**ORIGINAL RESEARCH**

**COL12A1 as a prognostic biomarker in HER2-enriched breast cancer and its association with immune infiltration**

Yi Hu¹, Xiaomei Huang²,†, Xiaohui Lu¹,*, Shan Lin³,*

1 Department of Clinical Laboratory, Xiamen Key Laboratory of Genetic Testing, the First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, 361003 Xiamen, Fujian, China
2 Department of Marine Biology, Xiamen Ocean Vocational College, 361100 Xiamen, Fujian, China
3 Department of Orthopedics, the First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, 361003 Xiamen, Fujian, China

*Correspondence lbx77823@sina.com
(Xiaohui Lu; 2018642046@xmu.edu.cn
(Shan Lin)
† These authors contributed equally.

**Abstract**

To inspect the expression, characteristics, clinical prognostic value and the level of immune infiltration of Collagen Type XII Alpha 1 Chain (COL12A1) in breast cancer (BC) using bioinformatics. We investigated the expression of COL12A1 in breast cancer (BC) and normal breast tissues using Tumor Immune Estimation Resource (TIMER2.0) and the University of Alabama at Birmingham Cancer data analysis Portal (UALCAN). The correlation between COL12A1 and overall survival (OS) was determined using Kaplan-Meier Plotter in BC. The correlation between COL12A1, immune infiltrating cells and immune checkpoints was detected via TIMER2.0. We found that the expression of COL12A1 was significantly higher in BC tissues than in normal tissues. A higher expression level of COL12A1 correlated with lymph node metastasis and worse OS in human epidermal growth factor receptor 2 (HER2)-enriched BC patients. Moreover, COL12A1 was positively correlated with M2 macrophages and immune checkpoint programmed cell death 1 ligand 2 (PDCD1LG2) but negatively correlated with activated natural killer (NK) cells and cluster of differentiation 8 (CD8)+ T cells in HER2-positive BC. COL12A1 might act as a biomarker indicating poor prognosis in HER2-enriched BC and correlate with immune infiltration in HER2-positive BC.

**Keywords**

COL12A1; Breast cancer; Immune infiltration; Bioinformatics

**1. Introduction**

Breast cancer (BC) is the most common cancer in women, with its morbidity and mortality ranking at the forefront of all cancers in 2020 [1]. Progress in gene expression profiling and molecular diagnostic techniques have revealed considerable heterogeneity in BC. Based on the expression of the human epidermal growth factor receptor 2 (HER2), progesterone receptor and estrogen receptor, BC can be grouped into basal-like/triple negative, luminal A, luminal B, HER2-enriched and normal-like intrinsic subtypes by Prediction Analysis of Microarray 50 (PAM50) gene expression assay [2]. Currently, growing data suggest several causes of biological heterogeneity containing DNA mutations, gene expression and immune microenvironment, may affect treatment outcomes, thus, affecting oncologists’ treatment decision process in BC [3].

Immune cells infiltration, especially anti-tumor lymphocytes infiltration, has shown promises in predicting improved clinical prognosis in all BC subtypes [4]. Moreover, higher levels of tumor-infiltrating lymphocytes (TILs) were reported to significantly associate with decreased distant recurrence rates in primary triple-negative BC and increased trastuzumab beneficial result in HER2-positive BC [5]. Thus, research on novel biomarkers has been a focus over the past few decades as they have the potential to identify novel therapeutic targets for optimizing the therapy of BC patients, especially those with HER2-positive metastatic BC [6].

Collagen Type XII Alpha 1 Chain (COL12A1), a major extracellular matrix (ECM) protein, is encoded by a gene chromosomally located at 6q12-q13 [7]. ECM in the tumor microenvironment has been proved to take part in the occurrence and progress of neoplasms [8]. Upregulation of COL12A1 and its association with prognosis have received increasing attention in several different cancers, including colorectal carcinoma [9], gastric carcinoma [10], pancreatic cancer [11], and breast cancer [12]. Due to the heterogeneity of BC, the role of COL12A1 in HER2 enriched BC needs further exploration.

