



Comprehensive Analysis of the Immune and Prognostic Implication of COL6A6 in Lung Adenocarcinoma

Yi Ma[†], Mantang Qiu^{*†}, Haifa Guo, Haiming Chen, Jiawei Li, Xiao Li and Fan Yang^{*}

Department of Thoracic Surgery, Peking University People's Hospital, Beijing, China

OPEN ACCESS

Edited by:

Yunfeng Feng,
Qinghai University Medical College,
China

Reviewed by:

Pasquale Pisapia,
University of Naples Federico II, Italy
Xi Yang,
Fudan University, China

*Correspondence:

Fan Yang
yangfan@pkuph.edu.cn
Mantang Qiu
qiumantang@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Molecular and Cellular Oncology,
a section of the journal
Frontiers in Oncology

Received: 25 November 2020

Accepted: 19 January 2021

Published: 26 February 2021

Citation:

Ma Y, Qiu M, Guo H, Chen H, Li J, Li X
and Yang F (2021) Comprehensive
Analysis of the Immune and
Prognostic Implication of COL6A6 in
Lung Adenocarcinoma.
Front. Oncol. 11:633420.
doi: 10.3389/fonc.2021.633420

Collagen type VI alpha 6 chain (*COL6A6*), a novel collagen, has been considered as a tumor suppressor and therapeutic target in several tumors. However, the functional role of *COL6A6* in immune cell infiltration and prognostic value in lung adenocarcinoma (LUAD) remains unknown. Here, we evaluated *COL6A6* expression and its impact on survival among LUAD patients from The Cancer Genome Atlas (TCGA) and several other databases. *COL6A6* was downregulated in LUAD tissues compared to normal tissues at both mRNA and protein levels. *COL6A6* expression was negatively associated with pathological stage, tumor stage, and lymph node metastasis. High *COL6A6* expression was a favorable prognostic factor in LUAD. Next, we explored the associations between *COL6A6* expression and immune cell infiltration. *COL6A6* expression was positively associated with the infiltration of B cells, T cells, neutrophils and dendritic cells. Additionally, the immune cell infiltration levels were associated with *COL6A6* gene copy number in LUAD. Consistently, gene set enrichment analysis showed that various immune pathways were enriched in the LUAD samples with high *COL6A6* expression, including pathways related to T cell activation and T cell receptor signaling. The impacts of *COL6A6* on immune activity were further assessed by enrichment analysis of 50 *COL6A6*-associated immunomodulators. Thereafter, using Cox regression, we identified a seven-gene risk prediction signature based on the *COL6A6*-associated immunomodulators. The resulting risk score was an independent prognostic predictor in LUAD. Receiver operating characteristic curve analysis confirmed that the seven-gene signature had good prognostic accuracy in the TCGA-LUAD cohort and a Gene Expression Omnibus dataset. Finally, we constructed a clinical nomogram to predict long-term survival probabilities, and calibration curves verified its accuracy. Our findings highlight that *COL6A6* is involved in tumor immunity, suggesting *COL6A6* may be a potential immunotherapeutic target in LUAD. The proposed seven-gene signature is a promising prognostic biomarker in LUAD.

Keywords: lung adenocarcinoma, *COL6A6*, tumor microenvironment, prognosis, signature, nomogram

INTRODUCTION

Lung cancer is the leading cause of cancer-related death globally and its incidence continues to increase (1). Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer, and lung adenocarcinoma (LUAD) is the major NSCLC subtype (2). The 5-year overall survival (OS) rate of lung cancer is approximately 15%, mainly as a result of late diagnosis and the limited treatment options (3). Therefore, novel treatment strategies and biomarkers are urgently needed.

In recent years, immunotherapy has emerged as a promising strategy for various cancers, especially lung cancer (4). The success of immunotherapy highlights the importance of the tumor microenvironment (TME). Immune cell infiltration of the TME has been shown to be associated with immunotherapy outcomes and survival among patients with solid tumors (5). For instance, tumors with T cell infiltration have a favorable response to anti-PD-1/PD-L1 therapy (6). However, only about 20% of NSCLC cases respond to immunotherapy. Further exploration of the interplay between immune cell infiltration and tumor cells will provide insights into the treatment of LUAD.

Collagen type VI alpha 6 chain (*COL6A6*) is a protein-coding gene. *COL6A6* belongs to the collagen VI (*COL6*) family which acts a vital role in the extracellular matrix. *COL6A6* encodes a 2,262-amino acid protein that contains multiple von Willebrand factor domains and forms a component of the basal lamina of epithelial cells (7). This protein may regulate epithelial cell-

fibronectin interactions and participate in cell adhesion (8, 9). Cell adhesion has a strong association with cancer metastasis (10). *COL6A6* overexpression significantly suppressed tumor growth and metastasis capacity in pituitary adenoma (11). The existence of *COL6* family members in the extracellular matrix indicates the potential function of these proteins in the TME. In addition, *COL6A6* was found to be highly expressed in various human organs, especially in lung. However, no previous studies have reported on *COL6A6* in lung cancer.

In this study, we explored the prognostic and immune implication of *COL6A6* in LUAD. We downloaded data from The Cancer Genome Atlas (TCGA) database for detailed analysis. The underlying roles and prognostic value of *COL6A6* in LUAD were confirmed through multiple bioinformatics analyses. Further, we systematically evaluated the association between *COL6A6* and immune cell infiltration, as well as the signaling pathways regulating the *COL6A6*-mediated immune response. Moreover, we identified an immune prognostic signature using the *COL6A6*-associated immunomodulators, and we then validated its prognostic accuracy in a LUAD dataset from the Gene Expression Omnibus (GEO) database. Finally, a nomogram was established integrating the immune signature and clinical pathological features.

MATERIALS AND METHODS

Acquisition of Gene Expression and Clinical Data

The mRNA sequencing data (HTSeq-FPKM) and associated clinical information for LUAD were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>, up to February 23, 2020) using TCGAAbiolinks R package (12), which involved 522 cases and 535 tumor tissues. Nine cases without corresponding tumor samples were eliminated. Subsequent processing excluded cases with insufficient or missing data on age, sex, race, overall survival time, local invasion, lymph node metastasis, distant metastasis, and TNM stage. Normalized gene expression data in the form of transcripts per million (TPM) were log₂ (TPM+1) transformed for survival analysis. Finally, 485 cases with eligible clinical information were included in the Cox regression analysis. Moreover, 535 tumor tissues were retained to perform Cell type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORTx) analysis to figure out the influence of *COL6A6* expression on the immune microenvironment. Furthermore, after creating a prognostic signature using the TCGA data, it was validated using mRNA expression data and related clinical data in the GSE26939 dataset from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). This dataset included 116 LUAD samples and the detailed clinical data are shown in **Supplementary Table 1**. Ethical approval and informed consent were not required in this study due to the data being publicly available in the TCGA and GEO databases.

