OPEN

The Prognostic Significance of Collagen VI in Pancreatic Ductal Adenocarcinoma

Ella Kesti, BM, Emmy Borgmästars, PhD,† Jaana Hagström, DDS, PhD,‡§ Harri Mustonen, PhD,*|| Hanna Seppänen, MD, PhD,*|| Caj Haglund, MD, PhD,*|| and Malin Sund, MD, PhD*†*

Objectives: Pancreatic ductal adenocarcinoma (PDAC) is characterized by poor prognosis and lack of biomarkers. A rich desmoplastic tumor stroma is considered a hallmark of PDAC and previous studies have indicated upregulated expression of collagen VI (COL6) in PDAC. COL6 is shown to associate with prognosis in many cancers but has been less extensively studied in PDAC.

Materials and Methods: The expression of COL6 was analyzed by immunohistochemistry in tissue microarrays containing resected tumor tissue samples from PDAC patients (n = 164). Significance of COL6 was estimated with Kaplan-Meier survival estimates and multivariable Cox regression analysis. COL6 protein and mRNA expression patterns were further investigated in publicly available datasets.

Results: There were no statistically significant ($P < 0.05$) differences in survival when comparing high and low protein expression of any of the analyzed COL6 α-chains (α1(VI): hazard ratio [HR] 0.90, 95% confidence interval [CI] 0.64–1.28; α2(VI): HR 1.28, 95% CI 0.86–1.89; α3(VI): HR 0.91, 95% CI 0.64–1.29). Similar results were obtained when assessing public data from the Cancer Proteome Atlas, Clinical Proteomic Tumor Analysis Consortium, and The Cancer Genome Atlas.

Conclusions: In contrast with previous studies and some other cancers, we did not find any association of COL6 tissue expression and PDAC survival.

Key Words: collagen VI, tumor stroma, microarray, pancreatic cancer

Abbreviations: CA19-9 - carbohydrate antigen 19-9, COL6 - collagen VI, CPTAC - Clinical Proteomic Tumor Analysis Consortium, DSS - disease-specific survival, ECM - extracellular matrix, mRNA - messenger RNA, NAT - neoadjuvant therapy, OS - overall survival, PDAC - pancreatic ductal adenocarcinoma, TCGA - The Cancer Genome Atlas , TCPA - The Cancer Proteome Atlas, TMA - tissue microarray

(*Pancreas* 2024;53: e729–e738)

Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer and accounts for more than 90% of pancreatic malignancies. PDAC is also a cancer characterized by poor prognosis and lack of effective treatment methods.[1] The survival rate of PDAC is one of the lowest among human cancers. The predicted 5-year survival of PDAC is less than 10%,[2] and this is partly due to detection at late disease stages. PDAC rarely shows any symptoms at the onset of the disease, thus making early diagnosis difficult. Symptoms such as abdominal pain, jaundice and weight loss usually occur at advanced stages with spread disease. Currently, the only possibly curative treatment method is surgical resection, often combined with neo-/adjuvant treatment. Because PDAC is often diagnosed at advanced stages, only 10–20% of the patients are surgically resectable,[3] and among those resected, the 5-year survival rate remains at approximately 20%.[1,4] Neoadjuvant therapy (NAT) combined with surgical resection is associated with prolonged survival, especially observed for patients with higher clinical stage (II-III) and poorly differentiated cancer.[5,6]

Known risk factors for pancreatic cancer are obesity, type 2 diabetes, smoking, and family history of pancreatic malignancies.[7] There are currently no effective biomarkers to detect PDAC early or to effectively monitor disease progress or treatment effects. Serum carbohydrate antigen 19-9 (CA19-9) is the gold standard biomarker for PDAC. CA19-9 holds high sensitivity (79–81%) and specificity (82–90%) in symptomatic patients but has a low positive predictive value of less than 1%.[8] Therefore, CA19-9 is not of clinical value in early detection of PDAC, although recent results from our group indicate that the levels increase already 2 years before clinical diagnosis in some patients.[9,10] A decrease of CA19-9 after NAT correlates with overall survival (OS), and thus it is of value when estimating prognosis and when assessing treatment effect.[11]

In addition to poor prognosis, PDAC is also characterized by an abundant and highly active desmoplastic tumor stroma. The excessive stroma can outnumber the amount of cancer cells in a PDAC tumor, which distinguishes PDAC from many other cancer forms. This heterogeneous stroma surrounding the cancer cells constitutes of various components such as fibroblasts, extracellular matrix (ECM), immune cells, and blood vessels.[12] The development of this dense fibrotic stroma is mainly due to cancer-associated fibroblasts that produce ECM components.[13] The dynamic tumor microenvironment of PDAC interacts continuously with cancer

From the *Department of Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; †Department of Surgical and Perioperative Sciences/Surgery, Umeå University, Umeå, Sweden; ‡Department of Pathology, HUS, HusLab, Helsinki, Finland; §Department of Oral Pathology and Radiology, University of Turku, Turku, Finland; and ||Translational Cancer Medicine Research Program, University of Helsinki, Helsinki, Finland.

Received for publication June 1, 2023; accepted February 15, 2024.

Address correspondence to: Ella Kesti, BM, University of Helsinki, Stenbäckinkatu 9, 00250 Helsinki, Finland (e-mail: ella.kesti@helsinki.fi).

This work was supported by grants from Finska Läkaresällskapet (to M.S. and C.H.); Medicinska understödsföreningen Liv & Hälsa (to M.S. and C.H.), federal VTR funding (to M.S.); Swedish Research Council (2019-01690 to M.S.); the Sjöberg Foundation (to M.S.); and the Swedish Cancer Society (19 0273 to M.S.).

The authors declare no conflict of interest.

Availability of data and materials: Data can be obtained by contacting the corresponding author.

Author contributions: EK; data collection and analysis, writing of the manuscript, EB; data analysis and writing the manuscript, JH; data collection, HM; statistical support and data curation, HS and CH; collection and maintenance of the original data, CH; production of the TMA series, MS; planning the study, conceptualization and administration. All authors have read and approved the final draft of the manuscript.

Ethical approval: This study was granted an approval by the Helsinki University Hospital Ethics committee (document number HUS 1223 2021, 16.6.2021). The Finnish National Supervisory Authority for Health and Welfare has granted this study an approval to use archived tissue samples without individually required consent from the patients (document number 10041/05.01.03.01/2012).

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.pancreasjournal.com).

