Identification of the collagen family as prognostic biomarkers in papillary thyroid carcinoma

Juntao Niu1 · Wenyu Guo1 · Yu-Zhou Chen2 · Ning Jiang1

Received: 26 February 2022 / Accepted: 16 August 2022 / Published online: 7 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Objective The aim of this study was to construct a collagen-related prognostic model for thyroid cancer and to investigate prognostic value of collagen family genes for thyroid cancer.
Methods A LASSO Cox regression model for thyroid cancer was developed based on the expression profiles of collagen-related genes. Kaplan–Meier survival analysis was performed for high and low risk groups. The ROC method was used to assess its predictive performance. Predictive independence was verified by multivariate Cox regression analysis. The relationship between this feature and immune cell infiltration was analyzed by tumor microenvironment. COL18A1 was validated by immunohistochemistry and RT-PCR in thyroid cancer tissues. The effect of COL18A1 on cell proliferation, migration and invasion ability of tumor cells were further valuated by CCK-8 assay and transwell assay. The effect of COL18A1 on the immune escape ability of tumor cells was further valuated by cytotoxicity assays.
Results A model including 4 collagen family genes was developed to predict thyroid cancer prognosis. Patients with high-risk score had a poorer prognosis than those with low-risk scores for 1-, 2-, 3-, and 5- year survival. The model independently predicted prognosis after adjusting for other prognostic factors. A nomogram combining risk score and age was constructed with high sensitivity and specificity. This feature was significantly associated with immune cell infiltration. COL18A1 was aberrantly over-expressed in thyroid cancer compared with control tissues and significantly increased proliferative capacity, migration capacity, invasion capacity, and immune escape ability of tumor cells.
Conclusion Our findings establish a signature associated with collagen family genes that can be a promising tool to predict the prognosis of thyroid cancer. High COL18A1 expression significantly correlates with the poor prognosis of patients and enhances the immune escape ability of tumor cells.

Keywords Collagen · Thyroid cancer · Prognosis · COL18A1

Introduction
Thyroid cancer is one of the malignancies with the fastest growth rate in recent years and is the most common malignant tumor of the endocrine system. Of these, papillary thyroid carcinoma is the most common type of pathology [1, 2]. Most papillary thyroid carcinomas have a good prognosis, with a cure rate of over 90% [3]. However, some patients will have a poor prognosis due to the occurrence of distant metastases and recurrence. There is a need to develop new predictive models that will in turn improve the diagnosis rate and prognosis of papillary thyroid cancer (PTC).

The extracellular matrix (ECM) forms the scaffolding of the tumor microenvironment and may regulate cancer behavior [4–6]. The structure and composition of the ECM is remodeled during cancer development, and changes in the expression pattern of ECM macromolecules can alter the motility, invasion and metastasis of cancer cells [7, 8]. In addition, it is recently reported that ECM-cell interactions can mediate cellular functional properties and cell phenotype [9–11], which regulate cellular signaling to promote cancer...
cell proliferation, migration, apoptosis, cell survival, angiogenesis and epithelial mesenchymal transition [8, 10]. As a major component of the extracellular matrix, collagen accounts for its primary function. Alterations in collagen in the tumor microenvironment have been found to correlate with cancer spread and prognosis [12–15]. Increased collagen density around cancer cells directs local invasion and metastasis [12]. Collagen alignment and orientation have been shown to be indicators of tumor metastasis in breast cancer [13, 14], glioblastoma [16], and prostate cancer [17]. However, relevant study about the role of collagen in PTC has rarely been reported [18]. The aim of this study was to investigate the impact of collagen family genes on the prognosis of PTC and to construct a predictive model for the prognosis of PTC patients. We obtained clinical data and mRNA expression levels of PTC patients from the TCGA and GEO databases. Differentially expressed genes (DGEs) associated with collagen proteins were used to construct a prognostic model. The prognostic model was then validated utilizing TCGA and ICGC database, providing a theoretical basis for clinical application. Besides, we examined the expression of COL18A1 in thyroid cancer and its effect on thyroid cancer by cytological assays. Due to the critical role of collagen family in cancer progression, they may be promising diagnostic markers and pharmacological targets.

**Materials and methods**

**Data collection**

RNA sequencing (RNA-seq) transcriptome data of PTC were downloaded from The Cancer Genome Atlas (TCGA) data portal (http://portal.gdc.cancer.gov/) and the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/gds/), including 509 PTC tissue samples and 58 normal thyroid tissue samples from TCGA and 60 thyroid tumor tissue samples and 45 normal thyroid tissue samples from GEO (GSE33630). ICGC database (https://dcc.icgc.org/) including 501 thyroid tumor tissue samples was utilized for prognostic model validation. 43 collagen family genes were chosen for further analysis (Table 1).