Bioinformatics Analysis of *COL6A6* Expression and Survival

Gene Expression Profiling Interactive Analysis 2 (GEPIA2) is an updated version of GEPIA for analyzing the RNA sequencing

Abbreviations: ADORA2A, adenosine A2a receptor; AUC, area under the curve; BP, biological process; BTLA, B and T lymphocyte associated; CC, cellular component; CD49LG, CD40 ligand; CIBERSORTx, Cell type Identification By Estimating Relative Subsets of RNA Transcripts; C-index, concordance index; CNA, copy number alteration; COL6, collagen VI; *COL6A6*, collagen type VI alpha 6 chain; CSF1R, colony stimulating factor 1 receptor; CTLA4, cytotoxic T-lymphocyte associated protein 4; CXCL12, C-X-C motif chemokine ligand 12; CXCR4, C-X-C motif chemokine receptor 4; ENTPD1, ectonucleoside triphosphate diphosphohydrolase 1; FDR, false discovery rate; GEO, Gene Expression Omnibus; GEPIA, Gene Expression Profiling Interactive Analysis; GO, Gene Ontology; GSEA, gene set enrichment analysis; GTEx, The Genotype-Tissue Expression; HAVCR2, hepatitis A virus cellular receptor 2; HHLA2: HERV-H LTR-associating 2; HPA, The Human Protein Atlas; HR, hazard ratio; ICOS, inducible T cell costimulatory; ICOSLG, inducible T cell costimulator ligand; IL2RA, interleukin 2 receptor subunit alpha; IL6, interleukin 6; IL6R, interleukin 6 receptor; IL10, interleukin 10; KDR, kinase insert domain receptor; KEGG, Kyoto Encyclopedia of Genes and Genomes; KLRK1, killer cell lectin like receptor K1; LCE, Lung Cancer Explorer; LGALS9, galectin 9; LTA, lymphotoxin alpha; LUAD, lung adenocarcinoma; MF, molecular function; NECTIN2, nectin cell adhesion molecule 2; NSCLC, non-small cell lung cancer; OS, overall survival; PD-1/PDCD1, programmed cell death protein 1; PDCD1LG2, programmed cell death 1 ligand 2; PD-L1, programmed cell death protein ligand-1; PI3K-Akt-mTOR: phosphatidylinositol 3-kinase-protein kinase B-mechanistic target of rapamycin kinase; PPI, protein-protein interactions; PVR, PVR cell adhesion molecule; ROC, receiver operating characteristic; STRING, Search Tool for the Retrieval of Interacting Genes; TCGA, The Cancer Genome Atlas; TGFBI, transforming growth factor beta 1; TIGIT, T cell immunoreceptor with Ig and ITIM domains; TILs, tumor-infiltrating lymphocytes; TIMER, Tumor Immune Estimation Resource; TNM, tumor node metastasis; TME, tumor microenvironment; TMEM173, transmembrane protein 173; TMIGD2, transmembrane and immunoglobulin domain containing 2; TNFRSF8/13B/13C/14/17/25, TNF receptor superfamily member 8/13B/13C/14/17/25; TNFSF9/13/13B/14/15, TNF receptor superfamily member 9/13/13b/14/15; TPM, transcripts per million; VSIR, V-set immunoregulatory receptor.

expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the Genotype-Tissue Expression (GTEx) projects, using a standard processing pipeline (<http://gepia2.cancer-pku.cn/#index>) (13). GEPIA2 was used to conduct expression and survival analysis of *COL6A6*. The “box plot” and “stage plot” modules allowed us to perform differential expression analysis and to compare *COL6A6* expression in different pathological stages, respectively. “Survival” module was used to evaluate the correlation of *COL6A6* expression with LUAD prognosis. Additionally, the expression and survival meta-analysis of *COL6A6* in LUAD was performed using the online web portal Lung Cancer Explorer (LCE) (<http://lce.biohpc.swmed.edu/lungcancer/>), which integrates 56 lung cancer datasets from TCGA, Gene Expression Omnibus (GEO), and other sources (14). Mutation data were retrieved from cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) (15).

Validation of COL6A6 Expression in Tissues and Cell Lines

Protein expression of *COL6A6* in LUAD and normal lung tissues were evaluated based on immunohistochemistry data from the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) (16). *COL6A6* expression in various LUAD cell lines were also analyzed using GENEVESTIGATOR 7.6.0, which enables analysis of deeply curated transcriptomic data from public repositories (17).

Construction of Protein-Protein Interaction Networks

PPI networks were constructed using the Search Tool for the Retrieval of Interacting Genes (STRING) database (version 11.0, <https://string-db.org/>) (18). The minimum required interaction score was set at 0.900 (highest confidence). The results were visualized in Cytoscape software (version 3.7.2) (19). Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of genes in the PPI networks were performed using the clusterProfiler R package (20). The GO annotation involved biological process (BP), molecular function (MF) and cellular component (CC) annotation. $P < 0.05$ was set as the cut-off criterion.

Gene Set Enrichment Analysis

GSEA was performed to identify the biological processes influenced by the genes in the prognostic signature (21). TCGA-LUAD samples with the top 25% and lowest 25% *COL6A6* expression were used as the high and low expression groups, respectively. GSEA was accomplished in GSEA software version 4.0.3 based on the Molecular Signatures Database version 7.1. C2 (curated gene sets) and C5 (GO gene sets) were searched to identify enriched KEGG pathways and GO terms associated with *COL6A6* expression. $P < 0.05$ and false discovery rate (FDR) < 0.05 were considered to indicate statistical significance.

Associations Between COL6A6 and Tumor-Infiltrating Immune Cells

Tumor Immune Estimation Resource (TIMER) is a comprehensive resource for systematically analysis of immune infiltrates across

diverse cancer types (<https://cistrome.shinyapps.io/timer/>) (22). The correlations between *COL6A6* expression and the abundances of six immune infiltrates (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells) were estimated using the “Gene” module. Additionally, the “SCNA” module was used to explore the correlations between somatic copy number alteration and the abundances of immune infiltrates.

We further determined the associations between *COL6A6* expression and the infiltration of 22 types of tumor-infiltrating lymphocytes (TILs) in LUAD using CIBERSORTx. CIBERSORTx (<https://cibersortx.stanford.edu/>) is the next-generation version of CIBERSORT, which supports deconvolution of bulk RNA-Seq data (23). Among the 535 LUAD tumor tissues with complete gene expression data in the TCGA database, samples with the top 25% and lowest 25% *COL6A6* expression were included as the high and low expression groups, respectively. One thousand permutations were used. $P < 0.05$ was set as the criterion to select the immune infiltrates that may be affected by *COL6A6* expression. In addition, the correlations between *COL6A6* expression and gene markers of TILs were also analyzed using the “correlation analysis” module in GEPIA2.

Analysis of COL6A6-Associated Immunomodulators

Immunomodulators associated with *COL6A6* were extracted from the integrated repository portal TISIDB (<http://cis.hku.hk/TISIDB/>) (24). This database aims to elucidate tumor and immune system interactions by integrating multiple heterogeneous data types. Immunostimulators and immunoinhibitors that were significantly correlated with *COL6A6* expression (Spearman correlation test, $P < 0.05$) were subjected to further analysis. More specifically, the *COL6A6*-associated immunomodulators were used to build a PPI network and were subjected to KEGG and GO analyses, as mentioned above.

Establishment of a Prognostic Signature and Survival Analysis

A prognostic immune gene signature was identified based on the *COL6A6*-associated immunomodulators. Forward stepwise variable selection was performed using the Akaike Information Criterion in Cox models (25). Based on the immune gene signature, the risk score was generated as follows: $risk\ score = \beta_1x_1 + \beta_2x_2 + \dots + \beta_ix_i$ where β_i is the coefficient of each gene derived from the Cox regression, and x_i is the expression level of each gene. Kaplan-Meier survival curve and log-rank test were used to estimate the associations of the signature and clinical features with OS. The time-dependent receiver operating characteristic (ROC) curve analysis was performed to appraise the prognostic accuracy of the risk score using the timeROC R package (26). The TCGA-LUAD dataset was used as the training dataset and the GSE26939 dataset was used as a validation dataset.

Construction and Evaluation of a Nomogram

As a widely used predictive model, a nomogram can provide an individualized prognostic risk assessment intuitively and visually (27). Based on clinical characteristics and the risk score, we constructed a nomogram to predict the probability of 1-, 3-, and 5- year OS using the rms R package. The concordance index (C-index) was calculated to assess the predictive accuracy of the nomogram based on a bootstrap method with 1,000 replicates. Calibration curves were plotted to compare the predicted OS with actual OS rates.

