Indian Journal of Experimental Biology
Vol. 62, March 2024, pp. 179-186
DOI: 10.56042/ijeb.v62i03.2203

# Systematic evaluation of collagen family gene transcripts: Implications for prognostic stratification in lung cancer patients

Sedef Hande Aktas[1,2]*

[1]Vocational School of Health Services, Eskisehir Osmangazi University, Eskisehir, Turkey

[2]Department of Biotechnology and Biosafety, Graduate School of Natural and Applied Sciences, Eskisehir Osmangazi University, Eskisehir, Turkey

*Received 03 June 2023*; *revised 23 October 2023*

Remodeling of the extracellular matrix (ECM) is expressed as one of the most powerful processes in the development of lung cancer (LC). Moreover, its alternative regulation showed significant relationship with a wide range of diseases and traits. Since the collagen family is one of the most important components of the ECM, in silico analyzes were performed to determine which collagen type among all collagens was significant for survival of lung cancer. Overall survival (OS) plots were generated for all collagen members and subsequently, the transcript levels of collagen members with significant OS values were examined using the GEPIA web tool. According to the survival graphs of 46 genes, the expressions of *COL6A4P2*, *COL6A6*, *COL7A1*, *COL11A1*, *COL19A1*, and *COL22A1* genes have a significant relationship with OS in Lung Adenocarcinoma (LUAD). *COL1A2*, *COL4A1*, *COL5A2*, *COL6A5*, and *COL28A1* genes have a significant relationship with OS for Lung Squamous Cell Carcinoma (LUSC) ($P \leq 0.05$). *COL6A6-001*, *COL6A6-002*, *COL11A1-001*, *COL11A1-007* transcripts for LUAD, and *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008*, *COL1A2-011*, *COL4A1-004*, *COL6A5-003*, *COL6A5-004* transcripts for LUSC were found to be significant. *COL6A6-002*, *COL11A1-001*, *COL11A1-007* and *COL1A2-005* transcripts were also found to be significant for stratifying pathological stages.

**Keywords:** Collagen, Lung cancer, Transcriptome

Lung cancer incidence (2480675) and mortality (1817469) rank at the top among all the cancer types[1]. To date, the optimal treatment strategy for lung cancer treatment has not been elucidated.

Many important events occur during the development of lung cancer. Remodeling of the extracellular matrix (ECM) is expressed as one of the most powerful processes in these events. The collagen family is one of the most important components of the ECM[2,3]. As the ECM structure is mainly responsible for the increased tissue pressure in the intercellular space and the rigid mechanical properties, ectopic expression of collagen family members can alter the properties of epithelial cells and reshape the morphology and function of epithelial cells[4,5]. This process enhances the separation of cancer cells from the primary tumor mass and increases cell invasiveness for the process leading to metastasis. In addition, it is known that the expression of different collagen families is evaluated as a biomarker in many types of cancer[6-10]. Furthermore, alternative regulation has been shown to significantly influence carcinogenesis, and differential isoform usage has been associated with a wide range of diseases and traits. Thus, the study of differential isoform usage has the potential to provide important information for understanding the biological mechanisms underlying disease[11,12].

The purpose of this research was to reveal significant prognostic collagen genes ($P \leq 0.05$) in lung cancer and to investigate the expression of the transcripts of these genes in detail. Thus, it may be possible to focus on a more specific gene/gene group in monitoring the prognosis of lung cancer and to use it for the diagnosis and treatment of lung cancer by selecting transcripts with significance.

In this study, we conducted *in silico* analyses to determine the collagen type that significantly affects lung cancer survival among all the collagens. We analysed the adenocarcinoma and squamous cell carcinoma subgroups of non-small cell lung cancer (NSCLC), which are the most common subtypes of lung cancer, for the collagen family. For this purpose, overall survival (OS) plots for all collagen members

*Correspondence:
E-Mail: sedefhande@gmail.com; shaktas@ogu.edu.tr
ORCID: 0000-0002-1091-6974

were performed using the GEPIA2 web tool. Subsequently, box plot expression analyses of the transcripts of these genes with significant OS were performed.

## Materials and Methods

### Data collection

The lung cancer (LC) data set, combined with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), was obtained from the cancer genome atlas (TCGA) database[13] (https://portal.gdc.cancer.gov/). Demographic and clinical data of the LC patients are summarized in Suppl Table S1 [*All supplementary data are available only online along with the respective paper at the journal website* (*http://ijeb.res.in*) *as well as NOPR repository at http://nopr.res.in*].

### Gene expression and overall survival analysis

Gene expression and overall survival analysis were performed with GEPIA2 web-based tool which provides functionalities for research based on TCGA and genotype-tissue expression (GTEx) data[14] (http://gepia2.cancer-pku.cn). A total of 483 lung adenocarcinoma samples and 347 normal samples were used to analyze the gene expression of the transcripts. Analyzing the gene expression of the transcripts, a total of 486 lung squamous cell carcinoma samples and 338 normal samples were used.

