

Prognostic Prediction, Immune Microenvironment, and Drug Resistance Value of Collagen Type I Alpha 1 Chain (COL1A1) in Pan-Cancer Analysis

Liu Yi

Chongqing Medical University

Jinmin Xue

Chongqing Medical University

Maoxi Zhong

Chongqing Medical University

Zhi Wang

Chongqing Medical University

Jie Li

Chongqing Medical University

Zhu Yuxi (✉ zhuyuxi@hospital.cqmu.edu.cn)

Chongqing Medical University <https://orcid.org/0000-0001-7563-505X>

Research Article

Keywords: COL1A1, pan-cancer, prognosis, tumor immune microenvironment, gastrointestinal cancer

DOI: <https://doi.org/10.21203/rs.3.rs-293163/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective

The protein encoded by collagen type I alpha 1 chain(COL1A1)is a component of the extracellular matrix, and critical for tumor microenvironments. However, it remains unclear that the expression level, prognostic prediction, and immunological value of COL1A1 in multiple cancers.

Methods

4 profiles and 6 hub genes were found by bioinformatics analysis. We further analyzed one of the hub genes, the COL1A1. And, we comprehensively analyzed gene expression and genetic alteration patterns among 33 types of malignancies from The Cancer Genome Atlas(TCGA). Besides, we explored the correlation of COL1A1 with cancer prognosis, immune infiltrates, tumor mutational burden (TMB)/ microsatellite instability status (MSI), pathway and drug sensitivity analysis of co-expressed genes.

Results

COL1A1 was highly expressed and associated with poor prognosis in the majority of cancer. On the other hand, COL1A1 expression was closely related to high TMB in THYM, LAML, ACC, KICH, PRAD, and LGG. High MSI was observed in TGCT, MESO, PRAD, COAD, SARC, and CESC. In addition, COL1A1 was positively correlated with the abundance of CAFs, macrophages, and tumor-infiltrating lymphocytes. However, it was negatively correlated with CD8 + T cells mainly in CESC, HNSC-HPV+, and SKCM.

Conclusion

COL1A1 can serve as a prognostic and immunological biomarker in multiple cancers.

1. Introduction

Esophageal, gastric, and colorectal cancer are the most common types of gastrointestinal cancer. And, global cancer statistics in 2018 showed that colon, gastric, rectal, and esophageal cancer ranked the fourth, sixth, eighth, and ninth respectively in the incidence of human malignant tumors, and their mortality rates were 5.8%, 8.2%, and 3.2%, 5.3%, respectively^[8]. However, it is found that gastrointestinal cancer patients might experience multiple primary tumors in the digesting tract. Yoshida et al. reported that 38% of esophageal squamous cell carcinoma patients experienced multiple primary cancers, of which 15% underwent gastrointestinal cancer, mainly in the stomach and colorectal^[53]. Moreover, Steffi et al. found that approximately one in 20 patients with primary esophageal squamous cell carcinoma had a second primary tumor in the upper digestive tract or stomach^[50]. Furthermore, it was reported that several

cases related to the synchronous or metachronous primary gastrointestinal tract malignancies^[3; 7; 23; 54]. Hence, we identified some hub genes including Collagen Type I Alpha 1 Chain (COL1A1) of multiple primary tumors in the gastrointestinal tract based on bioinformatics technology.

COL1A1 is the gene which encodes the pro-alpha1 chains of type I collagen whose triple helix comprises two alpha 1 chains and one alpha 2 chain^[39]. The protein encoded by which is an important component of the extracellular matrix (ECM). COL1A1 is known to be overexpressed in several cancers other than gastrointestinal cancer^[44; 52; 57], such as thyroid cancer^[18], lung cancer^[13], breast cancer^[31], renal cancer^[5], etc. However, there is no pan-cancer analysis to comprehensively elucidate the potential role of COL1A1 in various tumor types. Thus we expand our research from gastrointestinal cancer to pan-cancer analysis of COL1A1.

2. Materials And Methods

2.1 Microarray data of gastrointestinal cancer

We downloaded 4 gene expression profiles from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>)^[4]: the GSE20347 of esophageal squamous cell carcinoma (ESCC), the GSE92396 of esophageal adenocarcinoma (EAC), the GSE103236 of stomach adenocarcinoma (STAD), and the GSE110224 of colorectal adenocarcinoma (CRAC). GSE20347^[17] was based on GPL571 ([HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array) and contained 17 pairs of esophageal squamous cell carcinoma tissues and matched normal adjacent tissues. GSE92396 was based on GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]), and contained 12 esophageal adenocarcinoma tissues and 10 normal esophageal tissues. GSE103236^[11] was based on GPL4133 (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version)), which contained 10 pairs of cancerous and normal adjacent tissue from gastric adenocarcinoma patients. GSE110224^[51] was based on GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). It included 17 pairs of histologically confirmed colorectal adenocarcinoma tissues and normal adjacent tissues.