Statistical Analysis

Statistical analysis was performed in R software version 3.6.1 and IBM SPSS Statistics version 22.0. Logistic regression analysis was used to identify the associations between clinical characteristics and *COL6A6* expression. Univariate and multivariate Cox regression analyses were conducted to assess clinical factors associated with OS. $P < 0.05$ was considered statistically significant unless otherwise indicated.

RESULTS

Expression and Function of COL6A6

This study was carried out according to the flow chart shown in **Figure 1**. According to the GEPIA2 results (**Figure 2A**), the *COL6A6* mRNA expression in LUAD tumor tissues was significantly lower than that in normal tissues ($|\log_2(\text{fold change})$

> 2 , $P < 0.01$). The differential expression was also confirmed by a meta-analysis using the LCE web portal (**Figure 2B**). Furthermore, the *COL6A6* protein expression was explored using HPA database. The typical immunohistochemistry result revealed downregulated *COL6A6* expression in tumor tissues (**Figure 2C**). The protein concentration in the plasma was 490ng/L, which was detected by mass spectrometry and estimated from spectral counts in a publicly available dataset from the PeptideAtlas (**Supplementary Figure 1**). In addition, the GENEVESTIGATOR analysis showed that 92.7% (51 of 55) LUAD cell lines had low *COL6A6* expression (**Figure 2D**). According to cBioPortal, 13% of TCGA-LUAD patients had genomic *COL6A6* alterations, and missense mutation was the most common type (**Figure 2E**). As shown in **Figure 3A**, a PPI network of *COL6A6* was constructed using STRING, with the highest confidence threshold (0.900). GO annotation showed that the 11 genes in this PPI network were enriched in 29 GO terms (13 BP terms, 10 CC terms, and 6 MF terms). These terms indicated that *COL6A6* mainly plays a role in the extracellular matrix (**Figure 3B**). According to the KEGG analysis, *COL6A6* is involved in five pathways, including “Focal adhesion” and “PI3K-Akt signaling pathway” (**Figure 3C**).

Survival Analysis and Prognostic Factors

GEPIA2 survival analysis indicated that lower *COL6A6* expression was significantly associated with poor OS ($P = 0.014$, group cutoff=quartile; **Figure 4A**) and advanced pathological stage ($P = 0.015$; **Figure 4B**). Moreover, the prognostic value of *COL6A6* was also substantiated by the survival meta-analysis using LCE web portal [hazard ratio (HR)=0.83, **Figure 4C**]. As shown in **Table 1**, we performed Cox regression analysis to

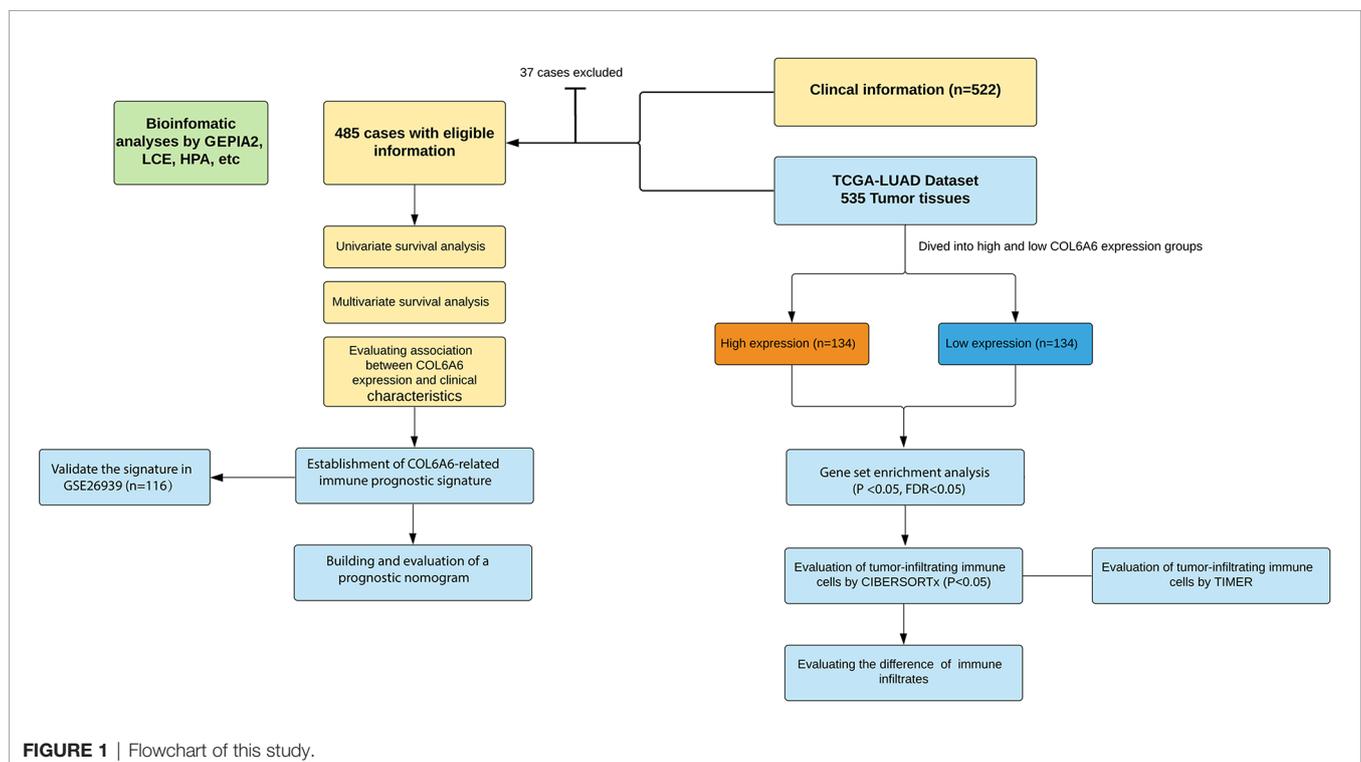
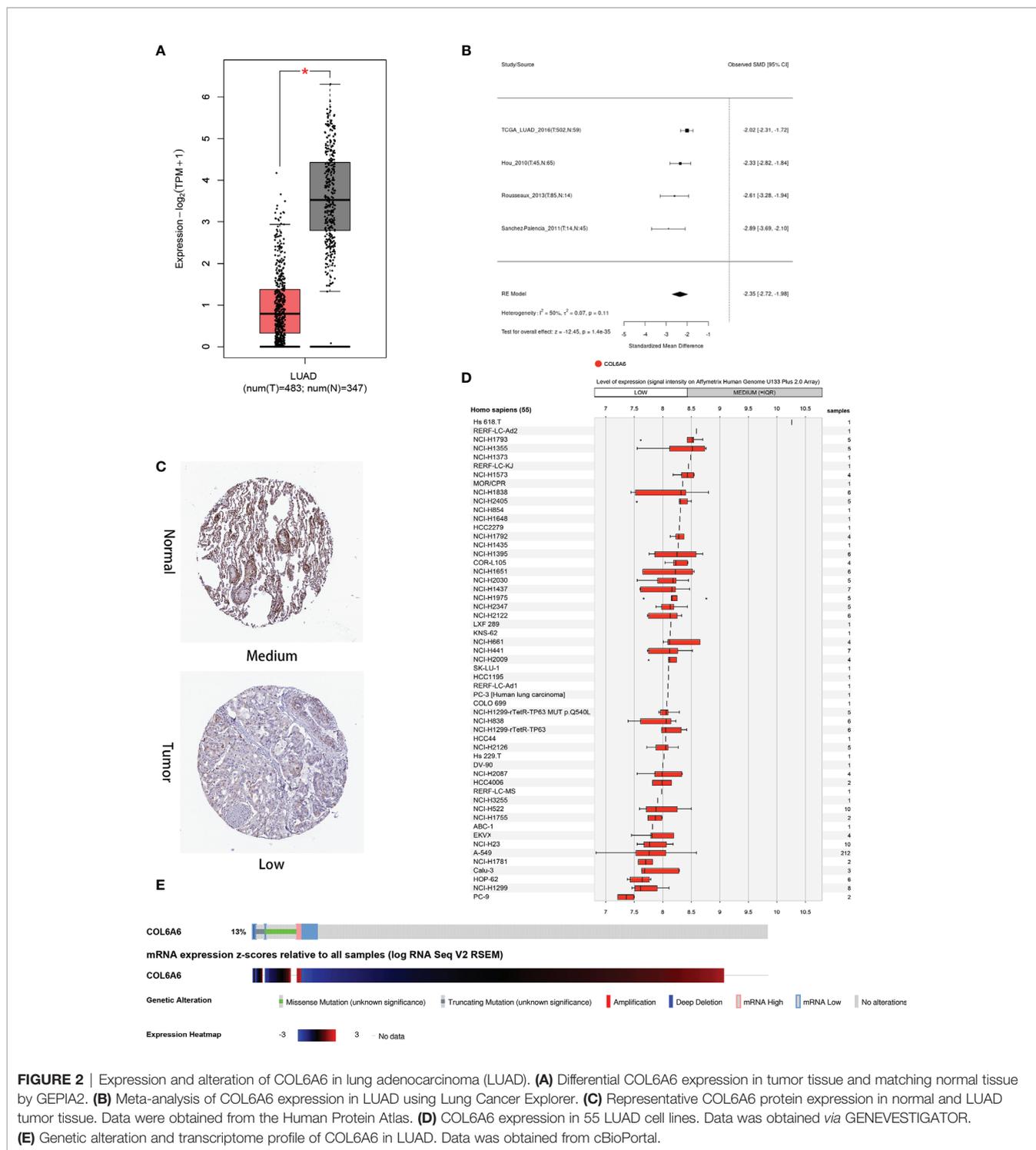