With advancements in sequencing technology, bioinformatics has been widely used in tumor pathogenesis, tumor immune infiltration and prognosis assessment. In this study, the expression characteristics of COL12A1 in BC, the correlation between COL12A1 and the prognosis of BC, and the level of immune infiltration correlated with COL12A1 in HER2-positive BC were analyzed using bioinformatics.

**2. Materials and methods**
Fig 1. The expression levels of COL12A1 in cancers, including BC. (A) The differential mRNA levels of COL12A1 between tumor and adjacent normal tissues in human cancers from the TCGA database analyzed using TIMER2.0. Gray columns indicate normal data were available. (B) The mRNA expression levels of COL12A1 were upregulated in major subtypes of BC in the TCGA database (114 normal breast samples, 566 luminal, 37 HER2-positive and 116 triple negative breast cancer samples), analyzed using UALCAN. (C) The protein expression levels of COL12A1 in BC from the CPTAC database (18 normal breast samples and 125 primary tumor samples) analyzed using UALCAN. * \( p < 0.05 \), ** \( p < 0.01 \), *** \( p < 0.001 \), **** \( p < 0.0001 \).

2.1 COL12A1 expression analysis

Two databases were used to investigate COL12A1 expression in BC. Tumor IMMune Estimation Resource (TIMER2.0) [13] was used to detect the expression level of COL12A1 in an array of tumors from The Cancer Genome Atlas (TCGA). The cancer exploration module and Gene-DE were used, and COL12A1 was selected for gene expression analysis. The mRNA and protein expression levels of COL12A1 in BC from the TCGA and Clinical Proteomic Tumor Analysis Consortium (CPTAC) were detected using UALCAN (ualcan.path.uab.edu) [14].

2.2 Survival analysis

The Kaplan-Meier Plotter [15] database was constructed relied on the gene chip and RNA-Seq databases of public databases as well as Gene Expression Omnibus (GEO) and TCGA. Start KM Plotter for BC was selected as the analysis module to investigate the clinical prognostic value of COL12A1 in BC. COL12A1 was selected as the Gene symbol, and the median cut-off value was identified. PAM50 was used to classify the cases into the following groups: all (n = 2976), basal (n = 309), luminal A (n = 1504), luminal B (n = 668), HER2 (n = 295), normal (n = 200), HER2 and positive lymph node status (n = 137), HER2 and negative lymph node status (n = 139), which were then used for overall survival (OS) analysis.
2.3 Correlation between COL12A1 and immune cells infiltration in HER2-positive BC

Here, the correlation between COL12A1 and the various immune cells infiltration was analyzed to investigate the potential regulatory function of COL12A1 in the immune microenvironment in HER2-positive BC using TIMER2.0 [13], a web service platform for comprehensive analysis of immune cell infiltration in tumor tissue based on RNA-Seq expression profiling data. Briefly, the immune association module and COL12A1 were selected, then NK cell, CD8+ T cell, CD4+ T cell and macrophage were selected as immune infiltrate. Purity adjustment and Spearman’s $\rho$ were selected (positive-correlation: $\rho > 0$, $p < 0.05$; negative-correlation: $\rho < 0$, $p < 0.05$). Eighty-two cases of breast invasive cancer (BRCA)-Her2 data from the TCGA were retrieved and analyzed using cell-type identification by estimating relative subsets of RNA transcript (CIBERSORT).