Using the “venn” package in R software, the obtained collagen-related DEGs were intersected with the prognosis-related collagen family genes to obtain the collagen-related DEGs associated with PTC prognosis, and venn plot was drawn. The “survival” package in R software was used to draw forest plot of the collagen-related DEGs associated with PTC prognosis. The “heatmap” package in R software was used to draw heatmap of the collagen-related DEGs associated with PTC prognosis. The protein interactions network of prognosis-associated DEGs was constructed by string online network tool, as well as correlation network, using the “igraph” package and “reshape2” package in R software.

| Number | Gene name |
|--------|-----------|
| 1      | COL9A2    |
| 2      | COL9A1    |
| 3      | COL8A2    |
| 4      | COL8A1    |
| 5      | COL7A1    |
| 6      | COL6A6    |
| 7      | COL6A5    |
| 8      | COL6A3    |
| 9      | COL6A2    |
| 10     | COL6A1    |
| 11     | COL5A3    |
| 12     | COL5A2    |
| 13     | COL5A1    |
| 14     | COL4A6    |
| 15     | COL4A5    |
| 16     | COL4A4    |
| 17     | COL4A3    |
| 18     | COL4A2    |
| 19     | COL4A1    |
| 20     | COL3A1    |
| 21     | COL2A1    |
| 22     | COL28A1   |
| 23     | COL27A1   |
| 24     | COL26A1   |
| 25     | COL25A1   |
| 26     | COL24A1   |
| 27     | COL23A1   |
| 28     | COL22A1   |
| 29     | COL21A1   |
| 30     | COL20A1   |
| 31     | COL1A2    |
| 32     | COL1A1    |
| 33     | COL19A1   |
| 34     | COL18A1   |
| 35     | COL17A1   |
| 36     | COL16A1   |
| 37     | COL15A1   |
| 38     | COL14A1   |
| 39     | COL13A1   |
| 40     | COL12A1   |
| 41     | COL11A2   |
| 42     | COL11A1   |
| 43     | COL10A1   |

Table 1 Collagen family genes from RNA sequencing transcriptome data of PTC
Construction and validation of the prognostic model

The least absolute contraction and selection operator (LASSO) penalized Cox regression analysis was performed to select optimal prognostic genes related to overall survival (OS). Prognostic risk scores were determined by a linear combination of regression coefficients ($\beta$) and gene expression levels in a LASSO Cox regression model. We calculated the risk score using the following formula: risk score = sum (the normalized expression level regarding each gene*its regression coefficient). PTC patients were divided into high and low risk groups according to the median of risk score. Kaplan–Meier (K–M) survival curves and time-dependent receptor operating characteristic (ROC) curves were used to assess the predictive ability of prognostic characteristics on OS. PCA and t-SNE analyses were performed using the “Rtsne” and “ggplot” packages in R software to visualize 4 collagen family genes by dimensionality reduction.

Construction and validation of the prognostic nomogram

Univariate and multivariate Cox regression analyses were used to determine whether prognostic models could predict OS in patients with PTC independently of other traditional clinical characteristics (including age, sex and TNM stage). Hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated for each variable. $P < 0.05$ was considered a statistically significant difference. Based on the results of the multivariate Cox regression analysis, a prognostic nomogram was constructed to predict the OS of PTC patients at 1, 2, and 5 years across the TCGA dataset. The predicted OS of nomogram against observed survival rates was plotted using the calibration curve.

Functional enrichment analysis

The “limma” package in R soft was used to identify collagen-related DGEs between the high-risk and low-risk groups using $|\log 2 FC| > 1$ and FDR < 0.05 as criteria. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the “clusterProfiler”.

Immune microenvironment and drug sensitivity analysis

The proportion of 20 tumor-infiltrating immune cells in patients with thyroid cancer was calculated based on the CIBERSORT score. ESTIMATE immune, stromal score for analysis of immune and stromal cell infiltration levels and tumor purity. In addition, the relationship between the four collagen-related DGEs and drugs was also explored. Drug data were obtained from the cellminer database (https://discover.nci.nih.gov/cellminer/loadDownload.do). Pearson’s correlation test was used in R soft to explore the correlation between the expression of collagen-related genes and drugs.

Immunohistochemistry

Human PTC tissues were collected from patients who underwent thyroidectomy at The Second Hospital of Tianjin Medical University. The study was approved by the Ethics Committee of The Second Hospital of Tianjin Medical University and written informed consent was obtained from these patients. The clinicopathological data of the patients are listed in Table 2, and a total of 119 patients were included in the study. 119 tumor tissue samples and 15 normal paracancerous tissue samples were selected for immunohistochemical staining. The tissues were fixed in 4% paraformaldehyde, embedded using paraffin, sectioned into 5 μm thick sections and dewaxed. After antigen repair and endogenous peroxidase disruption, sections were incubated with Anti-COL18A1(Abcam, ab275746) antibody overnight, followed by incubation with secondary antibody. Sections were treated with horseradish peroxidase and 3,3'-diaminobenzidine and then stained with hematoxylin and dehydrated. Light microscopy was used to capture images and immunohistochemical scoring using the 12-point scale [19].

Cell culture

All PTC cell lines purchased from the ATCC cell bank were cultured in modified medium (DMEM, Gibco, USA) supplemented with 10% fetal bovine serum (Gibco, USA), 100 KU/L penicillin and 100 mg/L streptomycin. Cells were maintained in an incubator at 37 °C and 5% CO2.

Transfection

Double-stranded siRNA targeting COL18A1 (siRNA, ID:130917) and negative control (NC) siRNA were synthesized by Thermo Fisher Scientific (Waltham, MA, USA). BCPAP and TPC-1 cells were transfected with 5 μL of siRNA and 5 μL of Lipofectamine 2000 reagent (Invitrogen, USA) and subsequent experiments were performed 48 h later.

Real time quantitative polymerase chain reaction (RT-qPCR)

TRIAZOL reagent (Invitrogen, USA) was used to extract all total RNA. cDNA from PTC tissues and cell lines was synthesized using a reverse transcriptase kit (Takara, Dalian, China). RT-qPCR was performed using a SYBR
Premix Ex Taq™ kit (Roche, Switzerland) on a Bio-Rad real-time PCR system (Bio-Rad, USA). Actin was used as an internal reference. The primers sequences of COL18A1 and Actin were as follows:

5'-GGAGAGATTGGCTTTCCTGGAC-3' and 5'-CCTCATGCCAAATCCAAGGCTG-3'; 5'-CACCATTGGCATGAGCGGTTC-3' and 5'-AGGTCTTTGCGGATGTCACGT-3'.