FIGURE 1 | Flowchart of this study.

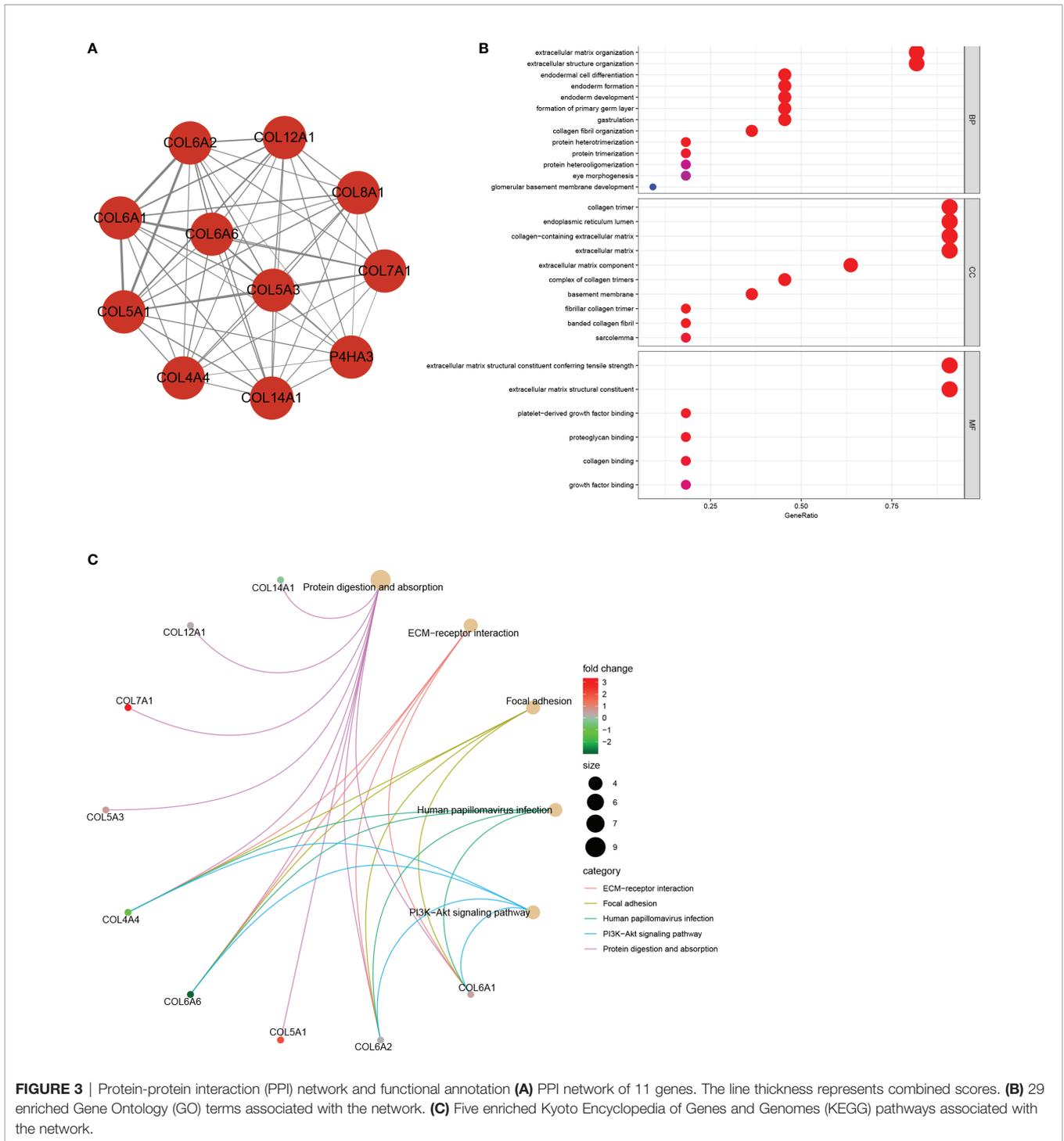


identify the prognostic factors. Univariate analysis showed that pathological stage (HR=1.604, $P < 0.001$), tumor (T) stage (HR=1.537, $P < 0.001$), lymph node (N) stage (HR=1.668, $P < 0.001$), and distant metastasis (M) stage (HR=1.955, $P = 0.02$), along with COL6A6 expression (HR=0.794, $P = 0.003$; **Figure 4D**) were significantly associated with OS. Multivariate analysis identified up-regulated COL6A6 expression and lower

pathological stage as independent prognostic factors that predicted favorable OS (**Figure 4E**).

Association Between COL6A6 Expression and Clinical Characteristics

We investigated the mechanism underlying the effect of COL6A6 expression in LUAD by analyzing its relationship with clinical



factors. The median expression value was used to create a categorical dependent variable based on COL6A6 expression. As shown in **Table 2**, univariate logistic regression analysis revealed that COL6A6 expression was significantly associated with age ($P=0.003$), pathological stage (III vs. I, $P=0.021$; III and IV vs. I and II, $P=0.034$), T stage (T2 vs. T1, $P<0.001$; T3 vs. T1, $P=0.013$; T4 vs. T1, $P=0.012$) and N stage (N1 vs. N0, $P=0.027$; N1&N2&N3 vs. N0, $P=0.011$).

