The KEGG pathway showed the enrichment signaling pathways of oxidative phosphorylation, phagosome, focal adhesion, and fatty acid metabolism, among which focal adhesion shared the same pathway with GSEA results. Interestingly, COL1A2 expression was positively correlated with the enrichment levels of the focal adhesion pathway. Focal adhesion plays a vital role in cancer metastasis, invasion, and drug resistance [37]. Ning et al. reported that the focal adhesion signaling pathway was crucial in the EMT process in pancreatic cancer [38]. The activation of the focal adhesion pathway alters cancer cell glycolysis and induces cisplatin resistance in cervical and breast cancer [39]. Similarly, the authors inferred that COL1A2 might promote metastasis, and induce drug resistance, thereby enhancing the progression of COAD via regulation of the focal adhesion pathway. For further mechanistic explorations, Pekow et al. reported that COL1A2 was predicted to be the target of miR-4728-3p. In the HCT116 human colon cancer cells, miR-4728-3p regulated the COL1A2 gene expression by targeting the wild type 3′untranslated region (3′UTR) of the COL1A2 gene in the focal adhesion pathway. Besides, they revealed that miR-4728-3p decreased focal adhesion intensity and area [40]. Taken together, the possible mechanism of COL1A2 involvement in COAD was shown in Figure 10, which should be verified in the future. Additionally, other molecules could regulate COL1A2 mRNA expression. Acute ulcerative colitis (UC)-like colitis was induced in *Chga*-C57BL/6-deficient (*Chga*$^{-/-}$) and wild-type (*Chga*$^{+/+}$) mice. COL1A2 mRNA expression was decreased in *Chga*$^{-/-}$ mice model [41]. It has also been demonstrated that patients with UC are at high risk to develop colonic neoplasia [42]. Therefore, the authors speculated that targeting CHGA might decrease the COL1A2 mRNA expression, inhibiting the progression of COAD.

In summary, our study comprehensively enlightens the role of COL1A2 in the initiation and development of COAD. COL1A2 might serve as a valuable prognostic biomarker and a potential therapeutic target for COAD. Moreover, COL1A2 might promote the progression of COAD by positive regulation of the focal adhesion pathway.

AUTHOR CONTRIBUTIONS

Conception and design: Jian-Jiang Jin and Ting Zheng; Collection and assembly of data: Xiao-Xia Xu and Lei Zheng; Data analysis and interpretation: Fang-Yuan Li, Xing-Xing Li and Li Zhou; Manuscript writing: all authors; Final approval of manuscript: all authors.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209–49.
https://doi.org/10.3322/caac.21660
PMID:33538338

2. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–7.
https://doi.org/10.1038/nature11252
PMID:22810696

3. Wang Q, Huang Y, Jia M, Lu D, Zhang HW, Huang D, Liu SH, Lv C. Safflower Polysaccharide Inhibits AOM/DSS-Induced Mice Colorectal Cancer Through the Regulation of Macrophage Polarization. Front Pharmacol. 2021; 12:761641.
https://doi.org/10.3389/fphar.2021.761641
PMID:34744741

4. Pan P, Skaer C, Yu J, Zhao H, Ren H, Oshima K, Wang LS. Berries and other natural products in the pancreatic cancer chemoprevention in human clinical trials. J Berry Res. 2017; 7:147–61.
https://doi.org/10.3233/JBR-170159
PMID:29367867

5. Jackson DN, Theiss AL. Gut bacteria signaling to mitochondria in intestinal inflammation and cancer. Gut Microbes. 2020; 11:285–304.
https://doi.org/10.1080/19490976.2019.1592421
PMID:30913966

6. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, McCullough ML, Patel AV, Ma J, Soerjomataram I, Flanders WD, Brawley OW, Gapstur SM, Jemal A. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin. 2018; 68:31–54.
https://doi.org/10.3322/caac.21440
PMID:29160902