Overall survival analysis of 46 collagen genes given in detail was performed with available patient data (Suppl. Fig. S1). In addition, the transcripts whose expressions were significant are presented and discussed in the article, and the expression analyses of all transcripts are presented in Suppl. Fig. S2A (Lung adenocarcinoma) and Suppl. Fig. S2B (Lung squamous cell carcinoma). Survival analysis according to the low and high mRNA expression levels of the study genes was performed via the web interface.

### Statistical analysis

All statistical analyzes were carried out on the Gene Expression Profiling Interactive Analysis (GEPIA) database[15]. Kaplan-Meier analyses were conducted to estimate the OS of LC. Low and high-expression groups were compared using the log-rank test. The p-values for the analyses were calculated automatically by the database, and p-values below 0.05 were considered statistically significant.

## Results and Discussion

### Overall survival analysis

According to the survival graphs of 46 genes, the expressions of *COL6A4P2*, *COL6A6*, *COL7A1*, *COL11A1*, *COL19A1* and *COL22A1* genes have a significant relationship with overall survival (OS) for lung adenocarcinoma (LUAD) (*COL6A4P2* p=0.043; *COL6A6* p=0.019; *COL7A1* p=0.0052; *COL11A1* p=0.014; *COL19A1* p=0.0018; and *COL22A1* p=0.0049). In addition, *COL1A2*, *COL4A1*, *COL5A2*, *COL6A5* and *COL28A1* genes have a significant relationship with OS for lung squamous cell carcinoma (LUSC) (*COL1A2* p=0.023; *COL4A1* p=0.024; *COL5A2* p=0.026; *COL6A5* p=0.036; and *COL28A1* p=0.051) (Fig. 1). Survival graphs of 46 genes are given in detail in Suppl. Fig. S1.

### Gene expression analysis of transcripts of selected genes

Transcripts of genes are given in detail in Table 1 and the expressions of all genes are presented in Suppl. Figs S2 A and B. The expression of *COL6A6-001*, *COL6A6-002*, *COL11A1-001* and *COL11A1-007* transcripts was found significant for LUAD. The *COL6A6-001* and *COL6A6-002* transcripts were found high in normal tissues than tumor tissues whereas *COL11A1-001* and *COL11A1-007* transcripts were found high in tumor tissues than normal tissues ($P \leq 0.05$) (Fig. 2). The expression of *COL1A2-001*, *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008*, *COL1A2-011*, *COL4A1-004*, *COL6A5-003* and *COL6A5-004* transcripts were found significant for LUSC. The *COL1A2-001* transcript was found high in tumor tissues than in normal tissues whereas *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008*, *COL1A2-011*, *COL4A1-004*, *COL6A5-003* and *COL6A5-004* were found high in normal tissues than tumor tissues ($P \leq 0.05$) (Fig. 3).

#### *Pathological stage plot analysis*

The transcripts have a significant relationship with OS also analysed for pathological stage. *COL6A6-001*, *COL6A6-002*, *COL11A1-001* and *COL11A1-007* transcripts were analysed for LUAD. *COL1A2-001*, *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008*, *COL1A2-011*, *COL4A1-004*, *COL6A5-003* and *COL6A5-004* transcripts were analyzed for LUSC. *COL6A6-002*, *COL11A1-001* and *COL11A1-007* transcripts were found significant to distinguish pathological stages for LUAD. On the other hand, only *COL1A2-005* transcript was found

