



# New insights into COL26A1 in thyroid carcinoma: prognostic prediction, functional characterization, immunological drug target and ceRNA network

Yulou Luo<sup>1#^</sup>, Yinghui Ye<sup>2#</sup>, Yuting Zhang<sup>3#</sup>, Lan Chen<sup>4</sup>, Ximing Qu<sup>5</sup>, Na Yi<sup>5</sup>, Jihua Ran<sup>6</sup>, Yan Chen<sup>5,7^</sup>

<sup>1</sup>Department of Breast Surgery, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, China; <sup>2</sup>Department of Laboratory Medicine, Peking University Shenzhen Hospital, Shenzhen, China; <sup>3</sup>Department of Breast Surgery, The First Affiliated Hospital, Jinan University, Guangzhou, China; <sup>4</sup>The Second Department of Gastroenterology, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China; <sup>5</sup>Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Xinjiang Medical University, Urumqi, China; <sup>6</sup>Clinical Laboratory Diagnostic Center, General Hospital of Xinjiang Military Region, Urumqi, China; <sup>7</sup>Xinjiang Key Laboratory of Molecular Biology for Endemic Diseases, Urumqi, China

*Contributions:* (I) Conception and design: Y Luo, Y Ye, Y Zhang, Y Chen; (II) Administrative support: Y Chen; (III) Provision of study materials or patients: L Chen, X Qu, N Yi, J Ran; (IV) Collection and assembly of data: L Chen, X Qu, N Yi, J Ran; (V) Data analysis and interpretation: Y Luo, Y Ye, Y Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work as co-first authors.

*Correspondence to:* Yan Chen, PhD. Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Xinjiang Medical University, 567 Shangde North Road, Urumqi 830011, China; Xinjiang Key Laboratory of Molecular Biology for Endemic Diseases, Urumqi, China. Email: yanchen@xjmu.edu.cn.

**Background:** Thyroid carcinoma (THCA) is one of the most commonly diagnosed malignancies. Collagen is the main component in extracellular matrix. Rising studies have determined the oncogenic effect of collagen in cancer progression, which is intriguing to be further explored. Collagen type XXVI alpha 1 chain (COL26A1) is a newly discovered collagen subtype, functions of which still remain poorly demonstrated in THCA.

**Methods:** Based on the transcriptome data from The Cancer Genome Atlas (TCGA) and other public databases, we conducted investigations of COL26A1 in THCA with respects to diagnostic/prognostic prediction, functional characterization, immune infiltration, chemical drug target and non-coding RNA regulatory network. Furthermore, quantitative real-time polymerase chain reaction (qRT-PCR) and western blot were used to verify the expression of COL26A1 in THCA.

**Results:** COL26A1 was significantly upregulated in THCA, and the high COL26A1 expression inferred poor prognosis [hazard ratio (HR) =4.76; 95% confidence interval (CI): 1.36–16.73; P=0.015]. The diagnostic area under the curve (AUC) of COL26A1 achieved 0.736 (95% CI: 0.669–0.802). COL26A1 was also identified as an independent prognostic predictor for THCA (HR =3.928; 95% CI: 3.716–4.151; P<0.001). Besides, logistic regression analysis indicated that age >45 years [odds ratio (OR) =1.532; 95% CI: 1.081–2.176; P=0.017], pathological stage III (OR =2.055; 95% CI: 1.314–3.184; P=0.001), tall cell subtype (OR =5.533; 95% CI: 2.420–14.957; P<0.001), residual tumor R1 (OR =1.844; 95% CI: 1.035–3.365; P=0.041) and extrathyroidal extension (OR =1.800; 95% CI: 1.225–1.660; P=0.003) were risk factors associated with high COL26A1 expression in THCA. Functional characterizations implied that COL26A1 was associated with immunological processes and oncogenic signaling pathways. High COL26A1 expression was accompanied by more abundant infiltration of immune cells and higher stromal/immune score. In addition, most immune checkpoints were significantly positively co-expressed with COL26A1, including

<sup>^</sup> ORCID: Yulou Luo, 0000-0002-3383-2079; Yan Chen, 0000-0002-4850-4318.

PD-1, PD-L1 and CTLA4. Drugs including trichloroethylene, acetamide and thioacetamide etc. that can decrease the expression of COL26A1 were also identified. The predicted long noncoding RNA (lncRNA)-microRNA (miRNA)-COL26A1 regulatory axes were successfully deciphered. qRT-PCR and western blot verified the upregulation of COL26A1 in THCA.

**Conclusions:** Our work has primarily appraised COL26A1 as a promising biomarker for diagnosis/prognosis and immuno/therapeutic target in THCA.

**Keywords:** Collagen type XXVI alpha 1 chain (COL26A1); thyroid carcinoma (THCA); prognosis; immune infiltration; competing endogenous RNA network (ceRNA network)

Submitted Feb 01, 2023. Accepted for publication Oct 08, 2023. Published online Dec 27, 2023.

doi: 10.21037/tcr-23-141

View this article at: <https://dx.doi.org/10.21037/tcr-23-141>

## Introduction

Thyroid carcinoma (THCA) has risen to the ninth most common malignant tumor type worldwide and the incidence of THCA is much higher in women, which is reported to be the fifth, according to global cancer statistics in 2020 (1). There are mainly four different histological subtypes of THCA, which are anaplastic thyroid carcinoma (ATC), medullary thyroid carcinoma (MTC), follicular thyroid carcinoma (FTC) and papillary thyroid carcinoma (PTC), ranging from poor prognosis to favourable outcome (2).

PTC and FTC are labelled as differentiated THCA according to the differentiation level of cancer cells, while ATC cells exhibit high invasiveness involving surrounding tissue and vessels (3). Although patients with PTC or FTC are well responsive to surgical treatment plus thyroid stimulating hormone (TSH) suppression treatment, lymph node metastasis remains to be common occurrence. Besides, characterized by rapid invasion, satisfying treatment still cannot be well achieved for patients with ATC. Therefore, it is getting necessary to scrutinize the sophisticated molecular mechanisms and exploit promising molecules in the development and progression of THCA. These molecules as pivotal regulators in tumorigenesis are potential biomarkers and drug targets for further therapy. However, relevant work has not been fully completed.

Collagen is a series of proteins commonly expressed in animal organisms and its most well-known function is to serve as a major component of extracellular matrix (ECM). Apart from that, oncogenic roles of collagen in affecting motility of immune cells, propelling metastasis and enhancing immune checkpoints blockade resistance have been continuously reported (4,5). There are plentiful subtypes in the collagen family, which are collectively characterized with a unique triple-helical structure formed by tandem repeats of the consensus sequence Xaa-Yaa-Gly (6). Collagen type XXVI alpha 1 chain (COL26A1) is a newly discovered member in collagen that maps at human chromosome 7q22(+) with a size of 196,150 base pairs (7). COL26A1 gene encodes a solid but pliable protein, which contains 441 amino acids, to constitute or provide support for cellular endoplasmic reticulum, golgi apparatus, plasma membrane, cytoskeleton and ECM (7). The potential roles of COL26A1 in human diseases remain poorly exploited hitherto. Diseases reported to be associated with COL26A1

### Highlight box

#### Key findings

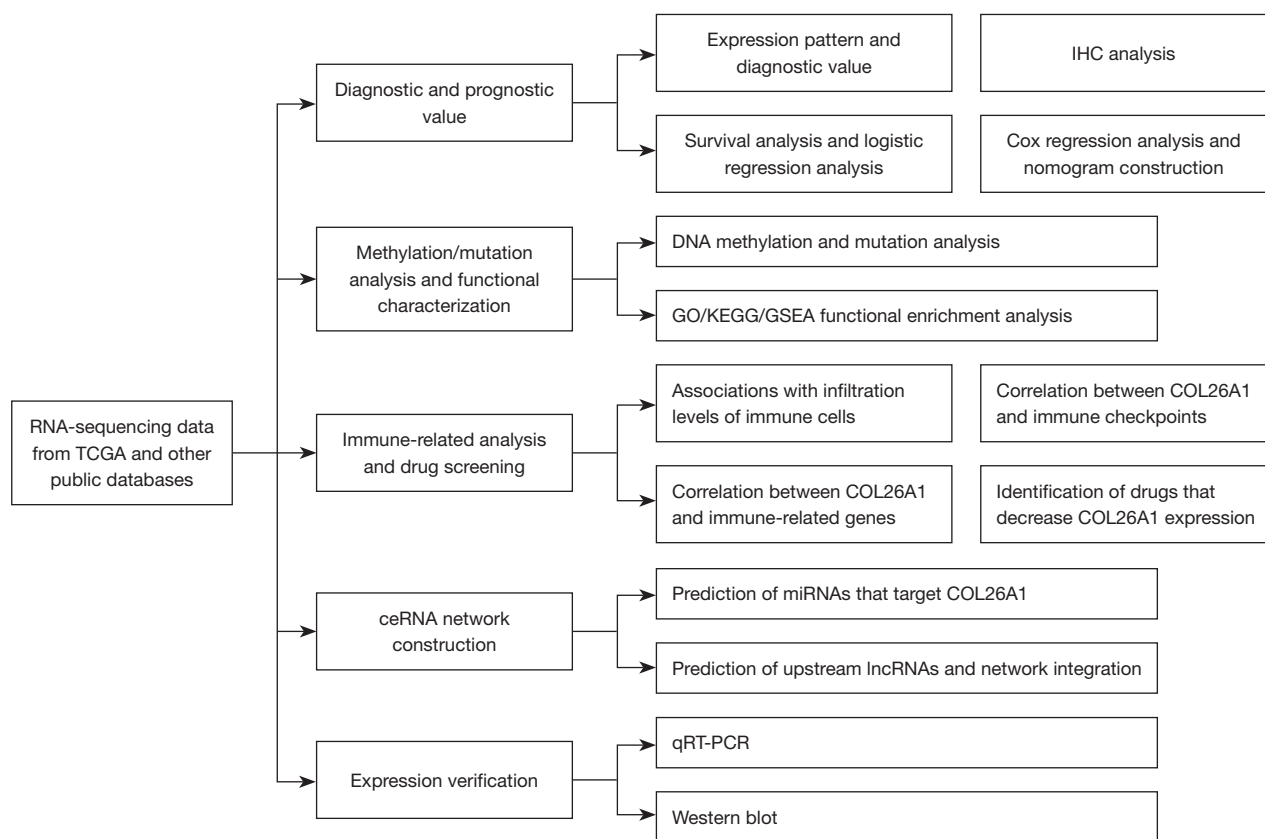
- Collagen type XXVI alpha 1 chain (COL26A1) is significantly upregulated in thyroid carcinoma (THCA).
- COL26A1 is an independent risky factor for THCA prognosis.
- COL26A1 significantly correlates with the tumor microenvironment of THCA.

#### What is known and what is new?

- Collagen dysfunction contributes to cancer progression and COL26A1 remains poorly demonstrated in cancer.
- We find that COL26A1 may serve as a biomarker for the diagnosis/prognosis of THCA. COL26A1 independently stratifies the prognosis of THCA.
- COL26A1 may be involved in several immune-related pathways during tumorigenesis, such as CTLA4 pathway, IL-10 signaling pathway and chemokine receptors bind chemokines etc.
- The drug interaction network and ceRNA network centered on COL26A1 are established.

#### What is the implication, and what should change now?

- The present study implies the role of COL26A1 in THCA for the first time that it can be applied as a promising biomarker for diagnosis/prognosis, as well as an immuno/therapeutic target.



**Figure 1** Framework of the present study. TCGA, The Cancer Genome Atlas; IHC, immunohistochemistry; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene set enrichment analysis; COL26A1, collagen type XXVI alpha 1 chain; ceRNA, competing endogenous RNA; lncRNA, long noncoding RNA; miRNA, microRNA; qRT-PCR, quantitative real-time polymerase chain reaction.

are merely acute ethmoiditis, asthma, nasal polyps and aspirin intolerance (7). COL26A1 may also participate in placental growth (8). In addition, no previous study has reported the recruitment of COL26A1 in cancer initiation and development.

Based on previously reported achievements concerning cancerous roles of collagen, we tried to primarily explore the role of COL26A1 in THCA via bioinformatics, in an attempt to fertilize relevant literature and serve as basis for further elucidation and experimental validation. In this present study, we carried out comprehensive investigations of the diagnostic/prognostic value of COL26A1 and its correlation with immune cells/genes in THCA. The drug interactive network and competing endogenous RNA (ceRNA) network centered on COL26A1 were also strengthened. Besides, the expression of PSMB8 in THCA was verified by quantitative real-time polymerase chain reaction (qRT-PCR) and western

blot. We present this article in accordance with the MDAR reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-141/rc>).

## Methods

### Genome and protein information

Framework of the present study is concluded in *Figure 1*. We obtained the genome annotation of the COL26A1 gene from the University of California Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu>) on Human Dec 2013 (GRCh38/hg28) assembly (9). The predicted 3D structure of COL26A1 protein was obtained from AlphaFold Protein Structure Database (<http://alphafold.ebi.ac.uk>) (10). The subcellular location of COL26A1 protein was retrieved from GeneCards (<http://www.genecards.org>) (7).

The study was conducted in accordance with the

Declaration of Helsinki (as revised in 2013).

