Medicine®

OPEN

A pan-cancer analysis of the oncogenic role of procollagen C-endopeptidase enhancer (PCOLCE) in human

Hui Gao, MM[a] , Qiuyun Li, MD[b],*

Abstract

There is no evidence showing that the expression of procollagen C-endopeptidase enhancer (PCOLCE) is associated with human tumors, and pan-cancer analysis is not available. Based on public databases such as the cancer genome atlas, we investigated the potential role of PCOLCE expression in 33 different human tumors. PCOLCE expression in 11 tumors was significantly correlated with tumor prognosis and was a prognostic predictor for pancreatic adenocarcinoma, thymoma and CES. We also found that PCOLCE expression correlated with the immune microenvironment of tumors and the level of cancer-associated fibroblast infiltration. PCOLCE is a potential predictor of small molecule targeted drugs and immune checkpoint inhibitors. Finally, we found by enrichment analysis that PCOLCE localizes to extracellular structures and the extracellular matrix and exerts substantial effects on tumors through the PI3K-Akt and AGE-RAGE signaling pathways. We have a preliminary and relatively comprehensive understanding of the role of PCOLCE in various tumors.

Abbreviations: ACC = adrenocortical carcinoma, COAD = colon adenocarcinoma, CPTAC = clinical proteomic tumor analysis consortium, DLBC = lymphoid neoplasm diffuse large b-cell lymphoma, ESCA = esophageal carcinoma, GBM = glioblastoma multiforme, GEO = gene expression omnibus, GO = gene ontology, HNSC = head and neck squamous cell carcinoma, KEGG = Kyoto encyclopedia of genes and genomes, KIRC = kidney renal clear cell carcinoma, KIRP = kidney renal papillary cell carcinoma, LGG = brain lower grade glioma, LIHC = liver hepatocellular carcinoma, LUSC = lung squamous cell carcinoma, OS = overall survival, OV = ovarian serous cystadenocarcinoma, PAAD = pancreatic adenocarcinoma, PCOLCE = procollagen C-endopeptidase enhancer, SARC = sarcoma, STAD = stomach adenocarcinoma, TCGA = the cancer genome atlas, TGCT = testicular germ cell tumors, THCA = thyroid carcinoma, THYM = thymoma, UCEC = uterine corpus endometrial carcinoma, UCS = uterine carcinosarcoma, UVM = uveal melanoma.

Keywords: gene expression, immune, mutation, signaling pathway

1. Introduction

The assessment of the expression of genes potentially associated with pan-cancer, an assessment of their relevance to clinical prognosis and their underlying molecular mechanisms are important steps in understanding the complex mechanisms of cancer development. We used publicly available databases for pan-cancer analysis to identify potential gene functions in different tumors. procollagen C-endopeptidase enhancer (PCOLCE), also known as Procollagen COOH-Terminal Proteinase Enhancer or PCEP, is a protein coding gene. Gene ontology (GO) annotations are related to heparin binding and metalloendopeptidase inhibitor activity. An important paralog of this gene is PCOLCE2. We analyzed the structure and function of PCOLCE from physiological and clinicopathological perspectives with the expectation of clinical benefit to patients. Previous studies have shown that PCOLCE may be present in some tumors and may have a potential association with patient prognosis, such as gastric cancer,[1] Osteosarcoma.[2]and bladder cancer.[3] To date, there is no large data showing a clear association between PCOLCE and pan-cancer, therefore, we first used databases such as the cancer genome atlas (TCGA) and gene expression omnibus (GEO) for pan-cancer analysis of PCOLCE. We explored this in a group of factors, such as gene expression, survival status, DNA methylation, genetic alteration, and immune infiltration, to explore the potential molecular mechanisms of PCOLCE in pathogenesis and clinical prognosis.

2. Materials and methods

2.1. Gene expression analysis

We retrieved the target gene PCOLCE through the "Gene_DE" module of TIMER2 (http://timer.cistrome.org/) and observed the difference in the expression of PCOLCE in different tumors or specific tumor subtypes in the TCGA project between tumor and normal tissues. For some tumors without normal tissues matching the tumor tissue, we used the GEPIA2 server "Expression Analysis-Box Plots" module to search the TCGA and GTEx databases for expression differences between tumor tissues and tissues resembling normal tissues. We set P value = .01, log2FC (fold change) set = 1. In addition, we analyzed the difference of PCOLCE expression in different stages (stage I, stage II, stage III, and stage IV) of TCGA tumors by the "Pathological Stage Plot" module of HEPIA2 and plotted the violin map. We performed protein expression analysis of potential target genes through the clinical proteomic tumor analysis consortium (CPTAC) dataset on the UALCAN web server (http://ualcan.path.uab.edu/analysis-prot.html), which is an interactive network dedicated to the analysis of cancer omics data. So here we explored the difference in protein expression of PCOLCE in tumor tissues and normal tissues. At last, we also used THE HUMAN PROTEIN ATLAS server to explore the differences in PCOLCE expression in tissue samples.

2.2. Survival prognosis analysis

We downloaded normalized pan-cancer data from the UCSC database and used R software to model the COX data to analyze the effect of PCOLCE expression on the prognosis of pan-cancer. We again used the "Survival Map" module of the GEPIA2 database to obtain overall survival (OS) and disease-free survival data for PCOLCE expression in TCGA tumors for validation. A cutoff value of (50%) was used to define the high and low expression groups. log-rank tests were used for hypothesis testing, and survival maps were also obtained by GEPIA2 survival analysis module. We further used ggplot2 package to extract and analyze tumor RNAseq data from TGCA and GTEx databases and drew ROC curves to judge the potential predictive role of PCOLCE expression in generalized cancer.

2.3. Genetic alteration analysis

We used "Quick Search" in the cBioPortal web server (https://www.cbioportal.org/) to retrieve the genetic alteration signature of PCOLCE associated with "TCGA Pan Cancer Atlas Studies." In the "Cancer Types Summary" module, we found that PCOLCE was modified in most of TCGA tumors in terms of frequency, type of mutation and DNA copy number. We used the "Mutations" module to view the PCOLCE mutation site information in the 3D schematic of the protein structure. We also used the "Comparison" module to obtain the data on the overall, disease-free, progression-free, and disease-free survival for the TCGA cancer altered with PCOLCE genetic expression. Kaplan–Meier plots with log-rank P value were generated. We also explored the methylation level of the PCOLCE promoter. We downloaded the uniformly normalized pan-cancer dataset: TCGA Pan-Cancer (PANCAN, N = 10535, G = 60499) from the UCSC (https://xenabrowser.net/) database, from which we further extracted the PCOLCE expression data in each sample, we extracted the expression data from the previous study obtained from DNAss tumor stemness scores calculated by methylation features for each tumor,[4] and we integrated the stemness index and gene expression data of the samples. We describe the differential expression of PCOLCE methylation in pan-cancer through UALCAN website and prognostic study of PCOLCE methylation on pan-cancer in TCGA database using MEXPRESS server. Finally, the expression of PCOLCE was analyzed against common methylation forms using the R package for correlation type.

2.4. Immune infiltration analysis

We explored the relationship between PCOLCE expression and immune infiltration in all TCGA tumors by using the "immune-gene" module of the TIMER2 website. We chose to use the "TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, XCELL, MCPCOUNTER and EPIC algorithms" method to do immune infiltration assessment, mainly assessing immune cells is cancer-associated fibroblasts cells. The P values and partial correlation (cor) values were obtained by the purity-adjusted Spearman's rank correlation test. The data were visualized as Circle Chart, heatmap and scatter plot. Then, we further correlated the expression of PCOLCE in the TGCA database with the immune environment, immune cells, and immune pathways using the R package estimate, psych etc.

2.5. Gene enrichment analysis of PCOLCE

We used STRING website (https://string-db.org/) to search for protein names PCOLCE with matching species selection "Homo sapiens." Subsequently, we set the following main parameters for the selection of potential target genes: the minimum required interaction score selection was set to "low confidence (0.150)," the network edge meaning selection was set to "evidence," the interaction to the maximum number of interactors to be displayed is set to "No more than 50 interactors in the first shell" and the active interaction source is selected "experiment." Finally, we obtained the exact PCOLCE-binding protein. Using the "Similar Gene Detection" module of GEPIA2, we obtained the top 100 target genes associated with tumor and normal tissues in all TCGA databases. We also performed gene Pearson association correlation analysis between PCOLCE and the selected genes by using the "correlation analysis" module of GEPIA2. The point plots use log2 TPM and give P values and correlation coefficients (R). In particular, we used the "Gene_Corr" block of TIMER2 to provide heatmap data for our selected genes, which contains partial correlation (cor) and P values from the Spearman's rank correlation test for purity adjustment. We used an interactive Venn diagram for cross-tabulation analysis comparing genes that bind to PCOLCE and interact with PCOLCE. In fact, we upload the list of genes to the Database, visualize and synthesize the discovery (DAVID) by setting the selected identifiers and species information and get the functional annotation map. The enriched pathways were finally visualized with the "tidyr" and "ggplot2" R packages. In addition, we applied the R language package [R-3.6.3, 64-bit] (https://www.r-project.org/) for GO and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis of PCOLCE, during which we used the cnetplot function to convert (circular = F, colorEdge = T, node_label = T) BP (Biological process), CC (Cellular component) and MF (Molecular function) data, data visualized with a 2-tailed $P < .05$ set to be considered statistically significant.