DOI: 10.1097/MPA.0000000000002360

cells and promotes tumor progression, cancer cell survival and metastasis.[14] This interaction goes both ways and cancer cells can support development and growth of stroma. The desmoplastic stroma has been shown to have a tumor protecting role and maintain a resistant for systemic treatments.[14,15] All in all, the PDAC stroma is under constant investigation in a hope of novel prognostic markers[16] or targets for possible treatments.[17]

Collagen VI (COL6) is an abundant ECM protein expressed in multiple tissues and found to be present in various cancers.[18] COL6 is one of the 28 known collagen proteins and it is mainly composed of three α-chains, α1(VI), α2(VI), and α3(VI) encoded by genes *COL6A1*, *COL6A2*, and *COL6A3*. Three other α-chains (α4(VI), α5(VI), and α6(VI)) also exist, but the gene encoding α4 (VI) is nonfunctional in humans.[19,20] As mentioned, COL6 can be found in many tissues such as nervous tissue, bone, and cartilage, but perhaps most importantly in skeletal muscles where COL6 acts as a crucial element of basement membrane linking muscle cells and ECM.[21]

In many cancers, COL6 has been shown to support tumorigenesis and modulate the cancer microenvironment.[18] In gastric cancer, COL6A3 has been shown to be coexpressed with multiple cancer related genes, which indicates role in ECM organization.[22] Owusu-Ansah et al. showed that knockdown of COL6A1 in PDAC cell models suppress the invasion and migration potential suggesting that COL6A1 having a role in epithelial to mesenchymal transition.[23] Nevertheless, the biological role of COL6 in PDAC still remains unclear. A prognostic value has been shown for upregulation of α1(VI) in cervical cancer[24] and clear cell renal carcinoma[25] as high expression of α1(VI) predicted poorer survival in both. Moreover, Svoronos et al. concluded that the α3 (VI) expression is higher in PDAC tissue compared to normal pancreas. Furthermore, they found that high expression of α3 (VI) was associated with factors such as poorly differentiated tumors and lymph node metastasis, which suggest negative prognosis.[26] The aim of the present project was to determine the prognostic value of COL6 expression in a large PDAC patient cohort. This was done by analyzing the individual COL6 α-chains in PDAC tissue samples on both protein and mRNA expression level.

MATERIALS AND METHODS

Patient Cohort and Tumor Microarrays

PDAC patients in the tissue microarray (TMA) cohort included in this retrospective cohort study were surgically operated with curative intent at the Department of Surgery at Helsinki University Hospital between years 2001–2011. The total number of patients was 228. Only patients with a histologically confirmed PDAC diagnosis were included. Patients who had undergone NAT treatment were excluded. Altogether 164 patients were included in this study after exclusions. The flow chart of the cohort used for protein expression analyses is shown in Figure 1.

We used a tissue microarray (TMA) containing PDAC tissues. These TMA had previously been assembled and contain 2–4 core biopsies from the resected tumor specimens of each patient. Assembly of the TMA has been previously described by Saukkonen et al.[27]

Immunohistochemistry

Four-μm sections were cut from the TMA blocks and stained by immunohistochemistry (IHC) for α1(VI), α2(VI), or α3(VI) chains. The staining of each α-chain was done separately. We used the following primary antibodies; anti-COL6A1 antibody (1:800, Rabbit polyclonal, HPA029401, Sigma-Aldrich, Stockholm, Sweden), anti-COL6A2 (1:400, Rabbit polyclonal, HPA007029, Sigma-Aldrich, Stockholm, Sweden), and anti-COL6A3 (1:100, Rabbit polyclonal, HPA010080, Sigma-Aldrich, Stockholm, Sweden). We performed IHC according to the manufacturers protocol on a LabVision Autostainer 480S (Thermo Fisher Scientific). Deparaffinized tissue slides that went through pretreatment protocol in retrieval solution of pH 9 for 15 minutes were used. To prevent nonspecific binding of antibodies to the tissue, we performed protein blocking protocol on an autostainer for 15 minutes with blocking reagent (EnVision FLEX peroxidase-blocking reagent, Agilent, Santa Clara, CA). Primary antibody was manually added, and the tissue slides were incubated overnight at +5°C. For antibody dilution we used Dako REAL Antibody Diluent S2022 (Agilent, Santa Clara, CA). Secondary antibody (EnVision Flex/HRP SM802, Agilent, Santa Clara, CA) application was programmed on the autostainer with 30-minute incubation and rinsing with washing buffer before and after application. Finally, diaminobenzidine tetrahydrochloride (EnVision FLEX DAB+ Chromogen, Agilent, Santa Clara, CA) was applied as chromogen and tissue slides were counterstained with hematoxylin (Hematoxylin DAKO Mayers, dilution 1:2).

Scoring of COL6 α-Chains Protein Expression

The immunopositivity in TMA samples was microscopically analyzed and scored by 2 separate researchers (E.K. and J.H.) and any differences in assessment of staining were solved by discussing and reviewing the sample. Both observers were blinded to the clinical data. The expression was initially scored both in terms of staining intensity and distribution. The staining intensity of stromal COL6 was scored into the following 4 categories: 0 = negative staining, 1 = low, 2 = moderate, and 3 = high. The distribution of stained stromal COL6 was scored into the following 4 categories: A = less than 10%, B_1 = 10–50%, B_2 = 50–90%, and C = more than 90% of the tumor stroma stained. Very low distribution of stained tissue (<10%) would lower the scoring one level (from 3A to 2 and from 2A to 1) and similarly very high distribution of stained tissue would lift the scoring one level (from 1C to 2, from 2C to 3). In category B, the distribution of 10–50% (B1) lowered the score 0.5 and distribution of 50–90% lifted the score 0.5. As a result, the combined score ranged from 0 to 4 and for statistical purposes the categories were merged so that scores 0–2 represent low staining and scores 2.5–4 represent high staining. Scoring and combination of different categories as well as number of patients at each category are presented in Supplemental Figure 1, http://links.lww.com/MPA/B185. TMAs included 2–4 core samples of each patient, which were all scored, and the maximum scoring value of each patient was used.

Public Data

The Cancer Proteome Atlas

Expression data of α1(VI), from The Cancer Proteome Atlas (TCPA) was downloaded from https://tcpaportal.org/tcpa/download.html and used according to TCPA recommendations.[28,29] Data had been acquired by assessing collagen VI protein levels by reverse phase protein arrays using a COL6A1 antibody (SC-20649, Santa Cruz). In our analysis, we used level 4 expression data, which represents the level of processing applied to the data. According to TCPA, level 4 means that expression data was obtained using a replicate-based method to combine reverse phase protein array data from different slides. TCPA-PAAD dataset consisted of 103 cancer patients (n = 103) that were included in survival analysis.