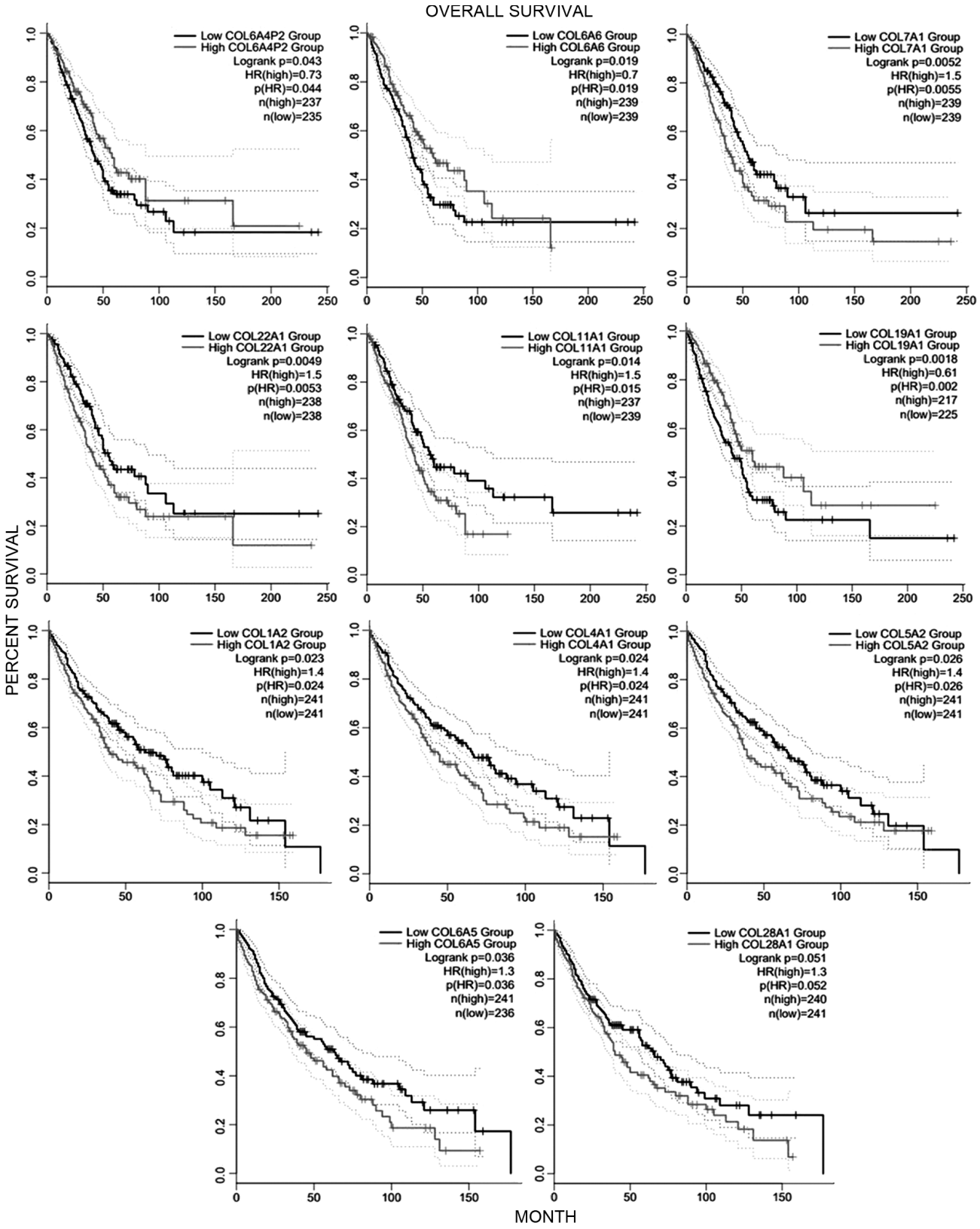


Fig 1 — Comparison of Kaplan-Meier OS curves of the high and low expression of genes have a significant relationship with OS for lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) among forty six collagen genes

significant to distinguish pathological stages for LUSC ($P \leq 0.05$) (Fig. 4). The transcripts found significantly stratifying patients pathological stages for LUAD are as follows in detail: *COL6A6-002* transcript; the corresponding p-value is 0.00596. As this value is well below the conventional 0.05 threshold for statistical significance, it implies that there is a statistically significant difference in the means across the four stages. Stages I & IV are notable for their sharply defined peaks, suggesting more uniform or specific measurements within these stages. This may indicate a more consistent observation or response in these groups. Stages II & III have broader peaks, suggesting a slightly wider spread of central data values compared to Stages I & IV. *COL11A1-001* transcript; the corresponding p-value is 0.03. Since this result is below the conventional significance threshold of 0.05, it indicates that there is a statistically significant difference in the means among the four stages. Stages I & IV have more pronounced double peaks in their distributions, indicating the possible presence of two distinct subgroups within these stages. The interquartile ranges (IQRs) are relatively consistent across all stages, suggesting a similar distribution of mean data values for each stage.

For *COL11A1-007* transcript the corresponding p-value was 0.0582. This result is slightly above the conventional 0.05 significance threshold, suggesting that the group means may not be statistically different at the 0.05 level. However, it is close enough to warrant further investigation or discussion of possible trends or practical significance. The medians across the stages appear to have shifted upward from Stage I to Stage IV. The IQRs for Stage I, II and IV are relatively similar, while Stage III has a slightly more extended IQR, suggesting a wider spread of mean data values.

The transcripts found significantly stratifying patients pathological stages for LUSC are as follows

Table 1 — Collagen gene transcripts and their overall survival statistical significance in LUAD and LUSC