**Western blot**

Equal amounts of protein samples (30 μg) were extracted with RIPA lysis buffer (Sigma, USA) and uploaded into 10% sodium dodecyl sulfate polyacrylamide gels for protein separation. The proteins were transferred to polyvinylidene difluoride membranes (Millipore, USA) and then sealing the membranes with 5% skimmed milk, the membranes were incubated with primary anti-COL18A1 (1:1000, Abcam, ab275746) and anti-rabbit secondary antibody (1:5000, Thermo Fisher). Immune binding was detected using an ECL detection system (Shanghai Peiqing Technology Co., Ltd., China).

**CCK-8 assay**

Cell viability was evaluated by Cell Counting Kit-8 (CCK-8) assay in BCPAP and TPC-1 cells. COL18A1 downregulated and control cells (1000 cells/well) were respectively seeded in 96-well plates. After indicated time, 10 μL CCK-8 solution was added to each well, and cultured at 37 °C for 2 h. The OD value of each well was measured at 450 nm. All values are repeated at least three times and presented as the mean ± standard deviation.

**Transwell assay**

The BCPAP and TPC-1 cell line were chosen for the cell migration and invasion assay. Cell invasion was assessed by using the Matrigel invasion system. Transwell chambers (8 μm pore size, polycarbonate filters, 6.5 mm diameter; Corning Costar) in 24-well plates were coated with 50 μL of polymerized Matrigel (BD Biosciences, Franklin Lakes, NJ, USA). To assess effect of COL18A1 on PTC cell lines migration and invasion ability, BCPAP and TPC-1 cells were placed into the upper chamber with 200 μL serum-free medium and loaded on the lower chamber with 500 μL complete medium. After incubation at 37 °C for 48 h, the invaded cells in the lower chamber were obtained by staining and counted. All values are repeated at least three times and presented as the mean ± standard deviation.

**Cytotoxicity assay**

We isolated peripheral blood mononuclear cells (PBMCs) using the Ficoll-Paque centrifugation method from normal person. The isolated PBMCs were activated by IL-2 and cultured for 3–5 days, and then co-cultured with CFSE(5,6-carboxyfluorescein diacetate, succinimidyl ester) -labeled tumor cells at a ratio of 1:5 for 10 h. Acquisition is performed using a FACSAría instrument with FACSDiVa software (BD Biosciences, San Jose, CA, USA).

**Statistical analysis**

SPSS version 22.0 (IBM Corporation, Armonk, NY) and GraphPad Prism version 8.0 (La Jolla, CA, USA) were used for statistical analyses. Chi-square test was performed for comparison of patients’ clinicopathological characteristics. Shapiro–Wilk test was performed to test data normality before using student’s t test for analysis. All data which was involved in student’s t test accorded with the normal distribution pattern. The 2-tailed Student t test was used for the two groups in the in vitro data, and each experiment was repeated three times. Cox regression was performed to analyze the relationship between the IHC markers and patients’ prognosis. Kaplan–Meier analysis was performed.

### Table 2 Clinicopathological characteristics of PTC patients

| Characteristics | N  | Percentage |
|-----------------|----|------------|
| Gender          |    |            |
| Male            | 30 | 25.2       |
| Female          | 89 | 78.4       |
| Age             |    |            |
| <55             | 46 | 38.7       |
| ≥55             | 73 | 61.3       |
| Tumor size      |    |            |
| ≤1 cm           | 97 | 81.5       |
| >1 cm           | 22 | 18.5       |
| T stage         |    |            |
| T1–T2           | 109| 91.6       |
| T3–T4           | 10 | 8.4        |
| N stage         |    |            |
| N0              | 69 | 58.0       |
| N1a-b           | 50 | 42.0       |
| Multifocality   |    |            |
| Absent          | 80 | 67.2       |
| Present         | 39 | 32.8       |
| TNM stage       |    |            |
| T1–2            | 114| 95.8       |
| T3–4            | 5  | 4.2        |
| BRAF mutation   |    |            |
| Absent          | 26 | 21.8       |
| Present         | 93 | 78.2       |

**PTC** Papillary thyroid microcarcinoma
to investigate the relationship between the IHC markers and patients’ prognosis. FlowJo version 10.0.7 (Tree Star, Inc., Ashland, OR, USA) was used to analyze cytotoxicity assay data. P-values indicating significant differences were shown as follows: ***P < 0.001, **P < 0.01, *P < 0.05.

Results

Identification of collagen-related DEGs with predictive value

We identified 33 and 32 collagen-related differentially expressed genes between tumor and adjacent normal tissues from the TCGA database and GEO database, respectively, of which 6 collagen-related DEGs were significantly associated with OS (Fig. 1A, P < 0.05). Forest plots showed the correlation between gene expression and OS (Fig. 1B). Heat maps showed that tumors and adjacent normal tissues could be distinguished by DEGs (Fig. 1C, from TCGA database; Fig. 1D, from GEO database). Protein-protein interaction network showed DEGs interactions at the protein level (Fig. 1E), Fig. 1F showed the association between these genes.