3. Results

3.1. Expression analysis of PCOLCE

In our study, we aimed to investigate the role PCOLCE within human tumors. We first analyzed the expression of PCOLCE in different tumor cells and paired non-tumor tissues or similar tissues in the TCGA database. We applied the TIMER2 approach to analyze the expression status of PCOLCE across various cancer types of TCGA. As shown in Figure 1A, PCOLCE's expression was lower in BLCA, breast cancer (BRCA), CESE, liver hepatocellular carcinoma (LIHC), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC) tumor tissues than in the corresponding non-tumor tissues or similar
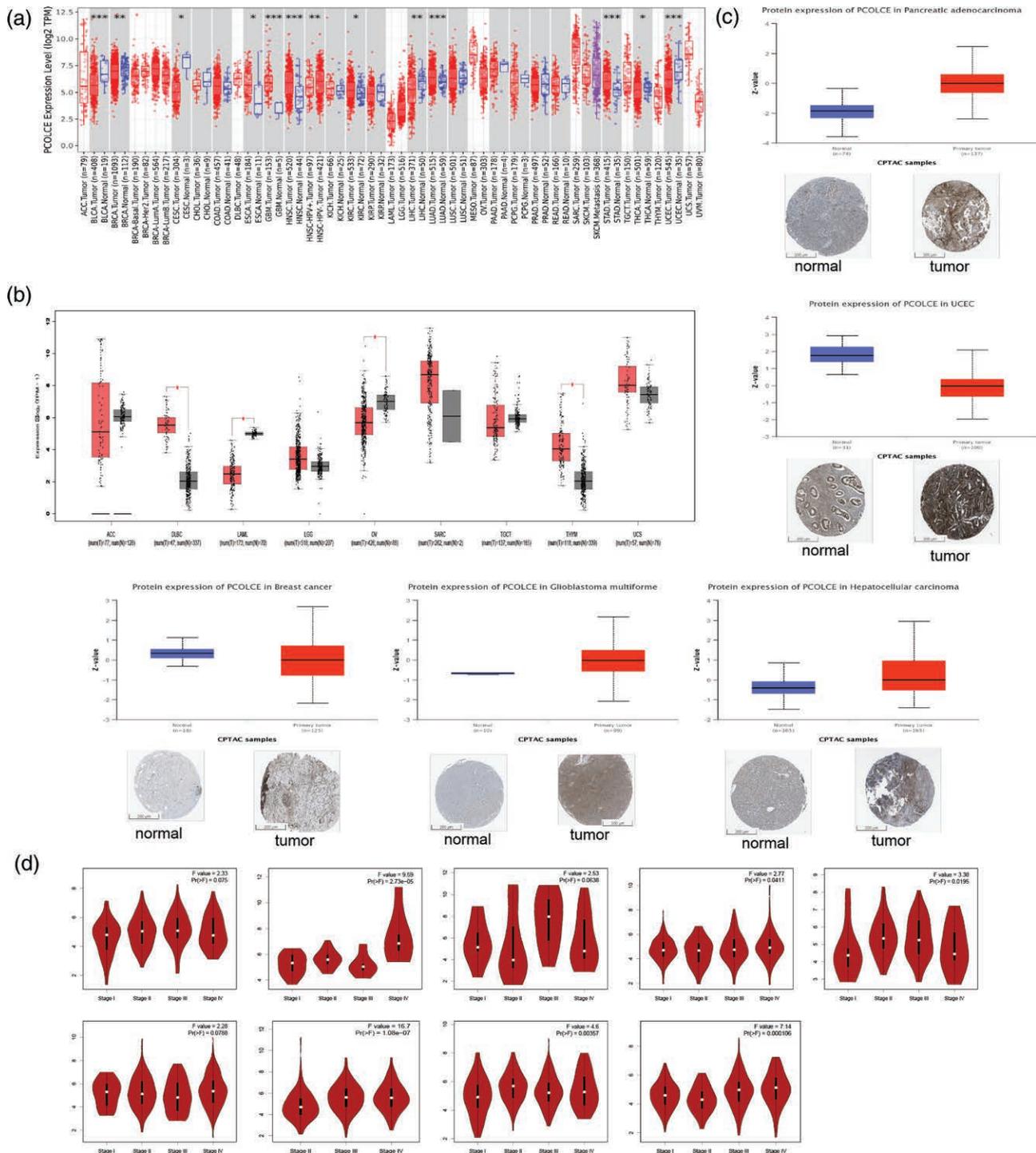
Figure 1. (A) Pan-cancer expression of PCOLCE in TCGA database was analyzed by TIMER2. (B) Differential PCOLCE expression was analyzed by GTEx. (C) Differential PCOLCE protein expression in GBM, LIHC, PAAD, BRCA and UCEC and microscopic tissue differences. (D) Differential PCOLCE expression in different tumor pathological stages. BRCA = breast cancer, GBM = glioblastoma multiforme, LIHC = liver hepatocellular carcinoma, PAAD = pancreatic adenocarcinoma, PCOLCE = procollagen C-endopeptidase enhancer, TCGA = the cancer genome atlas, uterine corpus endometrial carcinoma, UCEC = uterine corpus endometrial carcinoma.

tissues, and the expression was higher in esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), lung adenocarcinoma and stomach adenocarcinoma (STAD) tumor tissues. After including normal tissue data in the GTEx dataset as a control, we further explored the differences in PCOLCE expression between normal and tumor tissues in adrenocortical carcinoma (ACC), lymphoid neoplasm diffuse large

b-cell lymphoma (DLBC), acute myeloid leukemia, brain lower grade glioma (LGG), ovarian serous cystadenocarcinoma (OV), sarcoma (SARC), testicular germ cell tumors (TGCT), thymoma (THYM), and USC (Fig. 1B). We found that the expression level of PCOLCE in tumor tissues of DLBC, THYM was higher than its expression in normal tissues. In contrast, PCOLCE expression was significantly higher in OV non-tumor tissues than in tumor tissues. However, the expression of other tumors, we

did not find significant differences, for example ACC, LGG, SARC, TGCT and uterine carcinosarcoma (UCS). Analysis of the CPTAC dataset showed that total PCOLCE protein expression was lower in primary tumor tissues of GBM, LIHC, and Pancreatic adenocarcinoma (PAAD) compared to normal tissues (Fig. 1C, $P < .05$). In contrast, total protein expression of PCOLCE was higher in primary tumor tissues of Breast Cancer and UCEC than in normal tissues (Fig. 1C, $P < .05$). We also listed the differences in PCOLCE protein expression in different tumor tissues under the corresponding tumors under the microscope. See Figure 1C for details. We explored the correlation between PCOLCE expression and the pathological stage of different tumors including BLCA, WSCA, kidney chromophore, KIRC, STAD and THCA by using the "pathological stage map" module of HEPIA2 (Fig. 1D, all $P < .05$).