| Lung Adenocarcınoma (LUAD) | | | Lung Squamous Cell Carcınoma (LUSC) | | |
|---|---|---|---|---|---|
| Gene | Transcrıpts of the gene | *P* value | Gene | Transcrıpts of the gene | *P* value |
| *COL6A4P2* | | | *COL1A2* | | |
| | *COL6A4P2-001* | N, S | | *COL1A2-001* | $P \leq 0, 05$ |
| | *COL6A4P2-002* | N, S | | *COL1A2-002* | $P \leq 0, 05$ |
| | *COL6A4P2-003* | N, S | | *COL1A2-003* | N, S |
| | *COL6A4P2-004* | N, S | | *COL1A2-004* | $P \leq 0, 05$ |
| | *COL6A4P2-005* | N, S | | *COL1A2-005* | $P \leq 0, 05$ |
| | *COL6A4P2-006* | N, S | | *COL1A2-006* | N, S |
| *COL6A6* | | | | *COL1A2-007* | $P \leq 0, 05$ |
| | *COL6A6-001* | $P \leq 0, 05$ | | *COL1A2-008* | $P \leq 0, 05$ |
| | *COL6A6-002* | $P \leq 0, 05$ | | *COL1A2-009* | N, S |
| | *COL6A6-003* | N, S | | *COL1A2-010* | N, S |
| *COL7A1* | | | | *COL1A2-011* | $P \leq 0, 05$ |
| | *COL7A1-001* | N, S | | *COL1A2-012* | N, S |
| | *COL7A1-002* | N, S | | *COL1A2-201* | N, S |
| | *COL7A1-003* | N, S | *COL4A1* | | |
| | *COL7A1-004* | N, S | | *COL4A1-001* | N, S |
| | *COL7A1-005* | N, S | | *COL4A1-003* | N, S |
| | *COL7A1-006* | N, S | | *COL4A1-004* | $P \leq 0, 05$ |
| | *COL7A1-007* | N, S | | *COL4A1-005* | N, S |
| | *COL7A1-008* | N, S | | *COL4A1-006* | N, S |
| | *COL7A1-009* | N, S | *COL5A2* | | |
| | *COL7A1-010* | N, S | | *COL5A2-001* | N, S |
| *COL11A1* | | | | *COL5A2-002* | N, S |
| | *COL11A1-001* | $P \leq 0, 05$ | | *COL5A2-201* | N, S |
| | *COL11A1-002* | N, S | *COL6A5* | | |
| | *COL11A1-003* | N, S | | *COL6A5-001* | N, S |
| | *COL11A1-004* | N, S | | *COL6A5-002* | N, S |
| | *COL11A1-005* | N, S | | *COL6A5-003* | $P \leq 0, 05$ |
| | *COL11A1-006* | N, S | | *COL6A5-004* | $P \leq 0, 05$ |
| | *COL11A1-007* | $P \leq 0, 05$ | | *COL6A5-201* | N, S |
| | *COL11A1-008* | N, S | *COL28A1* | | |
| | *COL11A1-201* | N, S | | *COL28A1-001* | N, S |
| | *COL11A1-202* | N, S | | *COL28A1-002* | N, S |
| *COL19A1* | | | | *COL28A1-003* | N, S |
| | *COL19A1-001* | N, S | | *COL28A1-004* | N, S |
| | *COL19A1-002* | N, S | | *COL28A1-005* | N, S |
| | *COL19A1-003* | N, S | | *COL28A1-006* | N, S |
| | *COL19A1-004* | N, S | | | |
| | *COL19A1-006* | N, S | | | |
| *COL22A1* | | | | | |
| | *COL22A1-001* | N, S | | | |
| | *COL22A1-002* | N, S | | | |
| | *COL22A1-003* | N, S | | | |
| | *COL22A1-004* | N, S | | | |
| | *COL22A1-006* | N, S | | | |
| | *COL22A1-007* | N, S | | | |
| | *COL22A1-201* | N, S | | | |

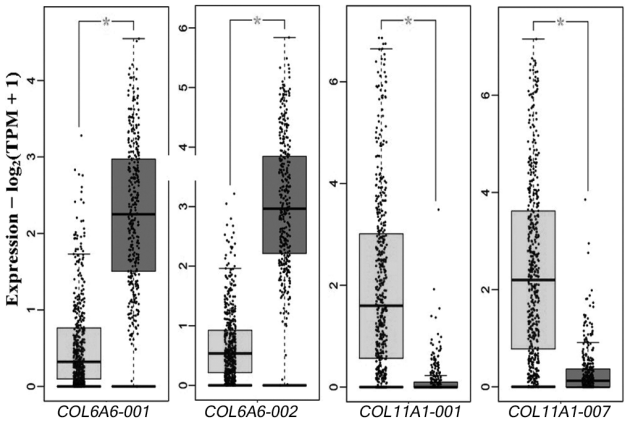


Fig 2 — Comparative analysis of the tissue-specific differential expression of *COL6A6-001 COL6A6-002 COL11A1-001* and *COL11A1-007* transcripts which were found significant among *COL6A4P2 COL6A6 COL7A1 COL11A1 COL19A1* and *COL22A1* genes in lung adenocarcinoma (LUAD) using GEPIA2, These fourty one box plots are based on 483 LUAD samples (marked in red) and 347 normal samples (marked in grey), (*indicates $P$ <0, 05)