Clinical Proteomic Tumor Analysis Consortium

Collagen VI protein expression data was downloaded from Clinical Proteomic Tumor Analysis Consortium (CPTAC) using

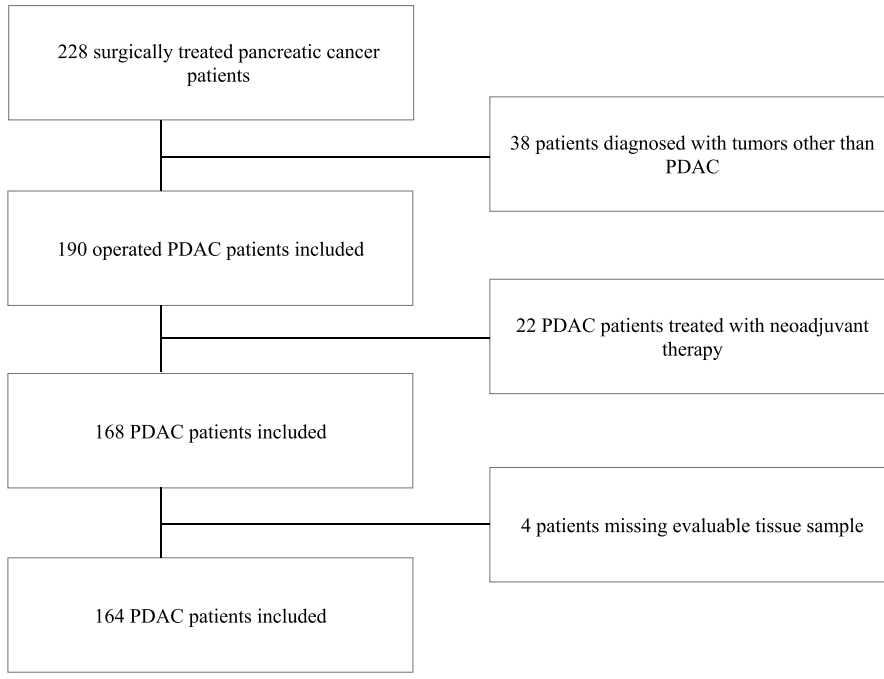

FIGURE 1. Flow chart displaying the inclusion and exclusion for patients in the TMA cohort. Altogether 164 surgically treated patients with histologically confirmed diagnosis of PDAC were included in the study. Patients with other form of malignancy were excluded. Our exclusion criteria also ruled out patients treated with NAT.

cptac python module 1.5.3 or 1.5.7 was used in python version 3.11.5 to retrieve umich proteomics data for PDAC.[30]

The Cancer Genome Atlas

Collagen VI mRNA expression data and corresponding clinical data was collected from The Cancer Genome Atlas (TCGA).[31] Publicly available log2(fpkm+1) RNA-seq data from GDC TCGA pancreatic cancer (PAAD), phenotype information, and survival data were downloaded from Xena browser (obtained 2022-04-23 from https://xenabrowser.net/).[32] TCGA-PAAD patients with the primary diagnosis 'Neuroendocrine CARCINOMA' (n = 8) were excluded from survival analysis resulting in a total of 168 patients (n = 168).

COL6 in Other Cancers

We further investigated COL6 protein and mRNA expression in other cancers, including breast invasive carcinoma (BRCA), lung adenocarcinoma, colon adenocarcinoma, and prostate adenocarcinoma. Data for these analyses were derived from TCPA, CPTAC, and TCGA databases and used similarly as described previously with PDAC data. CPTAC database included data on BRCA, lung adenocarcinoma, and colon adenocarcinoma only.

Statistical Analysis

The semiquantified data of COL6 expression in tissue samples was combined with clinical data and patient information. The impact of COL6 expression on prognosis was estimated with Kaplan-Meier survival analysis separately for each α-chain and statistical differences were analyzed with the log-rank test. The endpoint for disease-specific survival (DSS) was death from pancreatic cancer and for OS death from any cause censoring patients alive on the last date of the follow-up. Multivariable Cox regression analysis was used for the prognostic value of COL6 expression in relation to other variables, such as sex, clinical stage, histological tumor grade, and preoperative CA19-9 level. Clinical stages I-IIA were merged as one group and stages IIB-IV as one before statistical analysis. Assumptions for proportional hazards in Cox regression and interactions were tested with adding time-dependent covariates and plotting partial residuals over time. Correlations between COL6 α-chain expression and clinical parameters were estimated with Spearman correlation coefficient. Statistical tests concerning the TMA cohort were analyzed with SPSS statistics software for Mac, version 28 (IBM Corp. Armonk, NY).

Survival analyses of TCPA (COL6A1), CPTAC (COL6A1, COL6A2, and COL6A3), and TCGA (*COL6A1*, *COL6A2*, and *COL6A3* mRNA) cohorts were performed in R version 4.1.1 (R Core Team, 2021) using survival (Therneau 2022)[33] and survminer (Kassambara et al. 2021)[34] packages. The cohorts were separated by median, or the most optimal COL6 cutoff values as identified by the maximally selected rank statistics method implemented in the *surv_cutpoint* function in survminer R package. Cox regression analysis was performed with and without adjustment for age, sex, histological tumor grade, and clinical stage. One patient without a reported stage and one without reported histological tumor grade were excluded from Cox regression analysis in the TCGA cohort. Stages IA and IB were merged into stage I, IIA, and IIB into stage II before Cox regression analysis. Two individuals lacking information on tumor stage and one lacking information on histological grade were excluded from Cox regression analysis in the TCPA cohort. The *cox.zph* function in the survival R package was used to test the proportional hazards assumption for Cox regression model fit (Grambsch and Therneau, 1994). The *cox.zph* function correlates scaled Schoenfeld residuals with time and a nonsignificant result indicates that the assumption holds. Collagen VI variables from TCGA, TCPA, and CPTAC were scaled to unit before Cox regression analysis. All tests used were two-tailed and a P value below 0.05 was considered statistically significant. P values were corrected for multiple testing using the Benjamini-Hochberg method[35] and a false discovery rate (FDR) < 0.05 was considered statistically significant.