3.2. Survival analysis on PCOLCE

First, we performed a statistical analysis of PCOLCE expression in pan-cancer for overall survival in the UCSC database. As shown in Figure 2A, in GBMLGG ($P = 2.0e-61$), LGG ($P = 4.6e-18$), KIRC ($P = 8.4e-5$), kidney chromophobe ($P = 4.2e-4$), uveal melanoma (UVM) ($P = 1.2e-3$), KIPAN ($P = 1.4e-3$), skin cutaneous melanoma-P ($P = 7.9e-3$), GBM ($P = .01$), ACC ($P = .03$), STAD ($P = .04$) AND THCA ($P = .05$), PCOLCE expression was correlated with poor survival. And then we divided PCOLCE into high and low expression groups according to their expression levels in tumors, and then investigated the correlation between PCOLCE expression and clinical prognosis in different tumor patient groups using mainly TCGA and GEO datasets as a validation. As shown
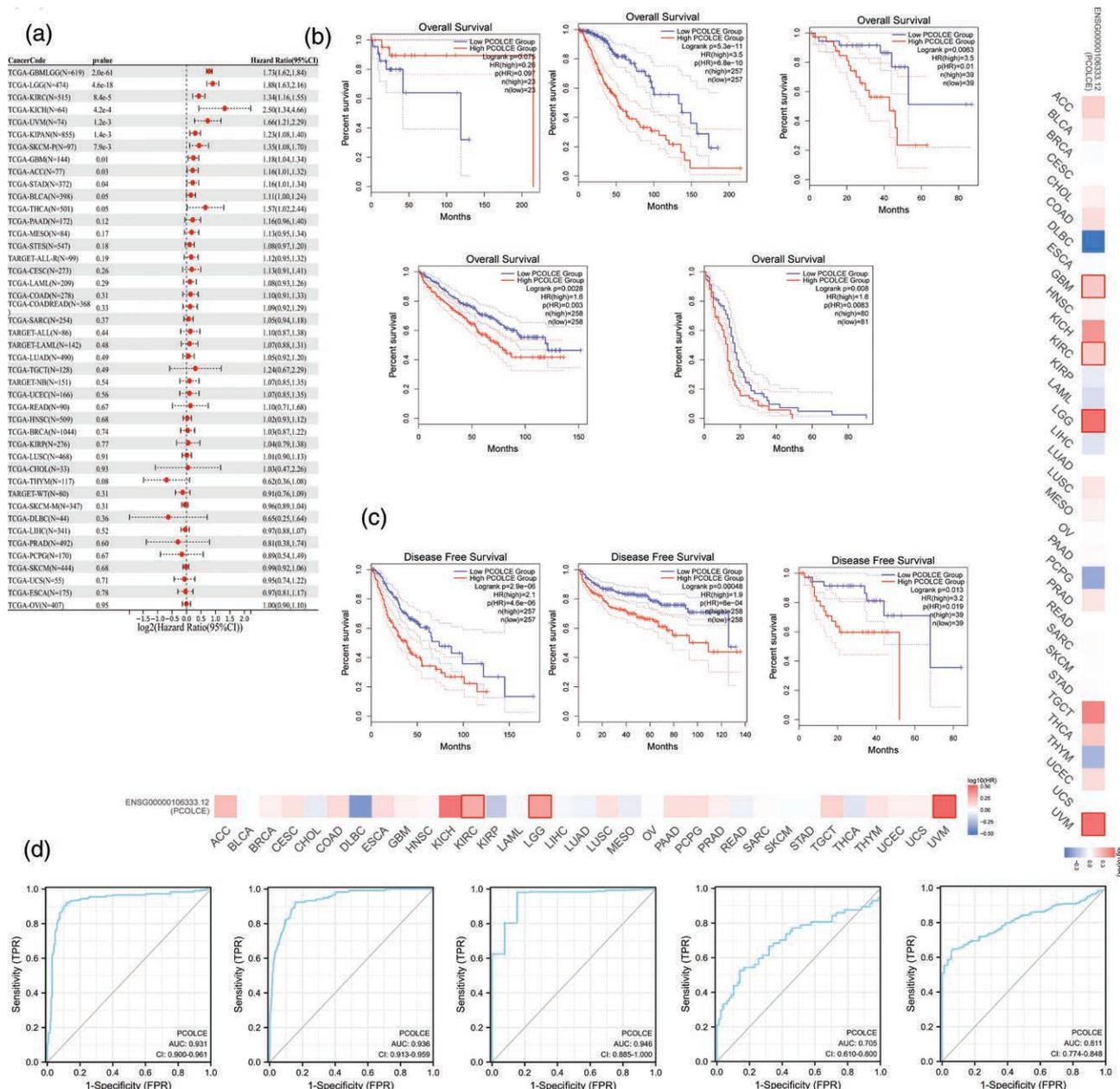
Figure 2. (A) Statistical analysis of PCOLCE expression and pan-cancer prognosis. (B) The relationship between PCOLCE expression and overall survival in GBM, KIRC, LGG, UVM and DLBC. (C) The relationship between PCOLCE expression and DFS in KIRC, LGG and UVM. (D) The diagnostic value of the gene was evaluated by ROC curve. DFS = disease-free survival, DLBC = lymphoid neoplasm diffuse large b-cell lymphoma, GBM = glioblastoma multiforme, brain lower grade glioma, KIRC = kidney renal clear cell carcinoma, LGG = brain lower grade glioma, PCOLCE = procollagen C-endopeptidase enhancer, UVM = uveal melanoma.

in Figure 2B, high PCOLCE expression was significantly associated with poor overall survival OS of the cancer, such as GBM ($P = .008$), KIRC ($P = .003$), LGG ($P = 0$) and UVM ($P = .006$) and within the TCGA project. Contrary to that higher expression levels of PCOLCE linked to good prognosis of OS for DLBC ($P = .075$). Disease-free survival analysis data in Figure 2C showed that a correlation between high PCOLCE expression levels and poor prognosis for the TCGA cases of KIRC ($P = .00$), LGG ($P = .000$), and UVM ($P = .013$). The above data suggest that PCOLCE expression is associated with significant survival prognostic differences in patients with different tumors. In addition, ROC analysis showed that the expressing of PCOLCE has excellent diagnostic value in PAAD (AUC = 0.931), THYM (AUC = 0.936), CESC (AUC = 0.946), and has general diagnostic value in UCS (AUC = O.705) and OV (AUC = 0.811) (Fig. 2D).

3.3. Genetic alteration analysis of PCOLCE

We found the altered status of the PCOLCE gene on different tumor samples in the TCGA dataset by a quick search in the Cbioportal server. As shown in Figure 3A, "Amplifying" mutations are most frequently seen in Esophageal Adenocarcinoma cases, with a frequency of alteration of nearly 9%, while in Stomach Adenocarcinoma and Diffuse Large B-cell Lymphoma they are more than 4%, and in Uterine Carcinosarcoma, Head and Neck squamous cell Carcinosarcoma and Lung Squamous