### **Raw data acquisition and preprocessing**

We downloaded and preprocessed the clinical information, RNA-sequencing data and microRNA (miRNA)-sequencing data of THCA from The Cancer Genome Atlas (TCGA) (<https://www.cancer.gov>), including 510 THCA samples (PTC) and 58 normal samples. Duplicated samples and samples without important clinical information (survival information, age and gender etc.) were excluded. We also downloaded the RNA-sequencing profiles across 33 cancer types from TCGA for pan-cancer analysis, including 10,363 tumor samples and 730 normal samples. We ranked the 510 THCA samples based on the expression level of COL26A1 thereby demarcating high/low COL26A1 expression subgroups according to the median value. Baseline characteristics of the 510 THCA samples were compared based on high/low COL26A1 subgroups using Chi-squared test. Logistic regression analysis was processed to determine risk factors associated with COL26A1 expression from several clinicopathological characteristics in THCA, low expression of COL26A1 was set as reference.

### **Gene expression analysis**

R software (version 4.0.3) and “ggplot2” package were used for expression analysis. We compared the differential expression level of COL26A1 across 33 cancer types and especially in THCA (tumor *vs.* normal). Then we checked out whether there was a significantly differential expression of COL26A1 in several subgroups based on clinicopathological characteristics, including T stage, N stage, M stage, pathological stage, histological type and tumor status. R software, “pROC” package and “ggplot2” package were used for diagnostic receiver operating characteristic (ROC) curve depiction. We acquired immunohistochemistry (IHC) results of COL26A1 protein in both THCA tissue and normal tissue from The Human Protein Atlas (HPA; version 21.0, <http://www.proteinatlas.org>). Image Pro Plus 6.0 software was utilized to quantify the integrated optical density (IOD) scores of IHC results, whereby we compared and visualized the difference of IOD scores between THCA tissue and normal tissue with paired *t*-test.

### **Survival analysis**

R software, “survminer” package and “survival” package

were used for survival analysis. Overall survival (OS) was evaluated between high/low COL26A1 expression subgroups. The survival probability of clinical subgroups (histological type and thyroid gland disorder history) was also evaluated based on COL26A1 expression. Univariate and multivariate Cox regression analysis were subsequently performed to determine independent prognostic factors for THCA. R software, “rms” package and “survival” package were used to draw a nomogram to predict survival probability of patients with THCA based on meaningful factors from Cox regression analysis. Corresponding area under the curve (AUC) of time dependent ROC curve ranging from 1- to 14-year was displayed via R software, “timeROC” package and “ggplot2” package.

### **DNA methylation analysis and functional characterization**

MEXPRESS (<http://www.mexpress.be>) is a robust tool for gene methylation exploration (11), whereby we performed DNA methylation analysis between COL26A1 gene of numerous probes (cg25290307, cg17361163 etc.) in THCA. We then obtained a COL26A1-centered protein-protein interaction (PPI) network from STRING database (<http://cn.string-db.org>) (12). R software and “DESeq2” package (13) were used to produce the volcano plot, which displayed the differentially expressed genes (DEGs) of high/low COL26A1 expression subgroups, low COL26A1 expression subgroup was set as reference. A gene with adjusted  $P \leq 0.05$  and  $|\text{Log}_2\text{FoldChange}| > 1$  was considered to be differentially expressed. The top fifty DEGs were selected for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis with R software, “clusterProfiler” package and “org.Hs.eg.db” package (14). R software and “clusterProfiler” package were used for Gene Set Enrichment Analysis (GSEA) of COL26A1 (14,15).

### **Mutation analysis**

cBioPortal (<http://cbioportal.org>) is a comprehensive open platform for cancer genomic data analysis (16), whereby we investigated genetic alteration of COL26A1. We selected “TCGA PanCancer Atlas Studies” and inserted “COL26A1” for analysis. Data including alteration frequency, structural variant, mutation and copy number alteration (CNA) was extracted. We then obtained mutated site summary of COL26A1 via the “Mutations” module. The “Thyroid Carcinoma (TCGA, PanCancer Atlas)” was re-selected to

conduct a survival analysis in OS with/without COL26A1 gene alteration via the “Comparison/Survival” module.

### *Immune-related analysis*

R software and “GSVA” package [single-sample GSEA (ssGSEA) algorithm] were used to check out the significance and correlation between COL26A1 expression and infiltration levels of 24 immune cells (17,18). R software and “ESTIMATE” package (ESTIMATE algorithm) were used to check out the significance and correlation between COL26A1 expression and stromal/immune score (19). Next, we processed co-expression analysis between COL26A1 and 47 immune checkpoints to evaluate their significance and correlation (Spearman correlation was chosen) with R software and “ggplot2” package. Moreover, we examined the significance and correlation (Spearman correlation was chosen) between COL26A1 and five immune-related gene clusters via co-expression analysis, each cluster contains certain genes that share similar functions in immune response. After excluding the genes with P value >0.05, we conducted least absolute shrinkage and selection operator (LASSO) regression analysis to identify prognostic risk factors from the five gene clusters.

### *Drug analysis*

We choose Comparative Toxicogenomics Database (CTD) database (<http://ctdbase.org>) to screen drugs that can decrease the mRNA expression of COL26A1. R software, “igraph” package and “ggraph” package were utilized to draw the drug network (20). Molecular structures of the selected drugs were retrieved from PubChem database (<https://pubchem.ncbi.nlm.nih.gov>) and displayed.

### *ceRNA network construction*

miRNet (<http://mirnet.ca/miRNet>) was utilized to predict miRNAs that may potentially target COL26A1 mRNA. R software and “ggplot2” package were used to display the negative correlation (Spearman correlation was chosen) between COL26A1 and predicted miRNA. We identified the potential binding site between COL26A1 mRNA and predicted miRNA via TargetScan (<http://targetscan.org>). starBase (<https://rnasysu.com/encori/>) and miRNet were utilized to predict upstream long noncoding RNAs (lncRNAs) that target the predicted miRNA, results of which were overlapped. The correlation and significance between lncRNAs and the predicted miRNA were determined by

starBase. R software and “ggalluvial” package were used to integrate ceRNA network.

### *Cell culture*

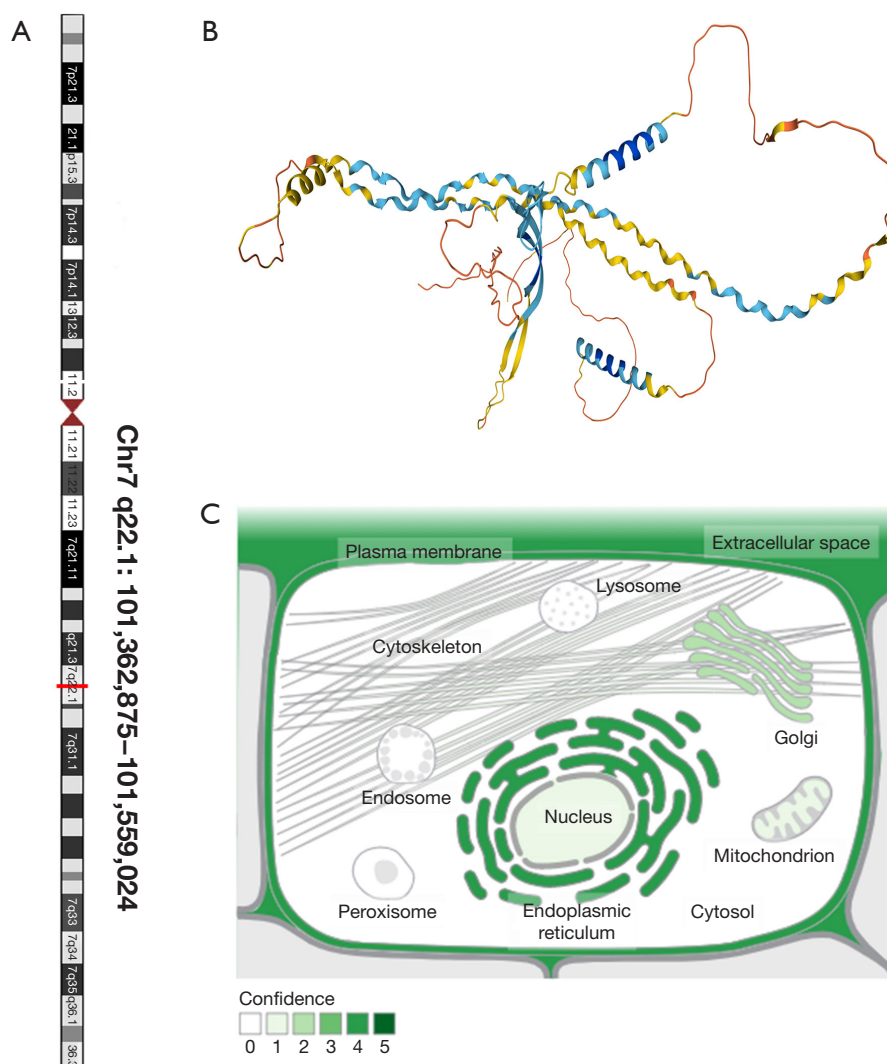
Human PTC cell lines (KTC-1 and B-CPAP) and human normal thyroid epithelial cell line (NY-OPI3-1) were purchased from Wuhan Procell Life Science and Technology Co., Ltd. (Wuhan, China). Each cell line was cultured in its dedicated medium (Wuhan Procell Life Science and Technology Co. Ltd.). Cells were cultured in RPMI-1640 (Gibco BRL, Grand Island, NY, USA), supplemented with 10% foetal bovine serum (Gibco FBS, Grand Island, NY, USA), 100 U/mL penicillin G and 100 µg/mL streptomycin.

### *qRT-PCR*

Total RNA was extracted from THCA cell lines and NY-OPI3-1 using TRIzol Reagent (Cat. No. P118-05, GenStar, Beijing, China) according to the manufacturer’s instructions. Total RNA was amplified by qRT-PCR using SYBR Green Master Mix (Cat#: C0006, TOPSCIENCE, Shanghai, China) according to the manufacturer’s instructions, and the mRNA levels of COL26A1 were detected. The experiment was replicated for three times. The primer pairs of COL26A1 were synthesized by Accurate Biology (Changsha, China). Forward (5’-3’): CTGCGCCAACCTCGTAAGG; reverse (5’-3’): TTCCTCATCACAGTTGCTCCC.

### *Western blot*

THCA cells and normal thyroid epithelial cells were harvested with radioimmuno precipitation assay (RIPA) buffer (Zhonghuihecai, Xi’an, China) and pelleted by centrifugation at 4 °C for 15 min, and the supernatant was discarded. Next, 1/5 sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) sample loading buffer, ×5 (Beyotime, Shanghai, China) was added to the supernatant and heated in a 100 °C metal bath for 10 min. The protein was separated on a 15% SDS-PAGE, transferred to a 0.22-mm polyvinylidene fluoride (PVDF) membrane (Sigma-Aldrich, Shanghai, China), placed in 5% skimmed milk, blocked for approximately 2 h, and incubated with specific antibodies. The experiment was replicated for three times. Antibodies used were as follows: the COL26A1 (Cat. No. PA5-37248, 1:2,000, Invitrogen, Carlsbad, USA) and β-actin (Cat. No. Ab6276, 1:100,000, Abcam, Shanghai, China). The protein bands were enhanced using a chemiluminescent kit (Vazyme,



**Figure 2** Genome and protein information of COL26A1. (A) Genome location of COL26A1 gene. (B) Predicted 3D structure of COL26A1 protein. Deep blue indicates very high prediction confidence, light blue indicates moderate prediction confidence, yellow indicates low prediction confidence and orange indicates very low prediction confidence. (C) Subcellular location of COL26A1 protein. COL26A1, collagen type XXVI alpha 1 chain.

Nanjing, China).

**Statistical analysis**

Bioinformatic analysis was all conducted by R 4.0.3. The comparison of Kaplan-Meier survival curve was achieved by Cox regression analysis. Difference of expression level between groups were compared by Wilcoxon rank sum test. Difference of composition rate between groups were compared by Chi-squared test. Spearman correlation was taken for correlation analysis.  $|r| > 0.1$  was considered to

be relevant and P value  $< 0.05$  was deemed as statistically significant. “\*” indicates P value  $< 0.05$ , “\*\*\*” indicates P value  $< 0.01$  and “\*\*\*\*” indicates P value  $< 0.001$  throughout this study.

**Results**

*Overview of COL26A1*

Genome information revealed that COL26A1 mapped at human chr7 (q22.1) (Figure 2A). The predicted 3D structure

of COL26A1 protein acquired from AlphaFold Protein Structure Database was showed (Figure 2B). Subcellular location indicated that COL26A1 protein was most abundant in ECM, plasma membrane and endoplasmic reticulum (Figure 2C).

### **COL26A1 was a promising diagnostic biomarker in THCA**

The expression pattern of COL26A1 among 33 cancer types was displayed (Figure 3A). Significantly differential expression of COL26A1 between tumor and normal samples was observed in 13 cancer types, among which COL26A1 was upregulated in six cancer types, including THCA, endometrioid cancer (UCEC), stomach cancer (STAD), colon cancer (COAD), head and neck cancer (HNSC) and prostate cancer (PRAD). Particularly, higher expression of COL26A1 in THCA compared to the normal was separately displayed (Figure 3B) ( $P < 0.001$ ). We did not observe statistical significance of COL26A1 expression based on subgroups from T stage, N stage and M stage (Figure 3C–3E). Significant upregulation of COL26A1 was found in pathological stage III compared to stage I ( $P < 0.001$ ) and stage II ( $P < 0.05$ ), revealing a potential role of COL26A1 in early pathological stage progression (Figure 3F). Based on histological subtypes, the expression of COL26A1 was significantly lower in follicular subtype in contrast to classical type ( $P < 0.01$ ) and tall cell type ( $P < 0.001$ ) (Figure 3G). There was no significant difference in COL26A1 expression based on subgroups from tumor status (Figure 3H). Next, we checked out the distinguishing capability of COL26A1 as a diagnostic biomarker for THCA by establishing a ROC curve (Figure 3I). The AUC of diagnostic ROC curve was calculated to be 0.736, suggesting a certain degree of accuracy for elevated expression of COL26A1 to distinguish THCA from the normal.