RESULTS

Patient Characteristics in TMA Cohort

A total of 164 PDAC patients were included in the final TMA study cohort and analyzed for COL6 protein expression. Characteristics of patients included in analysis are displayed in Table 1. The median follow-up time of the TMA cohort was 22 months. During the follow-up time, 140 (85%) of the 164 patients died of pancreatic cancer. Thirteen patients were reported to be alive and 11 died of other causes. Thus, pancreatic cancer deaths cover more than 90% of all observed deaths during follow-up time.

Tissue Expression of α1(VI), α2(VI), and α3(VI) and Survival in TMA Cohort

In the IHC stained TMA tissue samples, COL6 α-chains were expressed in tumor stroma with varying histological scoring results as demonstrated in Figure 2. After combining the scores for staining intensity and staining, distribution categories were merged to get 2 categories representing low and high COL6 protein expression (Fig. 2). The different α-chains, α1(VI), α2(VI), and α3(VI), were similarly expressed in PDAC tissue. Some of the cores also contained adjacent healthier pancreatic tissue as well, and here, α1(VI), α2(VI), and α3(VI) expressions were detected in epithelial basement membrane and vascular endothelium. Such expression was considered as an internal positive control for the staining.

Association between protein expression of each COL6 α-chain and clinical parameters such as age, sex, preoperative CA19-9 level, histological grade, and clinical stage were analyzed using Spearman correlation. Weak negative correlation was detected between α3 (VI) expression and age (Spearman rho −0.192; $P = 0.018$) and similarly weak positive correlation was shown between α3(VI) and sex (Spearman rho 0.168; $P = 0.039$). Other clinical parameters showed no statistically significant and nonnegligible correlations with the COL6 α-chains (Spearman correlation coefficients for each α-chain and clinical parameter shown in Supplemental Table I, http://links.lww.com/MPA/B186).

To detect potential correlation between DSS and expression of COL6, Kaplan-Meier survival analysis was performed for each α-chain separately. There was, however, no significant correlation between DSS and COL6 expression (Fig. 3). Benjamini-Hochberg correction was also performed to adjust the P values for multiple testing, but no significant results were found (Benjamini-Hochberg corrections for Kaplan-Meier analysis shown in Supplemental Table II, http://links.lww.com/MPA/B187, FDR 0.05 was considered significant).

The Cox regression analysis was first adjusted for age, sex, clinical stage, histological grade, and the preoperative CA19-9 level but after testing for the proportional hazards assumptions variables sex and histological grade were excluded because they violated the model. Results of the univariable and multivariable Cox regression analyses are shown in Table 2. No statistically significant correlation between α1(VI), α2(VI), or α3(VI) chains and survival was found. Additional Cox regression analyses were performed for COL6 stratifying by sex or histological grade. Statistically significant association was detected between survival and α3 (VI) expression in group histologic grade 1 (hazard ratio [HR] 6.65, confidence interval [CI] 95% 1.58–28.04). All the other stratified models showed no significant correlation between any of the COL6 α-chains and survival (Results for the stratified Cox regression analysis shown in Supplemental Table III, http://links.lww.com/MPA/B188). However, the multivariable analysis showed significant correlation for clinical stage and DSS (P = 0.013, HR 1.73, 95% CI (1.12–2.67)) and preoperative CA19-9 level and DSS (P = 0.006, HR 1.29, 95% CI (1.07–1.54)), when all the COL6 α-chains were included in analysis (Supplemental Table IV, http://links.lww.com/MPA/B189).

TABLE 1. Clinically Significant Characteristics of TMA Patient Cohort

PDAC Patients, n	164
Age, median (IQR, range)	64.6 (11.9, 39.2–83.7)
Sex, M/F (%)	91/73 (55.5/44.5)
Histological grade, n (%)	
G1	16 (9.8)
G2	110 (67.1)
G3	27 (16.5)
Unknown	11 (6.7)
Clinical stage, n (%)	
IA	9 (5.5)
IB	17 (10.4)
IIA	20 (12.2)
IIB	104 (63.4)
III	3 (1.8)
IV	7 (4.3)
Unknown	4 (2.4)
Stage	
T, n (%)	
T1	16 (9.8)
T2	103 (62.8)
T3	35 (21.3)
T4	1 (0.6)
Unknown	9 (5.5)
N, n (%)	
N0	45 (27.4)
N1	71 (43.3)
N2	41 (25.0)
Unknown	7 (4.3)
M, n (%)	
M0	155 (94.5)
M1	7 (4.3)
Unknown	2 (1.2)
CA19-9 (U/ml) (preoperatively), mean (range)	3267.0 (2.0–331,580.0)
Survival time in months, mean (range)	39.2 (2–206)

TCPA Data of α1(VI) Protein Expression and Survival

Protein expression data of α1(VI) was obtained from TCPA database and further analyzed to show possible correlation between α1(VI) protein levels and survival. However, Kaplan-Meier analysis showed no significant correlation (Supplemental Fig. 2, http://links.lww.com/MPA/B190). Cox regression analysis was performed as univariable analysis and as multivariable analysis (adjusted for sex, age, tumor stage, and histological grade), both failing to show statistically significant correlation (Table 2). None of the variables violated the proportional hazards assumption.

COL6 Expression in Proteomic Analysis and Survival

The collagen VI α-chains 1, 2, and 3 available in CPTAC pancreatic cancer were not associated with prognosis (Table 2).