cell Carcinoma they are nearly 4%. PCOLCE is predominantly altered as a "mutation" in Skin Cutaneous Melanoma and uterine Corpus Endometrial Carcinoma, with a frequency of about 3%. As shown in Figure 3B: we further explored the types, loci and number of cases of PCOLCE gene mutations. We found that the percentage of samples with somatic mutations in PCOLCE was 0.6%, where missense mutations were the main type of genetic alterations, with 51 missense mutations identified in the NM_013363 dataset, accounting for approximately 69.86%. The types, sites and case number of the PCOLCE genetic alteration are further presented in Figure 3B. We found that missense mutation of PCOLCE was the main type of genetic alteration, they were distributed in 10 samples and involved tumor types with Breast Invasive Carcinoma, Ovarian Serous Cystadenocarcinoma, Uterine Corpus Endometrial Carcinoma, Lung Squamous Cell Carcinoma, Kidney Chromophobe, Prostate Adenocarcinoma, Stomach Adenocarcinoma and Colon Adenocarcinoma. In addition, PCOLCE also undergoes truncating mutations in Stomach Adenocarcinoma, Uterine Corpus Endometrial Carcinoma and Pancreatic Adenocarcinoma eventually in 4 samples contributing to the change in PCOLCE protein expression. We can also observe the above mentioned site-specific mutational alterations in the 3D structure of the PCOLCE protein (Fig. 3C). Although the probability of mutation in PCOCLE is low from our current analysis, our further study found that mutation is still strongly associated with survival prognosis. We further
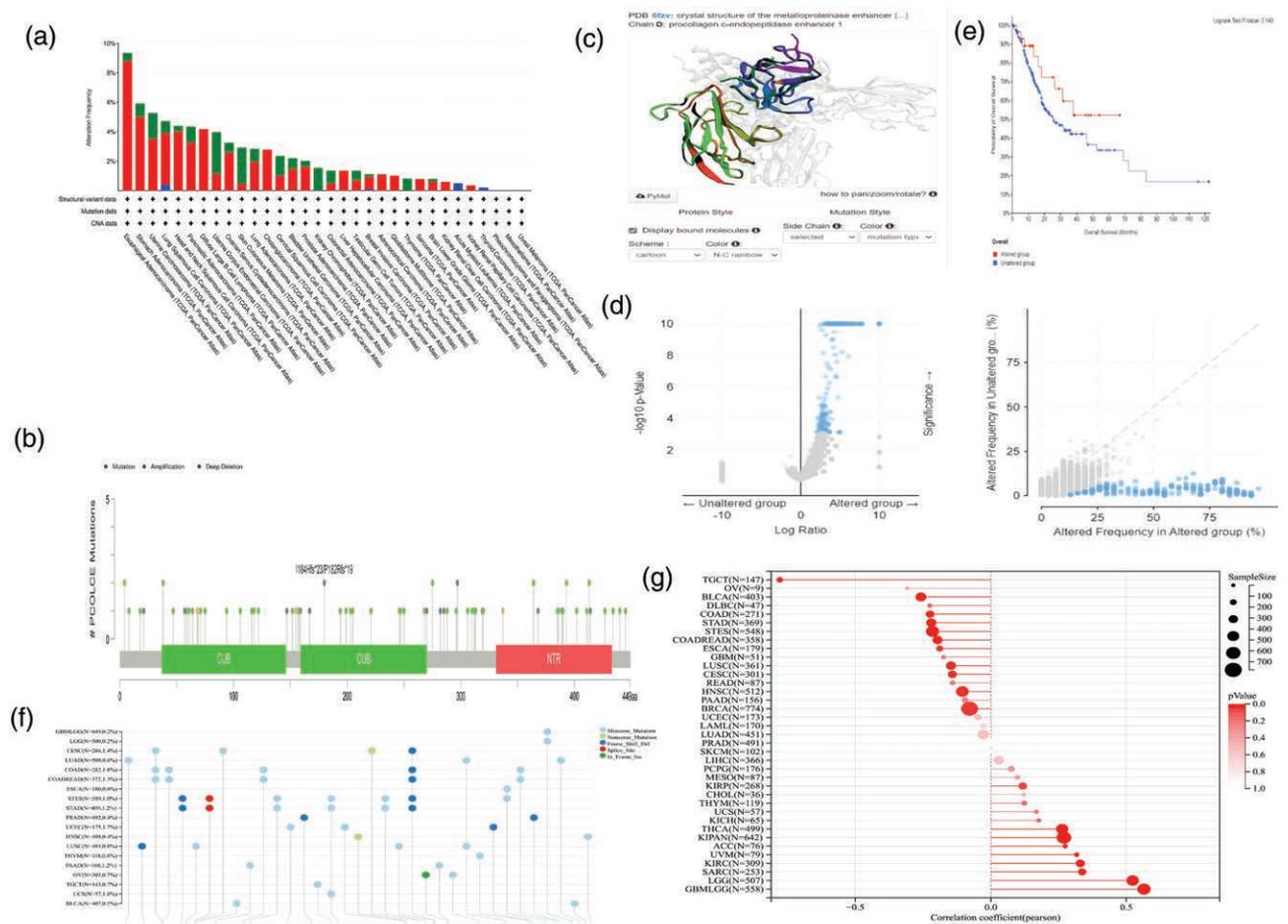
Figure 3. (A) PCOLCE gene mutation information in tumor samples. (B) Information on the main mutation sites, mutation types and mutation numbers of PCOLCE gene in tumor samples. (C) Information on mutation sites in the 3-dimensional structure of PCOLCE gene. (D) PCOLCE gene mutation frequency in tumor samples. (E) PCOLCE mutations are associated with prognosis in Esophageal Carcinoma. (F) The PCOLCE mutation status was reconfirmed in the pan-cancer data. (G) Expression of PCOLCE and tumor stemness score of different tumors. PCOLCE = procollagen C-endopeptidase enhancer.

explored the potential association between PCOLCE gene alterations in the clinical survival prognosis of cancer patients. As shown in Figure 3E, Esophageal Carcinoma patients with PCOLCE alterations had a better prognosis in terms of overall survival ($P < .05$) compared with patients without PCOLCE alterations. We also provide a detailed description of the mutation frequency of the mutation group (Fig. 3D). We further validated the mutational status of PCOLCE in TCGA database tumors using MuTect2 software. The probability of PCOLCE mutation was found to be extremely low, with the probability of mutation in all tumors being below 2% or even below 1% (Fig. 3F). Finally, we extracted PCOLCE expression data from the TCGA database and performed log2($x + 0.001$) transformation and found that PCOLCE expression was positively correlated with tumor stemness score in the following 8 tumor species, such as GBMLGG ($P = 2.11e-48$), LGG ($P = 5.54e-37$), SARC ($P = 3.93e-8$), KIPAN ($P = 3.63e-12$), KIRC ($P = 2.679e-9$), THCA ($P = 2.358e-9$), UVM ($P = .004$), ACC ($P = .016$). It was negatively correlated with the following 11 tumor stemness scores, such as CESC ($P = .0140$), colon adenocarcinoma (COAD) ($P = .000$), COAD rectum adenocarcinoma ($P = .000$), BRCA ($P = .029$), ESCA ($P = .011$), stomach and esophageal carcinoma (STES) ($P = 3.842e-7$), STAD ($P = .000$), HNSC ($P = .017$), lung squamous cell carcinoma (LUSC) ($P = .005$), TGCT ($P = 3.83e-31$), BLCA($P = 1.573e-7$) (Fig. 3G).

3.4. DNA methylation analysis of PCOLCE

We compared the differences in PCOLCE promoter methylation levels in normal tissues and primary tumor tissues. Using the TCGA dataset, 12 types of tumors (BLCA, BRCA, CHOL, COAD, HNSC, KIRC, kidney renal papillary cell carcinoma (KIRP), LIHC, LUSC, PAAD, TGCT and UCEC) were analyzed in Figure 4A. We found that the methylation levels of the PCOLCE promoter appeared significantly different in different tumors and corresponding non-tumor tissues. We found that the methylation level of the pcolce promoter was significantly higher in BRCA, CHOL, KIRP, UCEC tumors than in normal tissues. In contrast, the methylation level of the PCOLCE promoter was higher in normal tissues than in tumor tissues in BLCA, COAD, HNSC, KIRC, LIHC, LUSC, PAAD, TGCT. All the above P values were <.05. We used the MEXPRESS approach to investigate the potential association between PCOLCE DNA methylation and the pathogenesis of the tumor. Based on cases in the TCGA database, we found that PCOLCE methylation mutations were associated with a good prognosis for KIRC ($r = -0.123$, $P < .01$), PAAD ($r = -0.214$, $P < .01$) and UVM ($r = -0.247$, $P < .05$) (Fig. 4C). And far more than that, we observed a significant negative correlation of PCOLCE DNA methylation and gene expression at multiple probes of the nonpromoted region, such cg09326362 (RR = 0.250), cg13655570 (PR = 0.268), cg26100986 (PR = 0.289), cg25680486 (RR = 0.242) and cg22082800 (RR = 0.253), a significant positive correlation of PCOLCE DNA methylation and gene expression such as cg01706943 (PR = 0.354), cg26777475 (PR = 0.372), cg06402330 (PR = 0.401) in TGCT UVM. So, we further did in-depth analysis of several common RNA methylation forms of PCOCLE gene by R software, as shown in Figure 4B, in OV, the expression of PCOLCE showed positive correlation with common RNA methylation such as M6A, M5C, and M1A. However, no statistical effect of PCOLCE expression on M6A and M5C was found in DLBC.

