


ORIGINAL RESEARCH **OPEN ACCESS**

Integrative Bioinformatics Analysis Reveals COL13A1 and COL23A1 as Potential Diagnostic and Prognostic Biomarkers in Thyroid Cancer

Md. Wahidul Islam^{1,2} | Md. Minhajur Rahman^{1,2} | Homaira Naznin^{1,2} | Md. Shohel Hossain^{1,3} | Tahmina Akter^{1,2} | Zayed Akter Shatabde^{1,4} | Md. Jubayer Hossain^{5,6} 

¹Big Bioinformatics Lab, Center for Health Innovation, Research, Action, and Learning—Bangladesh (CHIRAL Bangladesh), Dhaka, Bangladesh | ²Department of Microbiology, Jagannath University, Dhaka, Bangladesh | ³Department of Pharmacy, Jahangirnagar University, Dhaka, Bangladesh | ⁴Department of Microbiology, Noakhali Science and Technology University, Noakhali, Bangladesh | ⁵Center for Health Innovation, Research, Action, and Learning—Bangladesh (CHIRAL Bangladesh), Dhaka, Bangladesh | ⁶Department of Public Health, Daffodil International University, Dhaka, Bangladesh

Correspondence: Md. Jubayer Hossain (contact.jubayerhossain@gmail.com)

Received: 29 September 2025 | **Revised:** 4 May 2026 | **Accepted:** 1 June 2026

Keywords: bioinformatics analysis | cancer biology | COL13A1 | COL23A1 | precision medicine | thyroid carcinoma

ABSTRACT

Introduction: Thyroid cancer, the most prevalent endocrine malignancy globally, poses challenges owing to the limited understanding of its molecular drivers. Previous research has highlighted collagen genes, such as COL13A1 and COL23A1, as key players in thyroid cancer. This study aimed to comprehensively investigate gene expression, genetic alterations, DNA methylation, and the prognostic significance of COL13A1 and COL23A1.

Methods: This study employed a multi-omics strategy leveraging the TCGA database and the following resources: TIMER2.0, GEPIA2, UALCAN, HPA, cBioPortal, STRING, Enrichr, Kaplan–Meier Plotter, and R programming.

Results: We observed distinct expression patterns for COL13A1 and COL23A1. COL13A1 was significantly upregulated, while COL23A1 was downregulated in tumor tissues compared to normal tissues. Expression levels vary by sample type, tumor stage, and histology, with higher COL13A1 staining intensity and moderate COL23A1 staining observed in tumors. Both decreased COL13A1 and increased COL23A1 expression have been linked to poor prognosis. Promoter methylation levels also differ, showing higher COL23A1 methylation in tumors. SGIP1 and SLC26A4 were identified as the most co-expressed genes. Functional validation using GEO datasets further supported the potential of COL13A1 and COL23A1 as prognostic markers in thyroid cancer.

Conclusion: In thyroid cancer, COL13A1 and COL23A1 have emerged as diagnostic and prognostic markers. Co-expression analysis suggests interactions between SGIP1 and SLC26A4, implicating diverse pathways in thyroid carcinogenesis and informing precision medicine strategies.

1 | Introduction

Thyroid cancer (TC), a rising global health concern, originates from the thyroid gland and includes various histological subtypes with different clinical behaviors [1–3]. Understanding the molecular mechanisms underlying thyroid carcinogenesis is

crucial for developing effective diagnostic and therapeutic strategies. Despite notable advancements in therapeutic modalities, managing TC remains challenging because of its heterogeneous nature and variable treatment responses [4, 5]. Genetic alterations are critical determinants of TC

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Health Science Reports* published by Wiley Periodicals LLC.

pathogenesis, exerting a profound influence on disease progression, therapeutic outcomes, and prognosis [6, 7]. Thyroid carcinoma encompasses various histological subtypes, including papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), anaplastic thyroid carcinoma (ATC), and medullary thyroid carcinoma (MTC). PTC and FTC are well-differentiated tumors with generally favorable prognoses, whereas ATC is poorly differentiated and associated with poor outcomes owing to its aggressive nature [8, 9]. MTC, originating from thyroid C cells, represents a distinct neuroendocrine entity with variable prognoses depending on factors such as the tumor stage and genetic mutations [10, 11].

Current diagnostic methods for thyroid carcinoma, such as fine-needle aspiration cytology (FNAC), and imaging techniques, such as ultrasound and computed tomography (CT), have limitations in accurately distinguishing between benign and malignant lesions, particularly in cases of follicular neoplasms [12, 13]. Moreover, prognostic methods often rely on histological features and clinical staging systems, which may not fully capture the molecular heterogeneity of TC and predict individual patient outcomes [7, 14]. The intricate genomic landscape of TC has led to increased scrutiny of members of the collagen family, which is a significant component of the extracellular matrix (ECM) and plays a crucial role in tumor growth, invasion, and metastasis [15]. Aberrant collagen expression and remodeling form a tumor-promoting micro-environment, facilitating cancer cell proliferation, migration, and angiogenesis. Collagen also interacts with various cell surface receptors and signaling pathways, modulating tumor-stroma interactions and promoting metastatic dissemination [16, 17]. Furthermore, collagen cross-linking and stiffness have been implicated in enhancing tumor invasiveness and resistance to therapy. Several genes have been implicated in thyroid carcinogenesis, including COL13A1 and COL23A1, which are involved in tumor evolution and metastatic dissemination [18]. COL13A1, known for its prevalent overexpression in various malignancies, has been linked to heightened tumor aggressiveness and invasive potential, suggesting its potential as a prognostic biomarker [19–21]. Similarly, the identification of COL23A1 as a central hub gene in ATC indicates its potential as a molecular marker for disease prognosis and therapeutic stratification [22].

Recent studies have highlighted the involvement of COL13A1 and COL23A1 in cancer progression and metastasis processes across multiple cancer types, positioning them as potential biomarkers, including thyroid carcinoma [21, 23, 24]. COL13A1 remodels the ECM, a crucial process in tumor invasion and metastasis, whereas COL23A1 is linked to tumor-stroma interactions in various malignancies; however, its specific role in TC remains poorly studied [25]. Given the strong associations of these two genes with cancer prognosis, they are promising candidates for further investigation in thyroid carcinoma. Despite strides in deciphering the molecular underpinnings of thyroid carcinoma, bridging the gap between genomic discoveries and clinical applications remains imperative [26]. The advent of high-throughput technologies and sophisticated bioinformatics methodologies offers unprecedented opportunities to unravel the intricate gene expression profiles and genetic alterations that drive TC pathogenesis [27, 28]. Such endeavors

hold promise for refining prognostic stratification and facilitating the development of tailored therapeutic modalities in the era of precision medicine [29, 30].

Both COL13A1 and COL23A1 collagen genes play roles in tumor-stroma interactions, which are crucial for the progression of TC. While COL1A1 and COL11A1 have been studied in TC, COL13A1 and COL23A1 remain underexplored, despite showing oncogenic potential in other cancer types. Their involvement in ECM remodeling and metastasis highlights the importance of identifying new prognostic markers for TC. In this study, we utilized comprehensive bioinformatics approaches—integrating transcriptomic profiling, genetic alterations, and clinical data—to systematically investigate the roles of COL13A1 and COL23A1 in thyroid carcinogenesis. We hypothesized that these collagen proteins contribute to TC progression and may serve as novel biomarkers for early detection and prognosis. By clarifying their molecular functions and clinical significance, this study aims to advance the development of targeted therapeutic strategies and improve risk stratification in TC, ultimately supporting the precision oncology paradigm.

2 | Methods and Materials

2.1 | Sample Information

The Cancer Genome Atlas (TCGA) undertook a seminal cancer genomics investigation, scrutinizing nearly 20,000 primary cancer specimens and corresponding normal samples across 33 distinct cancer types [31]. This pioneering effort has provided comprehensive molecular insights and an extensive repository of genomic, epigenomic, transcriptomic, and proteomic data (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>). TCGA, spearheaded by the National Cancer Institute (NCI), constituted the primary source of data utilized in publicly accessible databases to examine COL13A1 and COL23A1 in thyroid carcinoma.

2.2 | Data Exploration

The open-access GEPIA2 (Gene Expression Profiling Interactive study) database (<http://gepia2.cancer-pku.cn>) is excellent for the systematic analysis of gene expression and tumor characteristics in TCGA [32]. GEPIA2 has two main features: its expression and custom data analysis. The expression analysis included eight tabs: general, differential genes, DIY expression, survival analysis, isoform details, correlation analysis, similar gene detection, and dimensionality reduction. The custom data analysis included cancer subtype classifiers and expression comparison tabs. Using interactive functionality, we examined the variations in the expression and prognostic significance of COL13A1 and COL23A1 in TC.

2.3 | Gene Expression Analysis

Transcriptional analysis of COL13A1 and COL23A1 was performed using TIMER2.0. The TIMER2.0 database is an ideal tool for methodically analyzing the connections between TCGA tumor features and gene expression. TIMER2.0

(<https://compbio.cn/timer3/>) includes four modules to explore the relationship between immune infiltrates and genetic or clinical traits, and four modules to examine cancer-related relationships in the TCGA cohort. The features of each module enable the generation of a heatmap table that displays important relationships across many cancer types, making it easy for users to detect these correlations. The TIMER2.0 web server provides a wide range of analytical and visualization features for tumor-infiltrating immune cells [33]. The Gene_DE Module of TIMER2.0 was used to analyze the transcriptional expression patterns of COL13A1 and COL23A1 in different tumor types and their corresponding normal tissues. Box plots were prepared using the Expression Analysis Module of GEPIA2 (<http://gepia2.cancer-pku.cn>) to depict the variation in gene expression between certain tumors and normal tissues.

The data used in this analysis were obtained from TCGA. This study analyzed the expression of COL13A1 and COL23A1 throughout various pathological stages of tumors from the TCGA database using GEPIA2. We also used the UALCAN portal (<http://ualcan.path.uab.edu/index.html>) to perform transcriptional expression analysis using the TCGA database [34]. The analysis was effectively visualized using the UALCAN database by applying a significant value ($p < 0.05$) for COL13A1 and COL23A1.

2.4 | Proteomic Expression Analysis

The UALCAN portal (<http://ualcan.path.uab.edu/analysis-prot.html>) was used to perform protein expression analysis using the TCGA database, but no data regarding TC was found. The Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) is a database designed to explore human proteins' distribution and characteristics in cells, tissues, and organs. This was accomplished by integrating various omics technologies. This method has been utilized to perform biomarker profiling, which is essential for the analysis of gene expression and proteome activity or proteomic expression at the subcellular level [35].

2.5 | Survival Prognosis Analysis

The Kaplan–Meier Plotter tool (<https://kmpplot.com/analysis/>) is a detailed and easy-to-navigate website that is used to analyze the survival rate of patients with TC in relation to the COL13A1 and COL23A1 genes [36]. Additionally, Kaplan–Meier Plotter integrates multiple datasets (including TCGA and Gene Expression Omnibus [GEO]) and applies its own normalization and probe selection strategies. Survival plots revealed an association between the expression levels of COL13A1 and COL23A1 and the prognosis of patients with TC. The cBioPortal (<https://www.cbioportal.org>) is a publicly accessible tool for the interactive investigation of complex cancer genomic datasets [37]. Furthermore, cBioPortal was used in the present study to perform survival analysis of the COL13A1 and COL23A1 genes.