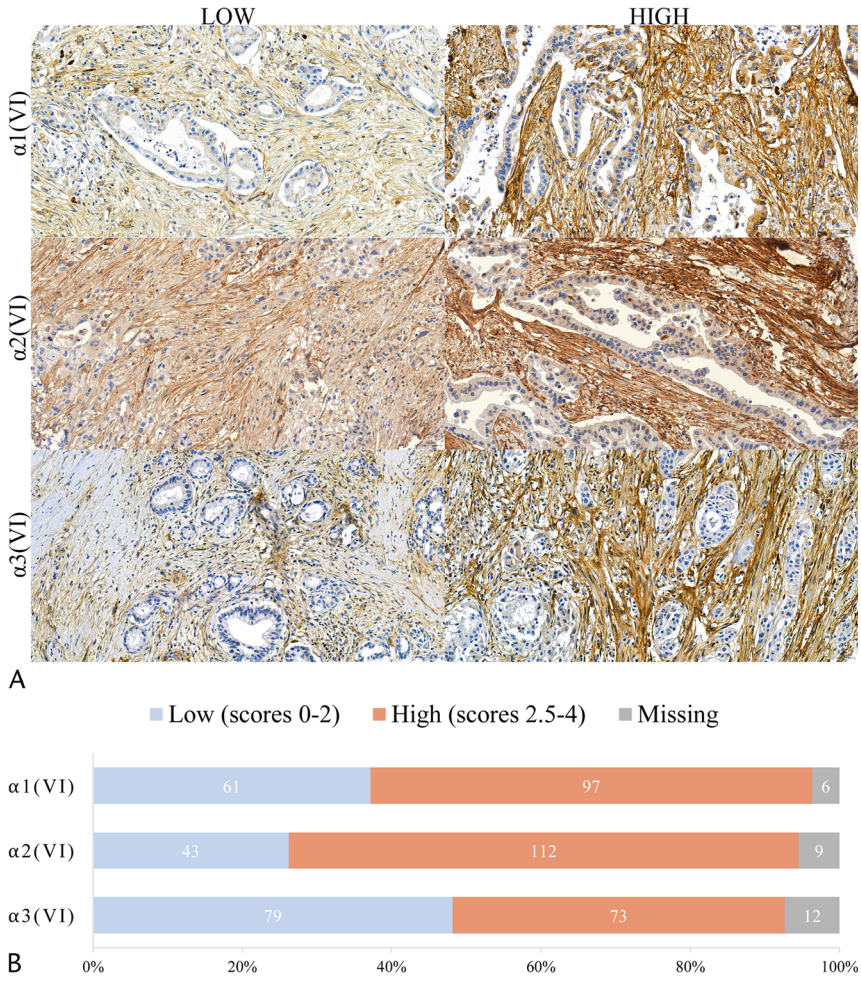

FIGURE 2. Collagen VI staining in the TMA. A, Histological staining patterns scored low or high for each COL6 α-chain displayed. Tissue slides were immunohistochemically stained with specific antibody to show α1(VI), α2(VI), and α3(VI). Hematoxylin was used as a counterstain. Magnification x20. B, Proportions of low or high COL6 protein expression in patient samples displayed for each α-chain. Exact number of patients in each category shown in bars while total number of patients was 164 (n = 164).

Because age violated the assumption of proportional hazards in Cox regression models, it was excluded from the cox regression models and instead collagen 6 alpha chains 1, 2, and 3 were analyzed separately in groups stratified by age (data not shown). The assumption of proportional hazard was met for all COL6 α-chains in all age groups except for COL6A3 and age ≤55.

mRNA Expression of COL6A1, COL6A2, and COL6A3 and OS

COL6 was also analyzed on mRNA expression level using publicly available data from the TCGA database. A Kaplan-Meier analysis was performed to assess correlation between *COL6A1, COL6A2,* and *COL6A3* mRNA expression and OS separating the cohort by optimal cutoff (Fig. 4) or by median cutoff (Supplemental Fig. 3, http://links.lww.com/MPA/B191). In the Kaplan-Meier analysis, we initially found that *COL6A1* had protective prognostic value. However, after correction with Benjamini-Hochberg, the significance was lost (FDR = 0.152) (corrected *P* values shown in Supplemental Table II, http://links.lww.com/MPA/B187).

Furthermore, a Cox regression analysis was performed, and survival was adjusted to relevant clinical parameters (age, sex, tumor stage, and histological grade) similarly as done for the protein expression. The covariates sex and histological grade violated the proportional hazards assumption and were excluded from the models. In addition, we performed Cox regression analyses on groups stratified by sex or histological grade. The histological grade 1 group violated the proportional hazards assumption for *COL6A1* and *COL6A2* mRNA. For the remaining combinations of strata and *COL6A1*, *COL6A2*, or *COL6A3*, none of the models were significant (stratified analysis shown in Supplemental Table III, http://links.lww.com/MPA/B188). Two individuals with a histologic grade of 4 were excluded from stratified analyses. In the Cox regression analysis, no prognostic value for COL6 was shown neither in the unadjusted or when adjusted for clinical parameters (Table 2). In summary, no prognostic value for *COL6A1, COL6A2*, or *COL6A3* mRNA expression in the TCGA-PAAD cohort was found.

Correlation Between COL6 Expression and Smoking or Alcohol Use in PDAC Patients

Data of smoking and alcohol use of PDAC patients was available in TCGA, TCPA, and CPTAC databases. Our additional analyses showed that COL6A1 protein expression had negative

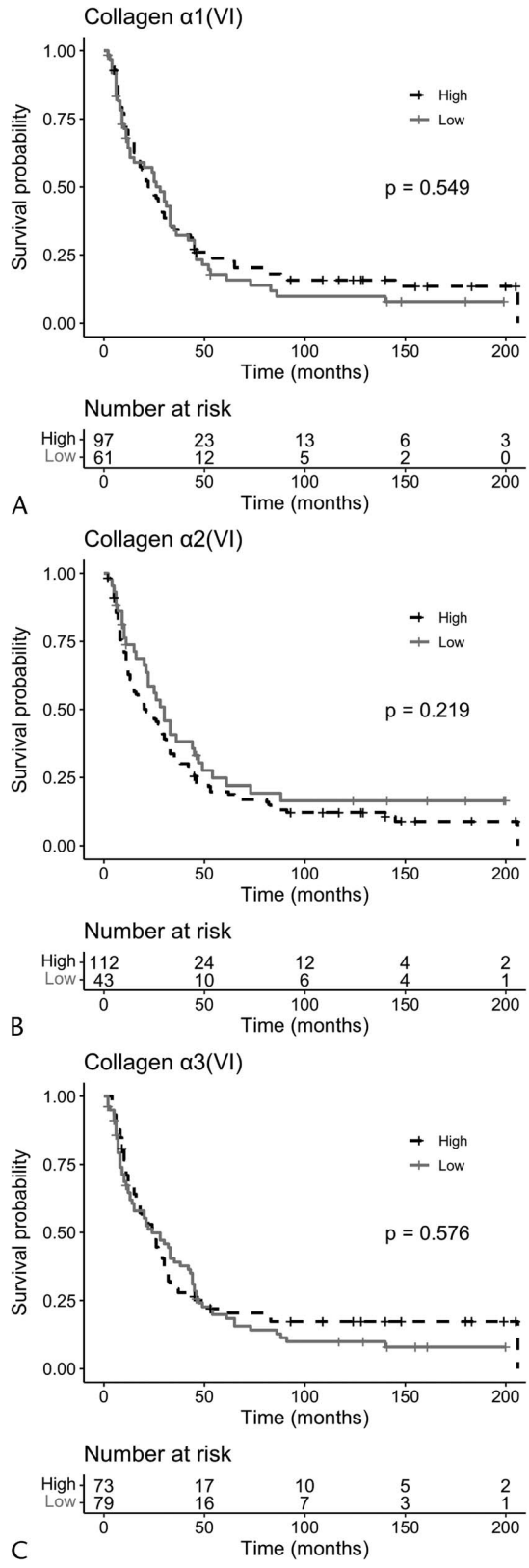

FIGURE 3. Disease-specific survival in relation to protein expression of collagen VI α-chains 1, 2, and 3 (A-C). Statistical analysis using Kaplan-Meier survival estimate declared no significant difference in survival when comparing low and high expression of each COL6 α-chain. *P* value <0.05 was considered significant. (A, n = 158; B n = 155; C n = 152).