7. He J, Zhao H, Liu X, Wang D, Wang Y, Ai Y, Yang J. Sevoflurane suppresses cell viability and invasion and promotes cell apoptosis in colon cancer by modulating

exosome‑mediated circ‑HMGCS1 via the miR‑34a‑5p/SGPP1 axis. Oncol Rep. 2020; 44:2429–42.
https://doi.org/10.3892/or.2020.7783
PMID:33125091

8. Sun X, Guo Y, Zhang Y, Zhao P, Wang Z, Wei Z, Qiao H. Colon Cancer-Related Genes Identification and Function Study Based on Single-Cell Multi-Omics Integration. Front Cell Dev Biol. 2021; 9:789587.
https://doi.org/10.3389/fcell.2021.789587
PMID:34901030

9. Nissen NI, Karsdal M, Willumsen N. Collagens and Cancer associated fibroblasts in the reactive stroma and its relation to Cancer biology. J Exp Clin Cancer Res. 2019; 38:115.
https://doi.org/10.1186/s13046-019-1110-6
PMID:30841909

10. Gelse K, Pöschl E, Aigner T. Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev. 2003; 55:1531–46.
https://doi.org/10.1016/j.addr.2003.08.002
PMID:14623400

11. Ponticos M, Abraham D, Alexakis C, Lu QL, Black C, Partridge T, Bou-Gharios G. Col1a2 enhancer regulates collagen activity during development and in adult tissue repair. Matrix Biol. 2004; 22:619–28.
https://doi.org/10.1016/j.matbio.2003.12.002
PMID:15062855

12. Prockop DJ, Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem. 1995; 64:403–34.
https://doi.org/10.1146/annurev.bi.64.070195.002155
PMID:7574488

13. Chen H, Xu H, Luo L, Qiao L, Wang Y, Xu M, Li Y, Zhu P, Yang B. Thalidomide Prevented and Ameliorated Pathogenesis of Crohn's Disease in Mice via Regulation of Inflammatory Response and Fibrosis. Front Pharmacol. 2019; 10:1486.
https://doi.org/10.3389/fphar.2019.01486
PMID:31920668

14. Qiao L, Fang L, Zhu J, Xiang Y, Xu H, Sun X, Chen H, Yang B. Total Flavone of Abelmoschus manihot Ameliorates TNBS-Induced Colonic Fibrosis by Regulating Th17/Treg Balance and Reducing Extracellular Matrix. Front Pharmacol. 2021; 12:769793.
https://doi.org/10.3389/fphar.2021.769793
PMID:35002710

15. van Kempen LC, Rijntjes J, Mamor-Cornelissen I, Vincent-Naulleau S, Gerritsen MJ, Ruiter DJ, van Dijk MC, Geffrotin C, van Muijen GN. Type I collagen expression contributes to angiogenesis and the development of deeply invasive cutaneous melanoma. Int J Cancer. 2008; 122:1019–29.
https://doi.org/10.1002/ijc.23147
PMID:17957794

16. Liang Y, Diehn M, Bollen AW, Israel MA, Gupta N. Type I collagen is overexpressed in medulloblastoma as a component of tumor microenvironment. J Neurooncol. 2008; 86:133–41.
https://doi.org/10.1007/s11060-007-9457-5
PMID:17653508

17. Mori K, Enokida H, Kagara I, Kawakami K, Chiyomaru T, Tatarano S, Kawahara K, Nishiyama K, Seki N, Nakagawa M. CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. Int J Oncol. 2009; 34:1593–602.
https://doi.org/10.3892/ijo_00000289
PMID:19424577

18. Yang B, Fu L, Xu S, Xiao J, Li Z, Liu Y. A nomogram based on a gene signature for predicting the prognosis of patients with head and neck squamous cell carcinoma. Int J Biol Markers. 2019; 34:309–17.
https://doi.org/10.1177/1724600819865745
PMID:31452437

19. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14:7.
https://doi.org/10.1186/1471-2105-14-7
PMID:23323831

20. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019; 20:590–607.
https://doi.org/10.1038/s41580-019-0159-6
PMID:31399642