2.2 Identification of the DEGs in Profile

GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r>) is a tool for analyzing differentially expressed genes in the GEO database, which can compare the expression of genes in tumor and normal samples. $\text{Adj.p} < 0.05$ and $|\log\text{FC}| > 1$ were set as the cut-off criterion to select DEGs for these microarray, respectively. Then the overlapping DEGs among the four datasets were identified by the online Venn diagram tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

2.3 Gene ontology (GO) and pathway enrichment analysis of DEGs

Database for Annotation, Visualization and Integrated Discovery (DAVIDv6.8 <https://david.ncifcrf.gov/>) provides a comprehensive set of functional annotation tools to understand the biological meaning

behind a large number of genes^[19]. DAVID was employed to carry out gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of DEGs. The GO and KEGG pathways were plotted by <http://www.bioinformatics.com.cn>, an online platform for data analysis and visualization.

2.4 Hub genes screening from the PPI network

The protein-protein interaction (PPI) network of differentially expressed genes was constructed based on the online website STRING (STRING; <http://string-db.org>) (version 11.0)^[47], and was further illustrated by the Cytoscape software^[43]. The MCODE plug-in in Cytoscape was utilized to identify key modules and hub genes. The preferred cut-off values were determined as degree cut-off values = 2, max. depth = 100, node score = 0.2 and the k-score = 2.

2.5 Gene expression analysis in pan-cancer

Oncomine(<https://www.oncomine.org>) database is currently the world's largest cancer gene-chip database and integrated data-mining platform, which can analyze differential gene expressions in normal and tumor tissues^[41]. Firstly, we used Oncomine to analyze the differential expression of COL1A1 between tumor tissues and normal tissues. Next, the “Gene_DE” module of TIMER2(tumor immune estimation resource, version 2) (<http://timer.cistrome.org/>) was employed to analyze the differential expression of COL1A1 in different tumors and normal tissues^[28]. For those tumors lack of normal or with highly limited normal tissues, the “Expression analysis-Box Plots” module of the GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2) (<http://gepia2.cancer-pku.cn/#analysis>)^[48] was used to obtain the expression difference between these tumor tissues and the corresponding normal tissues. In addition, we explored the COL1A1 expression difference in different stages by the “Pathological Stage Plot” module of GEPIA2.

2.6 Genetic alteration analysis

CBioportal(<https://www.cbioportal.org>) is an online database which provides visualization, analysis, and download of large-scale cancer genomics data^[9]. Herein, cBioportal was employed to obtain alteration frequency, mutation type of COL1A1 across all TCGA tumors. Then we explored the overall, disease-specific, progression-free, and disease-free survival differences with or without SND1 genetic alteration of the most alteration frequency type of tumor.

2.7 Methylation Profile of COL1A1

UALCAN is a comprehensive and interactive web resource for analyzing cancer OMICS data^[10]. We investigated COL1A1 promoter DNA methylation level in gastrointestinal cancer and some certain types of cancer by UALCAN. MEXPRESS is a web tool which can visualize DNA methylation, expression and clinical data^[25]. MEXPRESS was employed to determine the association between COL1A1 DNA methylation and clinical data.

2.8 Survival prognosis analysis

We used the “Survival Map” module of GEPIA2 to analyze the overall (OS) and disease-free survival (DFS) of COL1A1 among all tumors. Furthermore, the “Survival Analysis” module was used to further analyze the survival outcome of the specific type of tumor.

2.9 Immune infiltration analysis

The occurrence and development of tumors are closely related to the tumor immune microenvironment. We used the “Immune-Gene” module of the TIMER2 to analyze the relationship between COL1A1 expression and cancer-associated fibroblasts, CD8+ T cells, and macrophages.

TISIDB(<http://cis.hku.hk/TISIDB/index.php>) is an integrated repository portal for tumor and immune system interaction, which integrates multiple heterogeneous data types^[42]. We used TISIDB to analyze COL1A1 expression and tumor-infiltrating lymphocytes(TILs), immunoinhibitor, immunostimulator, and MHC molecule.

2.10 PD-L1, TMB and MSI in pan-cancer

PD-L1, TMB and MSI are important predictive markers of immunotherapy. We used the GEPIA to analyze the association of COL1A1 expression and PD-L1(CD272) among different cancers. RNA-seq data of 33 types of tumor was downloaded from the Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) data portal website(<https://portal.gdc.cancer.gov/>). TMB is derived from the article The Immune Landscape of Cancer published by Vestein Thorsson et al. in 2018^[49]; MSI is derived from the Landscape of Microsatellite Instability Across 39 Cancer Types article published by Russell Bonneville et al. in 2017^[6]. R software v4.0.3 was utilized for statistical analysis. If not otherwise stated, the rank-sum test detects two sets of data, and a P-value of <0.05 is considered statistically significant.

2.11 Genes co-expressed with COL1A1 in pan-cancer: pathway and drug sensitivity analysis

We used the STRING database (<https://string-db.org/>)^[47] to obtain the top 50 proteins most relevant to COL1A1 protein expression. The parameters are set as follows: minimum required interaction score ["Low confidence (0.150)"], meaning of network edges ("evidence"), max number of interactors to show ("no more than 50 interactors" in 1st shell), and active interaction sources ("Experiments"). Moreover, we used the GEPIA2 database to obtain the top 100 genes most correlated to COL1A1. Then we drew the Venn diagram to get the overlapping genes by the online Venn diagram tool(<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Next, we used GSCALite (<http://bioinfo.life.hust.edu.cn/web/GSCALite/>)^[30] to conduct pathway and drug sensitivity analysis on COL1A1 and the obtained genes.