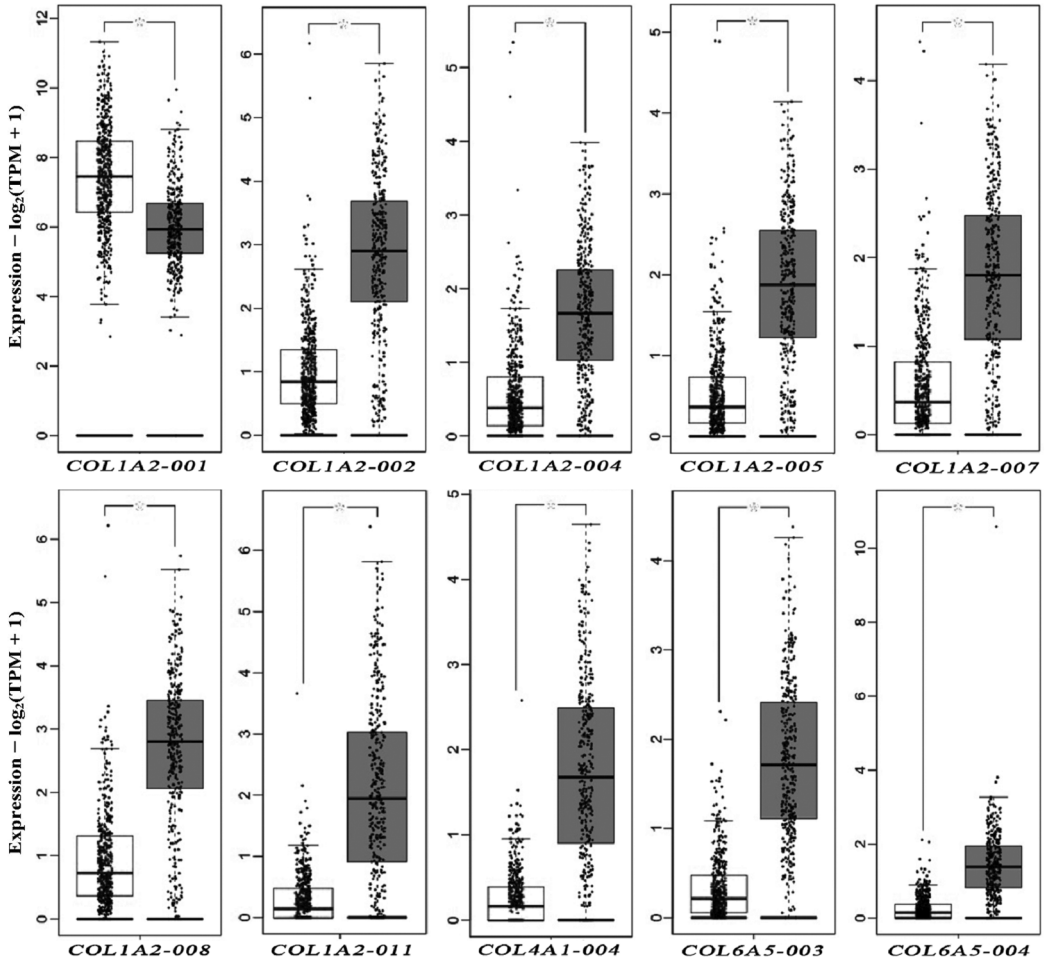


Fig 3 — Comparative analysis of the tissue-specific differential expression of *COL1A2-001 COL1A2-002 COL1A2-004 COL1A2-005 COL1A2-007 COL1A2-008 COL1A2-011 COL4A1-004 COL6A5-003* and *COL6A5-004* transcripts which were found significant among *COL1A2 COL5A2 COL6A5* and *COL28A1* genes in lung squamous cell carcinoma (LUSC) using GEPIA2, These fourty-one box plots are based on 486 LUSC samples (marked in red) and 338 normal samples (marked in grey), (*indicates *P* <0,05)

in detail; *COL1A2-005* transcript; a p-value of 0.0517 is just above the conventional significance level of <0.05. This means that the observed differences in the data across these stages border on statistical significance. The medians across stages appear to shift upward from Stage I to Stage IV. The IQRs for Stage I, II and IV are relatively similar, while Stage III has a slightly longer IQR, suggesting a wider spread of mean data values.

Although lung cancer treatment includes many treatment options such as mutation-based targeting, immunotherapy and radiotherapy with the development of today's technologies, it still cannot be treated effectively[16-18]. Collagen plays an important role in not only in providing tensile strength and elasticity to the lung tissue but also in maintaining a niche for cells for proper cellular functions, and regulating differentiation and proliferation processes. Therefore, the analysis of collagen for in lung cancer is critical. The current study raises some important points for future research. One of these is related to collagen VI which is found highly abundant in the extracellular matrix of the mammalian lung and is required for the establishment and maintenance normal lung structure and function[19]. To date, there is limited information about on the role of collagen VI in lung development or disease. Mereness *et al*.[20] have observed that the absence of collagen VI is associated with dysregulated autophagy and polarization. Increased apoptosis was also observed in pulmonary epithelial cells with collagen VI deficiency[16]. In addition to these findings, it has been elucidated that collagen VI promotes lung epithelial cell spreading and wound closure through phosphoinositide 3-kinase (PI3k) and cell division control 42 homolog (CDC42) downstream of the