21. Aparicio T, Bouché O, Taieb J, Maillard E, Kirscher S, Etienne PL, Faroux R, Khemissa Akouz F, El Hajbi F, Locher C, Rinaldi Y, Lecomte T, Lavau-Denes S, et al, and for PRODIGE 20 Investigators. Bevacizumab+chemotherapy versus chemotherapy alone in elderly patients with untreated metastatic colorectal cancer: a randomized phase II trial-PRODIGE 20 study results. Ann Oncol. 2018; 29:133–8.
https://doi.org/10.1093/annonc/mdx529
PMID:29045659

22. Wasan HS, Gibbs P, Sharma NK, Taieb J, Heinemann V, Ricke J, Peeters M, Findlay M, Weaver A, Mills J, Wilson C, Adams R, Francis A, et al, and FOXFIRE trial investigators, and SIRFLOX trial investigators, and FOXFIRE-Global trial investigators. First-line selective internal radiotherapy plus chemotherapy versus chemotherapy alone in patients with liver metastases from colorectal cancer (FOXFIRE, SIRFLOX, and FOXFIRE-Global): a combined analysis of three

multicentre, randomised, phase 3 trials. Lancet Oncol. 2017; 18:1159–71.
https://doi.org/10.1016/S1470-2045(17)30457-6
PMID:28781171

23. Rossert JA, Garrett LA. Regulation of type I collagen synthesis. Kidney Int Suppl. 1995; 49:S34–8. PMID:7674590

24. Zang S, Guo R, Xing R, Zhang L, Li W, Zhao M, Fang J, Hu F, Kang B, Ren Y, Zhuang Y, Liu S, Wang R, et al. Identification of differentially-expressed genes in intestinal gastric cancer by microarray analysis. Genomics Proteomics Bioinformatics. 2014; 12:276–83.
https://doi.org/10.1016/j.gpb.2014.09.004
PMID:25500430

25. Kiefer JA, Farach-Carson MC. Type I collagen-mediated proliferation of PC3 prostate carcinoma cell line: implications for enhanced growth in the bone microenvironment. Matrix Biol. 2001; 20:429–37.
https://doi.org/10.1016/s0945-053x(01)00159-7
PMID:11691583

26. Gan Q, Liu Q, Hu X, You C. Collagen Type I Alpha 2 (COL1A2) Polymorphism Contributes to Intracranial Aneurysm Susceptibility: A Meta-Analysis. Med Sci Monit. 2017; 23:3240–6.
https://doi.org/10.12659/msm.902327
PMID:28671939

27. Lim A, Greenspan DS, Smith BD. Expression of alpha 2 type I collagen in W8 cells increases cell adhesion and decreases colony formation in soft agar. Matrix Biol. 1994; 14:21–30.
https://doi.org/10.1016/0945-053x(94)90026-4
PMID:8061917

28. Patra R, Das NC, Mukherjee S. Exploring the Differential Expression and Prognostic Significance of the COL11A1 Gene in Human Colorectal Carcinoma: An Integrated Bioinformatics Approach. Front Genet. 2021; 12:608313.
https://doi.org/10.3389/fgene.2021.608313
PMID:33597969

29. Brooks M, Mo Q, Krasnow R, Ho PL, Lee YC, Xiao J, Kurtova A, Lerner S, Godoy G, Jian W, Castro P, Chen F, Rowley D, et al. Positive association of collagen type I with non-muscle invasive bladder cancer progression. Oncotarget. 2016; 7:82609–19.
https://doi.org/10.18632/oncotarget.12089
PMID:27655672

30. Gilles C, Polette M, Seiki M, Birembaut P, Thompson EW. Implication of collagen type I-induced membrane-type 1-matrix metalloproteinase expression and matrix metalloproteinase-2 activation in the metastatic progression of breast carcinoma. Lab Invest. 1997; 76:651–60.