2.4 Correlation between COL12A1 and immune checkpoints in HER2-positive BC

To further investigate the association of COL12A1 with immune checkpoint inhibitors in HER2-positive BC, the correlation between COL12A1 and immune checkpoints was analyzed using TIMER2.0. Briefly, the cancer exploration module and Gene-Corr were selected, then COL12A1 was selected as the gene of interest. PDCD1LG2 (PD-L2), CD274 (PD-L1) and PDCD1 (PD-1) were selected as gene expression. Lastly, purity adjustment and Spearman’s $\rho$ were selected (positive-correlation: $\rho > 0$, $p < 0.05$; negative-correlation: $\rho < 0$, $p < 0.05$). In all, the data of 82 cases of BRCA-Her2 from the TCGA were retrieved and analyzed.

3. Results

3.1 Expression of COL12A1 is increased in BC

The mRNA level of COL12A1 was overexpressed in human cancers, including BC, in the TCGA samples compared with corresponding normal tissue (Fig. 1A). In addition, COL12A1 was higher in all subtypes, including the HER2-positive subtype, compared with normal breast tissues (Fig. 1B, HER2 vs. Normal, ***$p < 0.001$). The protein expression level of COL12A1 in BC tissues ($n = 125$) was higher than in normal breast tissues in the CPTAC samples using the UALCAN platform ($n = 18$), and the difference was statistically significant (Fig. 1C, Primary tumor vs. Normal, ****$p < 0.0001$).
FIGURE 3. Correlation between COL12A1 and immune cell infiltration in HER2-positive BC by TIMER2.0. COL12A1 was found to be (A) uncorrelated with tumor purity, (B) uncorrelated with M1 macrophages, (C) positively correlated with M2 macrophages ($\rho = 0.452, p = 6.7 \times 10^{-5}$), (D) negatively correlated with CD8+ T cells ($\rho = -0.435, p = 1.37 \times 10^{-4}$), (E) uncorrelated with resting NK, and (F) negatively correlated with activated NK ($\rho = -0.43, p = 1.65 \times 10^{-4}$). (G) Naive CD4+ T cells were detected by CIBERSORT. (H) COL12A1 was uncorrelated with memory resting CD4+ T cells, and (I) memory activated CD4+ T cells. BRCA: breast invasive cancer; CD: cluster of differentiation; NK: natural killer; COL12A1: Collagen Type XII Alpha 1 Chain.

FIGURE 4. Correlation between COL12A1 and immune checkpoints in HER2-positive BC using TIMER2.0. COL12A1 was (A) uncorrelated with PDCD1. (B) uncorrelated with CD274. And (C) positively correlated with PDCD1LG2 (partial. $\rho = 0.449, p = 7.74 \times 10^{-5}$). partial. $\rho$, partial correlation coefficient. TPM: transcripts per million; BRCA: breast invasive cancer; COL12A1: Collagen Type XII Alpha 1 Chain; PDCD1LG2: programmed cell death 1 ligand 2; Her: human epidermal growth factor receptor.
3.2 COL12A1 is associated with the prognosis of HER2-enriched BC

The KM Plotter analysis showed a possible association between COL12A1 and OS in univariate analysis. Increased expression of COL12A1 was correlated with worse OS in BC and HER2-enriched BC subtype patients (Fig. 2A, \( p = 0.0016; \) Fig. 2F, \( p = 0.0041 \)). Moreover, the results indicated that COL12A1 expression was not associated with the OS of luminal A, luminal B and normal-like BC patients (Fig. 2B–E). A high expression level of COL12A1 was associated with poor OS in positive lymph node status and HER2-enriched BC patients (Fig. 2G, \( p = 0.0087 \)). In addition, COL12A1 expression was not associated with the OS of positive lymph node status and HER2-enriched BC patients (Fig. 2H, \( p = 0.025 \)).