Gene Set Enrichment Analysis

To elucidate the role of COL6A6 expression in the LUAD microenvironment, GSEA was used to compare the high and low expression groups. Each group included 134 samples (i.e., a quarter of the samples) from the TCGA-LUAD dataset. Various immune-related gene signatures were enriched in LUAD samples with high COL6A6 expression, such as B cell differentiation, B cell receptor signaling pathway, T cell receptor signaling pathway,

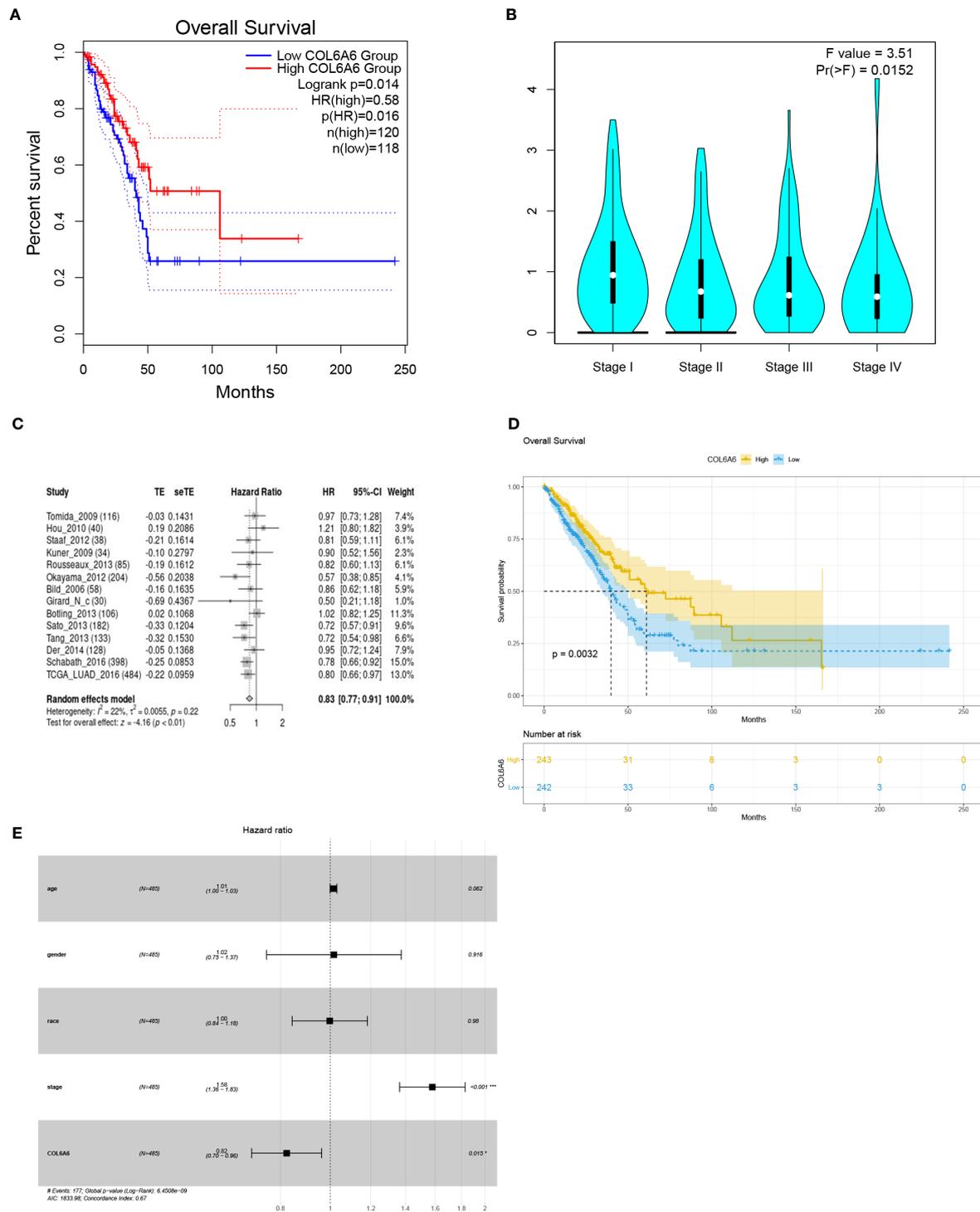


FIGURE 4 | Survival analyses of COL6A6 in lung adenocarcinoma (LUAD). **(A)** Association between COL6A6 expression and survival in LUAD based on GEPIA2 (group cutoff=quartile). **(B)** Association between COL6A6 expression and pathological stage by based on GEPIA2. **(C)** Survival meta-analysis of COL6A6 in LUAD by using the LCE web portal. **(D)** Kaplan–Meier survival curve for COL6A6 expression based on the TCGA-LUAD dataset (group cutoff=median). **(E)** Multivariate Cox analysis of COL6A6 expression and clinical characteristics. Up-regulated expression of COL6A6 and low pathological stage were independent predictors of favorable prognosis.

TABLE 1 | Cox regression results of clinicopathologic variables.

Clinicopathologic variables	HR (95% CI)	P-Value
A [†]		
Age	1.010 (0.994–1.025)	0.223
Gender	1.151 (0.857–1.546)	0.351
Race	0.974 (0.823–1.154)	0.763
Stage	1.604 (1.390–1.852)	0.000
T	1.537 (1.281–1.844)	0.000
N	1.668 (1.406–1.979)	0.000
M	1.955 (1.110–3.445)	0.020
COL6A6	0.794 (0.681–0.925)	0.003
B [‡]		
Stage	1.580(1.364–1.830)	0.000
COL6A6	0.823(0.703–0.962)	0.015

[†]univariate cox analysis; [‡]multivariate cox analysis.

lymphocyte differentiation, regulation of T cell activation, and chemokine signaling pathway (Figure 5). These results indicate that COL6A6 might play important roles in the tumor immune microenvironment.

Relationship Between COL6A6 Expression and Tumor Immune Infiltrates

Previous research has demonstrated that TILs act as independent predictors of survival and immunotherapy efficacy in lung cancer (28). Thus, we sought to determine whether COL6A6 expression was related to immune cell infiltration in LUAD. The “Gene” module of TIMER was used to roughly analyze the correlations. Upregulated COL6A6 was positively associated with the infiltration levels of all six immune cells assessed (B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, macrophages and dendritic cells) (Figure 6A, P<0.0001). As shown in Figure 6B, with chromosome arm-level deletion of COL6A6, infiltration levels of CD8⁺ T cells (P= 0.027), macrophages (P=0.001), neutrophils (P<0.001), and dendritic cells (P<0.001) were significantly lower. In addition, we also found COL6A6 expression varied among different immune subtypes (Figure 6C).

COL6A6 expression was highest in the inflammatory subtype, which indicates a better prognosis in LUAD (29).

To obtain a deeper understanding of the relationship between COL6A6 expression and tumor immune infiltrates, we further determined the proportions of 22 types of TILs in LUAD using CIBERSORTx. Among the 134 TCGA-LUAD samples in each group, 125 samples in the high expression group and 107 samples in the low expression group met the screening criterion. The differences in the proportions of the 22 subpopulations of immune cells in these two groups are shown in Figure 7. B cells naive, dendritic cells resting, eosinophils, macrophage M0, mast cells activated, mast cells resting, monocytes, neutrophils, natural killer (NK) cells activated, plasma cells, T cells CD4 memory resting, T cells follicular helper, and T cells regulatory (Tregs) were the main immune cells correlated with COL6A6 expression. Among them, there were higher proportions of B cells naive (P=0.041), dendritic cells resting (P=0.038), mast cells resting (P<0.0001), neutrophils (P=0.043) and T cells CD4 memory resting (P<0.0001) in the high expression group. In contrast, the proportions of macrophages M0 (P<0.0001), plasma cells (P<0.001), and Tregs (P<0.001) were lower in the high expression group. Furthermore, we analyzed the relationship between COL6A6 expression and gene markers of various TILs, including CD8⁺ T cells, CD4⁺ T cells, B cells, neutrophils, dendritic cells and eosinophils (Table 3). Taking these results together, it can be concluded that COL6A6 may play a potential role in modulating the abundance of B cells, T cells, neutrophils, and dendritic cells.