Construction of a prognostic model based on 6 collagen-related DGEs associated with PTC prognosis

Based on collagen-family DEGs associated with PTC prognosis, we developed a prognostic model for PTC by LASSO Cox regression analysis. Prognostic index (PI) = (0.031 * expression level of COL5A1) + (0.954 * expression level of COL18A1) + (−0.656 * expression level of COL13A1) + (0.203 * expression level of COL11A1). We used the median risk score as the standard for classifying patients into group with high risk and group with low risk (Fig. 2A). Patients in the high-risk group exhibited higher mortality rates relative to the low-risk group (Fig. 2B). Kaplan–Meier curves were plotted and OS was found to be lower for patients divided into the high-risk group than the low-risk group (P = 1.752e-03) (Fig. 2C). Figure 2D showed the OS predictive power of the risk score, with the area under the curve (AUC) of 0.901, 0.917, 0.812, and 0.788 at 1 year, 2 years, 3 years, and 5 years, respectively. PCA and t-SNE analysis showed that PTC patients were distributed between the two groups in two directions. The results of PCA and t-SNE analysis showed that this prognostic model could clearly distinguish patients in the high-risk group from those in the low-risk group (Fig. 2E, F). We used the same approach to analyze the data in the ICGC database to validate the findings and reached a consistent conclusion (Fig. 3A–F).

Construction and validation of the nomogram in the PTC cohort

Cox regression analysis was used to confirm whether risk scores independently predicted OS. Univariate Cox regression analysis found a significant association between risk scores and OS (HR = 3.117, 95% CI = 1.778–5.466, P < 0.001) (Fig. 4A). Consistently, multivariate Cox regression analysis found a significant association between risk score and OS (HR = 3.871, 95% CI = 1.865–8.035, P < 0.001) (Fig. 4B). Based on the 2 independent predictors, we constructed a prognostic nomogram to quantify the predicted individual survival probabilities of PTC patients at 1-, 2-, and 5- years (Fig. 4C). There was substantial agreement between the OS rates for 1-, 2-, and 5- years predicted by the nomogram and the actual observations (Fig. 4D). We used time-dependent ROC analysis for internal validation, and the results showed that the area under the curve (AUC) was 0.987, 0.993, and 0.916 for 1-, 2-, and 5- years, respectively (Fig. 4E). The nomogram further improved prediction accuracy over the use of a single gene model (Fig. 2D).

Functional analysis of collagen family DEGs in the TCGA

We performed GO enrichment and KEGG pathway analysis to identify the biological functions and pathways associated with risk scores. Figure 5A showed the top 10 BP terms, CC terms and MF terms. The main enriched GO terms were extracellular matrix organization, extracellular structure organization, collagen fibril organization, collagen-containing and extracellular matrix structural constituent. Figure 5B showed the first 18 KEGG pathways, mainly including Protein digestion and absorption, ECM – receptor interaction, PI3K – Akt signaling pathway and Focal adhesion. These functions play a key role in tumorigenesis and development. These results suggested that collagen-associated DEGs were closely associated with the development and biological features of cancer.

Correlation of collagen-related DEGs with tumor immune microenvironment and drug sensitivity

Collagen family genes are involved in extracellular matrix formation and are an important component of the tumor microenvironment. We evaluated the relationship between the expression of COL18A1, COL13A1, COL5A1, COL11A1, and the infiltration of immune cells in PTC. It was found that the expression of all the four genes significantly affected the infiltration of immune cells in different trends. The expression of COL11A1 and COL5A1 significantly downregulated CD8 cells infiltration.
and upregulated Treg cell infiltration (Fig. 6A). Collagen family genes plays a crucial role in the cancer ECM remodeling. We assessed the association of target genes with the presence of infiltrating stroma and immune cells by ESTIMATE analysis. Undoubtedly, COL18A1, COL5A1, COL11A1 were associated with increased matrix (all...
P < 0.05). Notably, COL5A1 (R = 0.4, \( P < 2.2e-16 \)) and COL11A1 (R = 0.49, \( P < 2.2e-16 \)) showed significant correlations with mesenchymal cells infiltration, followed by COL18A1 (R = 0.24, \( P < 6.1e-08 \)). In addition, COL5A1 (R = 0.4, \( P < 2.2e-16 \)) and COL11A1 (R = 0.49, \( P < 2.2e-16 \)) were significantly correlated with thyroid cancer immune
scores. ESTIMATE scores, a combination of stromal and immune scores, indicated that high expression levels of COL5A1, and COL11A1 were significantly associated with lower tumor purity (Fig. 6B). In addition, the relationship between four collagen family genes and drugs were explored (Fig. 6C). The top 16 most relevant ones were visualized.

Fig. 3 Prognostic analysis of 4 collagen-related genes models of the ICGC cohort. A The distribution and median value of the risk scores. B The distributions of OS status and risk score. C Kaplan–Meier curves of the OS in the high-risk group and the low-risk group. D Time-dependent ROC curve analysis for measuring the predictive performance on OS. E, F The results of PCA and t-SNE analysis.
Validation of the collagen-related genes in thyroid cancer cells

Among the four collagen-related DGEs, we selected COL18A1 with the highest risk on OS evaluation in all four genes to validate its function in PTC. First, we verified the expression of COL18A1 in PTC tissues and adjacent normal thyroid tissues. Immunohistochemical results showed that COL18A1 was differentially elevated in PTC tissues compared to normal thyroid tissues (Fig. 7A, B). High expression of COL18A1 were significantly related with larger tumor size, worse T stage and TNM stage (Table 3).
In addition, we examined the correlation between the COL18A1 immunohistochemistry score and OS in patients with PTC, suggesting that patients with higher immunohistochemistry scores have a worse prognosis (Fig. 7C). The results of the Cox proportional risk model analysis were shown in Table 4. In the multifactorial analysis, lymph node metastasis ($P = 0.019$, HR = 0.118), TNM stage ($P = 0.029$, HR = 10.862), and COL18A1 expression ($P = 0.012$, HR = 5.156) were the factors that significantly correlated with poor progression free survival of PTC patients. In addition, lymph node metastasis ($P = 0.010$, HR = 0.136), TNM stage ($P < 0.001$, HR = 33.508), and COL18A1 expression ($P = 0.004$, HR = 5.092) were the factors that significantly correlated with poor overall survival of PTC patients. These results indicated that lymph node metastasis, TNM stage and COL18A1 expression were independent risk factors for overall survival and progression free survival of PTC patients. We also examined the mRNA expression levels of COL18A1 in different thyroid cancer cell lines by RT-qPCR (Fig. 7D). The results showed that COL18A1 was significantly highly expressed in thyroid cancer tissues. In addition, the expression level of COL18A1 was significantly reduced by siCOL18A1 in TPC-1 and BCPAP cells (Fig. 7E). We also found that the killing ability of PBMCs was significantly enhanced after knockdown of COL18A1 in TPC-1 and BCPAP cells. These data suggested that COL18A1 might be involved in the progression of thyroid cancer (Fig. 7F). Then, the CCK-8 assay indicated that COL18A1 knockdown repressed the proliferation of thyroid cancer cells (Fig. 7G). Finally, we found the migration and invasion ability of thyroid cancer cell was significantly reduced after COL18A1 knockdown (Fig. 7H).

**Discussion**

Collagen family genes play an important role in the formation of the extracellular matrix. A growing number of studies have identified collagen family genes as playing an important role in the development of various cancers. Recent study showed that the distribution and structural characterization of collagen within the PTC nodule capsules could be used to differentiate PTC from follicular adenomas nodules [18]. However, it is not clear whether collagen family genes expression level affect the clinical prognosis of patients with PTC. Studying the molecular mechanisms and signaling pathways of collagen family genes can provide a theoretical basis for the development of new therapeutic targets for thyroid cancer. In this study, we constructed a 4 genes model associated with collagen family genes, which could be a promising tool to predict the prognosis of thyroid cancer.

We constructed a model of 4 genes associated with collagen family genes to predict thyroid prognosis by LASSO Cox regression analysis. For patients with thyroid cancer, the high-risk score suggested a lower prognosis than the low-risk score. ROC curve confirmed that the model had accurate and sensitive predictive efficacy. Multivariate COX regression analysis showed that the signature
Fig. 6 Correlation of four prognostic collagen-related DGEs with tumor immune microenvironment and drug sensitivity. A The proportion of 20 tumor-infiltrating immune cells in thyroid cancer. B Correlation of target gene expression with stromal score, immune score and ESTIMATE score, respectively. Spearman’s correlation analysis was used to calculate the association. C Correlation between gene expression levels and drugs. The 16 drugs most related to the COL18A1, COL5A1, COL11A1, and COL13A1 were visualized.
independently predicted survival in thyroid cancer. Of these, COL5A1, COL18A1, COL13A1, COL11A1 were risk factors for thyroid cancer prognosis. Collagen type V (COL5) is classified as a regulatory fibril-forming collagen and consists of a combination of three different chains, namely COL5A1, COL5A2, and COL5A3 [20, 21]. COL5
is found in most connective tissue stroma and has been found to play a functional role in a number of cancers, including breast, colon and pancreatic ductal adenocarcinoma [22–24]. Previous study found COL5A1 might be a new prognostic factor for patients with lung adenocarcinoma [25]. In addition, high levels of COL5A1 accelerate the growth and progression in renal cell carcinoma and breast cancer [26, 27]. COL18A1, also known as endothelial inhibitor, inhibits the migration and proliferation of endothelial cells [28, 29] and induces apoptosis [30]. It has been shown to inhibit the recruitment of peripheral cells into capillaries in vivo, thereby reducing the number of mature blood vessels [31]. It also inhibits growth factor-induced entry of endothelial progenitor cells into the circulatory system, thereby inhibiting angiogenesis [32]. Interestingly, our study on the role of COL18A1 in thyroid cancer has come to a different conclusion. We found that COL18A1 is activated in thyroid cancer, which was further confirmed by IHC assay and cell culture models. Collagen type XIII (COL13) is a transmembrane protein that is located at the cell-cell and cell-ECM junctions [33, 34]. Previous studies have found that downregulation of COL13A1 resulted in the observation of an E-calmodulin to N-calmodulin switch, suggesting an association between COL13A1 and the promotion of extracellular matrix. COL13A1 may be an activator of the PI3K/AKT pathway, which plays a key role in regulating a variety of cellular functions including metabolism, growth, proliferation, survival, transcription and protein synthesis [35]. COL11A1 is expressed and secreted primarily by a subset of cancer-associated fibroblasts and regulates tumor-stromal interactions and the mechanical