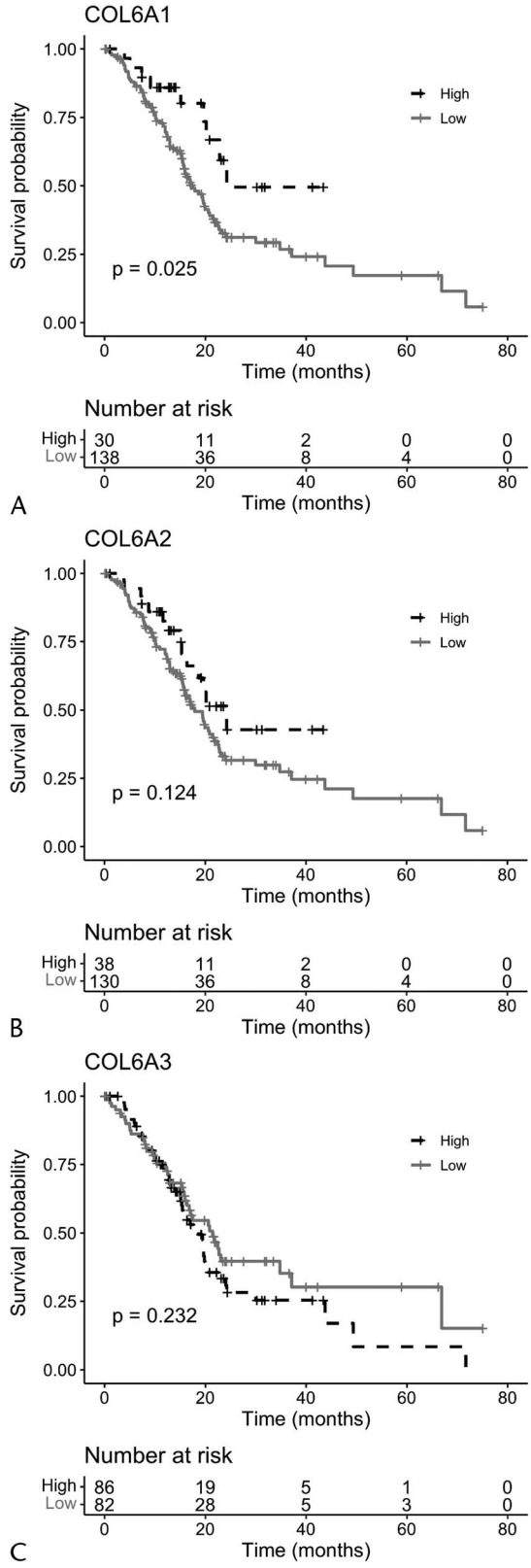

FIGURE 4. Overall survival in relation to mRNA expression of *COL6A1*, *COL6A2*, and *COL6A3* (A-C). Kaplan-Meier analysis of OS for TCGA-PAAD infiltrating duct carcinoma (n = 168) was performed separating the cohort using the optimal cutoff. Nominal *P* values are shown in the plots, and these were not significant after adjusting for multiple hypotheses testing using the Benjamini-Hochberg method.

TABLE 2. Prognostic Value of COL6 α-Chains 1, 2, and 3 Evaluated at mRNA and Protein Level

Cox Regression Analysis of Protein Expression in TMA Cohort						
	Crude Model			**Adjusted Model**		
Collagen chain	**HR (95% CI)**	***P* value**		**HR (95% CI)**	***P* value**	
α1(VI)	0.90 (0.64–1.28)	0.555		0.93 (0.64–1.34)	0.689	
α2(VI)	1.28 (0.86–1.89)	0.227		1.23 (0.82–1.86)	0.323	
α3(VI)	0.91 (0.64–1.29)	0.581		1.01 (0.70–1.47)	0.943	
Cox regression analysis of protein expression in TCPA-PAAD cohort (n = 103)						
	Crude model			**Adjusted model**		
Collagen chain	**HR (95% CI)**	***P* value**		**HR (95% CI)**	***P* value**	
α1(VI)	1.01 (0.79–1.27)	0.963		1.03(0.79–1.34)	0.814	
Cox regression analysis of protein expression in CPTAC-PDAC cohort (n = 138)						
	Crude model			**Adjusted model**		
Protein (ensemble ID)	**HR (95% CI)**	***P* value**	**FDR**	**HR (95% CI)**	***P* value**	**FDR**
COL6A1 (ENSP00000355180.3)	1.1 (0.91–1.34)	0.309	0.515	1.05 (0.85–1.30)	0.658	0.823
COL6A2 (ENSP00000300527.4)	1.12 (0.93–1.36)	0.228	0.515	1.08 (0.88–1.33)	0.479	0.798
COL6A2.1 (ENSP00000380870.1)	1.24 (1.01–1.53)	0.041	0.205	1.20 (0.94–1.52)	0.138	0.685
COL6A3 (ENSP00000295550.4)	1.07 (0.89–1.27)	0.484	0.605	1.00 (0.81–1.22)	0.965	0.965
COL6A3.1 (ENSP00000418285.1)	0.96 (0.79–1.17)	0.689	0.689	0.88 (0.71–1.10)	0.274	0.685
Cox regression analysis of mRNA expression in TCGA-PAAD cohort (n = 164)						
	Crude model			**Adjusted model**		
Gene	**HR (95% CI)**	***P* value**		**HR (95% CI)**	***P* value**	
COL6A1	1.02 (0.84–1.26)	0.808		1.01 (0.82–1.25)	0.931	
COL6A2	0.96 (0.79–1.17)	0.683		0.93 (0.76–1.15)	0.518	
COL6A3	1.09 (0.88–1.34)	0.426		1.07 (0.86–1.33)	0.532	

Analysis for TMA cohort was adjusted for age, clinical stage, and preoperative CA19-9, while sex and histologic grade were excluded because of violation of the proportional hazard assumption. In TCPA cohort, analysis was adjusted by age, sex, histological grade, and clinical stage. In CPTAC cohort, age violated the proportional hazard model and was excluded. Cox regression models for COL6 mRNA expression were adjusted for age and clinical stage. Confidence intervals of 95% and HRs shown for each model. In TCGA, TCPA, and CPTAC, analysis hazard ratio (HR) is estimated per standard deviation increase.

correlation with smoking exposure time measured in pack years (Spearman rho −0.49, $P = 0.003$) (Supplemental Table V, http://links.lww.com/MPA/B192).