We also explored the potential immune modulatory function of COL6A6 in LUAD. We identified 33 immunostimulators (CD27, VSIR, CD28, CD40, CD40LG, CD48, CD80, CD86, CXCL12, CXCR4, ENTPD1, HHLA2, ICOS, ICOSLG, IL2RA, IL6, IL6R, KLRK1, LTA, PVR, TMEM173, TMIGD2, TNFRSF13B, TNFRSF13C, TNFRSF14, TNFRSF17, TNFRSF25, TNFRSF8, TNFSF13, TNFSF13B, TNFSF14, TNFSF15, and TNFSF9) and 17 immunoinhibitors (ADORA2A, BTLA, CD160, CD244, CD274, CD96, CSF1R, CTLA4, HAVCR2, IL10, KDR, LGALS9, PDCD1,

TABLE 2 | Association between COL6A6 expression and clinical characteristics.

Clinical characteristics	Total (N)	Odds ratio in COL6A6 expression	P-Value
Age (continuous)	494	1.028 (1.009–1.047)	0.003
Gender female vs. male	513	1.234 (0.873–1.750)	0.233
Race black vs. white	439	0.810 (0.450–1.446)	0.476
others [†] vs. white	395	0.315 (0.046–1.386)	0.160
Stage II vs. I	401	0.723 (0.473–1.103)	0.133
III vs. I	359	0.558 (0.338–0.914)	0.021
IV vs. I	301	0.602 (0.261–1.351)	0.222
III and IV vs. I and II	511	0.630 (0.409–0.963)	0.034
T T2 vs. T1	444	0.470 (0.316–0.695)	0.000
T3 vs. T1	215	0.433 (0.222–0.833)	0.013
T4 vs. T1	187	0.270 (0.091–0.720)	0.012
T3 and T4 vs. T1 and T2	510	0.610 (0.357–1.029)	0.067
N N1 vs. N0	432	0.593 (0.372–0.938)	0.027
N2 vs. N0	411	0.649 (0.388–1.075)	0.095
N3 vs. N0	339	0.852(0.033–21.655)	0.910
N1 and N2 and N3 vs. N0	508	0.619 (0.426–0.897)	0.011
M M1 vs. M0	511	0.724 (0.318–1.600)	0.428

[†]Asian and American Indian or Alaska native.