3. Results

3.1 Identification of DEGs in gastrointestinal cancer

Based on the cut-off criterion, four mRNA expression profiles of esophageal squamous cell carcinoma, esophageal adenocarcinoma, gastric adenocarcinoma, and colorectal adenocarcinoma were extracted from the GEO database, including GSE20347, GSE92396, GSE103236, and GSE110224, respectively. After analyzing by the GEO2R tool and Venn diagram, a total of 21 DEGs were identified. And, 19 genes were up-regulated, and 2 genes were down-regulated. The Venn diagram was shown in the Figure.1a and Figure.1b.

3.2 GO and KEGG Analysis for the DEGs in gastrointestinal cancer

For all DEGs, extracellular regions, cell surfaces, proteinaceous extracellular matrix, and collagen trimer were mostly enriched in cellular components(CC). The biological process(BP) mainly included the collagen catabolic process, extracellular matrix organization and disassembly, cell adhesion, cartilage development, and skeletal system development. As for molecular function(MF), the DEGs were related to extracellular matrix structural constituent, serine-type endopeptidase activity, metalloendopeptidase activity, protein binding, and extracellular matrix binding. (Figure.1c) Furthermore, the KEGG pathway includes focal adhesion, ECM-receptor interaction, PI3K-Akt signaling pathway, protein digestion and absorption, and transcriptional misregulation in cancer(Figure. 1d).

3.3 PPI network construction and modules analysis

The PPI network of the DEGs was constructed through the STRING online website (Figure.1e). Thus the data was imported into Cytoscape software for visualization, and the MCODE plug-in was used to further screen the hub gene. Finally, a total of 6 hub genes were identified, namely SPP1, BGN, THBS2, MMP3, COL1A1, and TIMP1(Figure. 1f). All the hub genes were up-regulated in tumor tissues.

The above results showed that the hub genes were mainly enriched in the extracellular matrix by GO and KEGG. And the protein of COL1A1 is an important component of the extracellular matrix. Moreover, we found that COL1A1 was highly-expressed in other cancers in addition to gastrointestinal cancer^[5; 13; 18; 31]. Therefore, we aimed to explore COL1A1 in pan-cancer.

3.4 Gene expression analysis of COL1A1 in pan-cancer

Through the Oncomine database, we found that COL1A1 was highly expressed in a variety of tumors(Figure. 2a). Moreover, Through the TIMER2 database and GEPIA2 database, we found that COL1A1 was over-expressed mainly in breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon adenocarcinoma(COAD), esophageal adenocarcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), renal clear cell carcinoma (KIRC), hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid adenocarcinoma (THCA), diffuse large B lymphoma (DLBC), testicular germ cell tumors (TGCT), and thymoma (THYM)(Figure. 2b-c).

We further analyzed the expression of COL1A1 in different tumor stages. The results showed that COL1A1 expression is closely related to the late stage of ACC, BCLA, ESCA, KICH, KIRP, STAD, and THCA(Figure. 2d).

3.5 Genetic alteration analysis of COL1A1

We found COL1A1 mutation in various tumors by cBioportal. Among them, the highest mutation frequency was in melanoma(16.22%). Breast invasive carcinoma had the highest amplification frequency(6%) and sarcoma had the highest fusion frequency of COL1A1. It is noteworthy that all mesothelioma cases had copy number amplification(4.6%) and all uveal melanoma had deletion of COL1A1(1.25%) (Supplementary Figure.1a). Furthermore, the figure presented the types, sites, and case number of the COL1A1 genetic alteration. The most common alteration type was missense mutation(Supplementary Figure.1b). In addition, we analyzed the relationship between genetic changes and the prognosis of melanoma. The results showed that the alternations of COL1A1 are related to the better prognosis of melanoma with overall survival($p=0.0158$), progression-free survival($p=0.0385$), and disease-specific survival($p=0.0361$). Although these alternations showed a tendency of better disease-free survival, there is not statistically significant($p=0.395$) (Supplementary Figure.1c).

3.6 Methylation Profile of COL1A1

COL1A1 promoter methylation levels were significantly lower in primary tumors than in normal tissues in gastrointestinal cancer(Supplementary Figure.2a). The promoter methylation levels of other tumors were shown in supplementary Figures.3. Moreover, we investigated the detailed information of the correlation between COL1A1 expression level and DNA methylation, copy number and clinical data in STAD(Supplementary Figure.2b). The results showed that COL1A1 expression was significantly correlated with DNA methylation of CPG islands($p < 0.01$).

3.7 The relationship between survival and gene expression of COL1A1 in pan-cancer

In order to analyze the correlation between COL1A1 gene expression and prognosis, we used GEPIA2 to analyze the OS and DFS of COL1A1 among different cancers. The results indicated that high expression of COL1A1 is related to poor overall survival of KIRP (HR=2.2, $p = 0.011$), LGG (HR=2, $p = 0.00028$), MESO (HR=2.2, $p = 0.0014$), SKCM (HR=1.5, $p = 0.0032$), and STAD (HR=1.5, $p = 0.013$) (Figure. 3a). Moreover, high expression of COL1A1 is related to poor disease-free survival of CESC (HR=2, $p = 0.018$), COAD (HR=1.6, $p = 0.041$), ESCA (HR=1.6, $p = 0.045$), KIRP (HR=3.1, $p < 0.001$), LGG (HR=1.4, $p=0.024$), and PRAD (HR=2, $p=0.0011$) (Figure. 3b).