3.5. Immune infiltration analysis of PCOLCE

Tumor-infiltrating immune cells, as an important component of the tumor microenvironment, play a crucial role in tumorigenesis, progression and metastasis.[5,6] Earlier studies reported suggest that cancer-associated fibroblasts in the tumor microenvironment stroma can be involved in regulating the function of a variety of different tumor-infiltrating immune cells.[7,8] Therefore, we first evaluated the effect of PCOLCE expression on the immune microenvironment of the pan-cancer. As shown in Figure 5A, all tumor's immune environment is affected by PCOLCE expression. Next, we continued to investigate the effect of PCOLCE expression on the activity of common immune cells in various tumors. As shown in Figure 5B, we found that fibroblasts in almost all tumors were affected by PCOLCE expression. So, we went a step further, we used different web servers and different algorithms to evaluate the potential relationship of PCOLCE gene expression with the level of infiltration of different types of immune cells in the TCGA database, Furthermore, we observed a statistically positive correlation between PCOLCE expression and infiltrative valuation of cancer-associated fibroblasts in numerous TCGA tumors such as BLCA, BRCA, COAD, ESCA, HNSC, HNSC- hpv +/-, OV, PAAD, STAD, THCA, THYM and so on (Fig. 5D). As shown in Figure 5D, a scatter plot for assessing tumor infiltration obtained by us using a single algorithm. For example, according to the MCPCOUNTER algorithm, the expression level of PCOLCE in COAD was positively correlated with the level of infiltration of cancer-associated fibroblasts (Fig. 5D, cor = -0.361, $P = 5.34e-32$). Next, we did a comprehensive assessment of the correlation between PCOLCE expression and immune checkpoints. As shown in Figure 5C, we found a positive correlation between PCOLCE expression and immune checkpoint CD276 in all tumors in addition to mesothelioma, CHOL, UCS and ACC. In LGG, rectum adenocarcinoma, THCA and CHOL, PCOLCE expression was positively correlated with most of the immune checkpoints, but negatively correlated with most of the immune checkpoints in TCGA. Then, we evaluated the correlation between PCOLCE expression and common immune cells in pan-cancer, and found that central memory CD4 T cell, central memory CD8 T cell, Gamma delta T cell, Effector memory CD4 T cell, effector memory CD8 T cell, Macrophage Natural killer cell, Natural killer T cell and so on was positively correlated with PCOLCE expression in most tumors, while Activated B cell, Activated CD4 T cell, Activated CD8 T cell, CD56dim natural killer cell, Immature B cell, Neutrophil, Type 1 T helper cell and Activated dendritic cell was negatively correlated with B cells in TGCA (Fig. 5E). In order to further explore the relationship between PCOLCE expression and immune infiltration, we explored the possible relationship between PCOLCE expression and tumor mutation burden (TMB) microsatellite instability (MSI) and immune neoantigens. As shown in Fig5F, 5G and 5H, the expression of PCOLCE was positively correlated with TMB in LUSC, $P = .074$; in TGCT the expression of PCOLCE was positively correlated with MSI, and negatively correlated in UCEC ($P = .031$) and CHOL ($P = .006$); the expression of PCOLCE was positively correlated with neoantigen in COAD ($P = .041$), and negatively correlated in CESC ($P = .033$) and HNSC ($P = .095$).

3.6. Enrichment analysis of PCOLCE-related partners

To further investigate the molecular mechanism of the role of PCOLCE gene in tumorigenesis, we attempted to screen targeting PCOLCE binding protein and PCOLCE expression-related genes for relevant pathway enrichment analysis. In total, we obtained 45 PCOLCE binding proteins by STRING tool. Figure 6A shows the protein interaction network of our screen. We again used the GEPIA2 tool to identify the top 100 genes associated with PCOLCE expression in TCGA's database. As shown in Figure 6B, the PCOLCE expression level was positively correlated with that of P3H1 ($R = 0.64$), RCN3 ($R = 0.59$), MRC2 ($R = 0.57$), COL1A2 ($R = 0.55$) and FKBP7 ($R = 0.53$) genes (all $P = .000$). The corresponding heatmap data also showed that PCOLCE was positively associated with
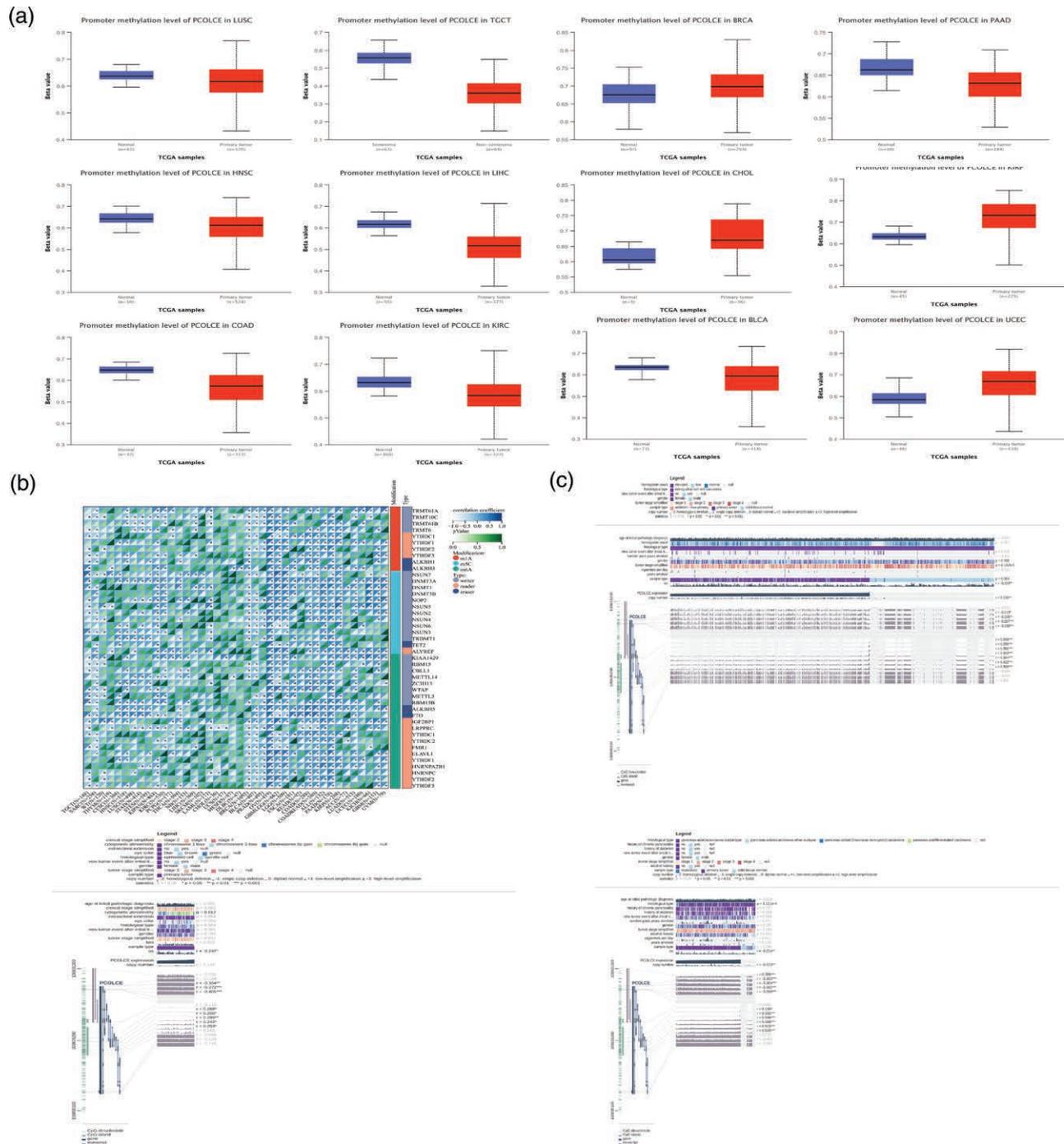
Figure 4. (A) Differential levels of promoter methylation of PCOLCE genes in 12 tumors. (B) Statistical differences between several major methylation types of PCOLCE and pan-cancer. (C) Relationship between PCOLCE methylation and clinical prognosis in KIRC PAAD and UVM. KIRC = kidney renal clear cell carcinoma, PAAD = pancreatic adenocarcinoma, PCOLCE = procollagen C-endopeptidase enhancer, UVM = uveal melanoma.

the above 5 genes in most TCGA tumors (Fig. 6C). An intersection analysis of the above 2 groups showed 12 common member genes (Fig. 6D). We integrated the 2 sets of appeal data for analysis and performed KEGG and GO enrichment analysis. The GO enrichment analysis of Figure 6E suggests that the main function of PCOLCE is located in "Extracellular Region" and "Extracellular Matrix." In contrast, KEGG enrichment analysis suggested a possible involvement of PCOLCE in the tumor pathway: "Focal adhesion, Human papillomavirus infection, PI3K-Akt signal pathway, ECM-receptor interaction, AGE-RAGE signal pathway, Relaxin signal pathway, Proteoglycans in cancer, Small cell lung cancer"

4. Discussion

It has been previously reported that PCOLCE can be expressed in multiple normal organs such as lung, liver, heart, brain etc, and there are few reports of diseases associated with it. Only a few papers have reported its possible involvement in cancer.[1,2,9] In our study, we further explored the expression of PCOLCE in various tumors and the corresponding clinical prognosis. Of course, whether PCOLCE can play a role in the pathogenesis of different tumors through some common molecular mechanism remains to be confirmed. Through literature search, we did not retrieve any literature on pan-cancer analysis of PCOLCE from a holistic tumor perspective.
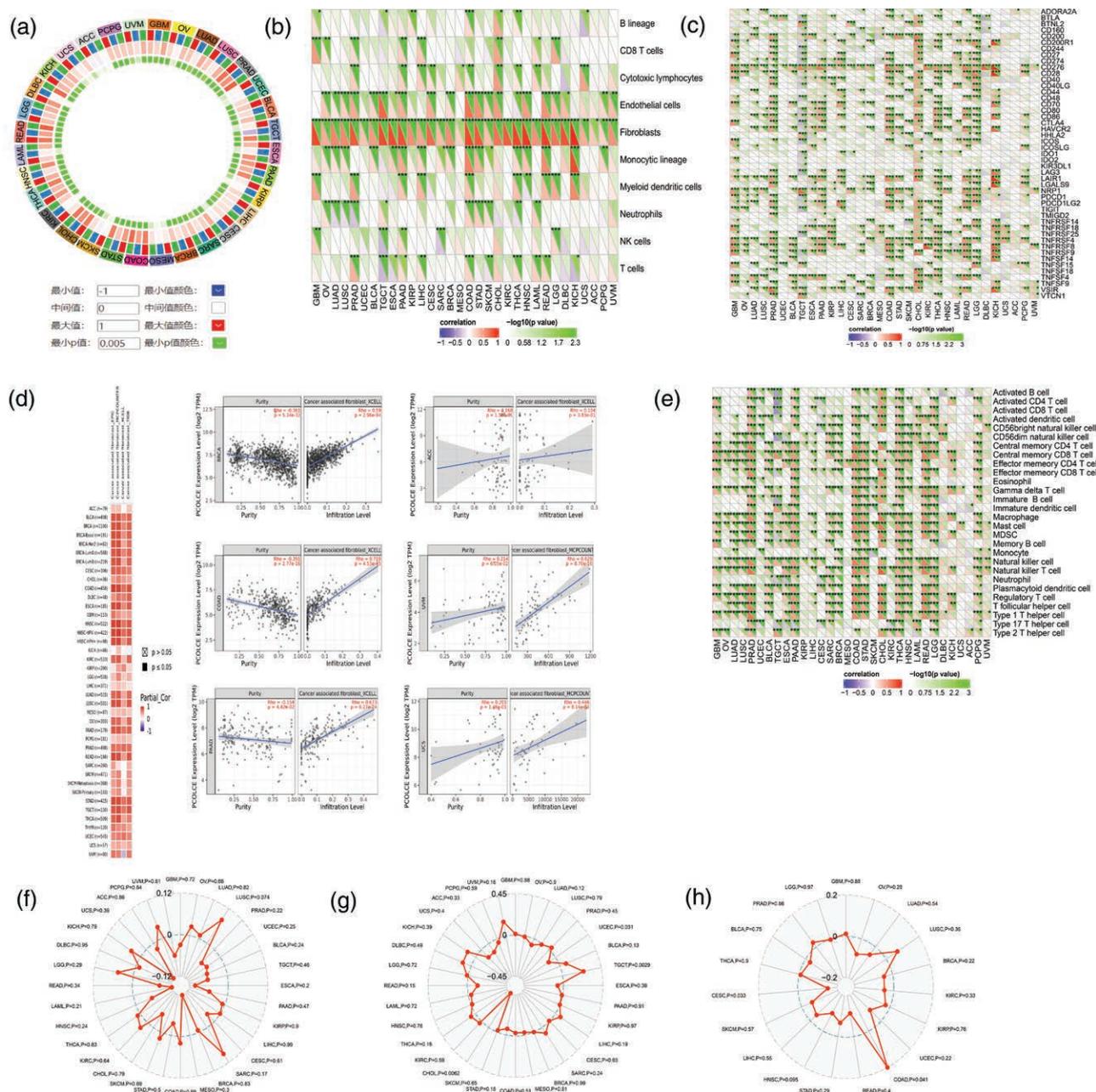
Figure 5. (A) Effects of PCOLCE on the pan-cancer immune environment. (B) Effects of PCOLCE expression on primary immune cells. (C) Relationship between PCOLCE expression and major immune checkpoints in pan-cancer. (D) PCOLCE expression correlates with cancer-associated fibroblasts in pan-cancer. (E) Effects of PCOLCE expression on immune pathways in pan-cancer. (F) Relationship between PCOLCE expression and tumor mutation burden (TMB) in pan-cancer. (G) Relationship between PCOLCE expression and microsatellite instability (MSI) in pan-cancer. (H) Relationship between PCOLCE expression and immune neoantigens in pan-cancer. PCOLCE = procollagen C-endopeptidase enhancer.

Therefore, we performed a series of assays based on data from TCGA, CPTAC and GEO databases to characterize the gene expression and gene alteration molecular profiles of PCOLCE in 33 different tumors. The analysis showed that PCOLCE was differentially expressed in 17 of 33 different tumors, and similarly its protein expression differed in 12 tumor species. We also list the differences in protein expression in different tumor or non-tumor tissues for your reference. In addition, we found that the expression of PCOLCE differed in 9 tumor types with different stages. As mentioned above, the presence of various differences predicts differences in disease prognosis. Meanwhile, ROC curve analysis showed that PCOLCE expression could be a good predictor of PAAD, THYM and CESC. Survival prognostic analysis data suggest different findings

in different tumors. In our study, by the UCSC database, we found that high expression of PCOLCE was associated with poorer prognosis in 11 tumors by statistical means. We using the GEPIA2 tool to find a statistical correlation between high PCOLCE expression and poorer overall survival prognosis in 4 tumors and better overall survival prognosis in 1 tumor. Different data processing or updated survival information may contribute to this result. Therefore, we validated the results we obtained using another web server, Oncolnc (http://www.oncolnc.org/), which performs Cox regression survival analysis using data from the TCGA database queue. As a result, we performed a survival analysis using the Kaplan-Meier plotter method and found that high PCOLCE expression in KIRC ($P = .00027$), KIRP ($P = .0093$), LGG ($P = .00043$) was
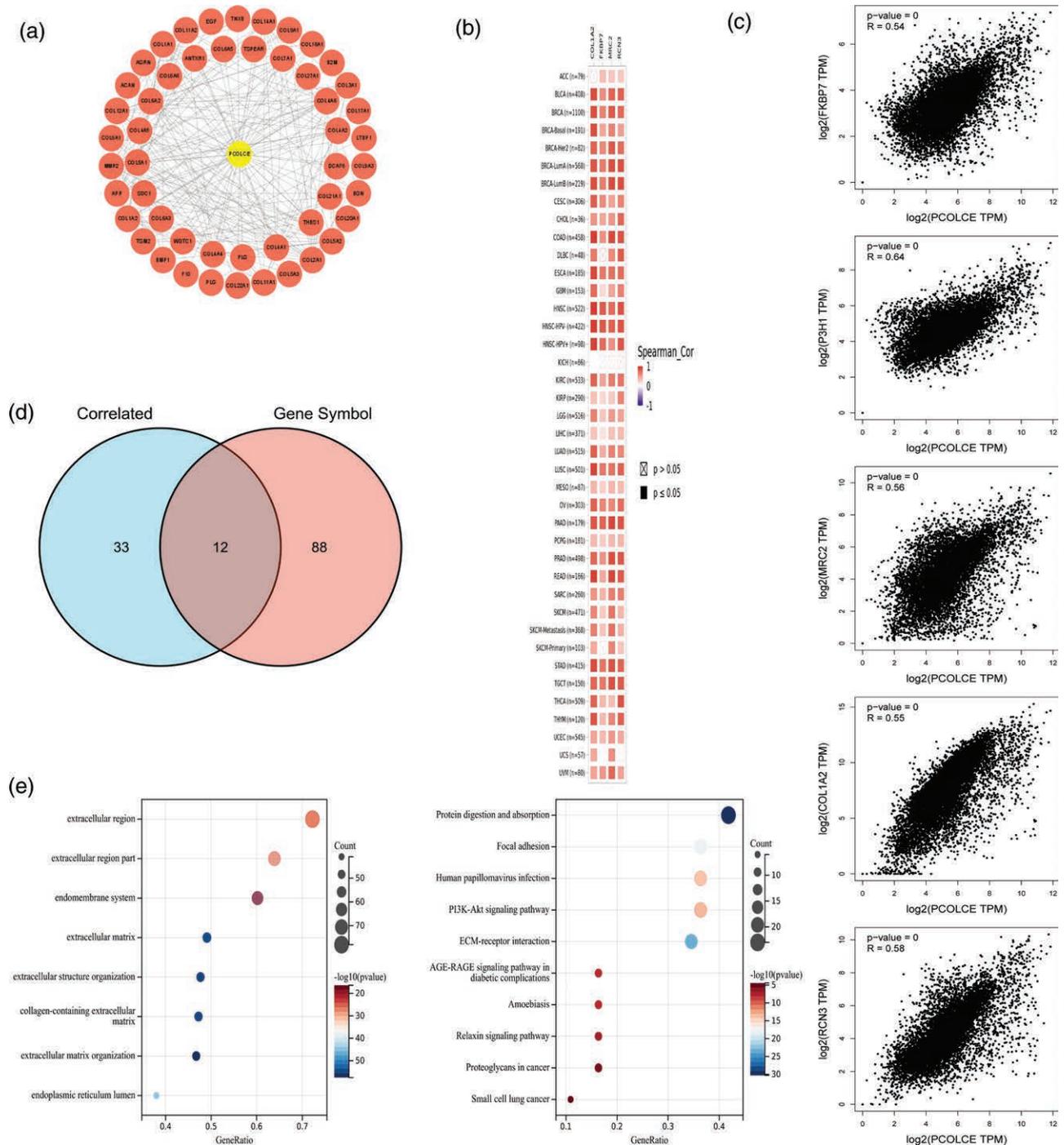
Figure 6. (A) Protein–protein interaction (PPI) networks analysis of PCOLCE. (B) Expression of genes positively related to PCOLCE in pan-cancer. (C) Heatmap of the relationship between PCOLCE and the top 5 positively correlated genes. (D) Analysis of PPI networks and correlated gene common gene sets. (E) KEGG and GO enrichment of PCOLCE-associated genes. GO = gene ontology, KEGG = Kyoto encyclopedia of genes and genomes, PCOLCE = procollagen C-endopeptidase enhancer.

statistically different from poor survival prognosis. This result is basically the same as our previous result. However, we did not find a statistical association between high expression of PCOLCE and good prognosis due to missing partial data.

In this paper, we also describe the PCOLCE mutation in detail. The mutation rate of PCOLCE is quite low compared to other oncogenes. However, the low mutation probability does not mean that the mutation is not oncogenic. An "amplifying" mutation in PCOLCE in 9% of esophageal cancers can alter the clinical prognosis. "Amplifying" mutations are the

most common oncogenic pathway, leading to the production of angiogenic factors, inflammatory cytokines and chemokines and modulating the biochemical properties of the extracellular matrix, altering cellular transport mechanisms, promoting the secretion of membrane proteins and promoting circulation, all of which contribute to the local and systemic pro-carcinogenic effects.[10,11] Genes can cause cancer through epigenetic alterations rather than changes in their own sequence leading to changes in the mode of inheritance, commonly methylation of cytosine residues. The study of DNA methylation is important

because it can be specifically detected and can be used as a detection tool for earlier cancers.[12,13] We found that PCOLCE methylation was detectable and differential in 12 tumor species and improved the prognosis of KIRC, PAAD and UVM. These findings could provide the basis for early prevention and detection of the 3 malignant tumors in the appeal.

However, previous studies have shown that PCOLCE-transcribed proteins are involved in the reorganization of extracellular and corneal repair as major components of the extracellular matrix. It has been reported that PCOLCE may be associated with the immune microenvironment of tumors, with implications for tumor prognosis and the potential to become a tumor marker, a finding that is consistent with ours.[1,3] Although our conclusions are consistent, the mechanism of action involved is still unclear. However, from our findings, the following points may be appropriate explanations: first, PCOLCE expression was positively correlated with the level of immune infiltration, and the data model Immune-based prognostic signature for OV (IPSOV) provided by previous studies validated in multiple independent datasets showed that high immune infiltration was associated with low survival probability was associated with statistically significant differences.[14] Our study showed that PCOLCE expression was significantly associated with the immune microenvironment of 33 tumors. The immune microenvironment is an extremely complex mechanism of action in which tumor cells and immune cells coexist, and is 1 of the key mechanisms of tumorigenesis and development.[15,16] By further digging we found that PCOLCE expression in various tumors is closely related to fibroblasts, which are the main components involved in the extracellular matrix, tumor microenvironment and cellular immunity, and can even have an impact on chemotherapy resistance.[17,18] Theoretically, PCOLCE is predictive of tumor chemotherapy resistance. The immune checkpoint CD276, which is closely associated with PCOLCE, belongs to the same family as PD-1 and has synergistic effects. CD276 is also an upstream regulatory molecule of the immune nova CTLA4, which can directly participate in the regulation of tumor cell immunity and thus affect the malignant biological behavior of tumor cells.[19,20] This further elucidates how PCOLCE affects the tumor microenvironment. Moreover, PCOLCE can be a major predictor of Immune checkpoints inhibitors (ICIs). Secondly, regardless of the previous studies or the results we obtained, PCOLCE expression acts mainly in the extracellular matrix, which is a major component of the cellular microenvironment. Therefore, abnormal expression of PCOLCE can lead to alterations in the cellular microenvironment, resulting in malfunctioning of cellular biological behavior.[21,22] The effects of the extracellular matrix on a variety of different tumors are well known,[23,24] and we will not dwell on them here.

Not only that. Through KEGG enrichment analysis, we also found that PCOLCE may act on tumors in different ways. For example: PI3K-Akt signal pathway and AGE-RAGE signal pathway, PI3K-Akt is an important signal pathway that transmits stimuli from extracellular stimulators into the cell, allowing the cell to adapt to the extracellular environment by transducing signals from the extracellular environment into the cell. Overstimulation or alteration of epigenetic modifications of this pathway is an important mechanism leading to cancer.[25] Several tumor inhibitors of this pathway are currently approved by the FDA and used in the clinic with good benefits.[26,27] Therefore, PCOLCE has the potential to be an important predictor of P small molecule targeted inhibitors. Activation of AGE-RAGE signal pathway can significantly affect various cell death mechanisms such as apoptosis, autophagy and necroptosis, as well as disrupt intracellular redox homeostasis and promote cancer cell survival.[28,29] We searched the PubMed database and found no reports of PCOLCE with AGE-RAGE signal pathway and PI3K-Aktsignal pathway. So how does PCOLCE act with the above signal pathways? This needs to be studied more deeply.

Relaxin is a highly potent hormone that promotes uterine and mammary gland growth and development in women and has significant effects on the male reproductive system. It has been shown that excessive relaxin expression can induce cell proliferation and increase tumor angiogenesis to promote breast and prostate cancer tumor cell growth and metastasis.[30,31] If it can be confirmed that PCOLCE expression promotes Relaxin signal pathways, PCOLCE has also been identified as an important target for the control of breast or prostate cancer. Proteoglycans, also an important component of the extracellular matrix, are widely involved in inflammatory and tumorigenic processes through structural tissue remodeling and cell signal in the extracellular matrix. Proteoglycans usually mediate intercellular (and also tumor cell) signal and control tumor cell properties, phenotype, and angiogenesis, and even tumor cell drug resistance.[32]

We then searched the PUMED database and found that earlier studies have found a clear correlation between serum albumin levels and survival of patients with malignant tumors, who may experience poorer survival if their serum albumin water is low.[33] Protein anabolism regulated by MAPK/ERK signal pathway, chromatin silencing and fibrinolytic phylosomes is the core biological process of HGSOC, which is closely related to the long-term overall survival of OV.[34] Transcriptome sequencing showed that ECM-receptor expression was affected by Mex3a, and RNA immunoprecipitation (RIP) assays showed that Mex3a directly bound to LAMA2 mRNA, increasing LAMA2 mRNA instability and that low expression of LAMA2 mRNA inhibited lung adenocarcinoma metastasis.[35] In the study of the mechanism of brefeldin A treatment for human epithelial ovarian malignant tumors, it was found that the lactone antibiotic brefeldin A was able to induce apoptosis-associated protein activity while inhibiting cell adhesion capacity and migration ability.[36] Therefore, PCOLCE can also inhibit tumorigenesis in tumors by inhibiting focal adhesion and thus tumor development. In other tumors, such as colorectal and gastric cancers, previous reported results are the same as those we obtained, that is PCOLCE expression is associated with a poorer prognosis.[37,38] However, we obtained other results that have not been reported in the literature, such as GBM, UVM and KIRC.

In the present study, we explored the evidence of a potential correlation with PCOLCE expression in all TCGA tumors for the first time. In addition, we collected information on genes related to PCOLCE binding and PCOLCE expression in all tumors, to determine the potential role of "protein metabolism," "ECM-receptor interactions" and "focal adhesion" in the etiology or pathogenesis of cancer, in addition to the extracellular matrix.