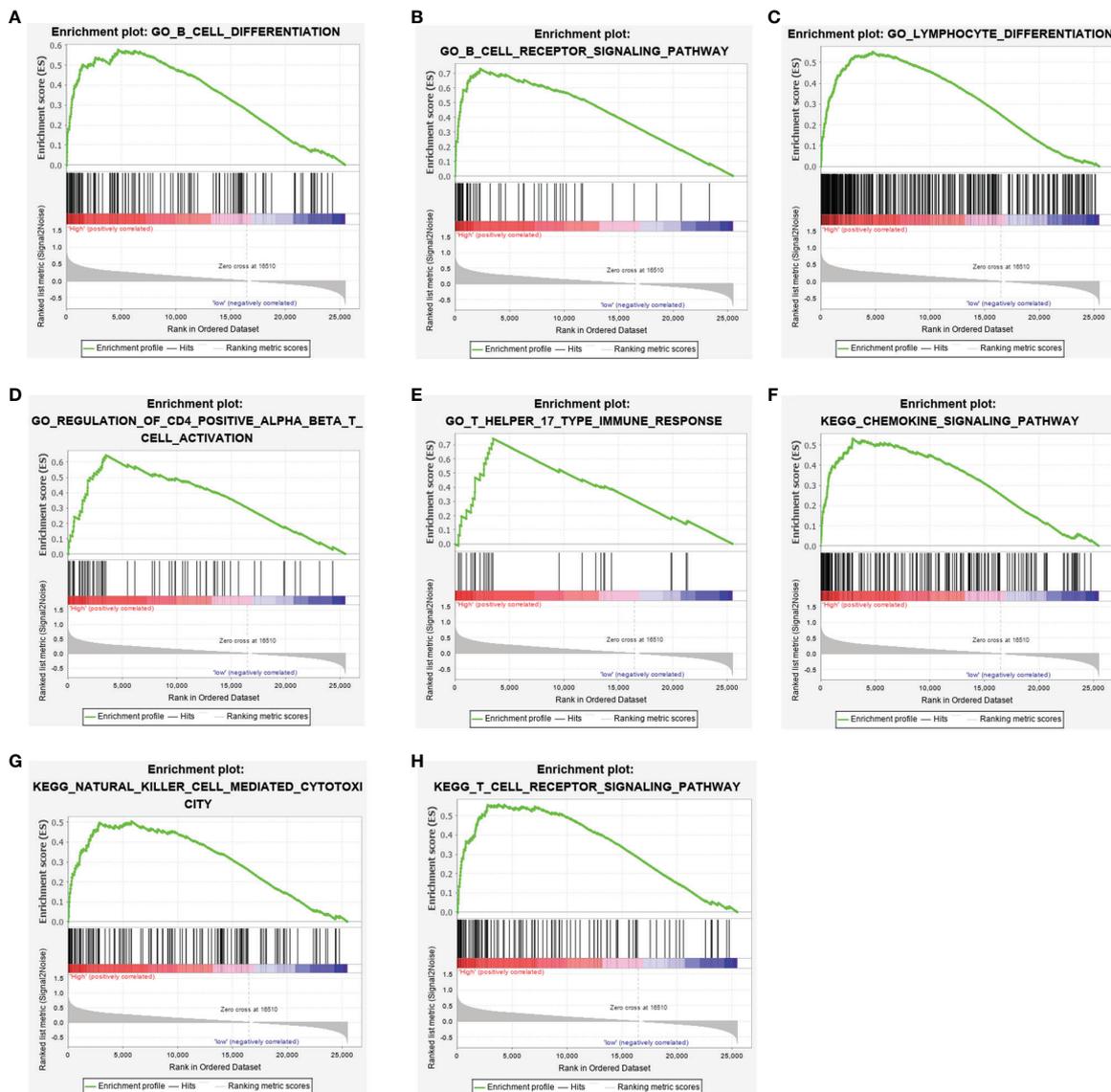


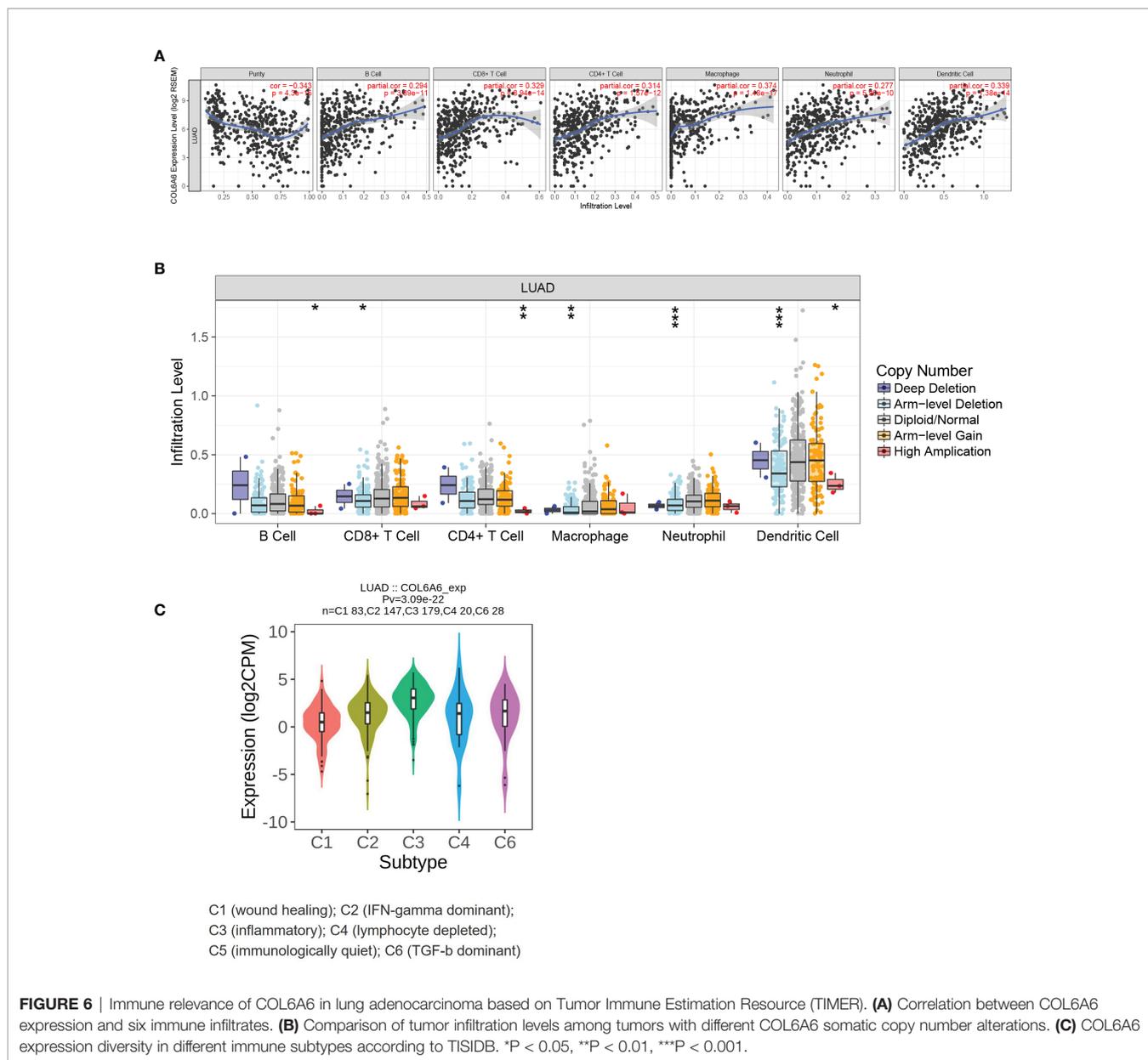
FIGURE 5 | Gene set enrichment analysis (GSEA) showed that COL6A6 is involved in the tumor immune microenvironment. **(A)** B cell differentiation. **(B)** B cell receptor signaling pathway. **(C)** Lymphocyte differentiation. **(D)** Regulation of CD4 positive alpha beta T cell activation. **(E)** T Helper 17 type immune response. **(F)** Chemokine signaling pathway. **(G)** Natural killer cell mediated cytotoxicity. **(H)** T cell receptor signaling pathway.

PDCD1LG2, *NECTIN2*, *TGFB1*, and *TIGIT*) that were significantly associated with *COL6A6* expression in LUAD (**Figure 8A**, **Supplementary Table 2**). A PPI network was constructed based on these 50 genes (**Figure 8B**). GO and KEGG pathway enrichment analyses of these genes indicated that regulation of lymphocyte activation, regulation of T cell activation, leukocyte cell–cell adhesion, and the T cell receptor signaling pathway were related to *COL6A6*-mediated immune events (**Figures 8C, D**).

Establishment and Validation of a Seven-Gene Signature

To determine the prognostic value of the *COL6A6*-associated immunomodulators in LUAD, we performed a stepwise Cox

multivariate regression analysis. A prognostic gene signature consisting of seven genes (*CD160*, *CD40LG*, *CD86*, *IL10*, *PVR*, *TNFSF13B*, and *TNFSF14*) was identified based on the TCGA-LUAD dataset. The biological functions of the seven genes and the risk coefficients are shown in **Table 4**. Univariate Cox regression analysis was performed to evaluate the associations between these genes and OS (**Supplementary Figure 2A**). The risk scores were then calculated based on the proposed formula. Patients in the training dataset were then divided into high- and low-risk groups according to the optimal risk score cut-off. The Kaplan–Meier survival curve revealed that the survival time of the high-risk group was significantly lower ($P < 0.001$) (**Figure 9A**). The area under the curve (AUC) values of the risk score and