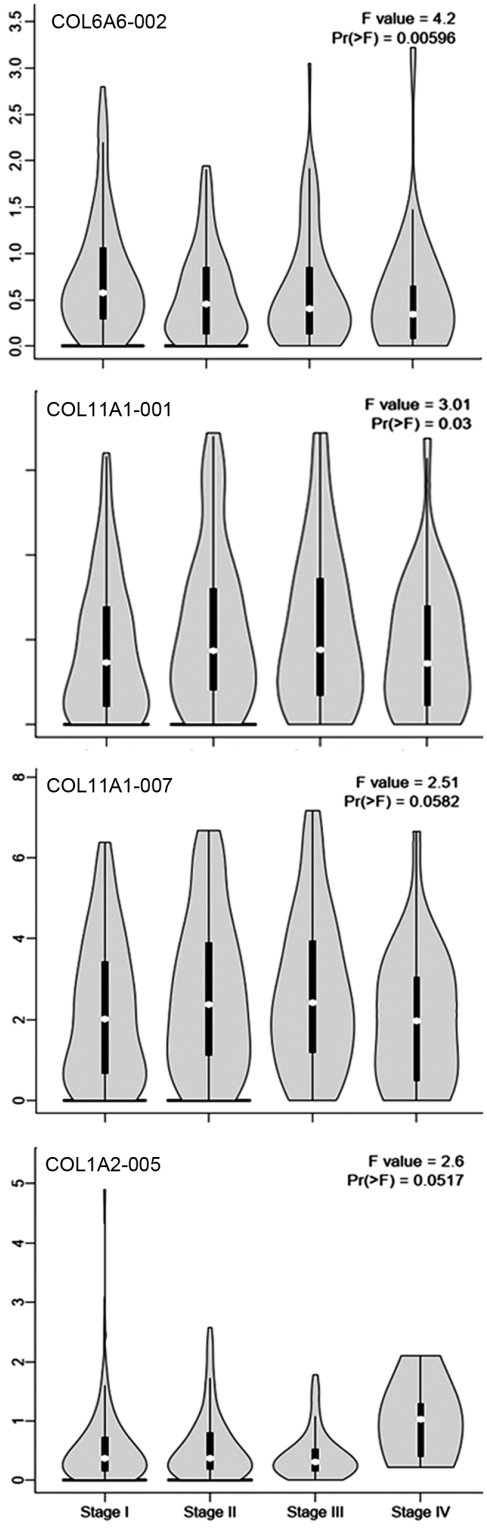


Fig 4 — (A) Comparative analysis of pathological stage expression of *COL6A6-002 COL11A1-001* and *COL11A1-007* transcripts which were found significant in lung adenocarcinoma (LUAD) ($p<0,05$), (B) Comparative analysis of pathological stage expression of *COL1A2-005* transcript which was found significant lung squamous cell carcinoma (LUSC) using GEPIA2 ($p<0,05$)

interaction with β1 integrins[20]. Our study demonstrated the importance of *COL6A4P2* and *COL6A6* genes in predicting overall survival in lung adenocarcinoma. Furthermore, detailed analysis revealed that transcript levels of *COL6A6-001* and *COL6A6-002* were higher in normal tissues than in lung adenocarcinoma tissues. Also, *COL6A6-002* transcript stratify patients with pathological stages. Both evaluating these data and the information that Collagen VI interacts and binds with other matrix components and is responsible for maintaining tissue organization and structure by connecting cells and matrix components with one another, can be quite significant to investigate the alteration of this collagen type in cancer tissues[19].

Another important finding in our study relates to collagen XI. Type XI collagen is thought to be important because of it contributes to fibrillogenesis by controlling the lateral growth of collagen II fibrils[21]. Type II collagen is the main component of airway cartilage and its expression is extremely important as it accounts for 95% of the total collagen present there[22]. Therefore, the status of type XI collagen, which helps type II collagen fibrillogenesis, could be important in type II collagen. Two transcripts of this collagen, *COL11A1-001* and *COL11A1-007*, were found to be higher in lung adenocarcinoma tissues compared to normal tissues according to our analysis. Since fibrils are of broad biomedical importance and play central roles in normal and disease-related conditions such as embryogenesis, tissue repair, fibrosis, and tumor invasion, performing detailed analysis of these *COL11A1-001* and *COL11A1-007* transcripts in lung adenocarcinoma may be clinically relevant[19]. In addition, *COL11A1-001* and *COL11A1-007* transcript levels stratify patients with pathological stages.

Regarding lung squamous cell carcinoma analysis, *COL1A2-001* transcript was highly more expressed in lung squamous cell carcinoma tissues, whereas other *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008* and *COL1A2-0011* transcripts were more highly expressed in normal tissues. Collagen I accumulation has been linked to desmoplasia which is associated with many malignancies[24-26]. In fact, the desmoplasia found in metastatic foci is known to facilitate metastases. In addition, increased expression of type I collagen and many of its modifying enzymes has been reported to correlate with the risk of metastasis[26]. Another

important role of fibrillar collagens including collagen I is to provide the basis for the accumulation and migration of various immune cells in the tumor. Besides being chemotactic for monocytes and neutrophils, type I collagen and collagen fragments have been shown to reduce cytotoxicity against cancer cells when macrophages were cultured on type I collagen[27]. Type I collagen has also been determined as a promoter of Epithelial-Mesenchymal Transition (EMT)[28]. Considering all these data, our findings especially related to lung squamous cell carcinoma analysis *COL1A2-001* transcript was found high in lung squamous cell carcinoma tissues. On the other hand, *COL1A2-002*, *COL1A2-004*, *COL1A2-005*, *COL1A2-007*, *COL1A2-008* and *COL1A2-0011* transcripts were found high in normal tissues. Among these transcripts, *COL1A2-005* also found significant to stratify patient's pathological stages. Ultimately, different isoforms of collagen may exert tumor-inhibitory or tumor-promoting effects. A critical question to be addressed in detail in future research is how these different isoforms of type I collagen exert their tumor-inhibitory or tumor-promoting effects.