3.8 Immune infiltration analysis of COL1A1 in pan-cancer

In addition, we observed a strong positive correlation between COL1A1 expression and cancer-associated fibroblasts in most tumors. Moreover, BLCA, ESCA, HNSC, COAD, READ, STAD showed a statistically positive correlation of COL1A1 expression and macrophages. However, CESC, HNSC-HPV+, and SKCM showed a negative relationship between COL1A1 expression and CD8+ T cells(Figure. 4a). Furthermore,

we presented the scatterplot with purity and infiltration level of COL1A1 in certain types of tumors(Figure. 4b). And the scatterplot showed consistent results with heatmap.

The results of TISIDB showed that COL1A1 expression is positively related to TILs in most tumors, especially THCA(Figure. 4c). Whereas the association of COL1A1 expression with MHC molecule(Figure. 4d), immunoinhibitor(Figure. 4e), and immunostimulator(Figure. 4f) are diverse. LGG and THCA showed positive relation in MHC molecule. THCA showed positive relation and TGCT showed negative relation in most kinds of immunoinhibitors. And TGCT showed negative relation in most kinds of immunostimulator.

3.9 The relationship of COL1A1 expression and PD-L1, TMB/MSI in pan-cancer

We found that COL1A1 expression was positive correlative to PD-L1 in COAD, DLBC, GBM, LAML, LGG, LIHC, LUAD, PAAD, PRAD, and THCA(Figure. 5a). Moreover, through analysis of COL1A1 expression and TMB/MSI, we found that COL1A1 is closely related to TMB and MSI in a variety of tumors, and can be used as a predictor of immunotherapy. Among them, COL1A1 expression was associated with high TMB in THYM, LAML, ACC, KICH, PRAD, and LGG(Figure. 5b). What's more, high MSI was observed in TGCT, MESO, PRAD, COAD, SARC, and CESC(Figure. 5c).

3.10 Genes co-expressed with COL1A1 in pan-cancer: pathway and drug sensitivity analysis

We obtained 50 experimentally verified proteins related to COL1A1 based on STRING(Supplementary Figure 4a). Besides, we obtained 100 genes related to COL1A1 expression from GEPIA2. Then the Venn diagram showed the overlapping genes(Supplementary Figure 4b), of which 8 genes were closely related to the expression of COL1A1, namely MMP2, SPARC, COL5A1, DCN, BMP1, BGN, COL1A2, and HTRA1. Heatmaps(Supplementary Figure 4c) and scatter plots(Supplementary Figure 4d) confirmed that these gene expressions were closely related to COL1A1 among tumors. Then, we performed pathway enrichment and drug sensitivity analysis of COL1A1 and these 8 genes in 33 tumors by GSCALite. The results showed that the pathway was mainly activated in epithelial-mesenchymal transition (EMT) (Supplementary Figure 5a), and high expression of HTRA1 was resistant to multiple drugs(Supplementary Figure 5b).

4. Discussion

Recently, reports have been gradually increasing in multiple primary tumors of the digestive tract. However, it is difficult to distinguish them from metastatic cancer, and the diagnosis is relatively difficult. In this study, we screened out 4 chips of esophageal, gastric, and colorectal cancer from the GEO database and identified 21 DEGs by using bioinformatics analysis. The proteins encoded by these genes were mainly enriched in the extracellular region, cell surface, proteinaceous extracellular matrix, and collagen trimer. Further PPI network analysis identified six hub genes, namely SPP1, BGN, THBS2, MMP3, COL1A1, and TIMP1. All of these genes were upregulated in gastrointestinal cancer.

SPP1, also known as osteopontin, is a secreted arginine glycine aspartic acid-containing phosphorylated glycoprotein. It could upregulate the expression of interleukin-10 (IL-10), interleukin-12 (IL-12), interferon- γ (IFN- γ), and is essential in the pathway that leads to type I immunity^[20; 24]. BGN is an important component of the extracellular matrix and plays a vital role in the regulation of epithelial cell morphology, growth, migration and differentiation. It has been confirmed that BGN is upregulated in gastrointestinal cancer and promotes resistance to chemotherapy^[16; 29; 60]. THBS2 belongs to the thrombospondin family and is a versatile glycoprotein that can be secreted by stromal fibroblasts, endothelial cells, immune cells, etc^[1]. It is related to angiogenesis and tumor metastasis^[21]. The protein encoded by TIMP1 is a natural inhibitor of matrix metalloproteinases (MMPs)^[27]. It can regulate tumorigenesis through ERK, JAK/STAT, and MAPK signaling pathways^[45]. Previous researches have confirmed that these genes were closely related to tumor invasion, metastasis, and drug resistance in esophageal, gastric, and colorectal cancer^[2; 16; 29; 45; 55]. Therefore, these hub genes could be viewed as the candidates as prognostic biomarkers for multiple primary tumors of the digestive tract.