We applied multiple immune deconvolution methods to observe a statistical positive correlation between PCOLCE expression and the immune infiltration level of cancer-associated fibroblasts in most of tumor in the TCGA database. The biological behavior of cancer-associated fibroblasts in malignant tumors is complex. In addition to being a major structural and functional component of the extracellular matrix, they are involved in numerous cancer-related mechanisms, such as tumor angiogenesis, immune infiltration and tumor metabolism. Therefore, the mechanism of action of cancer-associated fibroblasts is not well understood, but it is usually assumed to participate in and promote tumorigenesis and metastasis.[7,39] The enrichment analysis revealed that abnormal expression of PCOLCE affects the function and structure of the extracellular matrix. This clearly echoes our results and fully demonstrates the logic of our analysis.

We also found that methylation of PCOLCE DNA inhibits or promotes the expression of PCOLCE, and this finding can provide a good idea for our future research on the inhibition of PCOLCE. Additional evidence is merited for the potential role of PCOLCE DNA methylation in the tumorigenesis of TGCT. Previous promoter methylation is considered to be a marker of oncogene activation and oncogene silencing in malignant

tumors. The accumulation of epigenetic alterations due to promoter methylation has also been pointed out as a major factor in age-related diseases.[40] However, there are now studies that suggest that promoter hypermethylation may provide a survival benefit for cancer patients.[41,42] We found abnormal PCOLCE promoter methylation in 12 tumors, and it is unclear what impact this will have on patients.

In conclusion, our 1st pan-cancer analysis of PCOLCE showed a statistical correlation between PCOLCE expression and clinical prognosis, immune cell infiltration, tumor mutation, and further explored from its relevant pathways of action, which contributes to the understanding of the role of PCOLCE in tumorigenesis from the perspective of clinical tumor samples.

Author contributions

Conceptualization: Qiuyun Li.
Formal analysis: Hui Gao.
Methodology: Qiuyun Li.
Software: Hui Gao.
Writing – original draft: Hui Gao.
Writing – review & editing: Qiuyun Li.

References

[1] Xiang A, Lin X, Xu L, et al. PCOLCE is potent prognostic biomarker and associates with immune infiltration in gastric cancer. Front Mol Biosci. 2020;7:544895.

[2] Wang S, Zhong L, Li Y, et al. Up-regulation of PCOLCE by TWIST1 promotes metastasis in Osteosarcoma. Theranostics. 2019;9:4342–53.

[3] Song Y, Du Y, Qin C, et al. Gemcitabine-resistant biomarkers in bladder cancer are associated with tumor-immune microenvironment. Front Cell Dev Biol. 2021;9:809620.

[4] Malta TM, Sokolov A, Gentles AJ, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell. 2018;173:338–354.e15.

[5] DePeaux K, Delgoffe GM. Metabolic barriers to cancer immunotherapy. Nat Rev Immunol. 2021;21:785–97.

[6] Mao X, Xu J, Wang W, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer. 2021;20:131.

[7] Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov. 2019;18:99–115.

[8] Kwa MQ, Herum KM, Brakebusch C. Cancer-associated fibroblasts: how do they contribute to metastasis? Clin Exp Metastasis. 2019;36:71–86.

[9] Ligon AH, Scott IC, Takahara K, et al. PCOLCE deletion and expression analyses in uterine leiomyomata. Cancer Genet Cytogenet. 2002;137:133–7.

[10] Capaci V, Mantovani F, Del Sal G. Amplifying tumor-stroma communication: an emerging oncogenic function of mutant p53. Front Oncol. 2020;10:614230.

[11] Brady JJ, Chuang CH, Greenside PG, et al. An Arntl2-driven secretome enables lung adenocarcinoma metastatic self-sufficiency. Cancer Cell. 2016;29:697–710.

[12] Baylin SB, Jones PA. Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol. 2016;8:a019505.

[13] Morgan AE, Davies TJ, Mc Auley MT. The role of DNA methylation in ageing and cancer. Proc Nutr Soc. 2018;77:412–22.

[14] Shen S, Wang G, Zhang R, et al. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. EBioMedicine. 2019;40:318–26.

[15] Schulz M, Salamero-Boix A, Niesel K, et al. Microenvironmental regulation of tumor progression and therapeutic response in brain metastasis. Front Immunol. 2019;10:1713.

[16] Paluskievicz CM, Cao X, Abdi R, et al. T regulatory cells and priming the suppressive tumor microenvironment. Front Immunol. 2019;10:2453.

[17] Kalluri R. The biology and function of fibroblasts in cancer. Nat Rev Cancer. 2016;16:582–98.

[18] Biffi G, Tuveson DA. Diversity and biology of cancer-associated fibroblasts. Physiol Rev. 2021;101:147–76.

[19] Liu S, Liang J, Liu Z, et al. The role of CD276 in cancers. Front Oncol. 2021;11:654684.

[20] Wang C, Li Y, Jia L, et al. CD276 expression enables squamous cell carcinoma stem cells to evade immune surveillance. Cell Stem Cell. 2021;28:1597–1613.e7.

[21] Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 2014;15:1243–53.

[22] Walker C, Mojares E, Del Rio Hernandez A. Role of extracellular matrix in development and cancer progression. Int J Mol Sci. 2018;19:3028.

[23] Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. Nat Rev Cancer. 2014;14:430–9.

[24] Zanconato F, Cordenonsi M, Piccolo S. YAP and TAZ: a signalling hub of the tumour microenvironment. Nat Rev Cancer. 2019;19:454–64.

[25] Noorolyai S, Shajari N, Baghbani E, et al. The relation between PI3K/AKT signalling pathway and cancer. Gene. 2019;698:120–8.

[26] Alzahrani AS. PI3K/Akt/mTOR inhibitors in cancer: at the bench and bedside. Semin Cancer Biol. 2019;59:125–32.

[27] Costa RLB, Han HS, Gradishar WJ. Targeting the PI3K/AKT/mTOR pathway in triple-negative breast cancer: a review. Breast Cancer Res Treat. 2018;169:397–406.

[28] Waghela BN, Vaidya FU, Ranjan K, et al. AGE-RAGE synergy influences programmed cell death signaling to promote cancer. Mol Cell Biochem. 2021;476:585–98.

[29] Muthyalaiah YS, Jonnalagadda B, John CM, et al. Impact of Advanced Glycation End products (AGEs) and its receptor (RAGE) on cancer metabolic signaling pathways and its progression. Glycoconj J. 2021;38:717–34.

[30] Tashima LS, Mazoujian G, Bryant-Greenwood GD. Human relaxins in normal, benign and neoplastic breast tissue. J Mol Endocrinol. 1994;12:351–64.

[31] Silvertown JD, Summerlee AJ, Klonisch T. Relaxin-like peptides in cancer. Int J Cancer. 2003;107:513–9.

[32] Theocharis AD, Karamanos NK. Proteoglycans remodeling in cancer: underlying molecular mechanisms. Matrix Biol. 2019;75–76:220–59.

[33] Ayhan A, Gunakan E, Alyazici I, et al. The preoperative albumin level is an independent prognostic factor for optimally debulked epithelial ovarian cancer. Arch Gynecol Obstet. 2017;296:989–95.

[34] Wang L, Sun T, Li S, et al. Protein anabolism is key to long-term survival in high-grade serous ovarian cancer. Transl Oncol. 2021;14:100885.

[35] Liang J, Li H, Han J, et al. Mex3a interacts with LAMA2 to promote lung adenocarcinoma metastasis via PI3K/AKT pathway. Cell Death Dis. 2020;11:614.

[36] Lee SA, Kim YJ, Lee CS. Brefeldin a induces apoptosis by activating the mitochondrial and death receptor pathways and inhibits focal adhesion kinase-mediated cell invasion. Basic Clin Pharmacol Toxicol. 2013;113:329–38.

[37] Chen L, Lu D, Sun K, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene. 2019;692:119–25.

[38] Xu H, Wan H, Zhu M, et al. Discovery and validation of an epithelial-mesenchymal transition-based signature in gastric cancer by genomics and prognosis analysis. Biomed Res Int. 2021;2021:9026918.

[39] Chen Y, McAndrews KM, Kalluri R. Clinical and therapeutic relevance of cancer-associated fibroblasts. Nat Rev Clin Oncol. 2021;18:792–804.

[40] Ehrlich M. DNA hypermethylation in disease: mechanisms and clinical relevance. Epigenetics. 2019;14:1141–63.

[41] Moarii M, Boeva V, Vert JP, et al. Changes in correlation between promoter methylation and gene expression in cancer. BMC Genomics. 2015;16:873.

[42] Chen Y, Hu F, Zhou Y, et al. MGMT promoter methylation and glioblastoma prognosis: a systematic review and meta-analysis. Arch Med Res. 2013;44:281–90.