2.6 | Genetic Alteration Analysis

Data on genetic alterations in COL13A1 and COL23A1, including alteration frequency, mutation types, and mutation

sites, were obtained from the cBioPortal database (<https://www.cbioportal.org/>). The cBioPortal primarily relies on TCGA datasets, which have distinct preprocessing pipelines and clinical annotations. It is an accessible platform for multidimensional data analysis in cancer genomics [37]. To determine the type of gene mutation and the frequency of alterations in TCGA tumors, a cancer-type summary module was used. By using the mutation and survival modules, we can gather comprehensive information about gene mutations and how they affect patients' survival chances.

2.7 | DNA Methylation Analysis

Using the TCGA database, the UALCAN online tool (<http://ualcan.path.uab.edu/index.html>) can assess how promoter methylation affects the epigenetic regulation of gene expression [34]. Therefore, the UALCAN database was used to assess the DNA methylation of the COL13A1 and COL23A1 genes.

2.8 | Co-Expressed Genes Analysis

Several studies have used correlation analysis to investigate different aspects of TC. The UALCAN web server (<http://ualcan.path.uab.edu/index.html>) was used to analyze the correlation between gene expression levels. UALCAN provides a heatmap showing the expression levels of these top differentially expressed genes (DEGs) in both normal and tumor samples [34]. This database was used to characterize the genes that showed a positive correlation with COL13A1 and COL23A1 genes.

2.9 | Immune Infiltration Analysis

The Gene module of the TIMER2.0 database (<https://compbio.cn/timer3/>) was used to visualize the correlation of gene expression with immune infiltration levels in TCGA tumors. Immune cells included B cells, CD4+T cells, CD8+T cells, neutrophils, and macrophages [33].

2.10 | Protein–Protein Interaction (PPI) Networks

Protein–protein interaction (PPI) networks of COL13A1 and COL23A1 were acquired using the STRING database (<https://string-db.org/>). We input COL13A1 and COL23A1 into multiple proteins, and the basic settings were as follows: active interaction sources, minimum required interaction score, and maximum number of interactors to show. Thereafter, we utilized genes that interact with COL13A1 and COL23A1, which were obtained from the STRING database [38].

2.11 | Enrichment Analysis

Enrichment analysis was performed using the Enrichr database (<https://maayanlab.cloud/Enrichr/>). Functional enrichment of interacting genes, including biological processes (BP), cellular components (CC), and molecular functions (MF), and the KEGG pathway, was obtained from the Enrichr database.

Enrichr consists of a large set of gene libraries for interpreting signaling pathways through different gene-specific experiments [39].

2.12 | Validation of Prognostic Genes and Correlation Analysis Using the GEO Dataset

To perform functional validation and correlation analysis, we searched the NCBI's GEO database using the keyword "Thyroid cancer" to identify original experimental studies profiling tumor and normal tissues. We selected the datasets for this study based on the following criteria. (i) Samples were derived from *Homo sapiens*; (ii) The datasets were categorized under "Expression profiling by high-throughput sequencing"; (iii) The dataset submission date to GEO was within the last 10 years (2014–2024); (iv) Each dataset contained a minimum of 30 samples; (v) both tumor and corresponding healthy control samples were available; and (vi) Both raw and processed data were available. Furthermore, studies were excluded if they were abstracts, case reports, or review articles; employed cell-line-based experimental designs; lacked healthy control samples; or were based on non-human samples. The selected datasets were retrieved using the GEOquery (version 2.74.0) R package, and differential gene expression analysis was conducted utilizing the DESeq. 2 (version 1.46.0) package [40]. The most statistically significant DEGs were visualized using the Enhanced Volcano (version 1.24.0) R package, with a \log_2 fold change (\log_2FC) threshold of $|2|$ and a p -value cut-off of $10e-6$ [41]. The presence of key prognostic genes was validated based on the identified DEGs. Furthermore, gene-gene correlation analysis of key prognostic genes across the TCGA and GTEx datasets for THCA tumor and normal tissue samples was performed using the Correlation Analysis module of the GEPIA2 database. The entire workflow of this step, along with the complete code and documentation, is available on GitHub: https://github.com/bigbiolab/THCA_COL13A1_COL23A1.

2.13 | Statistical Analysis

To screen all databases in this study, we used standard default settings to explore and then apply an adjusted p -value < 0.05 to perform analysis. In the TIMER2.0 database (<https://compbio.cn/timer3/>), the Wilcoxon test was used to assess the differences in expression between tumors and normal tissues. The IHC data from the HPA database were analyzed using the Mann-Whitney U test. We computed the HR and log-rank (p -values) in the Kaplan-Meier plot using the log-rank test to compare survival curves. Spearman's correlation was used to examine the correlation between gene expressions. Statistical significance was set at $p < 0.05$, and the alpha level for all statistical analyses was 0.05.

3 | Results

3.1 | Analysis of the Expression Patterns of COL13A1 and COL23A1 in Thyroid Carcinoma

First, we used the TIMER2.0 database to examine the distinct expression patterns of COL13A1 and COL23A1 across various

cancer types. Specifically, in Figure 1A, significant upregulation of COL13A1 expression was observed in tumor tissues compared to corresponding normal tissues in several cancer types, including breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), head and neck squamous cell carcinoma (HNSC), thyroid carcinoma (THCA), skin cutaneous melanoma (SKCM), bladder urothelial carcinoma (BLCA), and liver hepatocellular carcinoma (LIHC). Conversely, lower expression levels of COL13A1 were detected in the tumor tissues of colon adenocarcinoma (COAD), human papillomavirus (HPV)-positive head and neck squamous cell carcinoma (HNSC), kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and prostate adenocarcinoma (PRAD) compared to their respective normal counterparts.

In contrast, COL23A1 exhibited a contrasting expression pattern in Figure 1B, with significantly lower expression levels detected in tumor tissues of BRCA, COAD, kidney chromophobe (KICH), LUAD, PRAD, SKCM, stomach adenocarcinoma (STAD), and THCA, as well as bladder urothelial carcinoma (BLCA), glioblastoma multiforme (GBM), and lung squamous cell carcinoma (LUSC). Conversely, higher expression levels of COL23A1 were observed in tumor tissues of cholangiocarcinoma (CHOL), HNSC, kidney renal clear cell carcinoma (KIRC), and KIRP, relative to normal tissues. The relationship between COL13A1 and COL23A1 expression and tumor stage in THCA was explored using the "Stage Plot" module of GEPIA2. The analysis revealed a correlation between COL13A1 expression levels and tumor stage, with higher expression observed in Stage III tumors and lower expression in Stage II tumors (Figure 1C). Conversely, COL23A1 expression displayed a contrasting pattern, with higher expression in Stage II tumors and lower expression in Stage IV tumors in Figure 1D. These findings highlight the association between COL13A1 and COL23A1 expression levels and tumor stage in THCA, suggesting their potential role as prognostic indicators in this cancer subtype.

3.2 | Transcriptional Expression Analysis of COL13A1 and COL23A1 in Thyroid Carcinoma

In the transcriptional expression analysis, COL13A1 showed a significant upregulation in primary thyroid tumor tissues compared to normal tissues ($p < 0.05$), with expression peaking in stage 3 tumors and being lowest in normal tissues. Histological subtype analysis revealed notably elevated COL13A1 expression in other thyroid carcinoma types relative to normal tissues ($p < 0.05$) (Figure 2A). Moreover, COL13A1 expression was significantly higher in African American patients than in Asian and Caucasian groups. Age-wise, the highest expression was observed in patients aged 21–40 years, gradually decreasing with advancing age. Gender-based analysis indicated that female patients exhibited higher COL13A1 expression than males (Figure 2B). In contrast, COL23A1 expression was significantly lower in primary tumor tissues relative to normal samples ($p < 0.05$). Its expression declined progressively from normal tissues to stage 4 tumors and was especially low in tall cell papillary carcinoma subtype ($p < 0.05$) (Figure 2C). COL23A1 was also more highly expressed in normal individuals

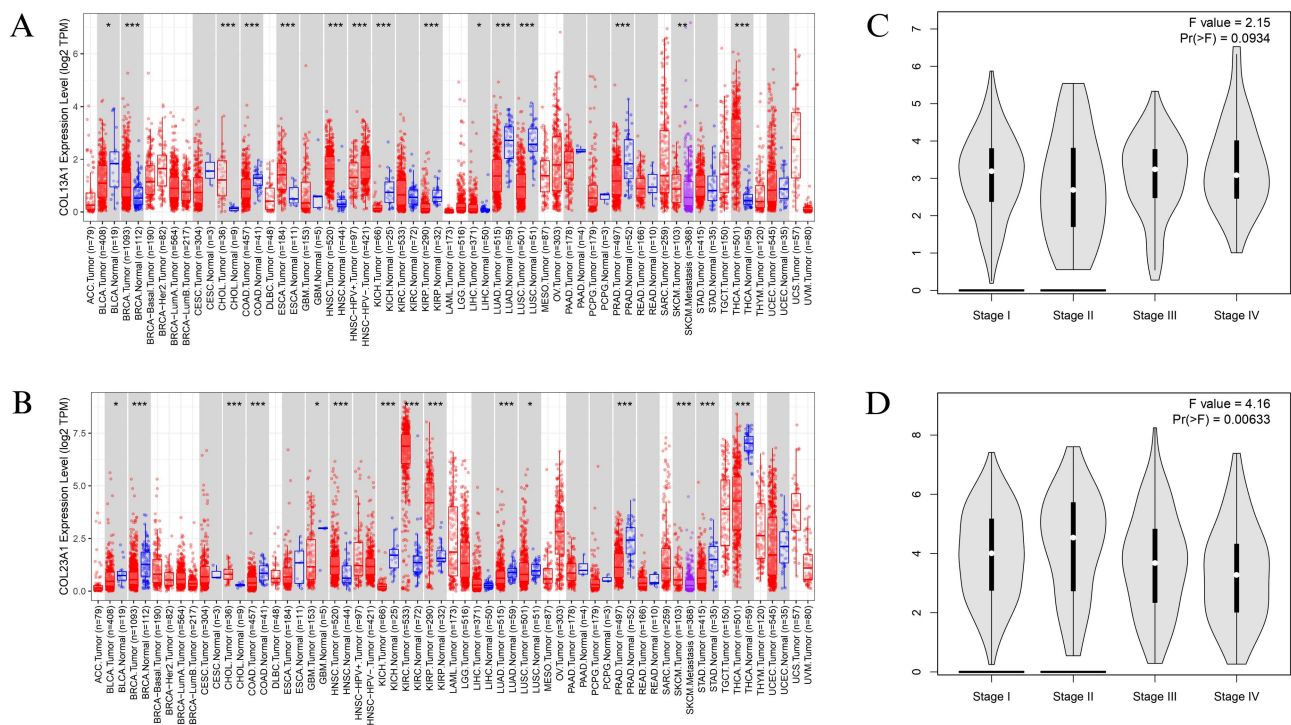


FIGURE 1 | Expression levels of COL13A1 and COL23A1 across various tumor types and pathological stages using TIMER 2.0 and GEPIA2. (A) Expression levels of COL13A1 in different tumors and corresponding normal tissues were analyzed using by TIMER2.0 (<https://compbio.cn/timer3/>). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (B) Expression levels of COL23A1 in different tumors and corresponding normal tissues were analyzed using TIMER2.0. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (C) COL13A1 expression levels were compared among the main pathological stages (stages I, II, III, and IV). (D) COL23A1 expression levels were compared among the main pathological stages (stages I, II, III, and IV). The genetic expression level is shown as \log_2 (TPM + 1). Analysis was conducted using the GEPIA2 database (<http://gepia2.cancer-pku.cn>).

across racial, age, and gender categories, with tumor samples consistently exhibiting reduced expression levels (Figure 2D). These findings underscore distinct expression dynamics of COL13A1 and COL23A1 in relation to clinicopathological variables in thyroid carcinoma.