COL1A1 is the main member of the type I collagen family, and the protein is a component of the extracellular matrix. Liu et al found that COL1A1 was closely related to Helicobacter pylori (+) gastric cancer through bioinformatics technology^[33]. Moreover, it is associated with serous membrane infiltration, lymph node metastasis, and hematogenous metastasis in colorectal cancer regardless of KRAS mutation^[57]. Through literature review, we found that it is not only highly expressed in gastrointestinal cancer, but also highly expressed in other tumors, and is involved in tumorigenesis, metastasis, and immune infiltration^[12; 34; 56]. Besides, it is the downstream molecule of multiple microRNAs. MiR-133b, miR-29b-3p, miR-133a-3p, and miR-196b-5p can inhibit tumor invasion and metastasis, and reverse drug resistance through targeting COL1A1^[14; 22; 52; 59]. Therefore, we expanded our research scope to pan-cancer.

The tumor microenvironment(TME) is a dynamic, complex, and heterogenic system which is mainly composed of ECM, blood vessels, tumor-infiltrating immune cells, CAFs, and cytokines, etc^[37; 40; 46]. It plays a critical role in tumorigenesis and metastasis and is responsible for cancer-target therapies. An important finding of this study was that the expression of COL1A1 was positively correlated with the abundance of CAFs in almost all tumors. Minna et al. found that CAFs coexisted with senescent thyroid cells at the tumor invasive front, accompanied by the deposition of COL1A1^[36]. This phenomenon needs further validation in other tumors. Besides, the expression of COL1A1 was found negatively correlated with CD8+ T cells in CESC, HNSC-HPV+, and SKCM, which may associated with deletion of CD8 + T cells induced by CAFs^[26]. Furthermore, we also found that the expression of COL1A1 was positively correlated with macrophage infiltration in gastrointestinal cancer, BLCA, and HNSC. In the future, more experiments could focus on whether targeting COL1A1 can we inhibit CAFs and macrophages, and promote tumor-specific CD8+ T cell differentiation, thus improving immunotherapy resistance. TMB and MSI are important predictors for the efficacy of cancer immunotherapy. By analyzing the information obtained from the TCGA database, we found that the high expression of COL1A1 was associated with high TMB

and MSI in many tumors. After literature review, we found that there was no research regarding the relationship between COL1A1 expression and TMB and MSI. Therefore, this study provided a good basis for predicting the efficacy of immunotherapy, but more evidence is needed to support this result.

In addition, we found that COL1A1 was closely related to activation of EMT through pathway enrichment analysis. EMT is a dynamic process that inactive polarized epithelial cells transit to active mesenchymal cells^[38]. Researches have found that knockdown of COL1A1 can inhibit the EMT process, thereby inhibiting the invasion and metastasis of liver cancer and bladder cancer^[34; 58]. Furthermore, through literature review, we found that the up-regulation of COL1A1 was related to radioresistance, and experiments have confirmed that knockdown of COL1A1 can increase the radiosensitivity of nasopharyngeal and cervical cancer^[15; 32]. Moreover, EMT was related to radioresistance^[35], which indicated that the radioresistance of COL1A1 may be related to EMT. More researches could be made to further verify whether the process of EMT can be down-regulated by inhibiting COL1A1, thus reducing tumor metastasis, and improving chemoresistance, radioresistance, and immunoresistance.

Nevertheless, although we employed multiple bioinformatics databases to analyze the role of COL1A1 across 33 tumors, this study still has some limitations. Firstly, the current study merely focused on COL1A1 at the genetic level, but lacked description of its protein transcriptome and phosphorylation level. Secondly, the results of our study lack external validation in other public datasets. Finally, the databases used in this study were mainly based on the gene-chip and sequencing data of tumor tissue from TCGA, so the cell-level analysis of immune cell markers could have introduced systematic bias. Future studies should focus more on using skills like single-cell sequencing technology.

5. Conclusion

In conclusion, we screened 4 chips, and found 6 hub genes including COL1A1, and firstly conducted a pan-cancer analysis to illustrate the gene expression, prognosis, and immunological role of COL1A1. The results demonstrated that COL1A1 can serve as a prognostic and immunological biomarker in multiple cancers.

Declarations

Funding This study was supported by the Natural Science Foundation Project of Chongqing Science and Technology Commission (CSTC), China (grant no. cstc2018jcyjAX0012).

Conflicts of Interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of data and material All data is available under reasonable request.

Code availability Not applicable.

Author Contributions YL wrote the initial manuscript. JX, MZ, ZW, and HE edited and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate Not applicable.

Consent for publication All authors consent to the publication of this study.

References

1. Adams, J. C., & Lawler, J. (2004). The thrombospondins. *Int J Biochem Cell Biol*, 36(6), 961-968. doi:10.1016/j.biocel.2004.01.004
2. Ao, R., Guan, L., Wang, Y., & Wang, J. N. (2018). Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. *J Cell Biochem*, 119(6), 4420-4434. doi:10.1002/jcb.26524
3. Arakawa, K., Hata, K., Yamamoto, Y., Nishikawa, T., Tanaka, T., Kiyomatsu, T., . . . Watanabe, T. (2018). Nine primary malignant neoplasms-involving the esophagus, stomach, colon, rectum, prostate, and external ear canal-without microsatellite instability: a case report. *BMC Cancer*, 18(1), 24. doi:10.1186/s12885-017-3973-2
4. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., . . . Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*, 41(Database issue), D991-995. doi:10.1093/nar/gks1193
5. Boguslawska, J., Kedzierska, H., Poplawski, P., Rybicka, B., Tanski, Z., & Piekliko-Witkowska, A. (2016). Expression of Genes Involved in Cellular Adhesion and Extracellular Matrix Remodeling Correlates with Poor Survival of Patients with Renal Cancer. *J Urol*, 195(6), 1892-1902. doi:10.1016/j.juro.2015.11.050
6. Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H. Z., . . . Roychowdhury, S. (2017). Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol*, 2017. doi:10.1200/po.17.00073
7. Bratislav, T., Branimir, N., Mihailo, B., Jovan, K., Veljović, M., & Dejan, Z. (2015). Synchronous gastric and colonic cancer—A case report. *Vojnosanit Pregl*, 72(7), 642-645.
8. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6), 394-424. doi:10.3322/caac.21492
9. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, 2(5), 401-404. doi:10.1158/2159-8290.Cd-12-0095

10. Chandrashekar, D. S., Bachel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., & Varambally, S. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*, 19(8), 649-658. doi:10.1016/j.neo.2017.05.002
11. Chivu Economescu, M., Necula, L. G., Dragu, D., Badea, L., Dima, S. O., Tudor, S., . . . Diaconu, C. C. (2010). Identification of potential biomarkers for early and advanced gastric adenocarcinoma detection. *Hepatogastroenterology*, 57(104), 1453-1464.
12. Dong, S., Zhu, P., & Zhang, S. (2020). Expression of collagen type 1 alpha 1 indicates lymph node metastasis and poor outcomes in squamous cell carcinomas of the lung. *PeerJ*, 8, e10089. doi:10.7717/peerj.10089
13. Grigoriu, M., Tagett, R., Draghici, S., Dima, S., Nastase, A., Florea, R., . . . Regnard, J. F. (2015). Gene-expression Profiling in Non-small Cell Lung Cancer with Invasion of Mediastinal Lymph Nodes for Prognosis Evaluation. *Cancer Genomics Proteomics*, 12(5), 231-242.
14. Guo, Y., Lu, G., Mao, H., Zhou, S., Tong, X., Wu, J., . . . Fang, F. (2020). miR-133b Suppresses Invasion and Migration of Gastric Cancer Cells via the COL1A1/TGF- β Axis. *Onco Targets Ther*, 13, 7985-7995. doi:10.2147/ott.S249667
15. Guo, Y., Zhai, J., Zhang, J., Ni, C., & Zhou, H. (2019). Improved Radiotherapy Sensitivity of Nasopharyngeal Carcinoma Cells by miR-29-3p Targeting COL1A1 3'-UTR. *Med Sci Monit*, 25, 3161-3169. doi:10.12659/msm.915624
16. Hu, L., Duan, Y. T., Li, J. F., Su, L. P., Yan, M., Zhu, Z. G., . . . Yang, Q. M. (2014). Biglycan enhances gastric cancer invasion by activating FAK signaling pathway. *Oncotarget*, 5(7), 1885-1896. doi:10.18632/oncotarget.1871
17. Hu, N., Clifford, R. J., Yang, H. H., Wang, C., Goldstein, A. M., Ding, T., . . . Lee, M. P. (2010). Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics*, 11, 576. doi:10.1186/1471-2164-11-576
18. Huang, C., Yang, X., Han, L., Fan, Z., Liu, B., Zhang, C., & Lu, T. (2019). The prognostic potential of alpha-1 type I collagen expression in papillary thyroid cancer. *Biochem Biophys Res Commun*, 515(1), 125-132. doi:10.1016/j.bbrc.2019.04.119
19. Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57. doi:10.1038/nprot.2008.211
20. Icer, M. A., & Gezmen-Karadag, M. (2018). The multiple functions and mechanisms of osteopontin. *Clin Biochem*, 59, 17-24. doi:10.1016/j.clinbiochem.2018.07.003
21. Iruela-Arispe, M. L., Luque, A., & Lee, N. (2004). Thrombospondin modules and angiogenesis. *Int J Biochem Cell Biol*, 36(6), 1070-1078. doi:10.1016/j.biocel.2004.01.025
22. Jia, R., & Wang, C. (2020). MiR-29b-3p Reverses Cisplatin Resistance by Targeting COL1A1 in Non-Small-Cell Lung Cancer A549/DDP Cells. *Cancer Manag Res*, 12, 2559-2566. doi:10.2147/cmar.S246625

23. Kim, S. H., Park, B. S., Kim, H. S., & Kim, J. H. (2017). Synchronous quintuple primary gastrointestinal tract malignancies: Case report. *World J Gastroenterol*, 23(1), 173-177. doi:10.3748/wjg.v23.i1.173
24. Klement, J. D., Paschall, A. V., Redd, P. S., Ibrahim, M. L., Lu, C., Yang, D., . . . Liu, K. (2018). An osteopontin/CD44 immune checkpoint controls CD8+ T cell activation and tumor immune evasion. *J Clin Invest*, 128(12), 5549-5560. doi:10.1172/jci123360
25. Koch, A., De Meyer, T., Jeschke, J., & Van Criekinge, W. (2015). MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics*, 16(1), 636. doi:10.1186/s12864-015-1847-z
26. Lakins, M. A., Ghorani, E., Munir, H., Martins, C. P., & Shields, J. D. (2018). Cancer-associated fibroblasts induce antigen-specific deletion of CD8 (+) T Cells to protect tumour cells. *Nat Commun*, 9(1), 948. doi:10.1038/s41467-018-03347-0
27. Lambert, E., Dassé, E., Haye, B., & Petitfrère, E. (2004). TIMPs as multifacial proteins. *Crit Rev Oncol Hematol*, 49(3), 187-198. doi:10.1016/j.critrevonc.2003.09.008
28. Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., . . . Liu, X. S. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*, 48(W1), W509-w514. doi:10.1093/nar/gkaa407
29. Liu, B., Xu, T., Xu, X., Cui, Y., & Xing, X. (2018). Biglycan promotes the chemotherapy resistance of colon cancer by activating NF- κ B signal transduction. *Mol Cell Biochem*, 449(1-2), 285-294. doi:10.1007/s11010-018-3365-1
30. Liu, C. J., Hu, F. F., Xia, M. X., Han, L., Zhang, Q., & Guo, A. Y. (2018). GSCALite: a web server for gene set cancer analysis. *Bioinformatics*, 34(21), 3771-3772. doi:10.1093/bioinformatics/bty411
31. Liu, J., Shen, J. X., Wu, H. T., Li, X. L., Wen, X. F., Du, C. W., & Zhang, G. J. (2018). Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. *Discov Med*, 25(139), 211-223.
32. Liu, S., Liao, G., & Li, G. (2017). Regulatory effects of COL1A1 on apoptosis induced by radiation in cervical cancer cells. *Cancer Cell Int*, 17, 73. doi:10.1186/s12935-017-0443-5
33. Liu, Y., Zhu, J., Ma, X., Han, S., Xiao, D., Jia, Y., & Wang, Y. (2019). ceRNA network construction and comparison of gastric cancer with or without Helicobacter pylori infection. *J Cell Physiol*, 234(5), 7128-7140. doi:10.1002/jcp.27467
34. Ma, H. P., Chang, H. L., Bamodu, O. A., Yadav, V. K., Huang, T. Y., Wu, A. T. H., . . . Lee, W. H. (2019). Collagen 1A1 (COL1A1) Is a Reliable Biomarker and Putative Therapeutic Target for Hepatocellular Carcinogenesis and Metastasis. *Cancers (Basel)*, 11(6). doi:10.3390/cancers11060786
35. Marie-Egyptienne, D. T., Lohse, I., & Hill, R. P. (2013). Cancer stem cells, the epithelial to mesenchymal transition (EMT) and radioresistance: potential role of hypoxia. *Cancer Lett*, 341(1), 63-72. doi:10.1016/j.canlet.2012.11.019
36. Minna, E., Brich, S., Todoerti, K., Pilotti, S., Collini, P., Bonaldi, E., . . . Borrello, M. G. (2020). Cancer Associated Fibroblasts and Senescent Thyroid Cells in the Invasive Front of Thyroid Carcinoma. *Cancers (Basel)*, 12(1). doi:10.3390/cancers12010112

37. Najafi, M., Goradel, N. H., Farhood, B., Salehi, E., Solhjoo, S., Toolee, H., . . . Mortezaee, K. (2019). Tumor microenvironment: Interactions and therapy. *J Cell Physiol*, 234(5), 5700-5721. doi:10.1002/jcp.27425
38. Nieto, M. A., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2016). EMT: 2016. *Cell*, 166(1), 21-45. doi:10.1016/j.cell.2016.06.028
39. Prockop, D. J. (1990). Mutations that alter the primary structure of type I collagen. The perils of a system for generating large structures by the principle of nucleated growth. *J Biol Chem*, 265(26), 15349-15352.
40. Quail, D. F., & Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat Med*, 19(11), 1423-1437. doi:10.1038/nm.3394
41. Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., . . . Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*, 6(1), 1-6. doi:10.1016/s1476-5586(04)80047-2
42. Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., . . . Zhang, J. (2019). TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics*, 35(20), 4200-4202. doi:10.1093/bioinformatics/btz210
43. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11), 2498-2504. doi:10.1101/gr.1239303
44. Shi, Y., Wang, J., Xin, Z., Duan, Z., Wang, G., & Li, F. (2015). Transcription factors and microRNA-co-regulated genes in gastric cancer invasion in ex vivo. *PLoS One*, 10(4), e0122882. doi:10.1371/journal.pone.0122882
45. Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T., . . . Wang, X. (2016). TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J Exp Clin Cancer Res*, 35(1), 148. doi:10.1186/s13046-016-0427-7
46. Spano, D., & Zollo, M. (2012). Tumor microenvironment: a main actor in the metastasis process. *Clin Exp Metastasis*, 29(4), 381-395. doi:10.1007/s10585-012-9457-5
47. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., . . . Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47(D1), D607-d613. doi:10.1093/nar/gky1131
48. Tang, Z., Kang, B., Li, C., Chen, T., & Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*, 47(W1), W556-w560. doi:10.1093/nar/gkz430
49. Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., . . . Shmulevich, I. (2019). The Immune Landscape of Cancer. *Immunity*, 51(2), 411-412. doi:10.1016/j.immuni.2019.08.004

50. van de Ven, S. E., Falger, J. M., Verhoeven, R. H., Baatenburg de Jong, R. J., Spaander, M. C., Bruno, M. J., & Koch, A. D. (2020). Increased risk of second primary tumours in patients with oesophageal squamous cell carcinoma: A nationwide study in a Western population. *United European Gastroenterol J*, 2050640620977129. doi:10.1177/2050640620977129
51. Vlachavas, E. I., Pilalis, E., Papadodima, O., Koczan, D., Willis, S., Klippel, S., . . . Chatziioannou, A. (2019). Radiogenomic Analysis of F-18-Fluorodeoxyglucose Positron Emission Tomography and Gene Expression Data Elucidates the Epidemiological Complexity of Colorectal Cancer Landscape. *Comput Struct Biotechnol J*, 17, 177-185. doi:10.1016/j.csbj.2019.01.007
52. Yin, Y., Du, L., Li, X., Zhang, X., & Gao, Y. (2019). miR-133a-3p suppresses cell proliferation, migration, and invasion and promotes apoptosis in esophageal squamous cell carcinoma. *J Cell Physiol*, 234(8), 12757-12770. doi:10.1002/jcp.27896
53. Yoshida, N., Eto, K., Kurashige, J., Izumi, D., Sawayama, H., Horinouchi, T., . . . Baba, H. (2020). Comprehensive Analysis of Multiple Primary Cancers in Patients with Esophageal Squamous Cell Carcinoma Undergoing Esophagectomy. *Ann Surg*. doi:10.1097/sla.0000000000004490
54. Yoshikawa, T., Aoki, K., Mitsuhashi, Y., Tomiura, S., Suto, A., Miura, T., . . . Endo, M. (2016). [A Case of Double Cancer of Initially Unresectable Sigmoid Colon Cancer and Advanced Gastric Cancer Treated with Curative Resection after mFOLFOX6 Therapy]. *Gan To Kagaku Ryoho*, 43(3), 365-368.
55. Z, Z., Y, W., J, Z., J, Z., & R, Y. (2018). COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Molecular medicine reports*, 17(4), 5037-5042. doi:10.3892/mmr.2018.8533
56. Zhang, C., Liu, S., Wang, X., Liu, H., Zhou, X., & Liu, H. (2021). COL1A1 Is a Potential Prognostic Biomarker and Correlated with Immune Infiltration in Mesothelioma. *Biomed Res Int*, 2021, 5320941. doi:10.1155/2021/5320941
57. Zhang, Z., Fang, C., Wang, Y., Zhang, J., Yu, J., Zhang, Y., . . . Zhong, J. (2018). COL1A1: A potential therapeutic target for colorectal cancer expressing wild-type or mutant KRAS. *Int J Oncol*, 53(5), 1869-1880. doi:10.3892/ijo.2018.4536
58. Zhu, H., Chen, H., Wang, J., Zhou, L., & Liu, S. (2019). Collagen stiffness promoted non-muscle-invasive bladder cancer progression to muscle-invasive bladder cancer. *Onco Targets Ther*, 12, 3441-3457. doi:10.2147/ott.S194568
59. Zhu, X., Rao, X., Yao, W., & Zou, X. (2018). Downregulation of MiR-196b-5p impedes cell proliferation and metastasis in breast cancer through regulating COL1A1. *Am J Transl Res*, 10(10), 3122-3132.
60. Zhu, Y. H., Yang, F., Zhang, S. S., Zeng, T. T., Xie, X., & Guan, X. Y. (2013). High expression of biglycan is associated with poor prognosis in patients with esophageal squamous cell carcinoma. *Int J Clin Exp Pathol*, 6(11), 2497-2505.

Supplementary Information

Supplementary Figure 1a. Mutation feature of COL1A1 in different tumors of TCGA by cBioportal.

b. mutation site of COL1A1, which contains 262 missense, 48 truncating, 17 fusion, 1 inframe. **c.** COL1A1 mutation status and overall, disease-specific, disease-free and progression-free survival of melanoma.

Supplementary Figure 2 a.Box plots of COL1A1 methylation level in normal and gastrointestinal cancer tissues by UALCAN. **b.**The detailed information of correlation between COL1A1 expression and DNA methylation, copy number and clinical data in STAD by MEXPRESS.

Supplementary Figure 3 Box plots of COL1A1 methylation level in normal and tumor tissues in BRCA, LUSC, LUAD, BLCA, CESC, and HNSC.

Supplementary Figure 4 The top 50 experimentally determined COL1A1-binding proteins using STRING. **b.** The Venn diagram showing overlapping of top 50 COL1A1-binding proteins and top 100 COL1A1-correlated genes in GEPIA2. The overlapping genes were MMP2, SPARC, COL5A1, DCN, BMP1, BGN, COL1A2, and HTRA1. **c.**The heatmap confirmed the association of COL1A1 and MMP2, SPARC, COL5A1, DCN, BMP1, BGN, COL1A2, and HTRA1 in pan-cancer. **d.**Scatterplot showed the association of COL1A1 and MMP2, SPARC, COL5A1, DCN, BMP1, BGN, COL1A2, and HTRA1, respectively.

Supplementary Figure 5. Pathway enrichment(**a**) and drug sensitivity analysis(**b**) analysis of COL1A1, MMP2, SPARC, COL5A1, DCN, BMP1, BGN, COL1A2 based on GSCALite. The Spearman correlation represent the gene expression correlates with the drug. The positive correlation means that the gene high expression is resistant to the drug, vise verse.