Comparative Cox Regression Analysis of COL6 in Other Cancers

Univariate Cox regression analysis was performed for COL6 expression in lung adenocarcinoma, colon adenocarcinoma, invasive BRCA, and prostate adenocarcinoma. Results are shown in Supplemental Table VI, http://links.lww.com/MPA/B193. Expression data of COL6 protein and mRNA expression were obtained from public databases TCGA, TCPA, and CPTAC. According to our survival analysis COL6A1 protein expression in BRCA predicted worse survival ($P = 0.004$, HR 0.74, 95% CI 0.61–0.91) (Supplemental Table VII, http://links.lww.com/MPA/B194).

DISCUSSION

In this retrospective cohort study, we analyzed the correlations between collagen VI expression and survival in PDAC patients. The protein expression of each COL6 α-chain was assessed both in tumor tissue samples from surgically resected patients who had not received neoadjuvant treatment and in data from the TCPA dataset. Additionally, COL6 expression was analyzed in cohort of PDAC patients from CPTAC database containing proteomic data. Gene expression data was analyzed using the TCGA dataset. The aim of this study was to validate previous findings of prognostic value of COL6 expression, but no statistically significant correlation between expression of the COL6 α-chains and survival was observed as an independent factor. Our results from analysis of both protein expression and protein data from TCPA and CPTAC databases, as well as analysis of COL6 mRNA data obtained from TCGA database all show similar results.

Despite extensive research efforts, pancreatic cancer still lacks prognostic factors and clinically applicable methods for early diagnosis. PDAC remains one of the deadliest cancers with limited treatment options. The need for new biomarkers is urgent. COL6 has previously been identified to play an active role in cancer stroma and cancer microenvironment[18] and PDAC tumors typically hold rich and active stromal components. Moreover, it has been shown that COL6 is abundantly expressed in PDAC tumor tissue when compared with normal pancreatic tissue.[23,26] PDAC patients also have high serum level of COL6A3 protein and mRNA compared to healthy correspondents.[36] A previous study by Owusu-Ansah[23] reported that *COL6A1* upregulation increases the metastatic potential in pancreatic cancer models. They additionally found α1(VI) protein expression in PDAC tissue to be correlated with poorer prognosis. This previous finding contradicts our present study, but it was detected in a smaller cohort (n = 65) using different antibody. The histological score was moreover stated as negative or positive rather than analyzing the intensity of staining.[23]

COL6 has however been linked to a number of other cancers with solid tumors as a prognostic factor. In salivary gland cancer, COL6 has been shown to correlate with malignancy and also to hold prognostic value, as higher tissue expression of COL6 was shown to correlate with poorer OS.[37] In cervical cancer, the α1 (VI) tissue expression is also found to correlate with advanced

clinical stage and poor prognosis.[24] When considering other gastrointestinal tract cancers, it has been reported that COL6 has possible prognostic value in colorectal carcinoma (CRC). Especially upregulation and expression of COL6A3 have been studied in CRC.[38,39] Qiao et al[38] found that COL6A3 is upregulated in CRC tumors in contrast to normal tissue using proteomics and that expression of α3(VI) in tissue correlated with poorer survival. In addition, they found that circulating level of COL6A3 was increased among CRC patients. It has been previously shown that COL6A3 expression is increased in gastric cancer.[22] These findings underline the potentiality of COL6 to act as a biomarker for other gastrointestinal malignancies. However, our additional analyses on public data of COL6 expression in relation to survival in other cancers suggested statistically significant prognostic value only for COL6A1 in breast cancer while no significance was found in lung cancer, prostate cancer, or colon cancer. In addition, the result of COL6A1 having positive prognostic value in breast cancer contradicts previous findings of highly expressed COL6A1 associating with poor prognosis.[40] All in all, the prognostic value of COL6 remains unclear.

As for this present study and the contradictory results, the possible weaknesses lie in the methods used. Handling large amount of tissue samples is practical and more efficient with TMA. TMA enable the IHC staining process to occur simultaneously for hundreds of tissue samples, which makes the staining results more comparable. Furthermore, TMA make the microscopical analysis of tissue samples faster and more viewer friendly. The fact that there is multiple tissue cores side by side makes a great reference point for scoring the staining. However, manual scoring is subjective and uncertain even at its best because the method relies on visual inspection. In the present study, 2 investigators, however, performed scoring, which could mitigate any bias from the subjectivity of tissue scoring. Another obvious weakness of using TMAs is that the tissue core biopsies represent only a small part of the entire tumor, even if there usually are at least 2 tissue cores from each patient.[41]

Nevertheless, the strength of this study lies in the rather large cohort of patients (n = 164) with minimal loss to follow-up and highly granular data regarding clinical variables. Moreover, we investigated multiple COL6 α-chains (α1(VI), α2(VI), and α3(VI)) and assessed each one separately. Finally, we evaluated the prognostic significance of COL6 α-chains both on protein and mRNA level. Further studies could include measuring all circulating COL6 α-chains, similar to what was previously done with COL6A3.[36] Additionally, it could be of benefit to assess prognostic value of COL6 combined with another biomarker or as a part of panel of biomarkers.

The aim of this study was to validate COL6 as a novel prognostic factor for PDAC. Our results do not support a prognostic value of COL6 in PDAC. The stroma remains a focus area in the quest to find future biomarkers for PDAC.

ACKNOWLEDGMENTS

The authors thank laboratory technician Pia Saarinen and research coordinator Maria Finne for their indispensable contribution to the project.

REFERENCES

1. Mizrahi JD, Surana R, Valle JW, et al. Pancreatic cancer. *Lancet*. 2020;395:2008–2020.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7–34.
3. Aaltonen P, Carpén O, Mustonen H, et al. Long-term nationwide trends in the treatment of and outcomes among pancreatic cancer patients. *Eur J Surg Oncol*. 2022;48:1087–1092.
4. American Cancer Society. *Cancer Facts & Figures 2022*. Atlanta: American Cancer Society; 2022.
5. Mirkin KA, Hollenbeak CS, Wong J. Survival impact of neoadjuvant therapy in resected pancreatic cancer: a prospective cohort study involving 18,332 patients from the National Cancer Data Base. *Int J Surg*. 2016;34:96–102.
6. Nurmi A, Mustonen H, Parviainen H, et al. Neoadjuvant therapy offers longer survival than upfront surgery for poorly differentiated and higher stage pancreatic cancer. *Acta Oncol*. 2018;57:799–806.
7. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med*. 2014;371:1039–1049.
8. Ballehaninna UK, Chamberlain RS. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: an evidence based appraisal. *J Gastrointest Oncol*. 2012;3:105–119.
9. Young MR, Wagner PD, Ghosh S, et al. Validation of biomarkers for early detection of pancreatic cancer: summary of the Alliance of Pancreatic Cancer Consortia for Biomarkers for Early Detection Workshop. *Pancreas*. 2018;47:135–141.
10. Mason J, Lundberg E, Jonsson P, et al. A cross-sectional and longitudinal analysis of pre-diagnostic blood plasma biomarkers for early detection of pancreatic cancer. *Int J Mol Sci*. 2022;23:12969.
11. Tzeng CWD, Balachandran A, Ahmad M, et al. Serum carbohydrate antigen 19-9 represents a marker of response to neoadjuvant therapy in patients with borderline resectable pancreatic cancer. *HPB (Oxford)*. 2014;16:430–438.
12. Feig C, Gopinathan A, Neesse A, et al. The pancreas cancer microenvironment. *Clin Cancer Res*. 2012;18:4266–4276.
13. Nielsen MFB, Mortensen MB, Detlefsen S. Key players in pancreatic cancer-stroma interaction: cancer-associated fibroblasts, endothelial and inflammatory cells. *World J Gastroenterol*. 2016;22:2678–2700.
14. Thomas D, Radhakrishnan P. Tumor-stromal crosstalk in pancreatic cancer and tissue fibrosis. *Mol Cancer*. 2019;18:14.
15. Neesse A, Bauer CA, Öhlund D, et al. Stromal biology and therapy in pancreatic cancer: ready for clinical translation? *Gut*. 2019;68:159–171.
16. Sánchez-Ramírez D, Medrano-Guzmán R, Candanedo-González F, et al. High expression of both desmoplastic stroma and epithelial to mesenchymal transition markers associate with shorter survival in pancreatic ductal adenocarcinoma. *Eur J Histochem [Internet]*. 2022;66:3360.
17. Ho WJ, Jaffee EM, Zheng L. The tumour microenvironment in pancreatic cancer — clinical challenges and opportunities. *Nat Rev Clin Oncol*. 2020;17:527–540.
18. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med*. 2013;19:410–417.
19. Gara SK, Grumati P, Urciuolo A, et al. Three novel collagen VI chains with high homology to the α3 chain. *J Biol Chem*. 2008;283:10658–10670.
20. Fitzgerald J, Holden P, Hansen U. The expanded collagen VI family: new chains and new questions. *Connect Tissue Res*. 2013;54:345–350.
21. Cescon M, Gattazzo F, Chen P, et al. Collagen VI at a glance. *J Cell Sci*. 2015;128:3525–3531.
22. Xie X, Liu X, Zhang Q, et al. Overexpression of collagen VI α3 in gastric cancer. *Oncol Lett*. 2014;7:1537–1543.
23. Owusu-Ansah KG, Song G, Chen R, et al. COL6A1 promotes metastasis and predicts poor prognosis in patients with pancreatic cancer. *Int J Oncol [Internet]*. 2019;55:391–404.
24. Hou T, Tong C, Kazobinka G, et al. Expression of COL6A1 predicts prognosis in cervical cancer patients. *Am J Transl Res*. 2016;8:2838–2844.

25. Wan F, Wang H, Shen Y, et al. Upregulation of COL6A1 is predictive of poor prognosis in clear cell renal cell carcinoma patients. *Oncotarget*. 2015; 6:27378–27387.

26. Svoronos C, Tsoulfas G, Souvatzi M, et al. Prognostic value of COL6A3 in pancreatic adenocarcinoma. *Ann Hepatobiliary Pancreat Surg*. 2020;24: 52–56.

27. Saukkonen K, Hagström J, Mustonen H, et al. Podocalyxin is a marker of poor prognosis in pancreatic ductal adenocarcinoma. *PloS One*. 2015; 10:e0129012.

28. Li J, Lu Y, Akbani R, et al. TCPA: a resource for cancer functional proteomics data. *Nat Methods*. 2013;10:1046–1047.

29. Li J, Akbani R, Zhao W, et al. Explore, visualize, and analyze functional cancer proteomic data using the cancer proteome atlas. *Cancer Res*. 2017; 77:e51–e54.

30. Lindgren CM, Adams DW, Kimball B, et al. Simplified and unified access to cancer proteogenomic data. *J Proteome Res*. 2021;20: 1902–1910.

31. https://cancergenome.nih.gov [Internet]. Available at: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga. Accessed April 28, 2022.

32. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38:675–678.

33. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000.

34. Kassambara, A., Kosinski, M., and Biecek, P. survminer: drawing survival curves using "ggplot2." 2021.

35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57:289–300.

36. Kang CY, Wang J, Axell-House D, et al. Clinical significance of serum COL6A3 in pancreatic ductal adenocarcinoma. *J Gastrointest Surg*. 2014; 18:7–15.

37. Angenendt L, Mikesch JH, Görlich D, et al. Stromal collagen type VI associates with features of malignancy and predicts poor prognosis in salivary gland cancer. *Cell Oncol (Dordr)*. 2018;41:517–525.

38. Qiao J, Fang CY, Chen SX, et al. Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. *Oncotarget*. 2015;6:29929–29946.

39. Liu W, Li L, Ye H, Tao H, He H. Role of COL6A3 in colorectal cancer. Oncol Rep [Internet]. 2018;39:2527–2536.

40. Lin Z, Zhu G, Tang D, Bu J, Zou J. High expression of COL6A1 correlates with poor prognosis in patients with breast cancer.

41. Nocito A, Kononen J, Kallioniemi OP, et al. Tissue microarrays (TMAs) for high-throughput molecular pathology research. *Int J Cancer*. 2001;94:1–5.