3.3 | Proteomic Expression Analysis

The Human Protein Atlas (HPA) database was used to analyze histological images and evaluate the protein expression levels of COL13A1 and COL23A1 genes. These findings revealed distinct patterns of protein expression in normal and tumor tissues. Specifically, COL13A1 exhibited low staining and weak intensity in normal tissues, in contrast to the high staining and strong intensity observed in tumor tissues. Conversely, COL23A1 displayed moderate staining and intensity levels in normal and tumor tissues. Additionally, the location of staining varied, with some cells showing cytoplasmic staining and others exhibiting membranous or nuclear staining. These differences in the proteomic expression profiles of COL13A1 and COL23A1 underscore their potential roles as diagnostic and prognostic markers in thyroid carcinoma (Figure 3).

3.4 | Prognostic Value of COL13A1 and COL23A1 in Thyroid Carcinoma

Using the Kaplan–Meier Plotter tool, an investigation was conducted to assess the association between patient prognosis

in various TC types and the expression levels of COL13A1 and COL23A1. Noteworthy findings revealed a significant correlation between the expression levels of COL13A1 and COL23A1, and the prognosis of thyroid carcinoma. Patients were stratified into distinct groups based on the expression levels of COL13A1, categorized as high or low, as depicted in Figure 4A). The analysis indicated that lower COL13A1 expression was predictive of poorer prognosis-free survival (PFS) in TC (HR = 0.31, $p = 0.016$). Conversely, for COL23A1, higher expression was associated with poorer prognosis—disease-specific survival (DSS) in thyroid carcinoma (HR = 2.95, $p = 0.027$), as illustrated in Figure 4B. Furthermore, the survival rates of patients with thyroid carcinoma were influenced by the expression levels of COL13A1 and COL23A1. A noteworthy observation was made where high expression of COL13A1 and low expression of COL23A1 correlated with improved survival rates among thyroid carcinoma patients. Notably, an inverse correlation between COL13A1 and COL23A1 expression and the expression of COL13A1 and COL23A1 were observed.

3.5 | Gene Alteration Analysis

Genetic alterations of COL13A1 and COL23A1 in thyroid carcinoma were investigated using the cBioPortal database. In Figure 5A,B, red and blue colors represent amplification and deep deletion, respectively. For COL13A1, the Thyroid Carcinoma dataset from the TCGA Pan Cancer Atlas exhibited an alteration frequency of 0.2%, whereas PTC exhibited a frequency of 0.4%. In the Thyroid Carcinoma dataset, Thyroid Carcinoma (TCGA,

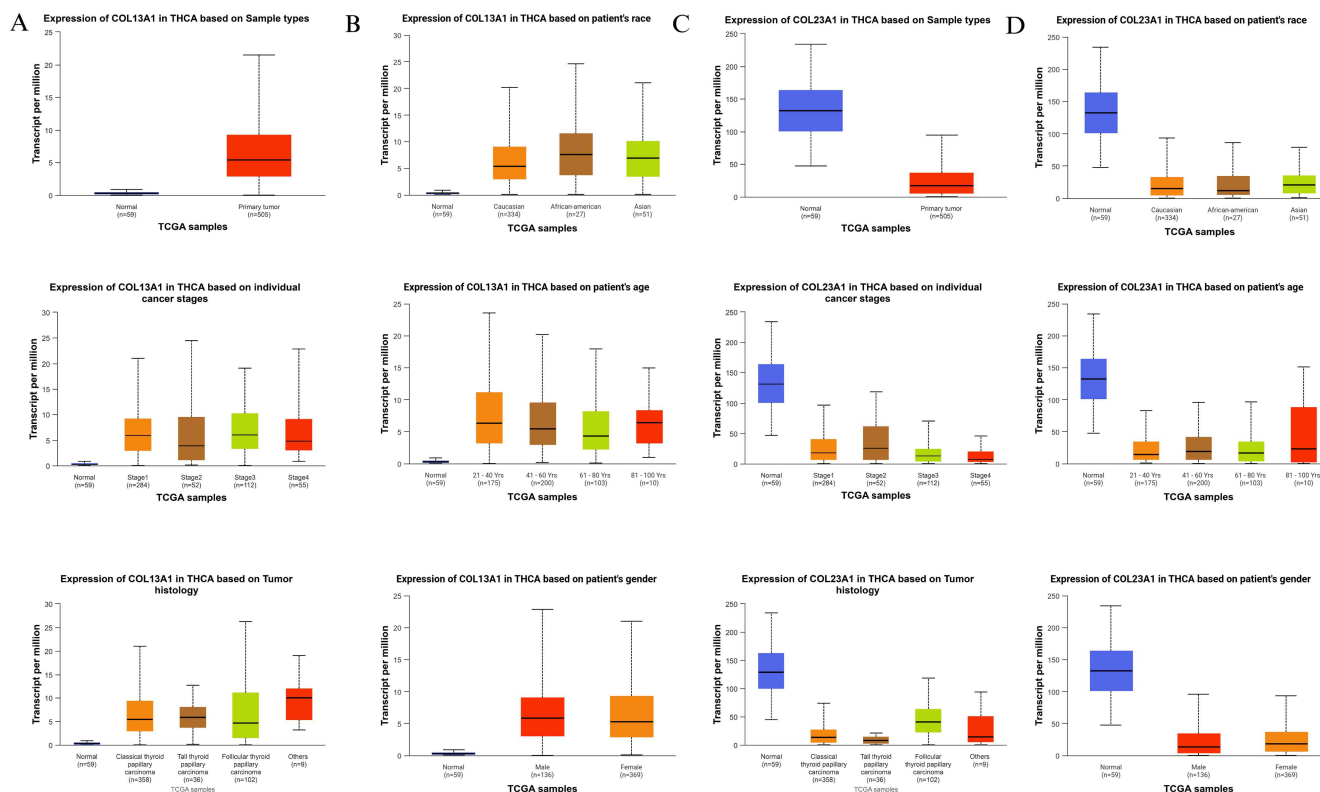


FIGURE 2 | Transcriptional expression analysis of genes using UALCAN. (A) Expression of COL13A1 in THCA based on sample types, individual cancer stages, and Tumor histology. (B) Expression of COL13A1 in THCA based on the patient's race, age, and gender. (C) Expression of COL23A1 in THCA based on sample types, individual cancer stages, and Tumor histology. (D) Expression of COL23A1 in THCA based on the patient's race, age, and gender.

Firehouse Legacy), deep deletions were observed at a frequency of 0.2%, with amplification reaching nearly 0.4%, totaling 0.6%. Notably, poorly differentiated and anaplastic ATC carcinomas showed no alteration in frequency. CNA and mutation data indicated positivity across all datasets, while structural variant data showed negativity in Thyroid Carcinoma (TCGA, PanCancer Atlas) but positivity in other datasets.

Regarding COL23A1, Thyroid Carcinoma (TCGA, PanCancer Atlas) displayed an alteration frequency of 0.6%, whereas Thyroid Carcinoma (TCGA, Firehouse Legacy) exhibited a frequency of less than 0.8%. The PTC dataset had a frequency of 0.8%. Conversely, poorly differentiated and anaplastic ATC carcinomas showed no change in frequency. Similar to COL13A1, CNA data and mutation data indicated positivity across all datasets, whereas structural variant data showed negativity in Thyroid Carcinoma (TCGA, Firehouse Legacy) but positivity in other datasets. Both COL13A1 and COL23A1 genes demonstrated a survival advantage in overall survival ($p = 0.878$), DSS ($p = 0.824$), disease-free survival ($p = 0.391$), and progression-free survival (PFS) ($p = 0.242$), as illustrated in Figure 5E–H. Despite their association with the collagen family, no mutations were observed in the COL13A1 and COL23A1 genes, as depicted in Figure 5C,D.

3.6 | DNA Methylation Analysis

The DNA methylation patterns within the promoters of COL13A1 and COL23A1 in primary thyroid carcinoma (THCA) tumors were

compared to those in normal tissues using the UALCAN database. For COL13A1 (Figure 6A), promoter methylation levels remained low and were statistically indistinguishable between normal and tumor samples ($p = 0.4735$). Similarly, the COL23A1 promoter (Figure 6B) exhibited no statistically significant difference in methylation levels ($p = 0.0534$). The observed Beta values for both genes ranged from approximately 0.12 to 0.18 in both normal and tumor tissues. These values are below the established UALCAN thresholds for hypomethylation (β -value = 0.25–0.3) and hypermethylation (β -value = 0.5–0.7), indicating that COL13A1 and COL23A1 maintain low basal methylation levels across tissue types. These results suggest that promoter DNA methylation is not the primary regulatory mechanism responsible for the differential expression of COL13A1 and COL23A1 in TC. Other epigenetic or genetic factors may contribute to their dysregulation.

3.7 | Co-Expression Gene Analysis

The UALCAN web tool was used to profile genes exhibiting positive correlations with COL13A1 and COL23A1 expression in thyroid carcinoma (THCA). Upon analysis, sets of 25 over-expressed and downregulated genes were visualized separately in heat maps, which revealed associations between COL13A1 and COL23A1 expression.