Finally, another important indication in the present study is related to collagen IV. Among the transcripts of collagen IV, the expression difference of *COL4A1-004* was found to be significant when comparing normal tissues with tumor tissues. According to the analysis, *COL4A1-004* was higher in normal tissues than in tumor tissues. As one of the main components of the basement membrane and the primary source of alveolar-capillary membrane strength, collagen IV plays an essential role in coordinating alveolar morphogenesis[6,8,29-32]. In addition to its fundamental roles, it induces the expression of the EMT transcription factors such as Snail and Slug[33]. Revert *et.al*.[34] concluded that 3-[4"-methoxy-3,2′-dimethyl-(1,1';4',1")terphenyl-2"-yl]propionic acid or T12, which targets the mesenchymal Goodpasture antigen-binding protein (GPBP) induces collagen IV expansion is a first-in-class drug candidate to treat cancer by selectively targeting the collagen IV. As several methods are reported in the literature to address the increased expression of collagen IV has been associated with chemoresistance and epithelial-to-mesenchymal transition (EMT), studying this collagen type is important, especially in order to properly address the role of the *COL4A1-004* transcript outlined above[31-35].

## Conclusion

The number of transcripts in cells provides information about gene activity and expression levels. Transcript information can be used to determine which specific gene expression changes are associated with cancer and which genes are active during cancer development. As not all transcripts that contribute to transcript diversity are equally likely to contribute to protein diversity, it is critical to underscore the importance of prioritizing target candidates in the diagnosis and treatment of cancer. As the result of our above analysis, the priority is COL6A6-001, COL6A6-002, COL11A1-001 and COL11A1-007 transcripts for LUAD, and COL1A2-002, COL1A2-004, COL1A2-005, COL1A2-007, COL1A2-008, COL1A2-011, COL4A1-004, COL6A5-003 and COL6A5-004 for for Lung squamous cell carcinoma (LUSC).

## Acknowledgement

The data used in our study are obtained from the public database the TCGA Research Network: https://www.cancer.gov/tcga. We thank the TCGA and GEPIA databases for data availability. In addition, statistical interpretation service was received for pathological stage plot analysis from Datasets Box, Izmir/Turkey.

## Conflict of Interest

Author declares no competing interests.

## References

1 Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I & Bray F. *Global Cancer Observatory: Cancer Today*. (International Agency for Research on Cancer, Lyon, France), 2024. Available from: https://gco.iarc.who.int/today, accessed.

2 Kular JK, Basu S & Sharma RI, The extracellular matrix: structure, composition, age-related differences, tools for analysis and applications for tissue engineering. *J Tissue Eng*, 5 (2014) 2041731414557112.

3 Muiznieks LD & Keeley FW, Molecular assembly and mechanical properties of the extracellular matrix: A fibrous protein perspective. *Biochim Biophys Acta*, 1832 (2013) 866.

4 Liang Y, Xia W, Zhang T, Chen B, Wang H, Song X, Zhang Z, Xu L, Dong G & Jiang F, Upregulated collagen COL10A1 remodels the extracellular matrix and promotes malignant progression in lung adenocarcinoma. *Front Oncol*, 10 (2020) 573534.

5 Li Z, Liu Z, Shao Z, Li C, Li Y, Liu Q & Liu Y, Identifying multiple collagen gene family members as potential gastric cancer biomarkers using integrated bioinformatics analysis. *Peer J*, 8 (2020) e9123.

6 Dong S, Zhu P & Zhang S, Expression of collagen type 1 alpha 1 indicates lymph node metastasis and poor outcomes in squamous cell carcinomas of the lung. *Peer J*, 8 (2020) e10089.

7 Thorlacius-Ussing J, Manon-Jensen T, Sun S, Leeming DJ, Sand JM, Karsdal M & Willumsen N, Serum type XIX collagen is significantly elevated in non-small cell lung cancer: a preliminary study on biomarker potential. *Cancers*, 12 (2020) 1510.