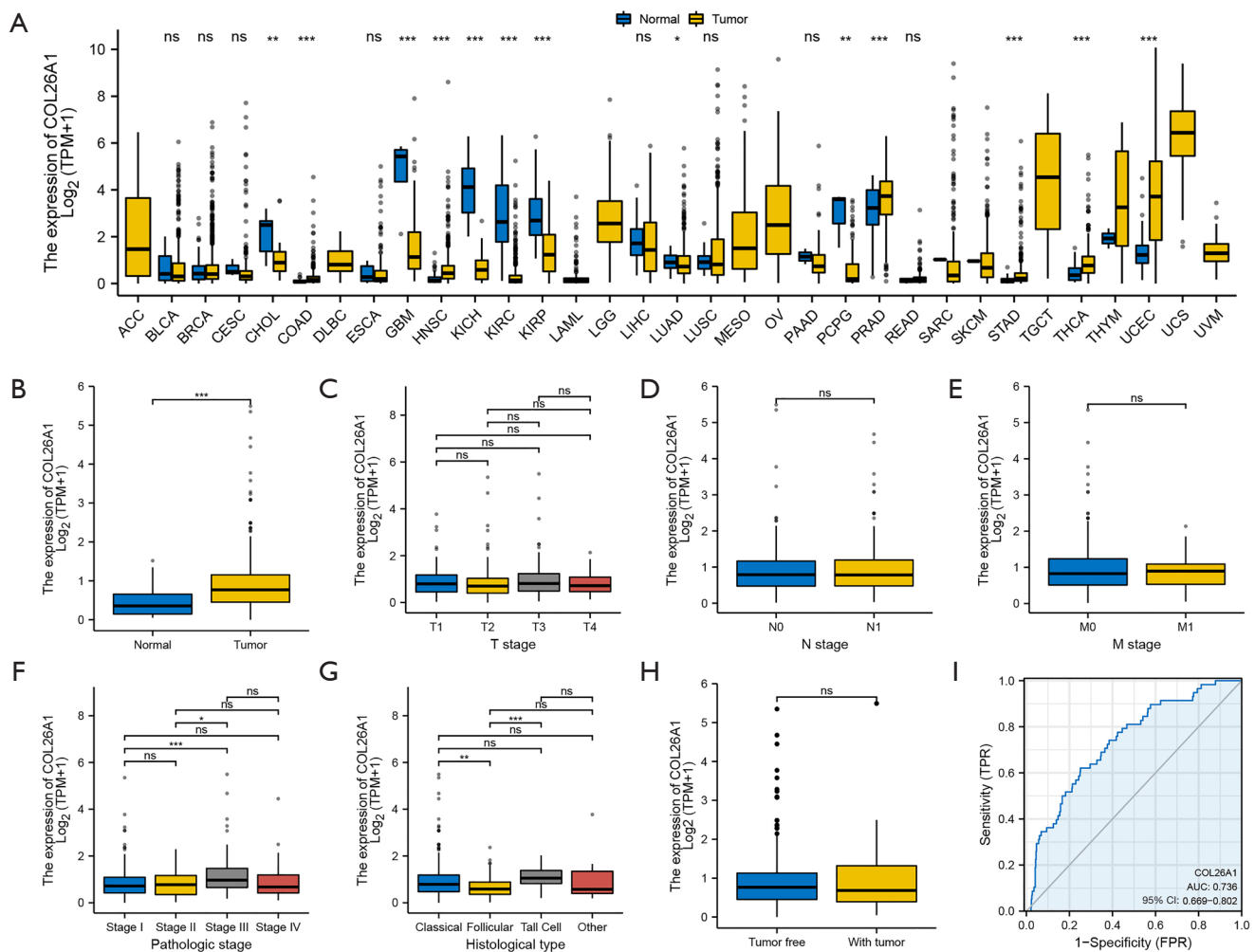
The IHC results of COL26A1 protein in both THCA tissue and normal tissue acquired from HPA database were displayed respectively (Figure 4A,4B). Two highlighted sites in THCA tissue and corresponding IOD scanning were demonstrated respectively (Figure 4C–4F). One highlighted site in normal tissue and corresponding IOD scanning were also demonstrated (Figure 4G,4H). Then we compared and visualized the difference between IOD scores from THCA tissue and IOD scores from normal tissue, which turned out to be significantly different (Figure 4I,  $P < 0.001$ ). This verified the upregulation of COL26A1 protein in THCA

tissues compared to the normal.

### **High COL26A1 expression inferred poor prognosis**

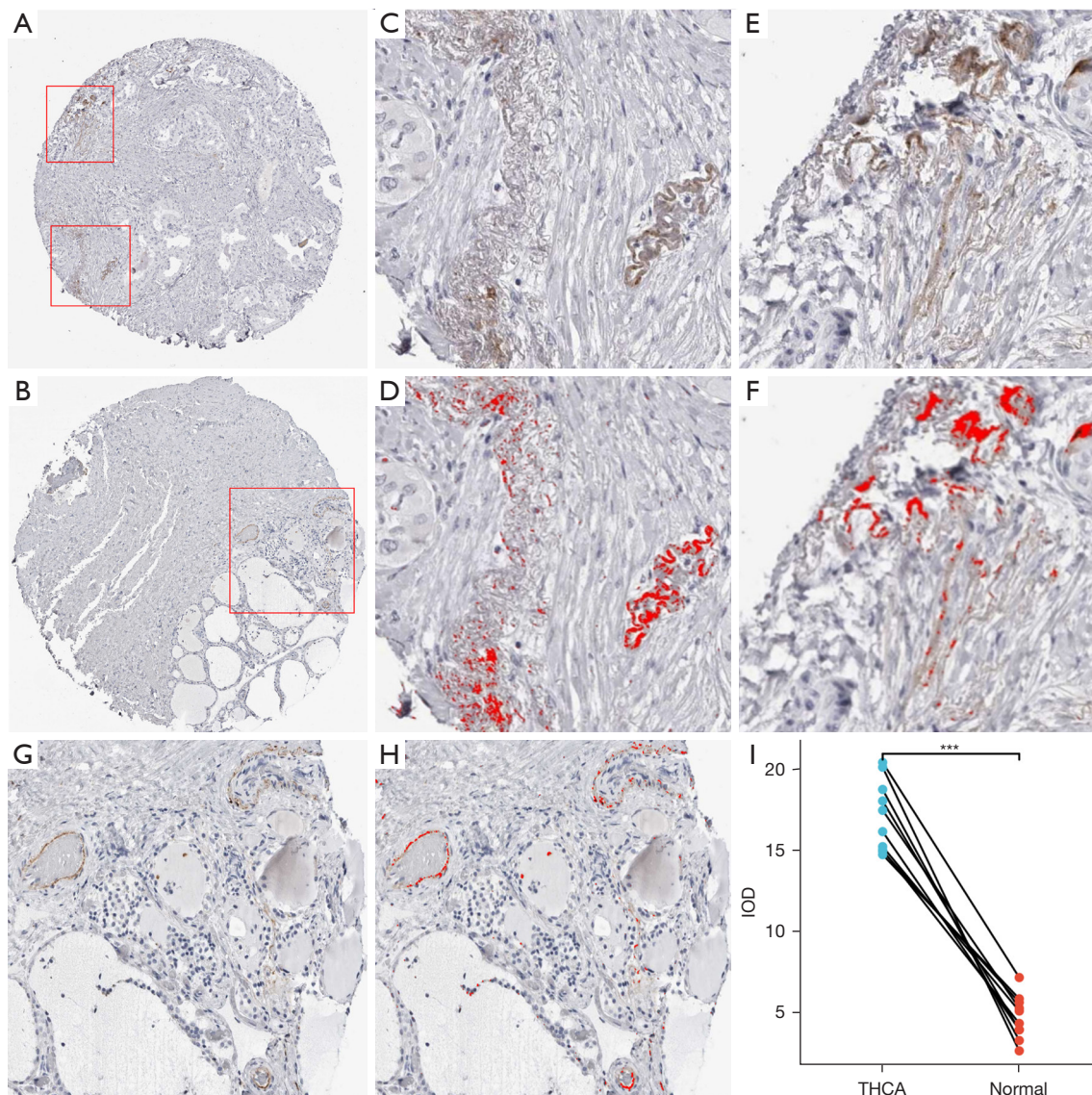
Clinicopathological characteristics of the 510 THCA samples are displayed in Table 1, based on high/low COL26A1 expression, including age, race, gender, T stage, N stage, M stage, pathological stage, histological type, tumor status, residual tumor, extrathyroidal extension, and thyroid gland disorder history. As a result, we found that patients with high expression of COL26A1 were accompanied by higher age ( $>45$  years,  $P < 0.025$ ), higher pathological stage (stage III,  $P < 0.005$ ), less follicular subtype ( $P < 0.005$ ), more tall cell subtype ( $P < 0.005$ ), wider residual tumor (R1,  $P < 0.05$ ) and extrathyroidal extension ( $P < 0.005$ ). We further included these characteristics to perform logistic regression analysis (Table 2). Age  $>45$  years [odds ratio (OR) =1.532; 95% confidence interval (CI): 1.081–2.176;  $P = 0.017$ ], pathological stage III (OR =2.055; 95% CI: 1.314–3.184;  $P = 0.001$ ), tall cell subtype (OR =5.533; 95% CI: 2.420–14.957;  $P < 0.001$ ), residual tumor R1 (OR =1.844; 95% CI: 1.035–3.365;  $P = 0.041$ ) and extrathyroidal extension (OR =1.800; 95% CI: 1.225–1.660;  $P = 0.003$ ) were determined as risk factors associated with high COL26A1 expression in THCA.

We next processed survival analysis to check out the effect of COL26A1 on prognosis of patients with THCA. Our results revealed that higher expression of COL26A1 inferred poorer OS for patients with THCA, patients with lower COL26A1 expression showed extremely favourable outcome even after 4,000 days (Figure 5A,  $P = 0.015$ ). We subsequently conducted subgroup survival analysis for an in-depth evaluation of the correlation between COL26A1 expression and several clinicopathological factors. It turned out that elevated COL26A1 expression was associated with worse OS in classical subgroup from histological subtype (Figure 5B,  $P = 0.013$ ) and normal subgroup from thyroid gland disorder history (Figure 5C,  $P = 0.013$ ). To further determine the prognostic value of COL26A1 expression, we performed uni/multivariate Cox regression analysis to identify independent prognostic predictors for THCA with  $P < 0.05$  in both univariate and multivariate analysis (Table 3). As a result, higher age ( $P < 0.001$ ), T1 & T2 & T3 stage ( $P < 0.001$ ), pathological stage I & II ( $P < 0.001$ ), with tumor status ( $P < 0.001$ ), residual tumor R1 ( $P = 0.046$ ) and higher COL26A1 expression ( $P < 0.001$ ) were identified as independent prognostic predictors for THCA. Nomogram was able to predict patients' survival probability using



**Figure 3** Expression pattern and diagnostic value of COL26A1. (A) COL26A1 expression levels across 33 cancer types. (B) Significant upregulation of COL26A1 in THCA. (C) COL26A1 expression based on T stage subgroups. (D) COL26A1 expression based on N stage subgroups. (E) COL26A1 expression based on M stage subgroups. (F) COL26A1 expression based on pathological stage subgroups. (G) COL26A1 expression based on histological subgroups. (H) COL26A1 expression based on tumor status subgroups. (I) Diagnostic ROC curve of COL26A1. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; ns, non-significant. COL26A1, collagen type XXVI alpha 1 chain; TPM, transcripts per million; ACC, adrenocortical cancer; BLCA, bladder cancer; BRCA, breast cancer; CESC, cervical cancer; CHOL, bile duct cancer; COAD, colon cancer; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal cancer; GBM, glioblastoma; HNSC, head and neck cancer; KICH, kidney chromophobe; KIRC, kidney clear cell carcinoma; KIRP, kidney papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, lower grade glioma; LIHC, liver hepatocellular cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian cancer; PAAD, pancreatic cancer; PCPG, pheochromocytoma & paraganglioma; PRAD, prostate cancer; READ, rectal cancer; SARC, sarcoma; SKCM, melanoma; STAD, stomach cancer; TGCT, testicular cancer; THCA, thyroid cancer; THYM, thymoma; UCEC, endometrioid cancer; UCS, uterine carcinosarcoma; UVM, ocular melanomas; AUC, area under the curve; CI, confidence interval; FPR, false positive rate; TPR, true positive rate; ROC, receiver operating characteristic.





**Figure 4** IHC results of COL26A1 protein in THCA. (A) IHC result of COL26A1 protein in THCA tissue. Ultra link: <https://www.proteinatlas.org/ENSG00000160963-COL26A1/pathology/thyroid+cancer>. (B) IHC result of COL26A1 protein in normal tissue. Ultra link: <https://www.proteinatlas.org/ENSG00000160963-COL26A1/tissue/thyroid+gland>. (C,D) One highlighted site in THCA tissue and corresponding IOD scanning. Magnification:  $\times 5$ . (E,F) Another highlighted site in THCA tissue and corresponding IOD scanning. Magnification:  $\times 5$ . (G,H) Highlighted site in normal tissue and corresponding IOD scanning. Magnification:  $\times 5$ . (I) Statistical difference between IOD scores in THCA tissue and normal tissue. The red box indicates positive staining of COL26A1, and the red area indicates quantitative scanning by software. \*\*\*,  $P < 0.001$ . IOD, integrated optical density; THCA, thyroid cancer; IHC, immunohistochemistry; COL26A1, collagen type XXVI alpha 1 chain.

integrated data from multiple types of variates, like gene expression level and clinicopathological factors. Therefore, based on results from Cox regression analysis, we constructed a novel nomogram containing six variables (age, pathological T stage, pathological stage, tumor status, residual tumor and COL26A1 expression) to predict the

survival probability of patients with THCA (Figure 5D). Concordance index (C-index) of our nomogram was calculated to be 0.989 (95% CI: 0.983–0.995), exerting robust prediction capability. AUCs from time-dependent ROC curve ranging from 1- to 14-year to verify prediction capability of this nomogram were calculated and displayed

**Table 1** Association between COL26A1 expression and clinicopathological characteristics

Clinicopathological characteristics	Low expression of COL26A1	High expression of COL26A1	$\chi^2$	P value
Age (years)	n=255	n=255		
≤45	134 (52.55%)	107 (41.96%)	5.73	<0.025
>45	121 (47.45%)	148 (58.04%)	5.73	<0.025
Race	n=193	n=222		
Asian	23 (11.92%)	28 (12.61%)	0.05	>0.05
Black	14 (7.25%)	13 (5.86%)	0.33	>0.05
White	155 (80.31%)	181 (81.53%)	0.10	>0.05
American Indian	1 (0.52%)	0	0.005	>0.05
Gender	n=255	n=255		
Female	185 (72.55%)	186 (72.94%)	0.005	>0.05
Male	70 (27.45%)	69 (27.06%)	0.005	>0.05
T stage	n=255	n=253		
T1	69 (27.06%)	74 (29.25%)	0.30	>0.05
T2	93 (36.47%)	74 (29.25%)	3.00	>0.05
T3	81 (31.76%)	94 (37.15%)	1.63	>0.05
T4	12 (4.71%)	11 (4.35%)	0.04	>0.05
N stage	n=223	n=237		
N0	110 (49.33%)	119 (50.21%)	0.01	>0.05
N1	113 (50.67%)	118 (49.79%)	0.01	>0.05
M stage	n=130	n=165		
M0	126 (96.92%)	160 (96.97%)	0.0005	>0.05
M1	4 (3.08%)	5 (3.03%)	0.0005	>0.05
Pathological stage	n=254	n=254		
Stage I	154 (60.63%)	132 (51.97%)	3.87	<0.05
Stage II	26 (10.24%)	26 (10.24%)	0	>0.05
Stage III	41 (16.14%)	72 (28.35%)	10.94	<0.005
Stage IV	33 (12.99%)	24 (9.45%)	1.60	>0.05
Histological type	n=255	n=255		
Classical	176 (69.02%)	188 (73.73%)	1.38	>0.05
Follicular	68 (26.67%)	33 (12.94%)	15.12	<0.005
Tall cell	6 (2.35%)	30 (11.76%)	17.22	<0.005
Other	5 (1.96%)	4 (1.57%)	0	>0.05
Tumor status	n=252	n=246		
With tumor	26 (10.32%)	21 (8.54%)	0.46	>0.05
Tumor free	226 (89.68%)	225 (91.46%)	0.46	>0.05

**Table 1** (continued)

**Table 1** (continued)

Clinicopathological characteristics	Low expression of COL26A1	High expression of COL26A1	$\chi^2$	P value
Residual tumor	n=225	n=223		
R0	203 (90.22%)	187 (83.86%)	4.03	<0.05
R1	20 (8.89%)	34 (15.25%)	4.27	<0.05
R2	2 (0.89%)	2 (0.90%)	0.24	>0.05
Extrathyroidal extension	n=244	n=248		
No	183 (75.00%)	155 (62.50%)	8.37	<0.005
Yes	61 (25.00%)	93 (37.50%)	8.37	<0.005
Thyroid gland disorder history	n=228	n=224		
Lymphocytic thyroiditis	40 (17.54%)	34 (15.18%)	0.46	>0.05
Nodular hyperplasia	35 (15.35%)	33 (14.73%)	0.03	>0.05
Normal	138 (60.53%)	147 (65.63%)	1.26	>0.05
Other	15 (6.58%)	10 (4.46%)	0.97	>0.05

COL26A1, collagen type XXVI alpha 1 chain.

**Table 2** Logistic regression analysis of COL26A1 in THCA

Clinicopathological characteristics	Total (N)	OR (95% CI)	P value
Age (>45 vs. ≤45 years)	510 (269 vs. 241)	1.532 (1.081–2.176)	0.017
Pathological stage (stage III vs. stage I & II & IV)	508 (113 vs. 395)	2.055 (1.314–3.184)	0.001
Histological type (tall cell vs. classical & follicular & other)	510 (36 vs. 474)	5.533 (2.420–14.957)	<0.001
Residual tumor (R1 vs. R0 & R2)	448 (54 vs. 394)	1.844 (1.035–3.365)	0.041
Extrathyroidal extension (yes vs. no)	492 (154 vs. 338)	1.800 (1.225–1.660)	0.003

COL26A1, collagen type XXVI alpha 1 chain; THCA, thyroid carcinoma; OR, odds ratio; CI, confidence interval.

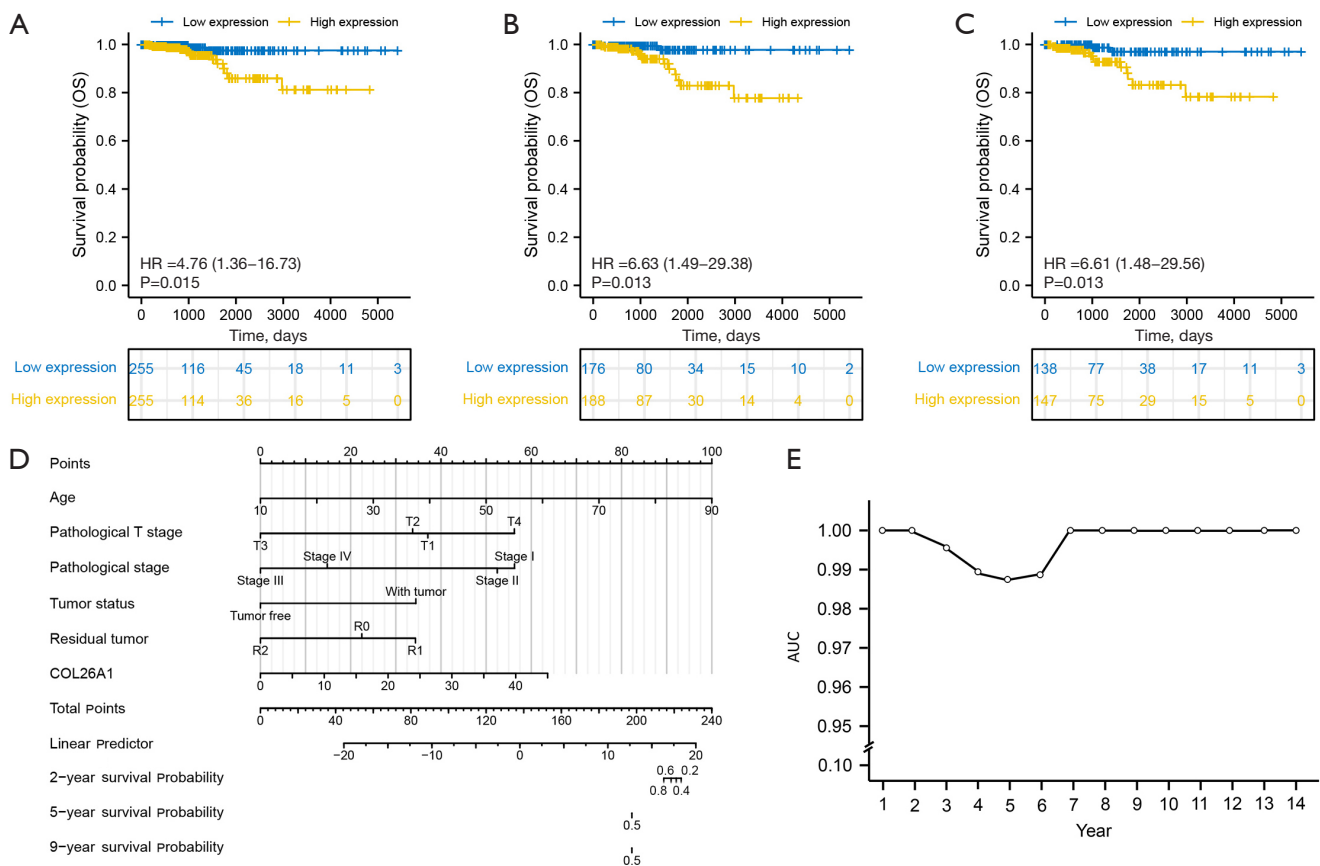
(Figure 5E).

### COL26A1 was involved in immune-related pathways in THCA

We investigated the potential correlation between COL26A1 DNA methylation and tumorigenesis in THCA via MEXPRESS. We observed significant differences in several probes (Figure 6A). In promoter region, significantly positive correlation was found in two probes: cg25290307 ( $r=1.190$ ,  $P<0.001$ ) and cg17361163 ( $r=0.115$ ,  $P<0.01$ ), while significantly negative correlation was found in one probe: cg18762727 ( $r=-0.087$ ,  $P<0.05$ ). In non-promoter region, significantly positive correlation was found in six probes: cg06054793 ( $r=0.149$ ,  $P<0.001$ ), cg07899060 ( $r=0.289$ ,  $P<0.001$ ), cg01380946 ( $r=0.102$ ,  $P<0.05$ ), cg06400590

( $r=0.293$ ,  $P<0.001$ ), cg17022667 ( $r=0.153$ ,  $P<0.001$ ), cg23657388 ( $r=0.192$ ,  $P<0.001$ ), while significantly negative correlation was found in four probes: cg14767790 ( $r=-0.092$ ,  $P<0.05$ ), cg27109461 ( $r=-0.376$ ,  $P<0.001$ ), cg15127702 ( $r=-0.181$ ,  $P<0.001$ ), cg29223895 ( $r=-0.250$ ,  $P<0.001$ ).

Limited to scarce literature of COL26A1 in human diseases, especially in cancers, the molecular mechanisms involved in COL26A1 functioning still remained obscure. Thus, we tried to primarily explore the functions that COL26A1 may be possibly annotated in tumorigenicity. Firstly, a COL26A1-centered PPI network was established via STRING database (Figure 6B). Then we managed to determine the DEGs (adjusted  $P\leq 0.05$  and  $|\text{Log}_2\text{FoldChange}| > 1$ ) between high and low COL26A1 expression subgroups and display these DEGs as a volcano plot (Figure 6C). Top fifty positive DEGs ( $\text{Log}_2\text{FoldChange}$



**Figure 5** Survival analysis of COL26A1 in THCA. (A) Correlation between COL26A1 expression and OS. (B) Correlation between COL26A1 expression and OS in classical subgroup from histological subtype. (C) Correlation between COL26A1 expression and OS in normal subgroup from thyroid gland disorder history. (D) Nomogram to predict survival probability of patients with THCA. (E) AUCs from time-dependent ROC curve of the nomogram. HR, hazard ratio; OS, overall survival; COL26A1, collagen type XXVI alpha 1 chain; AUC, area under the curve; THCA, thyroid cancer; ROC, receiver operating characteristic.

>1) in high COL26A1 subgroup were then thrown into GO/KEGG functional enrichment analysis. The most enriched functional annotations of COL26A1 including GO terms: biological process (BP), cellular component (CC), molecular function (MF) and KEGG terms were displayed (Figure 6D). These DEGs were mainly found to participate in cornification, humoral immune response, neutrophil chemotaxis, intermediate filament, keratin filament, protein-lipid complex and endopeptidase regulatory activity. No meaningful KEGG terms were determined. In terms of GSEA functional enrichment analysis of COL26A1, we selected top 10 statistically significant terms to display (Figure 6E,6F). Keratan sulfate degradation was checked out to be the most enriched functional annotation of COL26A1. Other functional annotations of COL26A1 were mostly immune-related,

such as LAIR pathway, cells and molecules involved in local acute inflammatory response, granulocyte pathway, CTLA4 pathway, interleukin (IL)-10 signaling pathway and chemokine receptors bind chemokines etc.

**Mutation analysis**

With the aid of cBioPortal database, we explored the genetic alteration status of COL26A1 across different cancer types from TCGA. Alteration frequency of COL26A1 in THCA (500 samples) was determined to be 0.2%, which indicated that genetic alteration was detected in only one THCA sample and its mutation component was deep deletion (Figure 7A). We also observed that amplification is the most common mutation component of COL26A1 across different cancer types. Detailed information of

**Table 3** Uni/multivariate Cox regression analysis of clinicopathological characteristics and COL26A1 expression

Clinicopathological characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value
Age	510	1.120 (1.078–1.164)	<0.001	1.330 (1.242–1.425)	<0.001
Race	415		0.967		
Black	27	Reference			
White	336	0.678 (0.153–2.999)	0.608	–	
Asian	51	0.000 (0.000–Inf)	0.998	–	
American Indian	1	0.000 (0.000–Inf)	1.000	–	
Gender	510		0.193		
Female	371	Reference			
Male	139	1.963 (0.710–5.428)	0.193	–	
Pathological T stage	508		<0.001		
T4	23	Reference			
T2	167	0.089 (0.022–0.359)	<0.001	0.000 (0.000–0.000)	<0.001
T1	143	0.087 (0.017–0.434)	0.003	0.011 (0.001–0.077)	<0.001
T3	175	0.139 (0.042–0.458)	0.001	0.000 (0.000–0.000)	<0.001
Pathological N stage	510		0.548		
N1	231	Reference			
N0	229	0.693 (0.226–2.123)	0.521	–	
NX	50	1.539 (0.407–5.818)	0.525	–	
Pathological M stage	509		0.083		
M0	286	Reference			
MX	214	0.691 (0.230–2.074)	0.510	0.973 (0.094–10.018)	0.981
M1	9	4.419 (0.948–20.605)	0.059	0.000 (0.000–Inf)	0.994
Pathological stage	508		0.005		
Stage IV	57	Reference			
Stage II	52	0.287 (0.054–1.519)	0.142	Inf (Inf–Inf)	<0.001
Stage I	286	0.053 (0.010–0.278)	<0.001	Inf (Inf–Inf)	<0.001
Stage III	113	0.519 (0.162–1.664)	0.270	0.288 (0.027–3.119)	0.306
Histological type	510		0.695		
Classical	364	Reference			
Follicular	101	0.289 (0.038–2.188)	0.229	–	
Tall cell	36	0.000 (0.000–Inf)	0.997	–	
Other	9	0.000 (0.000–Inf)	0.999	–	
Tumor status	498		<0.001		
Tumor free	451	Reference			
With tumor	47	24.280 (6.273–93.970)	<0.001	Inf (Inf–Inf)	<0.001

Table 3 (continued)

Table 3 (continued)

Clinicopathological characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value
Extrathyroidal extension	492		0.109		
Yes	154	Reference			
No	338	0.445 (0.165–1.199)	0.109	–	
Residual tumor	448		0.075		
R0	390	Reference			
R1	54	4.033 (1.214–13.402)	0.023	12.159 (1.045–141.460)	0.046
R2	4	0.000 (0.000–Inf)	0.998	0.000 (0.000–Inf)	0.997
Thyroid gland disorder history	452		0.954		
Other	25	Reference			
Nodular hyperplasia	68	Inf (Inf–Inf)	0.999	–	
Lymphocytic thyroiditis	74	0.995 (0.000–Inf)	1.000	–	
Normal	285	Inf (Inf–Inf)	0.999	–	
COL26A1	510	1.055 (1.007–1.106)	0.025	3.928 (3.716–4.151)	<0.001

COL26A1, collagen type XXVI alpha 1 chain; HR, hazard ratio; CI, confidence interval; Inf, infinite.

mutation types and frequencies of COL26A1 were shown (Figure 7B). Missense mutation of COL26A1 was the most common form (73/101, 72.28%). OS between COL26A1 altered and unaltered subgroups showed no statistical significance (Figure 7C,  $P>0.05$ ).

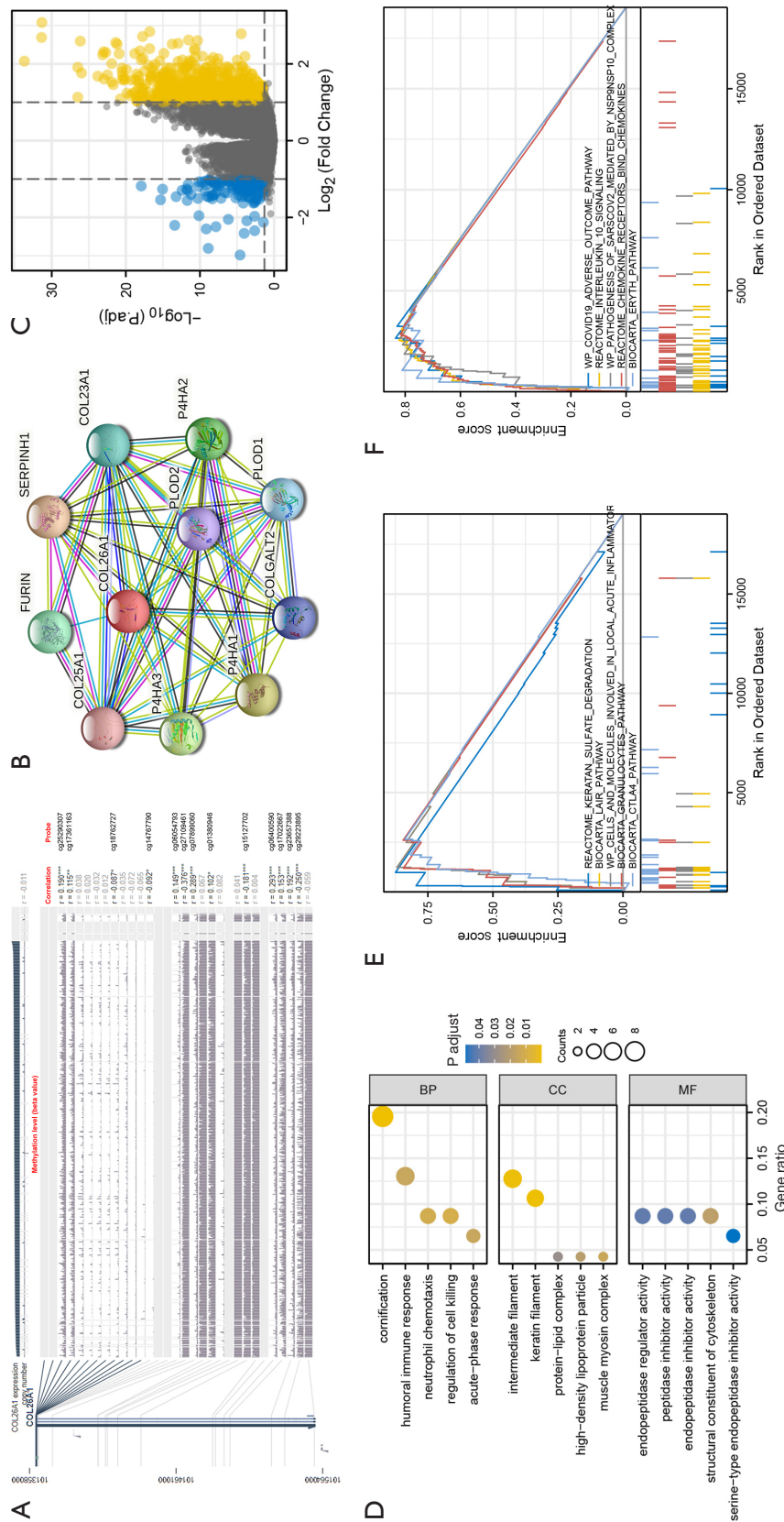
#### ***COL26A1 strongly correlated with immune infiltrating cells and immune checkpoints***

We have primarily shed a light on the potential roles of COL26A1 regarding immune via functional enrichment analysis above. Next, we tried to further explore the potential associations between COL26A1 and tumor immune microenvironment (TIME), as well as immune-related genes. Firstly, we examined the infiltration levels of 24 immune cells based on high and low COL26A1 expression subgroups (Figure 8A). More abundant infiltration levels of 19 immune cells (aDC, B cells, CD 8 T cells, cytotoxic cells and mast cells etc.) were identified in high COL26A1 expression subgroup, while the other five immune cells did not (pDC, NK cell, Th17, Tgd, and Tem, all  $P>0.05$ ). Results from ESTIMATE algorithm revealed that patients with higher COL26A1 expression are accompanied with higher stromal score and immune score (Figure 8B). The exact correlation between COL26A1

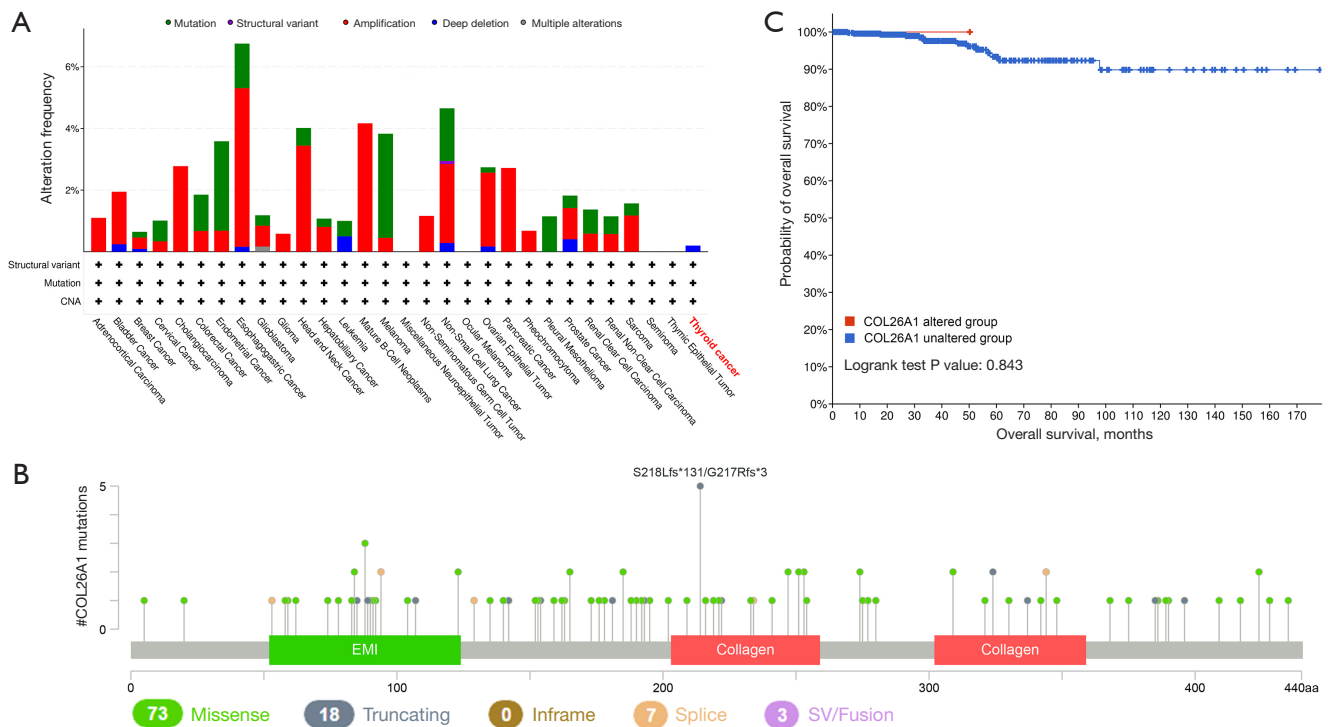
expression and infiltration levels of 24 immune cells in THCA was assessed and visualized (Figure 8C). A majority of immune cells' infiltration levels were significantly positively correlated with COL26A1 expression, while four immune cells did not (Th17 cell, NK cell, Tgd and pDC, all  $P>0.05$ ). Likewise, COL26A1 expression was significantly positively correlated with stromal score and immune score, respectively (Figure 8D). To further evaluate the application potential of COL26A1 as an immunological target, we carried out co-expression analysis between COL26A1 expression and 47 immune checkpoints to check out their correlation, results of which were visualized (Figure 8E,8F). The results revealed that a total of 43 immune checkpoints are significantly positively correlated with high COL26A1 expression, including putative CTLA4 ( $r=0.36$ ,  $P<0.001$ ), LAG3 ( $r=0.22$ ,  $P<0.001$ ), PD-1 ( $r=0.21$ ,  $P<0.001$ ), PD-L1 ( $r=0.15$ ,  $P<0.001$ ) and TIGIT ( $r=0.33$ ,  $P<0.001$ ). This may possibly indicate COL26A1 as a potential immunological target combined with immune checkpoints blockade.

#### ***Multiple immune-related genes correlated with COL26A1 were identified as prognostic factors in THCA***

Subsequently, we processed co-expression analysis between



**Figure 6** DNA methylation and functional enrichment analysis of COL26A1 in THCA. (A) Correlation between COL26A1 DNA methylation and THCA expressed as beta value, Pearson correlation coefficient (r) and Benjamini-Hochberg adjusted P value. (B) COL26A1-centered PPI network. (C) Volcano plot to show DEGs between high and low COL26A1 expression subgroups. Blue dots represent DEGs that are significantly upregulated in low COL26A1 expression subgroup, yellow dots represent DEGs that are significantly upregulated in high COL26A1 expression subgroup and black dots represent genes that are not differentially expressed. (D) GO/KEGG functional enrichment analysis of COL26A1. (E,F) GSEA functional enrichment analysis of COL26A1. \*, P<0.05; \*\*, P<0.01; \*\*\*, P<0.001. BP, biological process; CC, cellular component; MF, molecular function; COL26A1, collagen type XXVI alpha 1 chain; THCA, thyroid cancer; PPI, protein-protein interaction; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene Set Enrichment Analysis.