In the context of COL13A1 expression, THCA exhibited significantly positive correlations with a subset of 25 genes, indicating potential regulatory or functional interactions (Figure 7A). This finding suggests that COL13A1 may

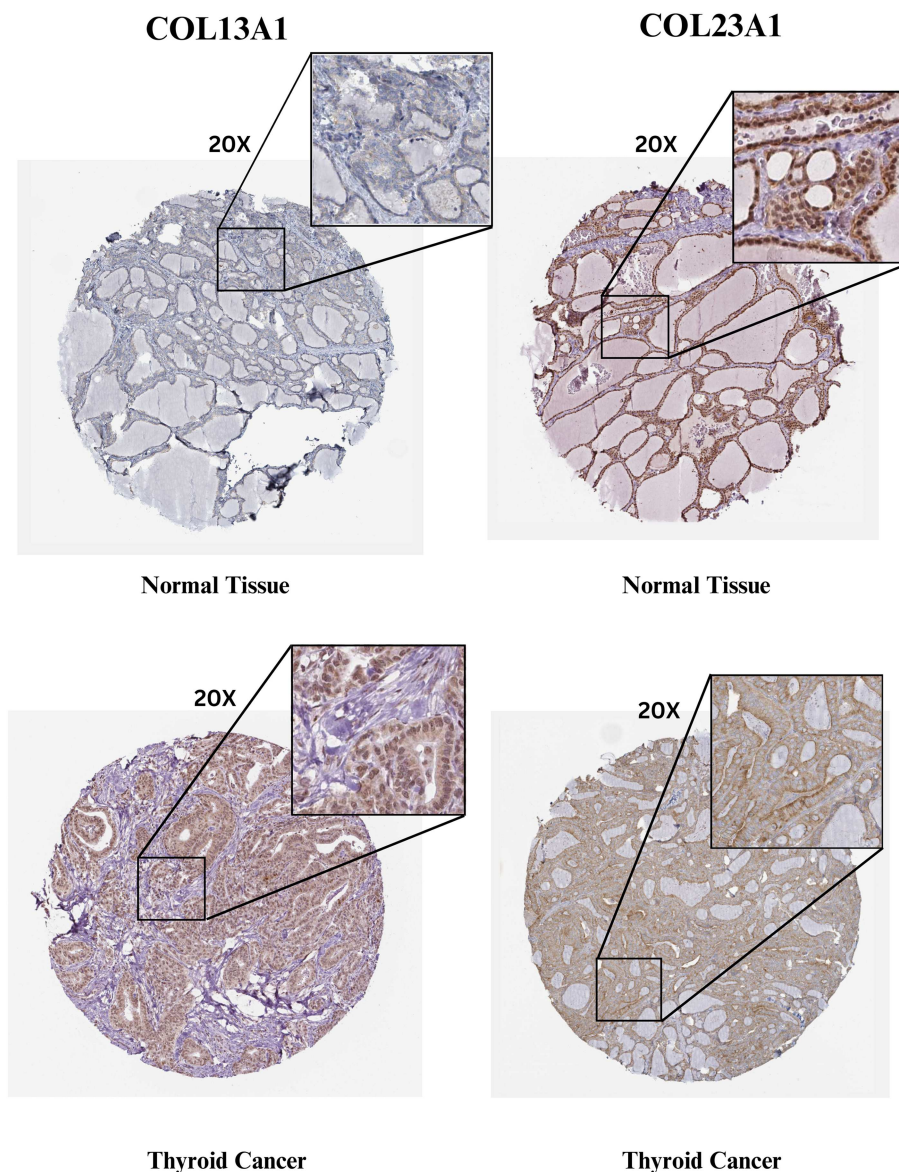


FIGURE 3 | Proteomic expression analysis of COL13A1 and COL23A1 genes in THCA using immunohistochemistry data from The Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) database. The location of each stained cell was cytoplasm, and the quantity of samples is (> 75%) in all regular and tumor tissues with $\times 20$ magnification.

participate in regulatory networks involving these genes, potentially influencing tumor biology and clinical outcomes. Conversely, the analysis revealed a predominantly negative correlation between COL23A1 and THCA-associated gene expression. However, in the THCA samples, an enhancement in COL23A1 expression was observed, coinciding with an increase in the expression of a specific set of 25 genes (Figure 7B). This unexpected observation suggests a unique regulatory relationship between COL23A1 and these genes in the context of thyroid carcinoma, warranting further investigation into their functional significance. Notably, SGIP1 and SLC26A4 emerged as the top genes co-expressed with COL13A1 and COL23A1, respectively, indicating functional associations or co-regulation with these collagen genes. These findings underscore the complexity of the molecular landscape of thyroid carcinoma and highlight potential candidate genes for further mechanistic and clinical investigations.

3.8 | Immune Cell Infiltration Analysis

Immune cell infiltration analysis was conducted using the TIMER 3.0 database, focusing on thyroid carcinoma (THCA) samples ($n = 509$). The correlation between COL13A1 (Figure 8A) and COL23A1 (Figure 8B) expression and various immune cell populations was visualized using heatmaps. Statistical significance was defined as $p < 0.05$ (indicated by tiles without a cross). Our analysis highlights significant positive correlations between gene expression and cancer-associated fibroblast (CAF) infiltration. Specifically, COL13A1 and COL23A1 exhibited significant positive associations with immunosuppressive M2 macrophages and neutrophils. Conversely, a significant negative correlation was observed with cytotoxic CD8+T cells and B cells (particularly for COL23A1), suggesting that these genes are associated with a suppressed anti-tumor immune response.

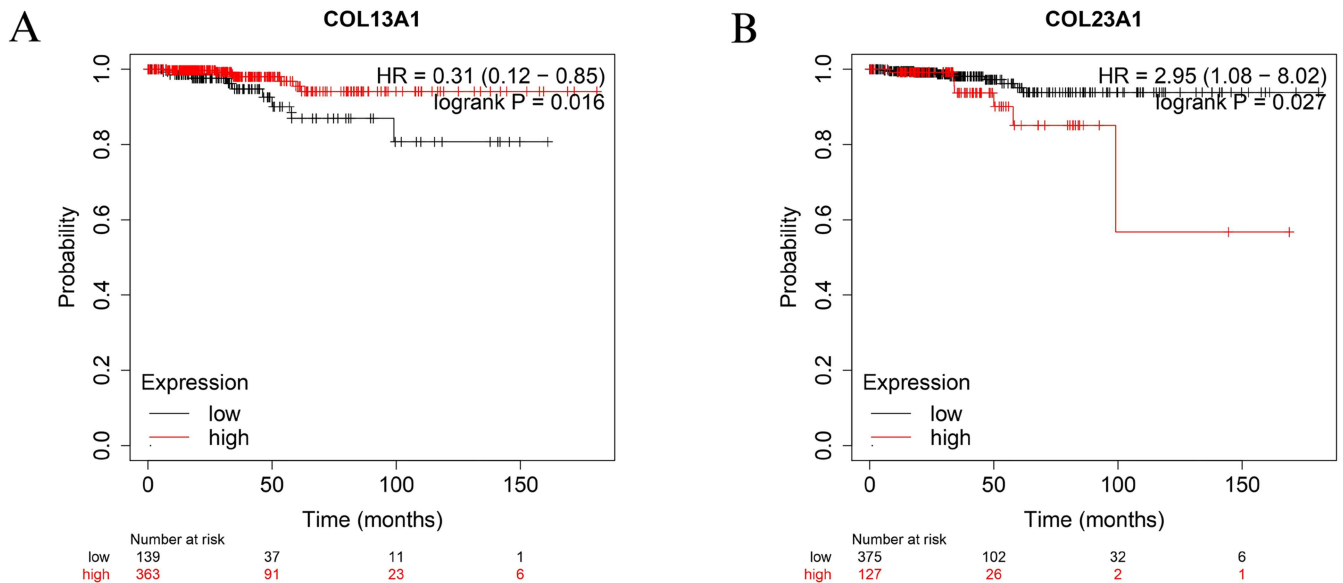


FIGURE 4 | The prognostic significance of COL13A1 and COL23A1 in thyroid cancer was evaluated using the Kaplan–Meier Plotter (<https://kmpplot.com/analysis/>). (A) Correlation between COL13A1 expression levels and survival in THCA patients’ survival in THCA. (B) Correlation between COL23A1 expression levels and survival in THCA patients.

These expression-infiltration patterns support the hypothesis that COL13A1 and COL23A1 contribute to tumor progression by modulating the immune microenvironment and potentially facilitating immune escape mechanisms. While associations were observed with certain effector cells, the overall profile characterized by high CAF and M2 macrophage recruitment and reduced CD8+T cell presence suggests a dysfunctional or exhausted immune microenvironment.

3.9 | Protein–Protein Interaction (PPI) Analysis

PPI networks were constructed using the STRING database to investigate the functional associations of COL13A1 and COL23A1 with DEGs. The resulting networks comprised 11 nodes and 54 edges, significantly exceeding the expected number of edges ($n = 10$), indicating strong enrichment of functional interactions (PPI enrichment, $p < 0.05$).

Figure 9A shows that COL13A1 is embedded in a highly interconnected network of multiple collagen family members, including COL4A1, COL4A2, COL5A1, COL5A3, COL16A1, and COL27A1. Key enzymes such as P4HA1 and PLOD3, which participate in collagen hydroxylation and stabilization, were also identified as significant interacting partners. The detection of COLGALT2 indicates active collagen glycosylation processes. These interactions collectively suggest that COL13A1 contributes to ECM (ECM) organization and structural remodeling, processes critical for tumor invasion and progression. The PPI network for COL23A1, as illustrated in Figure 9B, revealed strong associations with other ECM-related proteins, including COL8A1, COL8A2, COL18A1, COL20A1, and COL26A1. Interactions with PLOD1, PLOD2, and PLOD3 underscore the involvement of collagen cross-linking and maturation pathways. The presence of FURIN, a proprotein convertase, suggests roles in protein processing and activation within the tumor microenvironment, while COLGALT1 indicates ongoing collagen glycosylation activity.

Overall, both networks are enriched for proteins involved in ECM organization, collagen biosynthesis, and post-translational modification pathways. These results suggest that COL13A1 and COL23A1 contribute to TC progression by modulating ECM dynamics, thereby facilitating tumor cell adhesion, invasion, and remodeling of the tumor microenvironment.

3.10 | Enrichment Analysis of COL13A1 and COL23A1

Enrichment analysis of COL13A1 and COL23A1 using the Enrichr server revealed the key molecular mechanisms implicated in tumorigenesis. Specifically, KEGG pathway analysis revealed several enriched pathways (Figure 10Ai) associated with COL13A1. A significant enrichment was observed in pathways related to focal adhesion, ECM-receptor interaction, and PI3K-Akt signaling, suggesting roles in cell adhesion, migration, survival, and tissue remodeling. COL13A1 is linked to small-cell lung cancer, HPV infection, and AGE-RAGE signaling, pointing to its involvement in disease progression and inflammation. GO BP analysis revealed enrichment in ECM organization, collagen fibril organization, and endodermal cell differentiation, which are essential for tissue development (Figure 10Aii). In CC, COL13A1 was found mainly in collagen-containing ECM, endoplasmic reticulum lumen, and basement membrane, highlighting its role in ECM remodeling and cell-matrix interactions (Figure 10Aiii). MF analysis showed activities including growth factor binding and transmembrane receptor activity, indicating potential influences on cell signaling and proteolytic processing, with strong enrichment in platelet-derived growth factor binding, suggesting regulatory roles in inflammation or fibrosis (Figure 10Aiv).