8 Weng K, Huang Y, Deng H, Wang R, Luo S, Wu H & Hao W, Collagen family genes and related genes might be associated with the prognosis of patients with gastric cancer: an integrated bioinformatics analysis and experimental validation. *Transl Cancer Res*, 9 (2020) 6246.
9 Zhu H, Hu X, Feng S, Jian Z, Xu X, Gu L & Xiong X, The hypoxia-related gene COL5A1 is a prognostic and immunological biomarker for multiple human tumors. *Oxid Med Cell Longev*, 2022 (2022) 6419695.
10 Yang TH, Chiang YH, Shiue SC, Lin PH, Yang YC, Tu KC & Wu WS, Cancer DEIso: An integrative analysis platform for investigating differentially expressed gene-level and isoform-level human cancer markers. *Comput Struct Biotechnol J*, 19 (2021) 5149.
11 Zhang X, Hassan MA & Prendergast JG, Using machine learning to detect the differential usage of novel gene isoforms. *BMC Bioinformatics*, 23 (2022) 1.
12 National Cancer Institute, Genomic Data Commons. https://gdc.cancer.gov/ (Accessed: February 2024)
13 Gene Expression Profiling Interactive Analysis Web Tool, http://gepia2.cancer-pku.cn/#survival. (Accessed: February 2024).
14 Tang Z, Li C, Kang B, Gao G, Li C & Zhang Z, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*, 45 (2017) W98.
15 McLaughlin J, Berkman J & Nana-Sinkam P, Targeted therapies in non-small cell lung cancer: present and future. *Fac Rev*, 12 (2023)
16 Jayashree BS, Nigam S, Pai A, Patel HK, Reddy ND, Kumar N & Rao CM, Targets in anticancer research—A review. *Indian J Exp Biol*, 53 (2015) 489.
17 Zhang H, Berezov A, Wang Q, Zhang G, Drebin J, Murali R & Greene I, ErbB receptors: from oncogenes to targeted cancer therapies. *J Clin Investig*, 117 (2007) 2051.
18 Mereness JA & Mariani TJ, The critical role of collagen VI in lung development and chronic lung disease. *Matrix Biol Plus*, 10 (2021) 100058.
19 Mereness JA, Bhattacharya S, Ren Y, Wang Q, Anderson CS, Donlon K & Mariani TJ, Collagen VI deficiency results in structural abnormalities in the mouse lung. *Am J Pathol*, 190 (2020) 426.
20 Holmes DF, Lu Y, Starborg T & Kadler KE, Collagen fibril assembly and function. *Curr Top Dev Biol*, 130 (2018) 107.
21 Liu L, Stephens B, Bergman M, May A & Chiang T, Role of collagen in airway mechanics. *Bioengineering*, 8 (2021) 13.
22 Kadler KE, Hill A & Canty-Laird EG, Collagen fibrillogenesis: fibronectin integrins and minor collagens as organizers and nucleators. *Curr Opin Cell Biol*, 20 (2008) 495.
23 Armstrong T, Packham G, Murphy LB, Bateman AC, Conti JA, Fine DR & Iredale JP, (2004), Type I collagen promotes the malignant phenotype of pancreatic ductal adenocarcinoma. *Clin Cancer Res*, 10 (2004) 7427.
24 Shields MA, Dangi-Garimella S, Krantz SB, Bentrem DJ & Munshi HG, Pancreatic cancer cells respond to type I collagen by inducing snail expression to promote membrane type 1 matrix metalloproteinase-dependent collagen invasion. *J Biol Chem*, 286 (2011) 10495.
25 Zhang Z, Wang Y, Zhang J, Zhong J & Yang R, COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol Med Rep*, 17 (2018) 5037.
26 Egeblad M, Rasch MG & Weaver VM, Dynamic interplay between the collagen scaffold and tumor evolution. *Curr Opin Cell Biol*, 22 (2010) 697.
27 Zheng X, Liu W, Xiang J, Liu P, Ke M, Wang B & Lv Y, Collagen I promotes hepatocellular carcinoma cell proliferation by regulating integrin β1/FAK signaling pathway in nonalcoholic fatty liver. *Oncotarget*, 8 (2017) 95586.
28 Loscertales M, Nicolaou F, Jeanne M, Longoni M, Gould DB, Sun Y & Donahoe PK, Type IV collagen drives alveolar epithelial–endothelial association and the morphogenetic movements of septation. *BMC Biol*, 14 (2016) 1.
29 Tanjore H, & Kalluri R, The role of type IV collagen and basement membranes in cancer progression and metastasis. *Am J Pathol*, 168 (2006) 715.
30 Boudko SP, Danylevych N, Hudson BG & Pedchenko VK, Basement membrane collagen IV: Isolation of functional domains. *Methods in Cell Biology*, 143 (2018) 171.
31 Walter C, Davis JT, Mathur J & Pathak A, Physical defects in basement membrane-mimicking collagen-IV matrices trigger cellular EMT and invasion. *Integr Biol*, 10 (2018) 342.
32 Scott LE, Weinberg SH & Lemmon CA, Mechanochemical signaling of the extracellular matrix in epithelial-mesenchymal transition. *Front Cell Dev Biol*, 7 (2019) 135.
33 Revert F, Revert-Ros F, Blasco R, Artigot A, López-Pascual E, Gozalbo-Rovira R & Forteza J, Selective targeting of collagen IV in the cancer cell microenvironment reduces tumor burden. *Oncotarget*, 9 (2018) 11020.
34 Neira RE & Salazar EP, Native type IV collagen induces an epithelial to mesenchymal transition-like process in mammary epithelial cells MCF10A. *Int J Biochem Cell Biol*, 44 (2012) 2194.
35 Gonzàlez-Porta M, Frankish A, Rung J, Harrow J & Brazma A, Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol*, 14 (2013) 1.