PMID:9166284

31. Shintani Y, Maeda M, Chaika N, Johnson KR, Wheelock MJ. Collagen I promotes epithelial-to-mesenchymal transition in lung cancer cells via transforming growth factor-beta signaling. Am J Respir Cell Mol Biol. 2008; 38:95–104.
https://doi.org/10.1165/rcmb.2007-0071OC
PMID:17673689

32. Mika A, Pakiet A, Czumaj A, Kaczynski Z, Liakh I, Kobiela J, Perdyan A, Adrych K, Makarewicz W, Sledzinski T. Decreased Triacylglycerol Content and Elevated Contents of Cell Membrane Lipids in Colorectal Cancer Tissue: A Lipidomic Study. J Clin Med. 2020; 9:1095.
https://doi.org/10.3390/jcm9041095
PMID:32290558

33. Steinberg J. Collagen turnover and the growth state in 3T6 fibroblast cultures. Lab Invest. 1978; 39:491–6. PMID:215834

34. Steinberg J. The turnover of collagen in fibroblast cultures. J Cell Sci. 1973; 12:217–34.
https://doi.org/10.1242/jcs.12.1.217
PMID:4734737

35. Karpisheh V, Nikkhoo A, Hojjat-Farsangi M, Namdar A, Azizi G, Ghalamfarsa G, Sabz G, Yousefi M, Yousefi B, Jadidi-Niaragh F. Prostaglandin E2 as a potent therapeutic target for treatment of colon cancer. Prostaglandins Other Lipid Mediat. 2019; 144:106338.
https://doi.org/10.1016/j.prostaglandins.2019.106338
PMID:31100474

36. Vladoiu MC, Labrie M, Létourneau M, Egesborg P, Gagné D, Billard É, Grosset AA, Doucet N, Chatenet D, St-Pierre Y. Design of a peptidic inhibitor that targets the dimer interface of a prototypic galectin. Oncotarget. 2015; 6:40970–80.
https://doi.org/10.18632/oncotarget.5403
PMID:26543238

37. Liang R, Wang M, Zheng G, Zhu H, Zhi Y, Sun Z. A comprehensive analysis of prognosis prediction models based on pathway‑level, gene‑level and clinical information for glioblastoma. Int J Mol Med. 2018; 42:1837–46.
https://doi.org/10.3892/ijmm.2018.3765
PMID:30015853

38. Ning Z, Wang A, Liang J, Xie Y, Liu J, Yan Q, Wang Z. USP22 promotes epithelial-mesenchymal transition via the FAK pathway in pancreatic cancer cells. Oncol Rep. 2014; 32:1451–8.
https://doi.org/10.3892/or.2014.3354
PMID:25070659

39. Wang S, Xie J, Li J, Liu F, Wu X, Wang Z. Cisplatin suppresses the growth and proliferation of breast

and cervical cancer cell lines by inhibiting integrin β5-mediated glycolysis. Am J Cancer Res. 2016; 6:1108–17.
PMID:27294003

40. Pekow J, Hutchison AL, Meckel K, Harrington K, Deng Z, Talasila N, Rubin DT, Hanauer SB, Hurst R, Umanskiy K, Fichera A, Hart J, Dinner AR, Bissonnette M. miR-4728-3p Functions as a Tumor Suppressor in Ulcerative Colitis-associated Colorectal Neoplasia Through Regulation of Focal Adhesion Signaling. Inflamm Bowel Dis. 2017; 23:1328–37.
https://doi.org/10.1097/MIB.0000000000001104
PMID:28594651

41. Eissa N, Hussein H, Tshikudi DM, Hendy GN, Bernstein CN, Ghia JE. Interdependence between Chromogranin-A, Alternatively Activated Macrophages, Tight Junction Proteins and the Epithelial Functions. A Human and In-Vivo/In-Vitro Descriptive Study. Int J Mol Sci. 2020; 21:7976.
https://doi.org/10.3390/ijms21217976
PMID:33121008

42. Gondal B, Haider H, Komaki Y, Komaki F, Micic D, Rubin DT, Sakuraba A. Efficacy of various endoscopic modalities in detecting dysplasia in ulcerative colitis: A systematic review and network meta-analysis. World J Gastrointest Endosc. 2020; 12:159–71.
https://doi.org/10.4253/wjge.v12.i5.159
PMID:32477450