3.3 COL12A1 is associated with the infiltration of immune cell in HER2-positive BC

TIMER2.0 analysis indicated that COL12A1 was uncorrelated with the tumor purity of HER2-positive BC (Fig. 3A) and was closely correlated with the infiltration of immune cell in BC. COL12A1 was unrelated with M1 macrophages infiltration (Fig. 3B), but positively correlated with the M2 macrophages infiltration (\( \rho = 0.452, p = 6.7 \times 10^{-5} \), Fig. 3C). COL12A1 was negatively correlated with CD8+ T cells (\( \rho = -0.435, p = 1.37 \times 10^{-4} \), Fig. 3D). COL12A1 was unrelated with resting NK cells (Fig. 3E), but negatively correlated with activated NK cells (\( \rho = -0.43, p = 1.65 \times 10^{-4} \), Fig. 3F) infiltration in HER2-positive BC patients. In addition, COL12A1 was unrelated with CD4+ T cells (Fig. 3G–I).

3.4 COL12A1 is correlated with PDCD1LG2 in HER2-positive BC

COL12A1 was uncorrelated with immune checkpoints PDCD1 and CD274 (Fig. 4A–B). In addition, COL12A1 was correlated with immune checkpoint PDCD1LG2 (partial \( \rho = 0.449, p = 7.74 \times 10^{-5} \), Fig. 4C).

4. Discussion

COL12A1 was reported to be overexpressed in various cancers [9–12]. This study confirmed that COL12A1 was overexpressed in breast invasive carcinoma using the web tool TIMER2.0. Moreover, the mRNA and protein expression levels of COL12A1 analyzed by UALCAN in BC tissues were significantly higher than in normal tissues. Consistent with our findings, Xu et al. [12] found that COL12A1 was increased in breast cancer. COL12A1 was also shown to be significantly upregulated in gastric cancer and contributed to poor OS [9]. Together, the data suggested that COL12A1 might be involved in tumor progression.

Given that BC is a highly heterogeneous tumor, we further analyzed the association of COL12A1 with BC patients’ OS using the Kaplan-Meier Plotter. The results indicated that the OS of HER2-enriched BC patients with high level of COL12A1 was significantly lower than those with low level of COL12A1. In contrast, COL12A1 expression was uncorrelated with the OS of basal-like, luminal A, luminal B and normal-like BC. High level of COL12A1 was associated with lymph node metastasis and HER2-enriched BC. Altogether, these findings indicate that the prognostic value of COL12A1 in BC needs to pay close attention to its heterogeneity.

Another important source of HER2-positive BC heterogeneity might be related to its underlying tumor immune microenvironment (TIM) [16]. Immune cells and immune checkpoints in TIM are the executors of immune clearance, surveillance and tolerance of the body against tumors [17]. This finding gave a comprehensive understanding of COL12A1 on the infiltration of CD8+ T cells, NK cells and macrophages. COL12A1 in HER2-positive BC was positively correlated with the infiltration of M2 macrophages and negatively correlated with CD8+ T cells and activated NK cells. It suggested that COL12A1 may participate in the formation of tumor immunosuppressive microenvironment in HER2-positive BC. Consistent with our research report, it was previously reported that COL12A1 strongly impacted immune cell infiltrations [18].

The involvement of PDCD1LG2 in interferon (IFN)-\( \gamma \) production and T-cell proliferation was reported to be in a PDCD1-independent manner [19]. In this present study, we showed that COL12A1 was positively associated with PDCD1LG2, suggesting that increased COL12A1 might be involved in increased PDCD1LG2 infiltration in HER2-positive BC.

5. Conclusions

COL12A1 was overexpressed in all subtypes of BC. High COL12A1 expression may contribute to poor OS in HER2-enriched BC. COL12A1 may act as a potential poor prognostic biomarker in HER2-enriched BC patients. COL12A1 may correlate with immune infiltration in HER2-positive BC.

AUTHOR CONTRIBUTIONS

YH—Designing the study, Data analysis, Writing—Original Draft & Editing; XL—Data analysis, Writing—Original Draft & Editing; SL—Supervision, Writing—Original Draft/Review & Editing; SL—Supervision, Writing—Original Draft/Review & Editing.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

We thank the peer reviewers for their suggestions.

FUNDING

This research received no external funding.
CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES