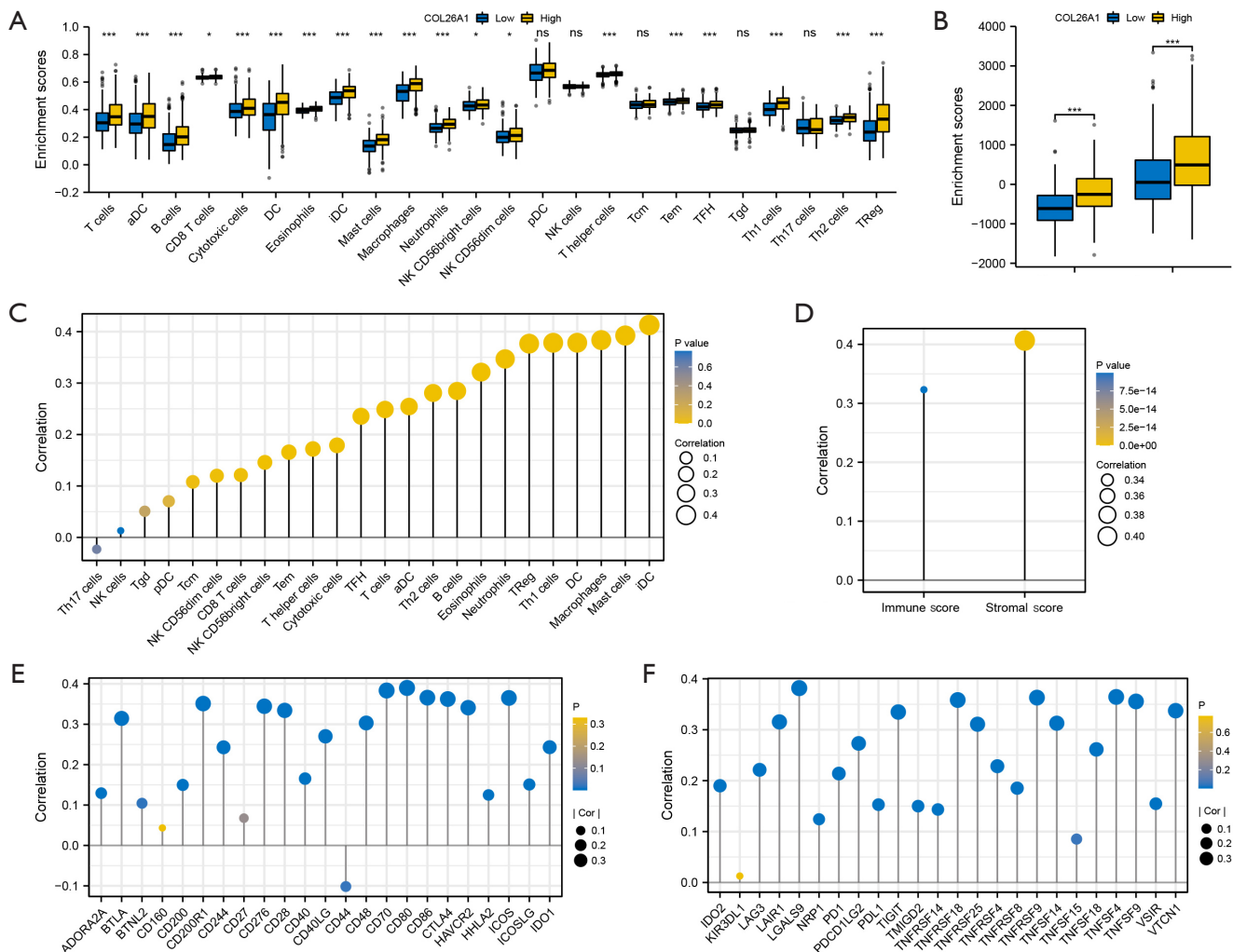


**Figure 7** Mutation features of COL26A1 across different cancer types. (A) Mutation frequency and mutation types of COL26A1. (B) Detailed information of mutation frequency and mutation types of COL26A1. (C) Effect of COL26A1 alteration on OS for patients with THCA. COL26A1, collagen type XXVI alpha 1 chain; OS, overall survival; THCA, thyroid cancer; CNA, copy number alteration; EMI, Elastin Microfibril Interface; SV, structural variant.

COL26A1 and five immune-related gene clusters to further identify risk genes correlated with COL26A1 expression, which were major histocompatibility complex (MHC) genes, chemokine genes, chemokine receptor genes, immune activation genes and immune suppression genes. For MHC genes, their correlations with COL26A1 expression were displayed (Figure 9A). All the MHC genes were significantly positively correlated with high COL26A1 expression. We then conducted LASSO regression analysis using the genes with  $P \leq 0.05$  and screened three most correlated risk genes out (TAPBP, B2M and HLA-E), as well as corresponding OS curve based on high and low risk score group (Figure 9B-9D). Patients in high risk score group suffering poorer prognosis were correlated with lower expression levels of TAPBP, B2M and HLA-E. For chemokine genes, their correlations with COL26A1 expression were displayed (Figure 9E). Most chemokine genes were significantly positively correlated with high COL26A1 expression. We then conducted LASSO regression analysis using the genes with  $P \leq 0.05$  and screened 15 most correlated risk genes out (CCL3, CCL5,

CCL11, CCL13, CCL17, CCL20, CCL25, CX3CL1, CXCL1, CXCL2, CXCL3, CXCL5, CXCL9, CXCL12 and CXCL14), as well as corresponding OS curve based on high and low risk score group (Figure 9F-9H). Patients in high risk score group suffering poorer prognosis were correlated with lower expression levels of CCL3, CCL5, CCL11, CCL13, CCL20, CCL25, CX3CL1, CXCL2, CXCL9 and CXCL14. For immune activation genes, their correlations with COL26A1 expression were displayed (Figure 9I). Most immune activation genes were significantly positively correlated with high COL26A1 expression. We then conducted LASSO regression analysis using the genes with  $P \leq 0.05$  and screened six most correlated risk genes out (CD40LG, CD70, CD80, ENTPD1, HHLA2 and TMIGD2), as well as corresponding OS curve based on high and low risk score group (Figure 9J-9L). Patients in high risk score group suffering poorer prognosis were correlated with lower expression levels of CD40LG, CD70, ENTPD1 and TMIGD2 and higher expression level of CD80. No correlated risk genes from chemokine receptor





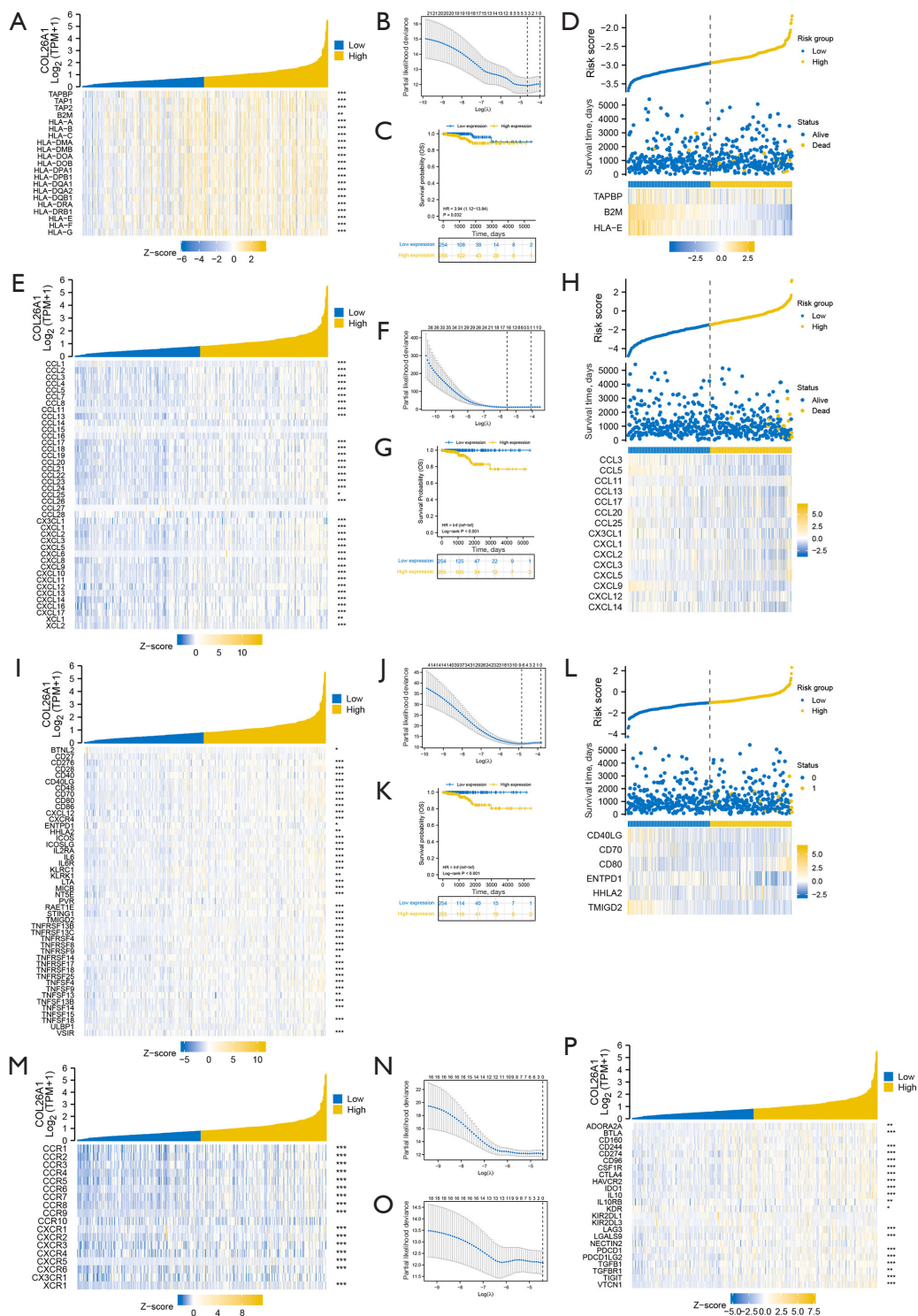
**Figure 8** Correlation of COL26A1 expression with immune cells, stromal/immune score and immune checkpoints in THCA. (A) Differential infiltration levels of immune cells between high and low COL26A1 expression subgroups. (B) Differential stromal score and immune score between high and low COL26A1 expression subgroups. (C) Correlation between COL26A1 expression and infiltration levels of immune cells. (D) Correlation between COL26A1 expression and stromal/immune score. (E,F) Correlation between COL26A1 expression and 47 immune checkpoints. \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ ; ns, non-significant. COL26A1, collagen type XXVI alpha 1 chain; THCA, thyroid cancer.

genes (Figure 9M,9N) and immune suppression genes (Figure 9O,9P) were filtered.

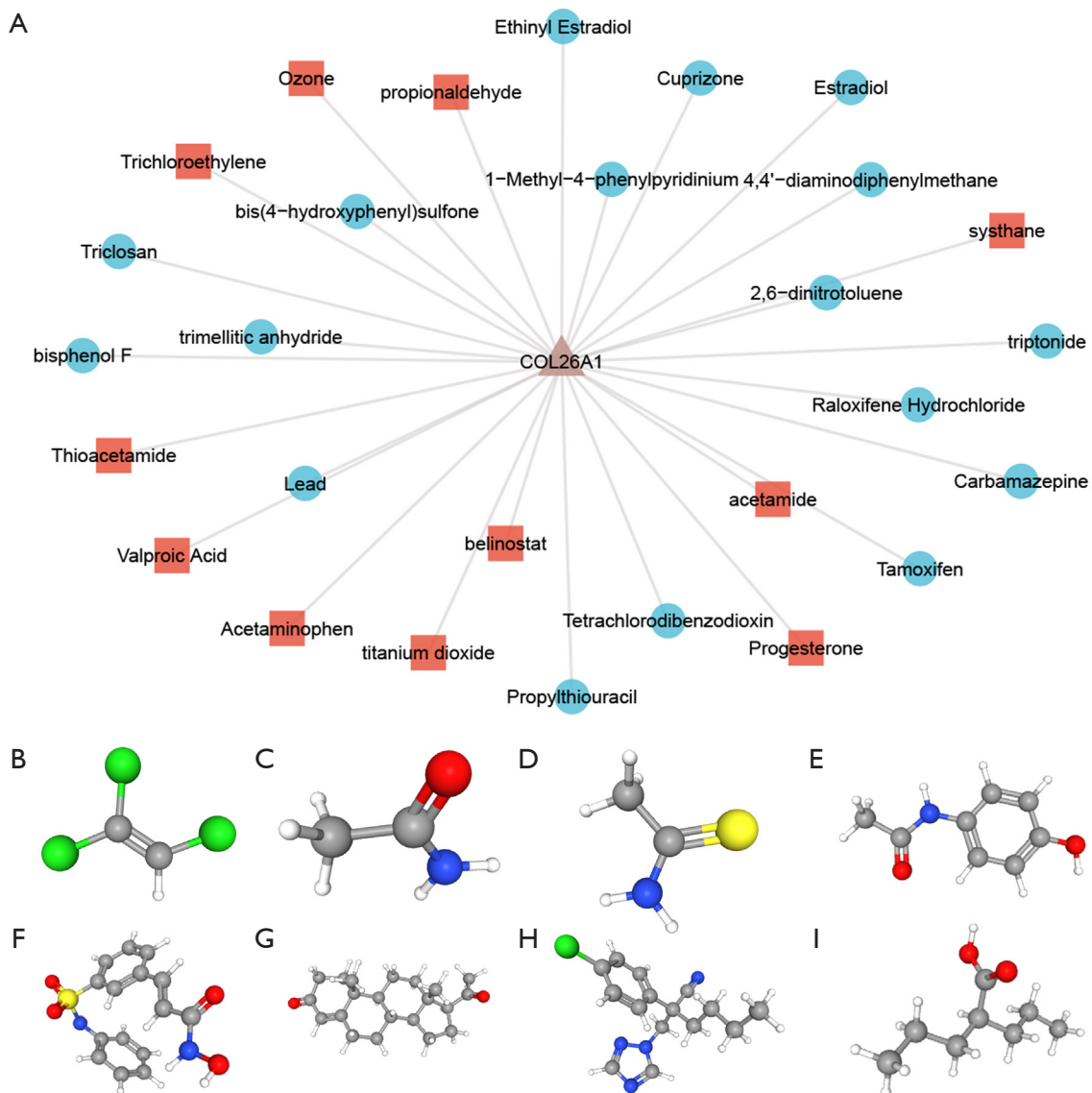
#### Several drugs were determined to downregulate the expression of COL26A1