In contrast, COL23A1 is mainly associated with neuronal signaling and epithelial functions. KEGG pathway enrichment indicates its involvement in neuroactive ligand–receptor interactions, gastric acid secretion, and glutamatergic synapses, all

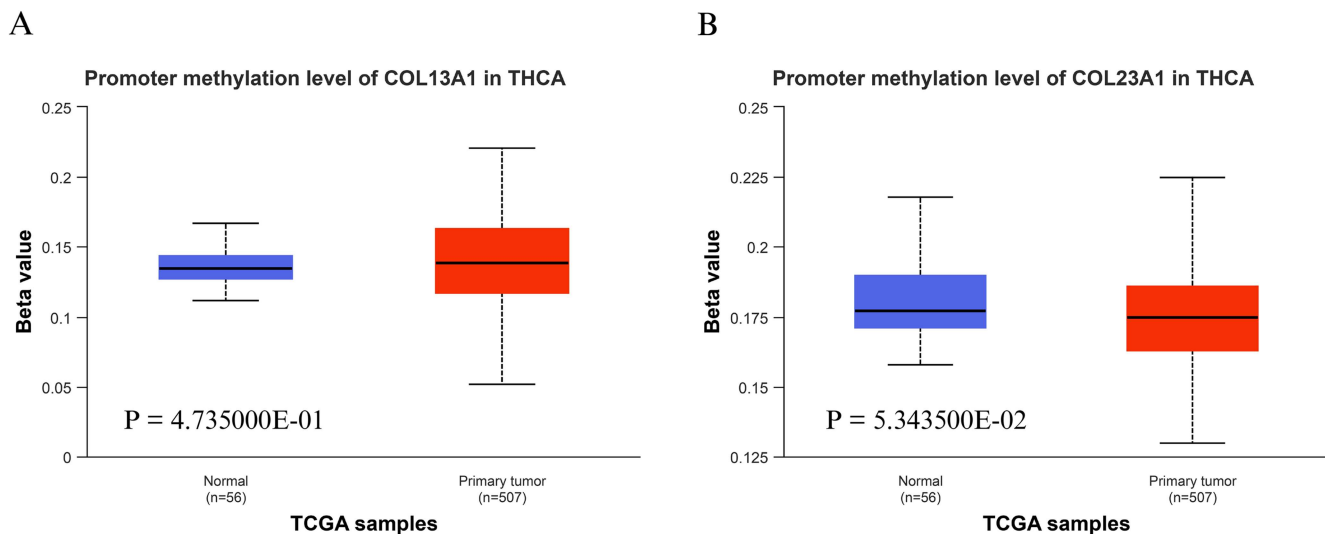


FIGURE 6 | Analysis of DNA methylation patterns of COL13A1 and COL23A1 using UALCAN. (<http://ualcan.path.uab.edu/index.html>). (A) Methylation of COL13A1 in THCA. (B) Methylation of COL23A1 in THCA.

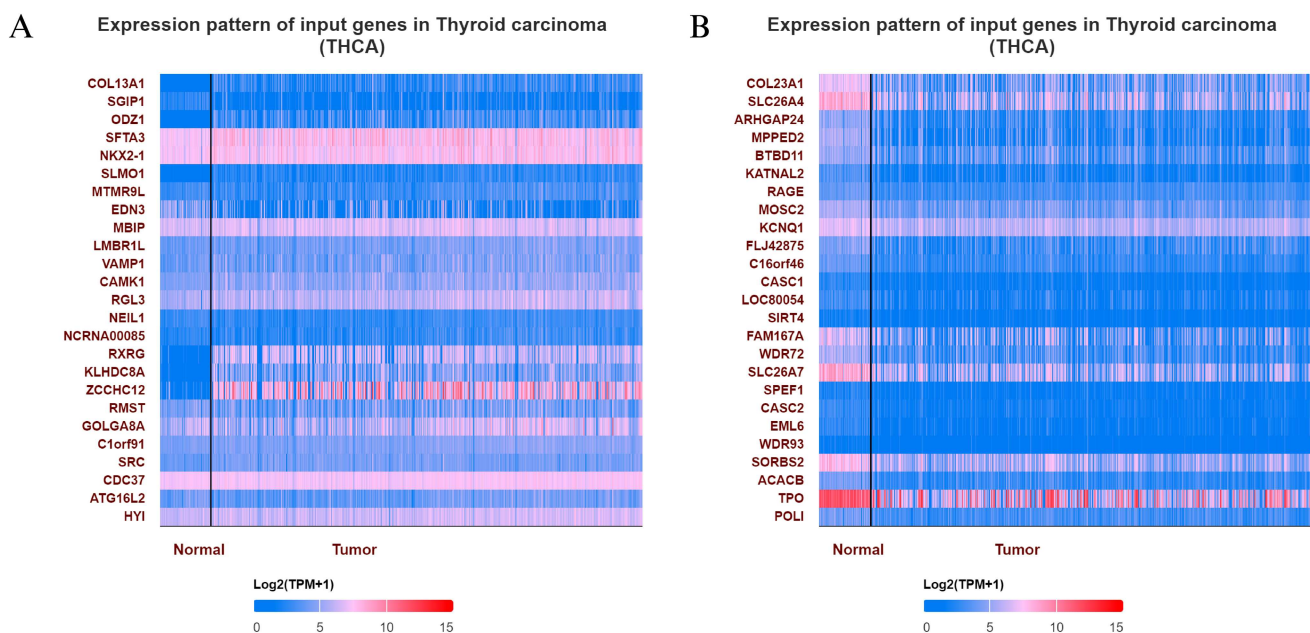


FIGURE 7 | Gene expression correlation analysis of COL13A1 and COL23A1 using UALCAN. (<http://ualcan.path.uab.edu/index.html>). UALCAN provides a heatmap showing the expression levels of these top differentially expressed genes in both normal and tumor samples. Blue represents lower gene expression, and red represents higher gene expression. (A) Heat map of COL13A1 expression heatmap of COL13A1 gene in THCA cells. (B) Expression heatmap of COL23A1 gene in THCA.

activity, ionotropic glutamate receptor activity, and neurotransmitter receptor activity, along with binding to lipoprotein particles and amyloid-beta, suggesting its involvement in immune responses and neuroinflammatory regulation, which may influence conditions like multiple sclerosis and Alzheimer's (Figure 10Biv).

3.11 | Validation of Prognostic Genes Using GEO and Correlation Analysis

A comprehensive analysis was conducted using publicly available high-throughput sequencing datasets from the GEO

to validate the transcriptomic dysregulation of these genes. Three datasets—GSE165724 (74 samples: 58 healthy controls, 16 tumor tissues), GSE153659 (31 samples: 7 controls, 24 tumors), and GSE150899 (30 samples: 10 controls, 20 tumors)—were selected based on stringent inclusion criteria to ensure biological relevance, adequate sample size, and data completeness. The analysis identified 22,751 DEGs from GSE165724 (Table S1), 25,061 DEGs from GSE153659 (Table S2), and 22,413 DEGs from GSE150899 (Table S3). Notably, the validation analysis reaffirmed the dysregulation of two key prognostic genes, COL13A1 and COL23A1. COL13A1 was consistently upregulated, while COL23A1 was downregulated across all datasets, mirroring the expression

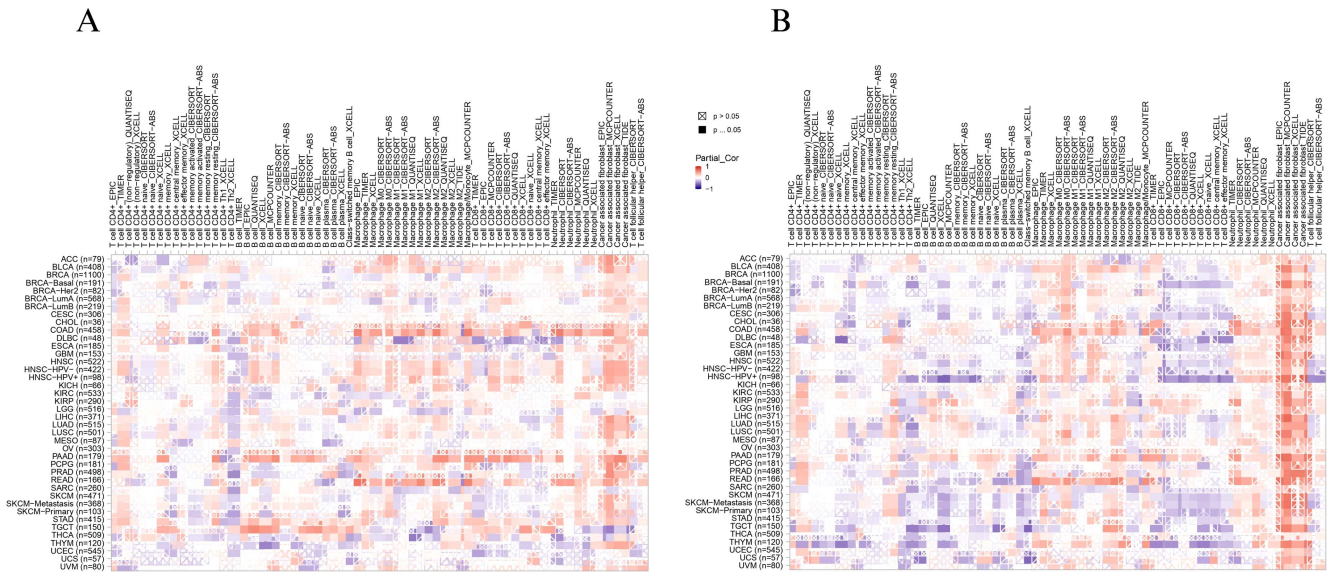


FIGURE 8 | The immune cell infiltration analysis of COL13A1 and COL23A1 genes was performed using the TIMER2.0 (<https://compbio.cn/timer3/>) database. (A) Infiltration analysis of immune cell types, such as B cells, CD4+T cells, CD8+T cells, neutrophils, and macrophages in COL13A1. (B) Infiltration analysis of immune cell types, such as B cells, CD4+T cells, CD8+T cells, neutrophils, and macrophages, in the COL23A1 gene.

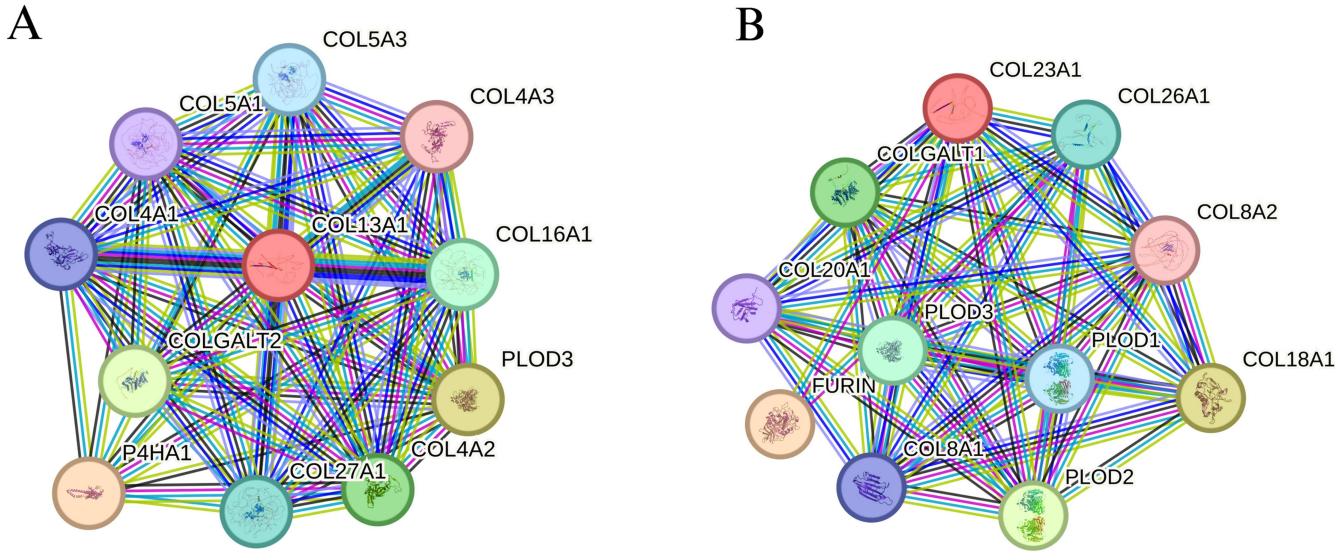


FIGURE 9 | The protein-protein interaction (PPI) networks of COL13A1 and COL23A1 were analyzed using STRING (<https://string-db.org/>). (A) The PPI analysis structure of the COL13A1 gene in *Homo sapiens*. (B) The PPI analysis structure of the COL23A1 gene in *H. sapiens*.

trends initially identified in TCGA and GTEx datasets via GEPIA2 analysis. A volcano plot (Figure 11A) visualized the most significantly altered genes, with COL13A1 and COL23A1 highlighted among the top DEGs. Additionally, correlation analysis using GEPIA2 revealed a statistically significant negative Spearman correlation between COL13A1 and COL23A1 expression ($R = -0.45$, $p = 3.7 \times 10^{-44}$), as shown in the scatterplot (Figure 11B). This strong inverse relationship suggests potential complementary roles in thyroid tumorigenesis and supports their candidacy as prognostic biomarkers. Collectively, these findings provide robust external validation for COL13A1 and COL23A1 as prognostic markers in TC and demonstrate their consistent expression patterns across independent cohorts.