| Characteristics | COL18A1 | \( \chi^2 \) value | \( P \) value |
|----------------|---------|----------------------|----------------|
| Gender         |         |                      |                |
| Male           | 24      | 6                    | 1.998          | 0.157          |
| Female         | 59      | 30                   |                |                |
| Age            |         |                      |                |                |
| <55            | 32      | 14                   | 0.001          | 0.973          |
| ≥55            | 51      | 22                   |                |                |
| Tumor size     |         |                      |                |                |
| ≤1 cm          | 73      | 24                   | 7.549          | 0.006          |
| >1 cm          | 10      | 12                   |                |                |
| T stage        |         |                      |                |                |
| T1–T2          | 79      | 30                   | 4.579          | 0.032          |
| T3–T4          | 4       | 6                    |                |                |
| N stage        |         |                      |                |                |
| N0             | 52      | 17                   | 2.453          | 0.117          |
| N1a-b          | 31      | 19                   |                |                |
| Multifocality  |         |                      |                |                |
| Absent         | 52      | 28                   | 2.608          | 0.106          |
| Present        | 31      | 8                    |                |                |
| TNM stage      |         |                      |                |                |
| T1–2           | 83      | 31                   | 12.033         | 0.001          |
| T3–4           | 0       | 5                    |                |                |
| BRAF mutation  |         |                      |                |                |
| Absent         | 20      | 6                    | 0.812          | 0.368          |
| Present        | 63      | 30                   |                |                |

**Table 3** Comparison of the clinicopathological data among two group of PTC patients with different COL18A1 expression level

**Table 4** Cox regression analysis of DSS or OS and clinicopathological features in PTC

| Characteristics          | DSS        | OS         |
|-------------------------|------------|------------|
|                         | \( P \) value | HR (95% CI) | \( P \) value | HR (95% CI) |
| Male                    | 0.543      | 0.625 (0.138–2.842) | 0.710 | 0.786 (0.222–2.788) |
| Age ≥55                 | 0.647      | 0.728 (0.188–2.827) | 0.210 | 0.478 (0.151–1.515) |
| Tumor size > 1 cm       | 0.384      | 1.880 (0.454–7.788) | 0.724 | 1.248 (0.364–4.281) |
| T stage                 | 0.449      | 1.998 (0.333–11.992) | 0.780 | 1.286 (0.219–7.535) |
| Lymph node metastasis   | 0.019      | 0.118 (0.020–0.703) | 0.010 | 0.136 (0.030–0.621) |
| Multifocality           | 0.903      | 1.085 (0.291–4.051) | 0.872 | 1.097 (0.357–3.368) |
| TNM stage               | 0.029      | 10.862 (1.278–92.321) | <0.001 | 33.508 (4.896–229.340) |
| BRAF mutation           | 0.424      | 1.914 (0.390–9.397) | 0.360 | 1.852 (0.495–6.927) |
| COL18A1 expression      | 0.012      | 5.156 (14.435–18.520) | 0.004 | 5.092 (1.706–15.202) |
properties of the extracellular matrix. COL11A1 also promotes cancer cell migration, metastasis and treatment resistance by activating pro-survival pathways and regulating tumor metabolic phenotypes [36]. High levels of COL11A1 are often associated with an aggressive tumor phenotype and poor prognosis in a variety of solid tumors including ovarian, breast, pancreatic and colorectal cancers [37, 38]. However, the function of these 4 collagen family genes in thyroid cancer remains to be clarified. Collagen I, code by COL1A1 and COL1A2, is the main component of stroma matrix. Collagen I processing and remodeling significantly affect tumor ECM topography to promote tumor progression and metastasis, and have the potential to contribute to predicting patient outcomes in various cancers, including lung adenocarcinoma, breast cancer, and pancreatic cancer [39, 40]. Our study also included COL1A1 and COL1A2 (Table 1) to analyze the potential clinical value of both genes. Although Cox analysis suggested that the HR of COL1A1 (HR = 1.165, 95% CI: 0.908–1.496, P = 0.230) and COL1A2 (HR = 1.267, 95% CI: 0.942–1.703, P = 0.118) was greater than 1, the correlation between the mRNA expression levels and the prognosis of TC patients was not significant. Future studies should include the structural parameters and accumulation level to further analyze the role of Collagen I in thyroid cancer progression.

Multivariate Cox regression analysis has shown that age is an independent risk factor for thyroid cancer. Consistently, age has been commonly applied as a risk factor for prognostic stratification of thyroid cancer [41]. To extend the clinical application of the model, we combined it with age to construct a nomogram that could be easily calculated for the expected survival of individual patients. Calibration curves confirmed that the 1-, 2-, and 5-year survival times predicted by the nomogram were close to the actual survival times. This suggests that the nomogram has the potential to be an important tool for predicting the prognosis of thyroid cancer in clinical practice.

Functional enrichment analysis of collagen-related differential genes revealed that extracellular matrix organization, extracellular structure organization, collagen fibril organization, collagen-containing and extracellular matrix structural constituent, as well as protein ECM-receptor interactions, PI3K-Akt signaling pathway and focal adhesion were all enriched. This is consistent with previous findings. Previous studies have found that collagen plays a regulatory role in tumor growth and invasion. Collagen, the most abundant component of the ECM, is excessively deposited in the extracellular matrix, forming the tumor microenvironment and regulating the polarity and migratory capacity of cancer cells [42]. The massive deposition of collagen enhances tumor cell proliferation and immune cell infiltration [43–45].

Many patients with thyroid cancer not associated with thyroiditis are found to have immune infiltration postoperatively, suggesting that alterations in the immune microenvironment are associated with the progression of thyroid cancer [46]. Our findings suggested that this model was significantly associated with aDC cell as well as NK cell infiltration. Immunotherapy has been a hot topic of research in tumor-targeted therapy in recent years, but its role in thyroid cancer is unclear. The relationship between collagen gene-related models and the tumor microenvironment needs further analysis.

In addition, we further analyzed the correlation between COL5A1, COL18A1, COL13A1, and COL11A1 gene expression and drugs. Most of the drugs related to collagen family genes expression are involved in PI3K-AKT pathway signaling inhibition, indicating a relationship between collagen family DEGs expression and AKT pathway activation, which also exist in KEGG analysis. It provides a theoretical basis for the combined application of drug therapy and collagen-targeted therapy. While we report the potential correlation between collagen family and AKT pathway, we cannot directly infer causality from above results. Future studies using cell culture or animal models would be useful in evaluating the impact of collagen family on mentioned drug’s efficacy and the directionality of associations between collagen family expression and AKT pathway activation.

In many types of cancer, stromal fibroblasts and/or myofibroblasts are the main cell types responsible for increased collagen production in solid tumors. Cancer cells produce factors that induce stromal cell proliferation and/or deposition of extracellular matrix. It has also been found that tumor cells, especially undifferentiated thyroid cancer cells, also secrete collagen and are associated with an aggressive biological behavior in recent years. In the univariate Cox regression analysis between prognostic collagen-related DEGs expression and OS, COL18A1 high expression significantly correlated with poor patient outcome and have the highest HR value. In the following LASSO regression analysis, COL18A1 has the largest weight as one of the candidate genes constituting the prediction model, and until now, the role and function of COL18A1 in thyroid cancer remains unclear, so we choose it for research. Our study is the first to analyze the expression of the COL18A1 gene and to demonstrate that abnormally high expression of this gene promotes tumor progression in PTC cells. We confirmed that COL18A1 was activated in thyroid cancer and predicted poor prognosis by immunohistochemistry assay. In addition, we determined the higher mRNA expression levels of COL18A1 in different thyroid cancer cell lines than thyroid cell line by RT-qPCR. Reducing the expression level of COL18A1 in PTC cell line significantly suppressed the proliferation ability,
migration ability, invasion ability and immune escape ability of PTC cells. The results were consistent with their mRNA levels in TCGA sequencing data, supporting our bioinformatic analysis. This study may provide new clues to explore the biological role and clinical significance of these collagen family genes. The results suggest that high expression of COL18A1 indicated poor prognosis and increases immune escape ability in thyroid cancer.

**Conclusion**

In conclusion, we constructed a 4 collagen family genes model and showed good predictive performance. The model was significantly correlated with the tumor immune microenvironment. The model combining the collagen risk score and patient age could offer new possibilities for individualized treatment of thyroid cancer patients. Collagen family genes, especially COL18A1, significantly predict patient prognosis and may be promising therapeutic targets for PTC treatment.

**Data availability**

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Author contributions** All authors contributed to the study conception and design. Data collection and analysis were performed by J.N., W.G., and N.J. The experiments in vitro were performed by J.N. and Y.C. The first draft of the manuscript was written by J.N. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This work was partially supported by grants from National Natural Science Foundation of China (Grant No. 81872079), Clinical Research Project of the Second Hospital of Tianjin Medical University (Grant No. KY2019K086-02).

**Compliance with ethical standards**

**Conflict of interest** The authors declare no competing interests.

**Ethical approval** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. And this study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Second Hospital of Tianjin Medical University.

**References**

1. C.D. Seib, J.A. Sosa, Evolving understanding of the epidemiology of thyroid cancer. Endocrinol. Metab. Clin. North Am., 48(1), 23-+ (2019)

2. R.C. Mould et al. Immune responses in the thyroid cancer microenvironment: making immunotherapy a possible mission. Endocr.-Relat. Cancer 24(12), T311–T329 (2017)

3. Y. Bai, K. Kakudo, Y. Li, Subclassification of non-solid-type papillary thyroid carcinoma identification of high-risk group in common type (vol 99, pg 1908, 2008). Cancer Sci. 99(12), 2548 (2008)

4. T.R. Cox, J.T. Erler, Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Dis. Models Mechanisms 4(2), 165–178 (2011)

5. P. Friedl, K. Wolf, Tube travel: The role of proteases in individual and collective a cancer cell invasion. Cancer Res. 68(18), 7247–7249 (2008)

6. P.G. Gritsenko, O. Ilina, P. Friedl, Interstitial guidance of cancer invasion. J. Pathol. 226(2), 185–199 (2012)

7. R.V. Iozzo, M.A. Gubbiotti, Extracellular matrix: The driving force of mammalian diseases. Matrix Biol. 71–72, 1–9 (2018)

8. N.K. Karamanos et al. Matrix modeling and remodeling: A biological interplay regulating tissue homeostasis and diseases. Matrix Biol. 75–76, 1–11 (2019)

9. N.K. Karamanos et al. Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. Chem. Rev. 118(18), 9152–9232 (2018)

10. N. Afratis et al. Glycosaminoglycans: key players in cancer cell biology and treatment. Febs J. 279(7), 1177–1197 (2012)

11. A.D. Theocharis, D. Manou, N.K. Karamanos, The extracellular matrix as a multitasking player in disease. Febs J. 286(15), 2830–2869 (2019)

12. P.P. Provenzano et al. Collagen density promotes mammary tumor initiation and progression. Bmc Med. 6, 11 (2008)

13. W.J. Han et al. Oriented collagen fibers direct tumor cell invavasation. Proc. Natl Acad. Sci. USA 113(40), 11208–11213 (2016)

14. M.W. Conklin et al. Collagen alignment as a predictor of recurrence after ductal carcinoma in situ. Cancer Epidemiol. Biomark. Prev. 27(2), 138–145 (2018)

15. M.W. Conklin et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am. J. Pathol. 178(3), 1221–1232 (2011)

16. K.B. Pointer et al. Association of collagen architecture with glioblastoma patient survival. J. Neurosurg. 126(6), 1812–1821 (2017)

17. M.F. Penet et al. Structure and function of a prostate cancer dissemination-permissive extracellular matrix. Clin. Cancer Res. 23(9), 2245–2254 (2017)

18. R. Hristu et al. Quantitative second harmonic generation microscopy for the structural characterization of capsular collagen in thyroid neoplasms. Biomed. Opt. Express 9(8), 3923–3936 (2018)

19. T. Li et al. CUL4 E3 ligase regulates the proliferation and apoptosis of lung squamous cell carcinoma and small cell lung carcinoma. Cancer Biol. Med. 17(2), 357–370 (2020)

20. K.M. Mak, C.Y.M. Png, D.J. Lee, Type V collagen in health, disease, and fibrosis. Anat. Rec.-Adv. Integr. Anat. Evolut. Biol. 299(5), 613–629 (2016)

21. T. Nardo et al. Adhesion to type V collagen enhances staurosporine-induced apoptosis of adrenocortical cancer cells. Tumor Biol. 35(10), 9939–9955 (2014)

22. S.H. Barsky et al. Increased content of Type V Collagen in desmoplasia of human breast carcinoma. Am. J. Pathol. 108(3), 276–283 (1982)

23. H. Fischer et al. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. Carcinogenesis 22(6), 875–878 (2001)

24. S. Berchtold et al. Collagen type V promotes the malignant phenotype of pancreatic ductal adenocarcinoma. Cancer Lett. 356(2), 721–732 (2015)
25. W.B. Liu et al. COL5A1 may contribute the metastasis of lung adenocarcinoma. Gene 665, 57–66 (2018)
26. G. Feng et al. Overexpression of COL5A1 promotes tumor progression and metastasis and correlates with poor survival of patients with clear cell renal cell carcinoma. Cancer Manag Res 11, 1263–1274 (2019)
27. W. Ren et al. Overexpression of collagen type V alpha1 chain in human breast invasive ductal carcinoma is mediated by TGF-beta1. Int J. Oncol. 52(5), 1694–1704 (2018)
28. M.S. O’Reilly et al. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88(2), 277–285 (1997)
29. N. Yamaguchi et al. Endostatin inhibits VEGF-induced endothelial cell migration and tumor growth independently of zinc binding. EMBO J. 18(16), 4414–4423 (1999)
30. M. Dhanabal et al. Endostatin induces endothelial cell apoptosis. J. Biol. Chem. 274(17), 11721–11726 (1999)
31. D.K. Skovseth et al. Endostatin dramatically inhibits endothelial cell migration, vascular morphogenesis, and perivascular cell recruitment in vivo. Blood 105(3), 1044–1051 (2005)
32. G. Schuch et al. Endostatin inhibits the vascular endothelial growth factor-induced mobilization of endothelial progenitor cells. Cancer Res 63(23), 8345–8350 (2003)
33. P. Hagg et al. Type XIII collagen is identified as a plasma membrane protein. J. Biol. Chem. 273(25), 15590–15597 (1998)
34. P. Hagg et al. Type XIII collagen: a novel cell adhesion component present in a range of cell-matrix adhesions and in the intercalated discs between cardiac muscle cells. Matrix Biol. 19(8), 727–742 (2001)
35. M. Yi, J.E. Schnitzer, Impaired tumor growth, metastasis, angiogenesis and wound healing in annexin A1-null mice. Proc. Natl Acad. Sci. USA 106(42), 17886–17891 (2009)
36. S. Nallanthighal, J.P. Heiserman, and D.J. Cheon, Collagen Type XI Alpha 1 (COL11A1): A Novel Biomarker and a Key Player in Cancer. Cancers (Basel), 13(5), (2021)
37. F. Vazquez-Villa et al. COL11A1/(pro)collagen 11A1 expression is a remarkable biomarker of human invasive carcinoma-associated stromal cells and carcinoma progression. Tumor Biol. 36(4), 2213–2222 (2015)
38. Z. Raglow, S.M. Thomas, Tumor matrix protein collagen XI alpha 1 in cancer. Cancer Lett. 357(2), 448–453 (2015)
39. P.M. Angel, S.C. Zambrycki, Predictive value of collagen in cancer. Adv. Cancer Res 154, 15–45 (2022)
40. A.C. Martins Cavaco et al. Collagen biology making inroads into prognosis and treatment of cancer progression and metastasis. Cancer Metastasis Rev. 39(3), 603–623 (2020)
41. X.P. Shen et al. Patient age-associated mortality risk is differentiated by BRAF V600E status in papillary thyroid cancer. J. Clin. Oncol. 36(5), 438–44 (2018)
42. T.R. Cox et al. The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase. Nature 522(7554), 106–110 (2015)
43. S. Kauppila et al. Aberrant type I and type III collagen gene expression in human breast cancer in vivo. J. Pathol. 186(3), 262–268 (1998)
44. M.A. Wozniak et al. ROCK-generated contractility regulates breast epithelial cell differentiation in response to the physical properties of a three-dimensional collagen matrix. J. Cell Biol. 163(3), 583–595 (2003)
45. G. Kaplan, G. Gaudernack, In vitro differentiation of human monocytes. Differences in monocyte phenotypes induced by cultivation on glass or on collagen. J. Exp. Med 156(4), 1101–14. (1982)
46. N.G. Nicolson et al. Immune cell infiltrate-associated dysregulation of DNA repair machinery may predispose to papillary thyroid carcinogenesis. Surgery 167(1), 66–72 (2020)

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.