Based on this kind of idea to exploit COL26A1 as a combined target with immune checkpoints blockade for prognosis improvement, we then determined the drugs that

can interact with COL26A1 mRNA to affect its expression via CTD database. A total of 28 drugs were included, among which eleven drugs can decrease COL26A1 mRNA expression, 13 drugs can increase COL26A1 mRNA expression and four drugs can affect COL26A1 mRNA expression with ambiguous effect. The interactions between these drugs and COL26A1 were illustrated (Figure 10A). Besides, molecular structures of the drugs that can decrease COL26A1 mRNA expression were displayed via PubChem



**Figure 9** Correlation of COL26A1 expression with five immune-related gene clusters in THCA. (A) Correlation between COL26A1 expression and MHC genes. (B) LASSO variables screening. (C) OS between high and low risk score groups. (D) Risk genes from MHC genes correlated with COL26A1 expression. (E) Correlation between COL26A1 expression and chemokine genes. (F) LASSO variables screening. (G) OS between high and low risk score groups. (H) Risk genes from chemokine genes correlated with COL26A1 expression. (I) Correlation between COL26A1 expression and immune activation genes. (J) LASSO variables screening. (K) OS between high and low risk score groups. (L) Risk genes from immune activation genes correlated with COL26A1 expression. (M) Correlation between COL26A1 expression and chemokine receptor genes. (N) LASSO variables screening of chemokine receptor genes. (O) LASSO variables screening of immune suppression genes. (P) Correlation between COL26A1 expression and immune suppression genes. \*, P<0.05; \*\*, P<0.01; \*\*\*, P<0.001. COL26A1, collagen type XXVI alpha 1 chain; TPM, transcripts per million; OS, overall survival; HR, hazard ratio; Inf, infinite; THCA, thyroid cancer; MHC, major histocompatibility complex; LASSO, least absolute shrinkage and selection operator.



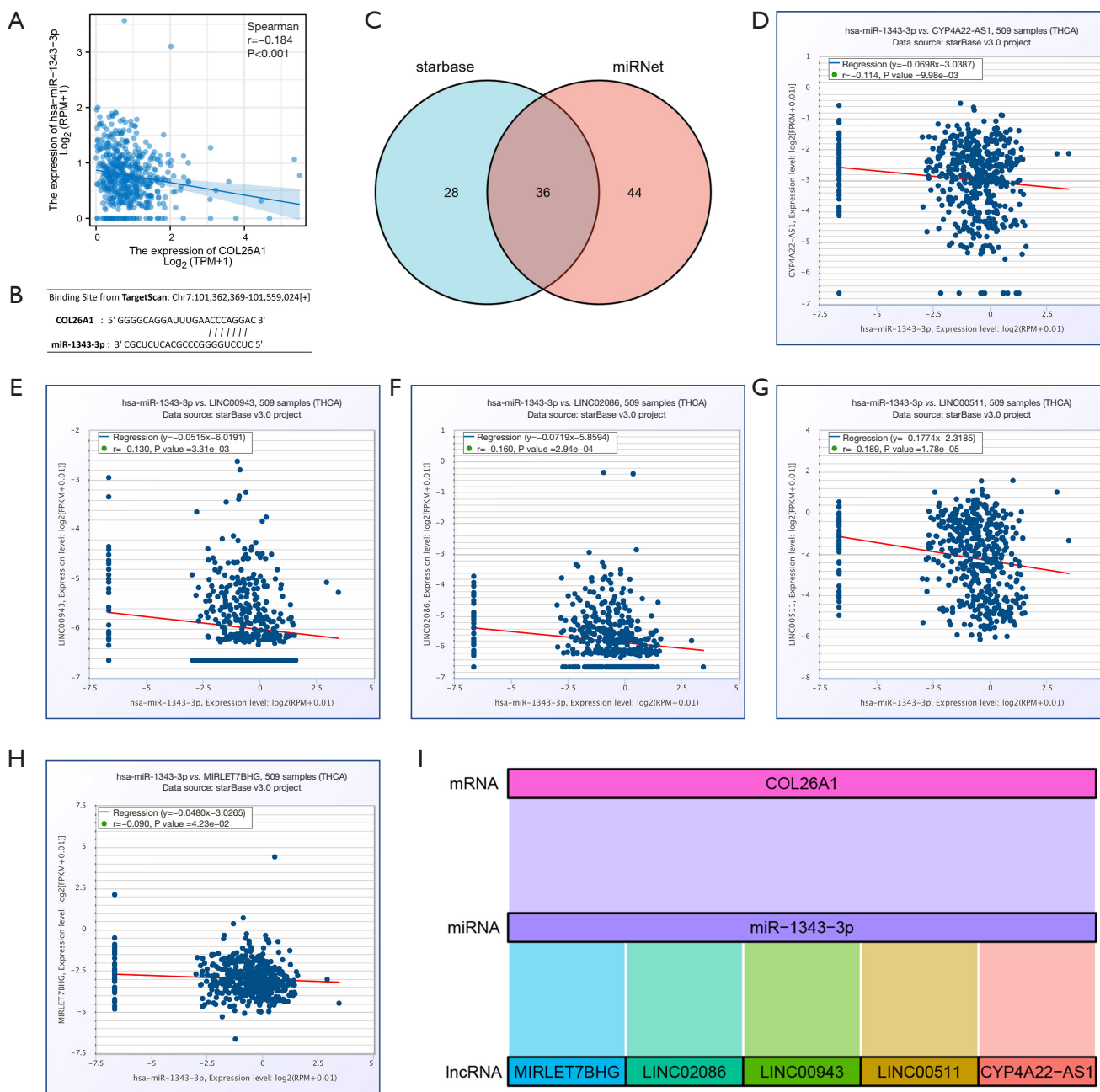
**Figure 10** Drugs that can interact with COL26A1. (A) Drugs that can affect the expression of COL26A1 mRNA. Red ones represent drugs that can decrease COL26A1 mRNA expression. Blue ones represent drugs that can increase COL26A1 mRNA expression. (B) Trichloroethylene. (C) Acetamide. (D) Thioacetamide. (E) Acetaminophen. (F) Belinostat. (G) Progesterone. (H) Myclobutanil. (I) Valproic acid. COL26A1, collagen type XXVI alpha 1 chain.

database (Figure 10B-10I).

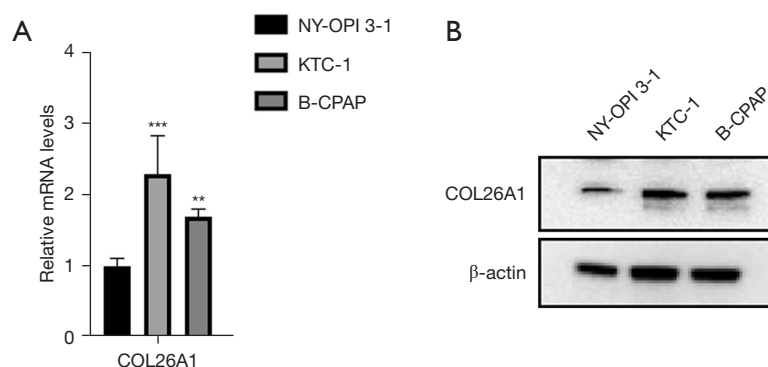
#### ceRNA network centered on COL26A1

lncRNA-miRNA-mRNA regulating axis is a kind of relatively mature regulatory mechanism in cellular signal transduction. Exploitation of ceRNA network will help to further understand the regulatory network of COL26A1 in THCA. We used online website miRNet to predict upstream

miRNAs that target COL26A1 mRNA. A total of four miRNAs were predicted to be able to bind with COL26A1 mRNA, but only miR-1343-3p was filtered as the candidate upstream miRNA via correlation analysis (Figure 11A,  $r=-0.184$ ,  $P<0.001$ ). Then we acquired the predicted binding site information between COL26A1 mRNA and miR-1343-3p from TargetScan (Figure 11B). Based on starBase and miRNet, we further identified upstream lncRNAs that target miR-1343-3p in THCA. A total of 64 and 80 lncRNAs were



**Figure 11** Construction of a COL26A1-related ceRNA network. (A) Correlation analysis between COL26A1 and miR-1343-3p. (B) The potential binding site between COL26A1 and miR-1343-3p. (C) lncRNA prediction intersection between starBase and miRNet. (D) Correlation between CYP4A22-AS1 and miR-1343-3p. (E) Correlation between LINC00943 and miR-1343-3p. (F) Correlation between LINC02086 and miR-1343-3p. (G) Correlation between LINC00511 and miR-1343-3p. (H) Correlation between MIRLET7BHG and miR-1343-3p. (I) lncRNA-miR-1343-3p-COL26A1 regulatory axes. COL26A1, collagen type XXVI alpha 1 chain; TPM, transcripts per million; RPM, reads per million; THCA, thyroid cancer; FPKM, fragments per kilobase of exon model per million; lncRNA, long noncoding RNA; miRNA, microRNA; ceRNA, competing endogenous RNA.



**Figure 12** Verification of COL26A1 expression in THCA. (A) qRT-PCR. (B) Western blot. The original data can be found in the Figure S1. \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . COL26A1, collagen type XXVI alpha 1 chain; THCA, thyroid cancer; qRT-PCR, quantitative real-time polymerase chain reaction.

predicted to target miR-1343-3p by starBase and miRNet, respectively, 36 lncRNAs were commonly predicted by the two websites (Figure 11C). We next managed to determine five candidate lncRNAs from the 36 lncRNAs by starBase, which were significantly negatively correlated with miR-1343-3p, they were CYP4A22-AS1 ( $r = -0.114$ ,  $P = 9.98E-03$ ), LINC00943 ( $r = -0.130$ ,  $P = 3.31E-03$ ), LINC02086 ( $r = -0.160$ ,  $P = 2.94E-04$ ), LINC00511 ( $r = -0.189$ ,  $P = 1.78E-05$ ) and MIRLET7BHG ( $r = -0.090$ ,  $P = 4.23E-02$ ) (Figure 11D-11H). Ultimately, we were able to construct the ceRNA network to vividly demonstrate the potential regulatory axes in THCA (Figure 11I).

### COL26A1 was significantly upregulated in THCA

qRT-PCR revealed that the mRNA expression levels of COL26A1 are significantly higher in THCA cells compared with normal thyroid epithelial cells (Figure 12A). Western blot also verified that the protein levels of COL26A1 are elevated in THCA cells compared with normal thyroid epithelial cells (Figure 12B).

## Discussion

Roles of COL26A1 remain scarcely demonstrated in addition to its role as a member of collagen family, especially under the background of cancer. Until now, we merely acquired that COL26A1 is associated with acute ethmoiditis, asthma, nasal polyps, aspirin intolerance and placental growth (5,7,8). Tons of work await being accomplished to further understand the influence of COL26A1 exerting on human diseases. Therefore, we

tried to primarily explore the potential roles of COL26A1 in THCA, attempting to lay a preliminary foundation for subsequent researches and experimental verification.

It is understandable for COL26A1 as a collagen subtype to be located in ECM, plasma membrane and endoplasmic reticulum for constitutive support. But other unknown functions need further research. Different tumor samples were not always accompanied by high COL26A1 expression, only tumor samples from six cancer types showed significantly high COL26A1 expression compared to the normal, including THCA. We inferred that it is pathological stage that upregulation of COL26A1 may propel in THCA, because significantly elevated COL26A1 expression is observed in pathological stage III subgroup compared to pathological stage I or stage II subgroup. This indicated that examination of elevated COL26A1 expression in THCA tissues may contribute to assess early stage progression. Quantified IOD scores of IHC results showed significant higher COL26A1 protein expression in THCA tissues compared to the normal, which can be taken as verification of COL26A1 expression in protein level. High COL26A1 expression infers poorer OS for patients with THCA. However, there is no corresponding previous study for reference. Clinical correlation analysis (both baseline analysis and logistic regression analysis) indicated the correlation of higher COL26A1 expression with higher age (>45 years), higher pathological stage (stage III), tall cell subtype, wider residual tumor (R1) and extrathyroidal extension. Thus, it was possible that patients under clinical conditions above may have relatively unsatisfying prognosis. Patients with higher COL26A1 expression in histological subtype-classical subgroup or thyroid gland

disorder history-normal subgroup may also suffer poorer OS according to our subgroup survival analysis. More and larger clinical cohort is required for testing though.

Higher age ( $P < 0.001$ ), T1 & T2 & T3 stage ( $P < 0.001$ ), pathological stage I & II ( $P < 0.001$ ), with tumor status ( $P < 0.001$ ), residual tumor R1 ( $P = 0.046$ ) and higher COL26A1 expression ( $P < 0.001$ ) were identified as independent prognostic factors for THCA via uni/multivariate Cox regression analysis, whereby a novel nomogram was successfully established to predict the survival probability for patients with THCA. Perhaps due to the low rate of THCA-caused death events itself, the C-index (0.989; 95% CI: 0.983–0.995) and AUCs (1- to 14-year) of the nomogram were extremely ideal, validation from a larger or clinical cohort may provide better verification. In conclusion, work above have primarily determined COL26A1 as a novel and reliable biomarker to predict poor prognosis for patients with THCA with a certain extent of accuracy via bioinformatics analysis.

No previous studies have reported cellular functional characterization of COL26A1. A series of functional enrichment analysis processed by us via GO/KEGG/GSEA primarily determined several functional annotations of COL26A1 in THCA. As a member of collagen, it was predictable for COL26A1 to participate in cornification, intermediate filament and keratin filament etc. Some other novel functional annotations were also labelled to COL26A1: inflammatory processes including humoral immune response, regulation of cell killing and acute phase response; cell movement including neutrophil chemotaxis and muscle myosin complex; others including protein-lipid complex and endopeptidase activity etc. More worthwhile to mention is that, the top 10 GSEA functional annotations may further contribute to the potential immunological role of COL26A1 in THCA. Most of them were immune-related and could be mainly concluded as: cytokines (CKs) and inflammatory response including cells and molecules involved in local acute inflammatory response and IL-10 signaling pathway; chemokine-related including chemokine receptors bind chemokines; immunological processes including LAIR pathway, granulocytes pathway and CTLA4 pathway. ECM is a highly dynamic environment according to extracellular signals and interactions with cells in it (21). As a major extracellular protein, dynamics of collagen itself is originally chemotactic for cancer cells (22,23). Thus, holistically viewing these functional annotations, on the one hand, we primarily inferred that extracellular COL26A1 protein may also render chemotactic for immune cells and THCA cells in cancer development and progression and

this process may be mediated by CKs including ILs and chemokines. On the other hand, we inferred that constitutive engagement in cellular signaling pathways is another critical functional characterization of COL26A1, possibly LAIR signaling pathway, CTLA signaling pathway and IL-10 signaling pathway etc. It was certain that tons of experiments are required to be done to verify our inferences concerning COL26A1's functional characterizations. From another perspective, a majority of CKs and signaling pathways mentioned above have been identified to tightly correlate with human cancers (24-31), which may help to further evidence the potential associations between COL26A1 and tumorigenesis. In addition, Yadav *et al.* (32) reported that microtubule (one type of cytoskeleton) polymerization can act as downstream of chemokines (CXCL1 and fMLP) to propel chemotaxis and invasion of neutrophils. According to this observation plus cell movement-related functional annotations of COL26A1 (structural constituent of cytoskeleton and muscle myosin complex), we hypothesize that COL26A1 may also promote cell movement of cancer cells contributing to invasion and metastasis triggered by chemokines, which requires more experimental evidences.

TIME has been proven to deeply regulate cancer progression and immunotherapy response (33-35). As part of ECM, phenotypes of TIME can be affected by collagen via manipulating immune cell infiltration levels (4,5,36). Herein, we tried to determine the potential associations between COL26A1 and TIME in THCA, attempting to provide more probabilities for immunotherapy. We observed that COL26A1 expression is significantly positively correlated with most immune cells' infiltration levels, and high COL26A1 expression is commonly accompanied by more abundant infiltration levels of immune cells in THCA. COL26A1 expression was also significantly positively correlated with stromal score and immune score, high COL26A1 expression was accompanied by higher stromal score and immune score. Furthermore, co-expression analysis revealed that a majority of immune checkpoints are significantly positively correlated with high COL26A1 expression in THCA. Mariathasan *et al.* (37) identified exclusion of CD8<sup>+</sup> T cells in peritumoural stroma from patients with metastatic urothelial cancer that showed tolerance to PD-L1 blockade. Peng *et al.* (5) reported that PD-1/PD-L1 blockade resistance in both murine and human lung tumors are resulted from decreased total CD8<sup>+</sup> T cells and increased exhausted CD8<sup>+</sup> T cell subpopulations, and reduction of tumor collagen deposition can increase infiltration levels of T cells. This indicated that decrease of collagen in ECM may contribute to restore or elevate

the infiltration levels of T cells, especially CD8<sup>+</sup> T cells to promote immunotherapy response. Therefore, it is possibly reasonable to combine COL26A1 targeting and immune checkpoints blockade as a novel strategy to improve the OS for patients with THCA in advanced period. In addition, our co-expression analysis between COL26A1 and five immune-related gene clusters further evidenced the immunological intimacy of COL26A1 and strongly suggested its application value as a novel immunological biomarker for the development of immunosuppressants. Using LASSO regression analysis, we screened out several immune-related genes as potential risk genes with significant correlation with COL26A1.

Combined targeting for both cancer cells and TIME/ECM is becoming an emerging strategy for cancer therapy (38). In terms of collagen's relatively conserved structure, it is more feasible for targeted drugs to be commonly applied in multiple solid tumors. Several agents utilized to target collagen for cancer therapy have entered clinical trials (39). We identified eleven molecules that are capable to downregulate the expression levels of COL26A1 mRNA via CTD database for the first time. Structures of the eight drugs among the eleven molecules are displayed via PubChem website. This has offered potential candidate drugs for further researches.

lncRNA-miRNA-mRNA axis is an important regulatory method in cellular signal transduction. Exploration of COL26A1-related ceRNA network will contribute to better understanding of the cellular roles of COL26A1. We have identified miR-1343-3p as the single upstream miRNA that targets COL26A1 mRNA in THCA. Five different lncRNAs were identified to target miR-1343-3p in THCA, they were CYP4A22-AS1, LINC00943, LINC02086, LINC00511 and MIRLET7BHG. Thus, a total of five lncRNA-miRNA-mRNA targeting flows were determined in THCA, which commonly shared the same exit. More experimental evidence is required for further validation of these possible targeting relationships.

Certainly, there are several limitations awaiting further promotion based on the present study. Firstly, verification of COL26A1 expression in THCA using clinical specimen will be more convincing. Secondly, the predicted five lncRNA-miRNA-COL26A1 regulatory axes require more experimental evidence. Thirdly, applying COL26A1 for prognosis prediction in a large, prospective real-world trial may further confirm its clinical value.

## Conclusions

In this present study, the potential roles of COL26A1 in

THCA have been primarily explored for the first time, which mainly attributes to four aspects: prognostic/diagnostic prediction, functional characterization, immunological/therapeutic target and ceRNA network. The aberrant expression of COL26A1 in THCA has been verified by qRT-PCR and western blot. Our work has appraised COL26A1 as a promising biomarker for diagnosis/prognosis and immuno/therapeutic target in THCA. The potential mechanism of COL26A1 and TIME and relevant immunotherapy can be the focus of future research.

## Acknowledgments

We would like to show gratitude to the preprint website Research Square ([www.researchsquare.com](http://www.researchsquare.com)) for its online display of our work. Besides, we want to thank Xiantao platform ([www.xiantao.love](http://www.xiantao.love)) for integrated R analysis. Xinjiang Key Laboratory of Molecular Biology for Endemic Diseases also provided kind support.

*Funding:* None.

## Footnote

*Reporting Checklist:* The authors have completed the MDAR reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-141/rc>

*Peer Review File:* Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-141/prf>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-141/coif>). The authors have no conflicts of interest to declare.

*Ethical Statement:* 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. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the

original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
- Gambardella C, Offi C, Patrone R, et al. Calcitonin negative Medullary Thyroid Carcinoma: a challenging diagnosis or a medical dilemma? *BMC Endocr Disord* 2019;19:45.
- Wang W, Shen C, Zhao Y, et al. Identification and validation of potential novel biomarkers to predict distant metastasis in differentiated thyroid cancer. *Ann Transl Med* 2021;9:1053.
- Kaur A, Ecker BL, Douglass SM, et al. Remodeling of the Collagen Matrix in Aging Skin Promotes Melanoma Metastasis and Affects Immune Cell Motility. *Cancer Discov* 2019;9:64-81.
- Peng DH, Rodriguez BL, Diao L, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8(+) T cell exhaustion. *Nat Commun* 2020;11:4520.
- Sorushanova A, Delgado LM, Wu Z, et al. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. *Adv Mater* 2019;31:e1801651.
- COL26A1 Gene - Collagen Type XXVI Alpha 1 Chain. Accessed 13 Apr 2022. Available online: <http://www.genecards.org/cgi-bin/carddisp.pl?gene=COL26A1&keywords=COL26A1>
- Koppes E, Shaffer B, Sadovsky E, et al. Klf14 is an imprinted transcription factor that regulates placental growth. *Placenta* 2019;88:61-7.
- Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* 2019;47:D853-8.
- AlphaFold Protein Structure Database. Collagen alpha-1(XXVI) chain. Accessed 13 Apr 2022. Available online: <http://alphafold.ebi.ac.uk/entry/Q96A83>
- Koch A, Jeschke J, Van Criekinge W, et al. MEXPRESS update 2019. *Nucleic Acids Res* 2019;47:W561-5.
- Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017;45:D362-8.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-7.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
- Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401-4.
- Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
- Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 2013;39:782-95.
- Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
- Mora A, Donaldson IM. iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. *BMC Bioinformatics* 2011;12:455.
- Karamanos NK, Theocharis AD, Piperigkou Z, et al. A guide to the composition and functions of the extracellular matrix. *FEBS J* 2021;288:6850-912.
- Hayasaka H, Yoshida J, Kuroda Y, et al. CXCL12 promotes CCR7 ligand-mediated breast cancer cell invasion and migration toward lymphatic vessels. *Cancer Sci* 2022;113:1338-51.
- Azimizade Y, Saberi AA, Sahimi M. Regulation of migration of chemotactic tumor cells by the spatial distribution of collagen fiber orientation. *Phys Rev E* 2019;99:062414.
- Rallis KS, Corrigan AE, Dadah H, et al. IL-10 in cancer: an essential homeostatic regulator between homeostatic immunity and inflammation - a comprehensive review. *Future Oncol* 2022;18:3349-65.
- Yan J, Smyth MJ, Teng MWL. Interleukin (IL)-12 and IL-23 and Their Conflicting Roles in Cancer. *Cold Spring Harb Perspect Biol* 2018;10:a028530.
- Karin N. Chemokines and cancer: new immune checkpoints for cancer therapy. *Curr Opin Immunol* 2018;51:140-5.
- Morein D, Erlichman N, Ben-Baruch A. Beyond Cell



- Motility: The Expanding Roles of Chemokines and Their Receptors in Malignancy. *Front Immunol* 2020;11:952.
28. Marotta V, Sciammarella C, Capasso M, et al. Germline Polymorphisms of the VEGF Pathway Predict Recurrence in Nonadvanced Differentiated Thyroid Cancer. *J Clin Endocrinol Metab* 2017;102:661-71.
  29. Ramos MIP, Tian L, de Ruiter EJ, et al. Cancer immunotherapy by NC410, a LAIR-2 Fc protein blocking human LAIR-collagen interaction. *Elife* 2021;10:e62927.
  30. Ghorbaninezhad F, Masoumi J, Bakhshivand M, et al. CTLA-4 silencing in dendritic cells loaded with colorectal cancer cell lysate improves autologous T cell responses in vitro. *Front Immunol* 2022;13:931316.
  31. Gambardella C, Offi C, Romano RM, et al. Transcutaneous laryngeal ultrasonography: a reliable, non-invasive and inexpensive preoperative method in the evaluation of vocal cords motility—a prospective multicentric analysis on a large series and a literature review. *Updates Surg* 2020;72:885-92.
  32. Yadav SK, Stojkov D, Feigelson SW, et al. Chemokine-triggered microtubule polymerization promotes neutrophil chemotaxis and invasion but not transendothelial migration. *J Leukoc Biol* 2019;105:755-66.
  33. Wang W, Shen C, Zhao Y, et al. The Role of m6A RNA Methylation-Related lncRNAs in the Prognosis and Tumor Immune Microenvironment of Papillary Thyroid Carcinoma. *Front Cell Dev Biol* 2022;9:719820.
  34. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res* 2019;79:4557-66.
  35. Lei X, Lei Y, Li JK, et al. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Lett* 2020;470:126-33.
  36. Maller O, Drain AP, Barrett AS, et al. Tumour-associated macrophages drive stromal cell-dependent collagen crosslinking and stiffening to promote breast cancer aggression. *Nat Mater* 2021;20:548-59.
  37. Mariathasan S, Turley SJ, Nickles D, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 2018;554:544-8.
  38. Wang W, Bai N, Li X. Comprehensive Analysis of the Prognosis and Drug Sensitivity of Differentiation-Related lncRNAs in Papillary Thyroid Cancer. *Cancers (Basel)* 2022;14:1353.
  39. Xu S, Xu H, Wang W, et al. The role of collagen in cancer: from bench to bedside. *J Transl Med* 2019;17:309.

**Cite this article as:** Luo Y, Ye Y, Zhang Y, Chen L, Qu X, Yi N, Ran J, Chen Y. New insights into COL26A1 in thyroid carcinoma: prognostic prediction, functional characterization, immunological drug target and ceRNA network. *Transl Cancer Res* 2023;12(12):3384-3408. doi: 10.21037/tcr-23-141