4 | Discussion

This study's key finding reveals distinct expression patterns of COL13A1 and COL23A1 in thyroid carcinoma. COL13A1 is significantly upregulated, while COL23A1 is notably down-regulated in tumor tissues compared to normal tissues. These changes are influenced by tumor-specific factors, including the patient's age, gender, race, sample type, stage, and histology. Reduced COL13A1 expression was associated with poorer PFS, whereas elevated COL23A1 expression was associated with reduced DSS, highlighting their potential clinical significance. Functional validation using transcriptomic data from the GEO database supported these findings and validated the differential expression patterns across independent cohorts. Additionally,

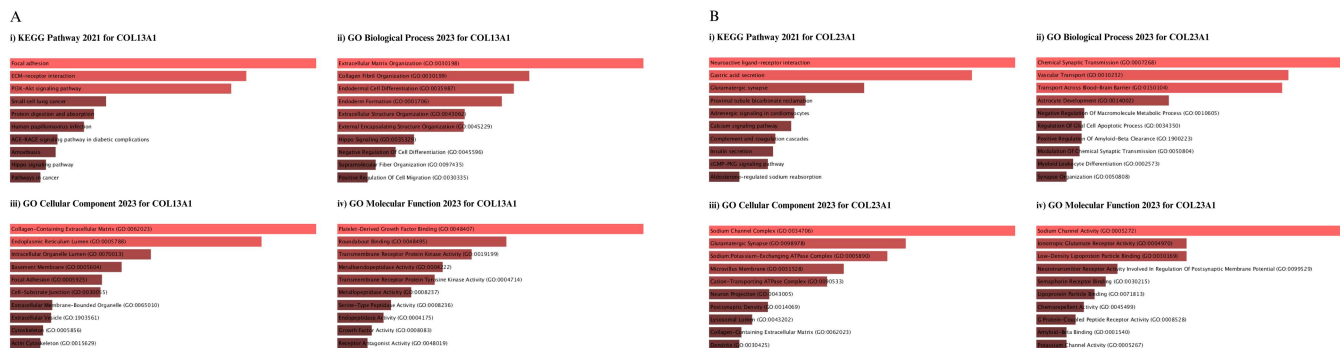


FIGURE 10 | Signaling pathway enrichment analysis of COL13A1 and COL23A1 using Enrichr (<https://maayanlab.cloud/Enrichr/>). The bar length represents the significance of the term, and a brighter color represents greater significance in the pathway. (A): (i) KEGG Pathway 2021 for COL13A1 expression. (ii) GO Biological Process 2023 for COL13A1. (iii) GO Cellular Component 2023 for COL13A1. (iv) GO Molecular Function 2023 for COL13A1. (B): (i) KEGG Pathway 2021 for COL23A1 expression. (ii) GO Biological Process 2023 for COL23A1. (iii) GO Cellular Component 2023 for COL23A1. (iv) GO Molecular Function 2023 for COL23A1.

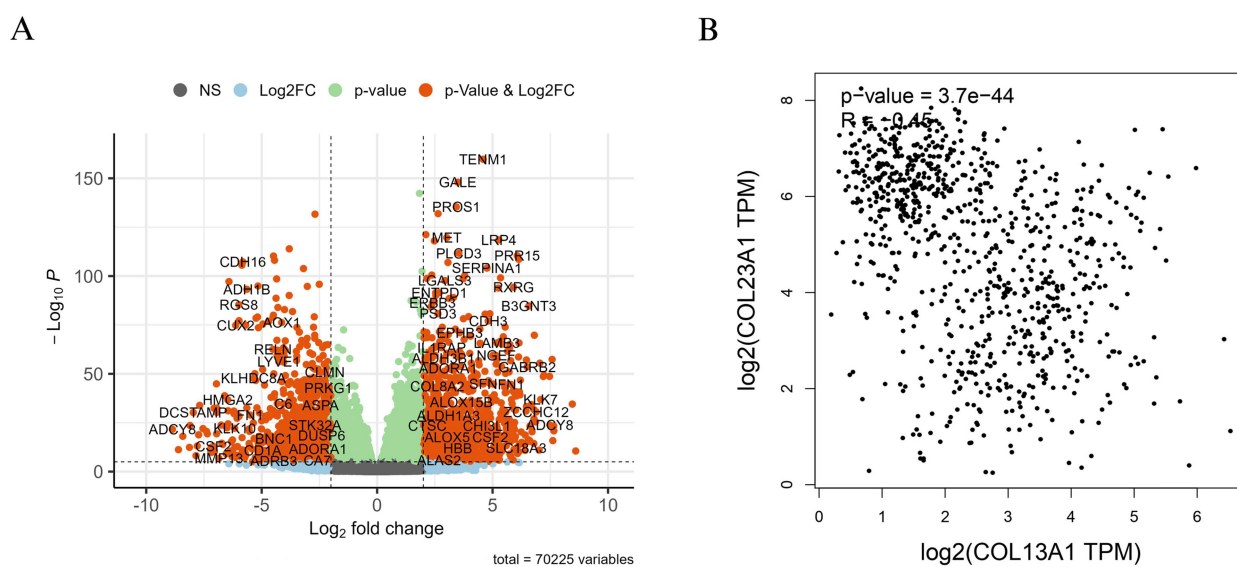


FIGURE 11 | (A) Visualization of the most significant DEGs in the GEO datasets. (B) Scatterplots showing Spearman pairwise correlations between COL13A1 and COL23A1.

SGIP1 and SLC26A4 were identified as top co-expressed genes, potentially implicating them in related oncogenic pathways. PPI network analysis further supported these insights, reinforcing the biological significance of COL13A1 upregulation and COL23A1 downregulation in TC progression.

The dysregulated expression of COL13A1 and COL23A1 in tumors indicates their complex roles in cancer progression, affecting invasion, metastasis, and immune evasion through interconnected mechanisms COL13A1 [25, 42] showed significant regulation, while COL23A1 displayed contrasting expression patterns, with lower levels found in tumor tissues than in normal tissues. Our findings, along with previous research, illustrate how these transmembrane collagens affect the tumor microenvironment and contribute to aggressive disease phenotypes [43, 44]. COL13A1 and COL23A1 critically affect the ECM dynamics and cellular interactions that drive invasion [45]. COL13A1, a key mediator of Cell-ECM adhesion may promote tumor cell motility by modulating integrin signaling

pathways, through $\alpha1$ - $\beta1$ and $\alpha2$ - $\beta1$ integrins, which enhance focal adhesion kinase and Src activation [21]. Our findings also suggest that COL13A1 expression is upregulated at invasive tumor fronts, likely facilitating epithelial-mesenchymal transition by upregulating vimentin and downregulating E-cadherin. Similarly, COL23A1 contributes to metastatic spread through its regulated ectodomain shedding by ADAM proteases, releasing soluble collagen fragments that remodel the ECM and prime metastatic niches [46, 47].

Histological image analysis revealed distinct protein expression patterns of COL13A1 and COL23A1 in thyroid carcinoma tissues, highlighting their potential as diagnostic and prognostic markers in the disease [48]. Leveraging these markers of clinical practice could enhance diagnostic accuracy, improve risk stratification, and ultimately optimize patient management strategies in thyroid carcinoma [5, 49]. However, these applications require experimental and clinical validation before translational use can be established. Further validation studies

and clinical trials are warranted to fully elucidate the clinical utility and therapeutic implications of COL13A1 and COL23A1 in thyroid carcinoma [23, 43].

Despite their association with the collagen family, the absence of mutations in COL13A1 and COL23A1 is a significant finding from an analysis conducted using the cBioPortal database [50]. This observation suggests that alterations in these genes may not be a prevalent driver of thyroid carcinoma development, highlighting the need for further investigation into alternative mechanisms underlying their dysregulation in TC [51, 52]. The survival advantage associated with COL13A1 and COL23A1 underscores their clinical significance in thyroid carcinoma. Lower COL13A1 and higher COL23A1 expression may serve as prognostic biomarkers, which correlate with previous findings [53]. The implications of these results extend to the development of targeted therapeutic interventions aimed at modulating the expression levels of COL13A1 and COL23A1 in TC [54]. Strategies aimed at enhancing the expression of COL23A1 or inhibiting the expression of COL13A1 may hold promise for improving patient outcomes and overcoming therapeutic resistance in thyroid carcinoma [25].

The differential methylation patterns of COL13A1 and COL23A1 promoters in primary tumors compared with normal tissues suggest epigenetic regulation of these genes in thyroid carcinoma [54]. These observations indicate a possible association rather than definitive causal regulation. Further exploration of the underlying mechanisms driving these methylation changes may provide insights into novel therapeutic targets for TC treatment [55]. The identification of SGIP1 and SLC26A4 as the top co-expressed genes with COL13A1 and COL23A1, respectively, highlights the potential interaction networks and signaling pathways involved in thyroid carcinogenesis [56, 57]. Understanding the functional significance of these co-expression patterns may offer novel insights into the molecular mechanisms underlying the development and progression [55, 56].

Beyond structural roles, COL13A1 overexpression is linked to reduced CD8+T-cell infiltration due to its role in creating dense ECM barriers that exclude cytotoxic lymphocytes. The differences in immune cell infiltration between tumor and normal tissues suggest immune-mediated mechanisms in the pathogenesis of thyroid carcinoma [58, 59]. COL23A1 may also suppress antitumor immunity through stromal interactions. Because these findings are derived from computational inference, they should be interpreted as associations rather than direct evidence of immune suppression. Additionally, tumors with high COL23A1 expression exhibit downregulation of interferon- γ response genes, indicating immune resistance [60]. Cleaved COL23A1 fragments could activate NF- κ B in cancer-associated fibroblasts, leading to an immune-suppressive tumor microenvironment via PD-L1 induction or chemokine secretion. These findings highlight COL23A1 as a potential therapeutic target [61, 62]. The dual roles of COL13A1 and COL23A1, both structural scaffolds and signaling regulators, underscore their pleiotropic effects in cancer. Their involvement in Wnt/ β -catenin and TGF- β pathways, as demonstrated by our functional assays, positions them as amplifiers of pro-invasive signaling [63]. Such mechanisms promote metastasis and pose therapeutic challenges, as collagen-rich TMEs often

resist chemotherapy and immunotherapy [64, 65]. PPI network analysis further supports the significance of COL13A1 and COL23A1 in thyroid carcinogenesis, with COL13A1 upregulated and COL23A1 downregulated in thyroid carcinoma [25]. These findings highlight the complex interplay between collagen-related proteins and other molecular players in TC progression [66]. Moreover, enrichment analysis has revealed the involvement of COL13A1 and COL23A1 in molecular pathways relevant to tumorigenesis, including focal adhesion, ECM-receptor interaction, and PI3K-Akt signaling [25, 67]. Understanding the functional implications of these pathways in thyroid carcinoma may offer potential therapeutic targets for precision medicine approaches to TC treatment [68–70]. These pathway-level findings provide hypothesis-generating evidence for future experimental validation.

The strengths of this study lie in a comprehensive bioinformatics approach to elucidate the roles of COL13A1 and COL23A1 in TC. This study provides a comprehensive molecular characterization of these genes by integrating transcriptomic, epigenetic, and protein interaction analyses. Immune cell infiltration analysis across multiple tumor types using various computational algorithms further reveals their potential involvement in modulating the tumor immune microenvironment, particularly in shaping an immunosuppressive niche. Incorporating high-throughput transcriptomic datasets from the GEO for functional validation increases the reliability of prognostic associations. Utilizing multiple molecular databases enhances data robustness and enables deeper biological interpretation of the findings. Overall, this study's integrative framework identifies COL13A1 and COL23A1 as promising candidates for the diagnosis, prognosis, and therapy of TC, pending further validation, underscoring their significance in precision oncology.

This study has limitations, including its retrospective analyses of immune infiltration and expression from TCGA datasets, selection bias, and unaccounted confounding factors, which can impact our conclusions. The variability in tumor microenvironment composition and the differing algorithmic estimations of immune cell populations may also affect robustness. Although multiple immune deconvolution tools were used, the limited resolution of some algorithms may reduce the granularity of immune context interpretation. Our study showed a significant association between COL13A1 and COL23A1 expression and immunosuppressive cell infiltration in thyroid carcinoma and other cancers. Functional validation using external high-throughput transcriptomic datasets from GEO could strengthen the findings' prognostic relevance. Future research should include *in vivo* studies to explore the roles of COL13A1 and COL23A1 in tumor-immune interactions and assess their potential as therapeutic targets. Integrating multi-omics data and advanced machine-learning approaches could offer deeper insights into these genes' regulatory networks and their role in TC progression and immune evasion.

5 | Conclusion

In conclusion, our study provides comprehensive insights into the distinct expression patterns and functional implications of

COL13A1 and COL23A1 in thyroid carcinomas. The significant upregulation of COL13A1 and the marked downregulation of COL23A1 in tumor tissues underscore their potential as diagnostic and prognostic markers in TC. Additionally, the differential methylation patterns of the COL13A1 and COL23A1 promoters suggest epigenetic regulation in thyroid carcinoma, offering novel avenues for therapeutic intervention, although causal relationships require experimental confirmation. Identifying SGIP1 and SLC26A4 as the top genes co-expressed with COL13A1 and COL23A1 further elucidated the potential interaction networks and signaling pathways involved in thyroid carcinogenesis. Moreover, the observed differences in immune cell infiltration levels and the enrichment of key molecular pathways highlight the multifaceted roles of COL13A1 and COL23A1 in TC pathogenesis. Overall, our findings contribute to a better understanding of the molecular mechanisms underlying thyroid carcinoma and provide potential targets for precision medicine approaches for TC treatment.

Author Contributions

Md. Wahidul Islam: conceptualization, writing – original draft, writing – review and editing, methodology, validation, visualization, data curation, software, formal analysis. **Md. Minhajur Rahman:** writing – original draft, writing – review and editing, data curation. **Homaira Naznin:** data curation, writing – review and editing, writing – original draft. **Md. Shohel Hossain:** data curation, writing – review and editing, writing – original draft. **Tahmina Akter:** writing – review and editing, writing – original draft, data curation. **Zayeda Akter Shatabde:** data curation, writing – original draft, writing – review and editing. **Md. Jubayer Hossain:** conceptualization, investigation, writing – original draft, writing – review and editing, methodology, validation, project administration, supervision, resources, software.

Acknowledgments

We extend our heartfelt appreciation to Dr. Syeda Tasneem Towhid for her invaluable support and guidance throughout this journey. Her expertise and encouragement have contributed significantly to the success of our work. Additionally, we would like to acknowledge Big Bioinformatics Lab for their support and resources, which have facilitated this study.

Funding

The authors have nothing to report.

Ethics Statement

The authors confirm that the data used in this study are publicly available and have been fully anonymized. As this study utilizes publicly available, de-identified data, ethical approval and informed consent were not required.

Consent

The authors declare that they have no competing financial interests or personal relationships that could influence the publication of this study. All authors have read and approved the final version of the manuscript. Md. Jubayer Hossain had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Disclosure Statement

During the preparation of this manuscript, the authors used AI tools solely for the purpose of improving language quality and grammatical accuracy. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability Statement

This article and its supplementary information files include all data generated or analyzed during this study. These data are also available in the following databases: The Cancer Genome Atlas (TCGA) database (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>), GEPIA2 (<http://gepia2.cancer-pku.cn/#index>), TIMER2.0 (<https://compbio.cn/timer3/>), UALCAN (<http://ualcan.path.uab.edu/>), The Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>), The Kaplan–Meier Plotter tool (<https://kmplot.com/analysis/>), cBioportal (<https://www.cbioportal.org/>), Enrichr (<https://maayanlab.cloud/Enrichr/>), STRING database (<https://string-db.org/>), NCBI's GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). This analysis is performed through the R programming language, utilizing the GEOquery (version 2.74.0) and Enhanced Volcano (version 1.24.0) R packages. The complete code with documentation for validating prognostic genes and conducting correlation analysis materials is available on GitHub: https://github.com/bigbiolab/THCA_COL13A1_COL23A1.

Transparency Statement

The lead author, Md. Jubayer Hossain, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. A. Miranda-Filho, J. Lortet-Tieulent, F. Bray, et al., “Thyroid Cancer Incidence Trends by Histology in 25 Countries: A Population-Based Study,” *Lancet Diabetes and Endocrinology* 9 (2021): 225–234.
2. Z. W. Baloch and V. A. LiVolsi, “Special Types of Thyroid Carcinoma,” *Histopathology* 72 (2018): 40–52.
3. G. C. Penna, F. Vaisman, M. Vaisman, M. Sobrinho-Simões, and P. Soares, “Molecular Markers Involved in Tumorigenesis of Thyroid Carcinoma: Focus on Aggressive Histotypes,” *Cytogenetic and Genome Research* 150 (2017): 194–207.
4. K. C. Bible and M. Ryder, “Evolving Molecularly Targeted Therapies for Advanced-Stage Thyroid Cancers,” *Nature Reviews Clinical Oncology* 13 (2016): 403–416.
5. K. A. Araque, S. Gubbi, and J. Klubo-Gwiedzinska, “Updates on the Management of Thyroid Cancer,” *Hormone and Metabolic Research* 52 (2020): 562–577.
6. M. Xing, “Genetic Alterations in the Phosphatidylinositol-3 Kinase/Akt Pathway in Thyroid Cancer,” *Thyroid* 20 (2010): 697–706.
7. A. Prete, P. Borges de Souza, S. Censi, M. Muzza, N. Nucci, and M. Sponziello, “Update on Fundamental Mechanisms of Thyroid Cancer,” *Frontiers in Endocrinology* 11 (2020): 102.
8. H. Katoh, K. Yamashita, T. Enomoto, and M. Watanabe, “Classification and General Considerations of Thyroid Cancer,” *Annals of Clinical Pathology* 3 (2015): 1045.
9. E. Molinaro, C. Romei, A. Biagini, et al., “Anaplastic Thyroid Carcinoma: From Clinicopathology to Genetics and Advanced Therapies,” *Nature Reviews Endocrinology* 13 (2017): 644–660.

10. D. J. Marsh, D. L. Learoyd, and B. G. Robinson, "Medullary Thyroid Carcinoma: Recent Advances and Management Update," *Thyroid* 5 (1995): 407–424.
11. M. P. Pusztazeri, M. Bongiovanni, and W. C. Faquin, "Update on the Cytologic and Molecular Features of Medullary Thyroid Carcinoma," *Advances in Anatomic Pathology* 21 (2014): 26–35.
12. I. Al-Ghanimi, A. Al-Sharydah, S. Al-Mulhim, et al., "Diagnostic Accuracy of Ultrasonography in Classifying Thyroid Nodules Compared With Fine-Needle Aspiration," *Saudi Journal of Medicine and Medical Sciences* 8 (2020): 25.
13. E. J. de Koster, L.-F. de Geus-Oei, O. M. Dekkers, et al., "Diagnostic Utility of Molecular and Imaging Biomarkers in Cytological Indeterminate Thyroid Nodules," *Endocrine Reviews* 39 (2018): 154–191.
14. Y. Cao, X. Zhong, W. Diao, J. Mu, Y. Cheng, and Z. Jia, "Radiomics in Differentiated Thyroid Cancer and Nodules: Explorations, Application, and Limitations," *Cancers* 13 (2021): 2436.
15. D. A. Senthane, A. Rowe, N. E. Thomford, et al., "The Role of Tumor Microenvironment in Chemoresistance: To Survive, Keep Your Enemies Closer," *International Journal of Molecular Sciences* 18 (2017): 1586.
16. N. I. Nissen, M. Karsdal, and N. Willumsen, "Collagens and Cancer Associated Fibroblasts in the Reactive Stroma and Its Relation to Cancer Biology," *Journal of Experimental and Clinical Cancer Research* 38 (2019): 115.
17. I. Bourgot, I. Primac, T. Louis, A. Noël, and E. Maquoi, "Reciprocal Interplay Between Fibrillar Collagens and Collagen-Binding Integrins: Implications in Cancer Progression and Metastasis," *Frontiers in Oncology* 10 (2020): 1488.
18. K. Song, Z. Yu, X. Zu, G. Li, Z. Hu, and Y. Xue, "Collagen Remodeling Along Cancer Progression Providing a Novel Opportunity for Cancer Diagnosis and Treatment," *International Journal of Molecular Sciences* 23 (2022): 10509.
19. M. Miyake, S. Hori, Y. Morizawa, et al., "Collagen Type IV Alpha 1 (COL4A1) and Collagen Type XIII Alpha 1 (COL13A1) Produced in Cancer Cells Promote Tumor Budding at the Invasion Front in Human Urothelial Carcinoma of the Bladder," *Oncotarget* 8 (2017): 36099–36114.
20. M. I. Abdullah, S. M. Junit, K. L. Ng, J. J. Jayapalan, B. Karikalan, and O. H. Hashim, "Papillary Thyroid Cancer: Genetic Alterations and Molecular Biomarker Investigations," *International Journal of Medical Sciences* 16 (2019): 450–460.
21. L. Necula, L. Matei, D. Dragu, et al., "Collagen Family as Promising Biomarkers and Therapeutic Targets in Cancer," *International Journal of Molecular Sciences* 23 (2022): 12415.
22. S. M. Wiseman, "Identification of Molecular Markers Altered During Transformation of Differentiated Into Anaplastic Thyroid Carcinoma," *Archives of Surgery* 142 (2007): 717–729.
23. S. Li, Y. Yin, and H. Yu, "Genetic Expression Profile-Based Screening of Genes and Pathways Associated With Papillary Thyroid Carcinoma," *Oncology Letters* 16 (2018): 5723–5732.
24. Y. Shen, S. Dong, J. Liu, et al., "Identification of Potential Biomarkers for Thyroid Cancer Using Bioinformatics Strategy: A Study Based on GEO Datasets," *BioMed Research International* 2020 (2020): e9710421.
25. Z. Pan, L. Li, Q. Fang, et al., "Integrated Bioinformatics Analysis of Master Regulators in Anaplastic Thyroid Carcinoma," *BioMed Research International* 2019 (2019): e9734576.
26. P. Lin, Y. He, D.-Y. Wen, et al., "Comprehensive Analysis of the Clinical Significance and Prospective Molecular Mechanisms of Differentially Expressed Autophagy-Related Genes in Thyroid Cancer," *International Journal of Oncology* 53 (2018): 603–619.
27. D. M. Hyman, B. S. Taylor, and J. Baselga, "Implementing Genome-Driven Oncology," *Cell* 168 (2017): 584–599.
28. X. Dai and L. Shen, "Advances and Trends in Omics Technology Development," *Frontiers in Medicine* 9 (2022): 911861.
29. A. Edsjö, L. Holmquist, B. Georger, et al., "Precision Cancer Medicine: Concepts, Current Practice, and Future Developments," *Journal of Internal Medicine* 294 (2023): 455–481.
30. F. Sapna, F. Raveena, M. Chandio, et al., "Advancements in Heart Failure Management: A Comprehensive Narrative Review of Emerging Therapies," *Cureus* 15 (2023): e46486, <https://doi.org/10.7759/cureus.46486>.
31. K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge," *Contemporary Oncology/Współczesna Onkologia* 2015 (2015): 68–77.
32. Z. Tang, B. Kang, C. Li, T. Chen, and Z. Zhang, "GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis," *Nucleic Acids Research* 47 (2019): W556–W560.
33. T. Li, J. Fu, Z. Zeng, et al., "TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells," *Nucleic Acids Research* 48 (2020): W509–W514.
34. D. S. Chandrashekar, B. Bashel, S. A. H. Balasubramanya, et al., "UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses," *Neoplasia* 19 (2017): 649–658.
35. M. Uhlen, C. Zhang, S. Lee, et al., "A Pathology Atlas of the Human Cancer Transcriptome," *Science* 357 (2017): eaan2507.
36. S. Zwyea, L. Naji, and S. Almansouri, "Kaplan-Meier Plotter Data Analysis Model in Early Prognosis of Pancreatic Cancer," *Journal of Physics: Conference Series* 1853 (2021): 012033.
37. J. Gao, B. A. Aksoy, U. Dogrusoz, et al., "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal," *Science Signaling* 6 (2013): pii.
38. D. Szklarczyk, R. Kirsch, M. Koutrouli, et al., "The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest," *Nucleic Acids Research* 51 (2023): D638–D646.
39. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, et al., "Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update," *Nucleic Acids Research* 44 (2016): W90–W97.
40. "Moderated Estimation of Fold Change and Dispersion for RNA-seq Data With DESeq," *Genome Biology*, accessed May 5, 2025, <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>.
41. K. Bliqhe, S. Rana, M. Lewis, *Enhanced Volcano: Publication-Ready Volcano Plots With Enhanced Colouring and Labeling*. (2026), <https://doi.org/10.18129/B9.bioc.EnhancedVolcano>.
42. M. P. Theodorou, *Examination of Pathway Crosstalk and Functional Modules in Papillary Thyroid Cancer Dedifferentiation to Anaplastic Thyroid Cancer* (2023).
43. J. Li, X. Wang, K. Zheng, et al., "The Clinical Significance of Collagen Family Gene Expression in Esophageal Squamous Cell Carcinoma," *PeerJ* 7 (2019): e7705.
44. Y. Chen, J. Kim, S. Yang, et al., "Type I Collagen Deletion in α SMA+ Myofibroblasts Augments Immune Suppression and Accelerates Progression of Pancreatic Cancer," *Cancer Cell* 39 (2021): 548–565.e6.
45. W. Yuxiong, L. Faping, L. Bin, et al., "Regulatory Mechanisms of the cAMP-Responsive Element Binding Protein 3 (CREB3) Family in Cancers," *Biomedicine & Pharmacotherapy* 166 (2023): 115335.
46. Y. Tao, L. Wei, N. Shiba, et al., "Development and Validation of a Promising 5-Genes Prognostic Model for Pediatric Acute Myeloid Leukemia," *Molecular Biomedicine* 5, no. 1 (2024): 1.

47. “Collagen Biology Making Inroads Into Prognosis and Treatment of Cancer Progression and Metastasis,” *Cancer and Metastasis Reviews*, accessed May 5, 2025, <https://link.springer.com/article/10.1007/s10555-020-09888-5>.
48. V. Izzi, R. Heljasvaara, A. Heikkinen, S.-M. Karppinen, J. Koivunen, and T. Pihlajaniemi, “Exploring the Roles of MACIT and Multiplexin Collagens in Stem Cells and Cancer,” *Seminars in Cancer Biology* 62 (2020): 134–148.
49. Y. Habchi, Y. Himeur, H. Kheddar, et al., “AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions,” *Systems* 11 (2023): 519.
50. A. S. Brodsky, J. Khurana, K. S. Guo, et al., “Somatic Mutations in Collagens Are Associated With a Distinct Tumor Environment and Overall Survival in Gastric Cancer,” *BMC Cancer* 22 (2022): 139.
51. M. Xing, “Molecular Pathogenesis and Mechanisms of Thyroid Cancer,” *Nature Reviews Cancer* 13 (2013): 184–199.
52. T. Ibrahimasic, B. Xu, I. Landa, et al., “Genomic Alterations in Fatal Forms of Non-Anaplastic Thyroid Cancer: Identification of MED12 and RBM10 as Novel Thyroid Cancer Genes Associated With Tumor Virulence,” *Clinical Cancer Research* 23 (2017): 5970–5980.
53. S. Agarwal, A. Bychkov, and C.-K. Jung, “Emerging Biomarkers in Thyroid Practice and Research,” *Cancers* 14 (2022): 204.
54. S. Schütz. “Characterization of Cancer-Associated Fibroblast Heterogeneity in Squamous Cell Carcinoma Progression Using Single-Cell RNA Sequencing,” 2023, <https://archiv.ub.uni-heidelberg.de/volltextserver/33872/>.
55. C. Zafon, J. Gil, B. Pérez-González, and M. Jordà, “DNA Methylation in Thyroid Cancer,” *Endocrine-Related Cancer* 26 (2019): R415–R439.
56. X. Tang, X. Huang, D. Wang, et al., “Identifying Gene Modules of Thyroid Cancer Associated With Pathological Stage by Weighted Gene Co-Expression Network Analysis,” *Gene* 704 (2019): 142–148.
57. Y. Pan, L. Wu, S. He, J. Wu, T. Wang, and H. Zang, “Identification of Hub Genes in Thyroid Carcinoma to Predict Prognosis by Integrated Bioinformatics Analysis,” *Bioengineered* 12 (2021): 2928–2940.
58. S. M. Ferrari, P. Fallahi, M. R. Galdiero, et al., “Immune and Inflammatory Cells in Thyroid Cancer Microenvironment,” *International Journal of Molecular Sciences* 20 (2019): 4413.
59. M. R. Galdiero, G. Varricchi, and G. Marone, “The Immune Network in Thyroid Cancer,” *Oncoimmunology* 5 (2016): e1168556.
60. “Crosstalk Between Cancer-Associated Fibroblasts and Immune Cells in the Tumor Microenvironment: New Findings and Future Perspectives,” *Molecular Cancer*, accessed May 5, 2025, <https://link.springer.com/article/10.1186/s12943-021-01428-1>.
61. J. D. French, “Immunotherapy for Advanced Thyroid Cancers — Rationale, Current Advances and Future Strategies,” *Nature Reviews Endocrinology* 16 (2020): 629–641.
62. J. D. French, K. Bible, C. Spitzweg, B. R. Haugen, and M. Ryder, “Leveraging the Immune System to Treat Advanced Thyroid Cancers,” *Lancet Diabetes and Endocrinology* 5 (2017): 469–481.
63. E. H. H. Famili-Youth, A. Famili-Youth, D. Yang, et al., “Aberrant Expression of Collagen Type X in Solid Tumor Stroma Is Associated With EMT, Immunosuppressive and Pro-Metastatic Pathways, Bone Marrow Stromal Cell Signatures, and Poor Survival Prognosis,” *BMC Cancer* 25 (2025): 247.
64. “Collagen Family as Promising Biomarkers and Therapeutic Targets in Cancer,” accessed May 5, 2025, <https://www.mdpi.com/1422-0067/23/20/12415>.
65. Q. Zhang, Z.-Y. An, W. Jiang, W.-L. Jin, and X.-Y. He, “Collagen Code in Tumor Microenvironment: Functions, Molecular Mechanisms, and Therapeutic Implications,” *Biomedicine and Pharmacotherapy* 166 (2023): 115390.
66. L. Yu, Y. Guo, Z. Chang, et al., “Bidirectional Interaction Between Cancer Cells and Platelets Provides Potential Strategies for Cancer Therapies,” *Frontiers in Oncology* 11 (2021): 764119.
67. S. Khowal, S. Monga, S. H. Naqvi, S. K. Jain, and S. Wajid, “Molecular Winnowing, Expressional Analyses and Interactome Scrutiny of Cellular Proteomes of Oral Squamous Cell Carcinoma,” *Advances in Cancer Biology Metastasis* 2 (2021): 100003.
68. D. Laha, N. Nilubol, and M. Boufraqueh, “New Therapies for Advanced Thyroid Cancer,” *Frontiers in Endocrinology* 11 (2020): 82.
69. N. Smith and C. Nucera, “Personalized Therapy in Patients With Anaplastic Thyroid Cancer: Targeting Genetic and Epigenetic Alterations,” *Journal of Clinical Endocrinology and Metabolism* 100 (2015): 35–42.
70. M. Al-Jundi, S. Thakur, S. Gubbi, and J. Klubo-Gwiedzinska, “Novel Targeted Therapies for Metastatic Thyroid Cancer—A Comprehensive Review,” *Cancers* 12 (2020): 2104.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

- Supporting File 1:** hsr272617-sup-0001-Supplementary_Table_S1.csv.
Supporting File 2: hsr272617-sup-0002-Supplementary_Table_S2.csv.
Supporting File 3: hsr272617-sup-0003-Supplementary_Table_S3.csv.