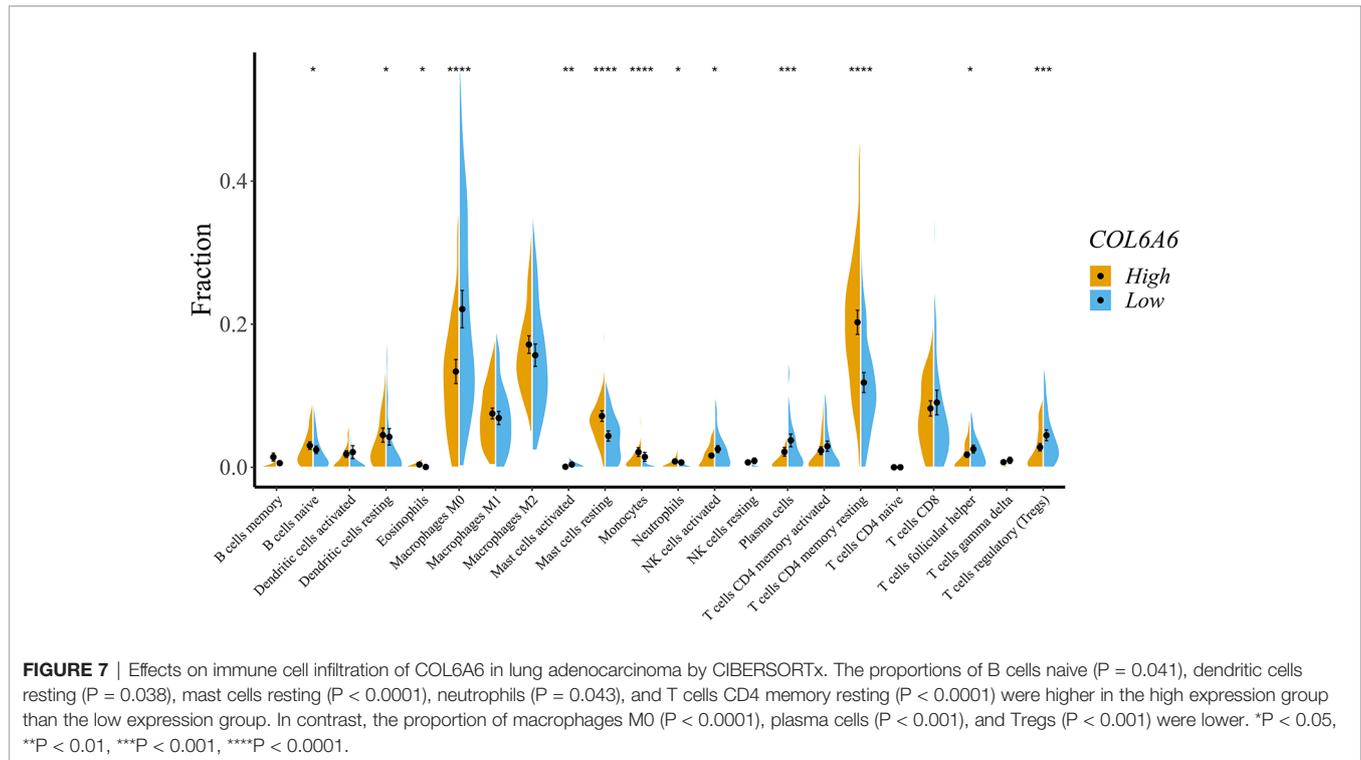


pathological stage were 0.721 and 0.683, respectively. Combining the risk score and pathological stage, an AUC of 0.77 was achieved (**Figure 9C**). The risk scores, survival status distribution of the patients, and gene expression profiles related to the seven-gene signature in the training dataset are shown in **Figure 9E**. In addition, univariate Cox regression analysis showed that the risk score was significantly associated with OS (HR= 2.718, 95% CI=2.143–3.448, P<0.001). As shown in **Supplementary Figure 2B**, the risk score was an independent predictor of prognosis in the multivariate Cox regression model (HR=2.462, 95% CI=1.905–3.183, P<0.001) after adjusting for age, gender, pathological stage, T, N, and M stage. The GSE26939 dataset was used to validate the performance of the seven-gene signature. A risk score for each patient was calculated using the same method. The optimal cut-off was also determined and

patients in the validation dataset were then divided into high- and low-risk groups accordingly. Consistently, the patients in the high-risk group had notably poorer prognosis (P =0.014) (**Figure 9B**). In the validation dataset, the AUCs for 1-, 3-, and 5-year OS prediction based on the gene signature were 0.65, 0.625, and 0.666, respectively (**Figure 9D**). The distribution of the risk scores and the gene expression profiles are shown in **Figure 9F**. Collectively, these results indicated that the seven-gene signature had good performance in predicting the OS of LUAD patients.

Construction and Evaluation of a Prognostic Nomogram

A prognostic nomogram to predict the 1-, 3-, and 5-year OS of LUAD patients was established based on the TCGA-LUAD



dataset using Cox regression (**Figure 10A**). Risk score, age, gender and stage were features that were included in the nomogram. By calculating the score of each feature for each patient, we can predict the 1-, 3-, and 5-year OS probability, contributing to personalized precision treatment. The C-index of the prognostic nomogram was 0.723. The calibration curves further revealed that the nomogram had good performance in predicting the OS of LUAD patients (**Figures 10B–D**).

DISCUSSION

Lung cancer is an increasingly prevalent disease that threatens global health. Late-stage diagnosis and the limitations of classical therapies lead to a poor OS. Recent evidence suggests that immunotherapy is a powerful tool for lung cancer treatment (4–6). Therefore, the identification of novel biomarkers and promising immune-related therapeutic targets for lung cancer has become imperative in clinical practice. In this study, we found that the expression of *COL6A6* was downregulated at both mRNA and protein levels in LUAD tissues. High *COL6A6* expression was an independent predictor of a favorable prognosis. Additionally, *COL6A6* expression was significantly associated with several clinical characteristics, including pathological stage, T stage, and N stage. Moreover, *COL6A6* was found to be associated with immune cell infiltration levels and various immune pathways in LUAD. Furthermore, we identified a seven-gene risk signature based on *COL6A6*-associated immunomodulators using stepwise Cox regression model. This signature showed good predictive performance and a prognostic nomogram was constructed based on the signature for clinical application.

Previous studies revealed that *COL6A6* expression was distinctly higher in lung tissue than in other tissues (**Supplementary Figure 3**) (30, 31). However, as a new member of the *COL6* family, the role of *COL6A6* in lung cancer is unclear. We found that *COL6A6* expression in LUAD tissues was significantly lower than in normal tissues. Additionally, we found that upregulated expression of *COL6A6* was associated with a favorable prognosis. Logistic regression analysis revealed that *COL6A6* expression was related to pathological stage, T stage, and N stage. Meta-analysis effectively combines statistical strength from multiple data sets which allows for greater precision compared to when using any of the single studies. The survival meta-analysis in this study suggested that high *COL6A6* expression was a protective factor for LUAD patients. Therefore, we speculated that *COL6A6* may act as a tumor suppressor in LUAD.

Another important finding of this study was the effects of *COL6A6* on immune cell infiltration in LUAD. The TIMER analysis showed close correlations between *COL6A6* expression and abundances of B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells. Decreased copy number may lead to reduced infiltration of CD8+ T cells, macrophages, neutrophils and dendritic cells. The above-mentioned evidence indicated the potential role of *COL6A6* in the immune microenvironment. As some immune infiltrates, such as macrophages, have been found to exist in both antitumorigenic and protumorigenic forms, we used CIBERSORTx for further analysis. Consistent with the TIMER results, we found that the proportions of several types of B cells, T cells, neutrophils, and dendritic cells were apparently increased in the high expression group while the proportions of macrophages M0 and Tregs were lower. It has been reported

TABLE 3 | Correlation between COL6A6 expression and gene markers in GEPIA2.

Immune celltypes	Gene markers	LUAD			
		Tumor		Normal	
		R	P	R	P
B cell	CD19	0.33	1.4e-13	0.053	0.69
	CD79A	0.27	1.5e-09	0.15	0.26
T cell (general)	CD2	0.41	1e-20	0.14	0.3
	CD3E	0.43	1e-23	0.29	0.027
CD8 ⁺ T cell	CD8A	0.34	1.9e-14	0.21	0.11
	CD8B	0.28	5.5e-10	0.055	0.68
CD4 ⁺ T cell	CD4	0.57	5.4e-43	0.37	0.0035
Natural killer cell	KIR2DL1	0.21	2.7e-06	0.25	0.054
	KIR2DL3	0.23	5.8e-07	0.36	0.0057
	KIR2DL4	-0.025	0.59	0.15	0.27
	KIR3DL1	0.25	3.9e-08	0.34	0.0078
	KIR3DL2	0.18	9.6e-05	0.31	0.015
	KIR3DL3	-0.083	0.07	0.04	0.76
	KIR2DS4	0.15	0.0012	0.25	0.054
Neutrophils	CD66b	0.37	1.9e-174	0.07	0.6
	CD11b	0.42	4e-22	0.086	0.52
	CCR7	0.52	5.7e-35	0.14	0.29
Th1	T-bet	0.4	2.7e-20	0.45	0.00032
	STAT4	0.38	5.3e-18	0.29	0.028
Th2	TNF- α	0.25	2.3e-08	0.17	0.21
	GATA3	0.31	5.2e-12	0.45	0.00034
	STAT6	0.32	1e-12	0.6	5.8e-07
Tfh	STAT5A	0.49	3.1e-30	0.34	0.0092
	IL13	0.27	9.3e-10	0.25	0.058
	BCL6	0.26	9.4e-09	0.51	4.5e-05
Th17	STAT3	0.29	1.4e-10	0.54	8.5e-06
	IL17A	0.13	0.0045	0.12	0.37
Mast cells	TPSB2	0.45	3.1e-25	0.33	0.011
	TPSAB1	0.49	4.3e-31	0.42	0.001
	CPA3	0.55	9.1e-40	0.45	4e-04
	MS4A2	0.6	5.3e-49	0.5	4.8e-05
	HDC	0.57	1.2e-42	0.43	0.00061
Eosinophils	IL5	0.17	2e-04	-0.038	0.77
	CD125	0.43	7.2e-23	0.055	0.68
Monocytes	CD86	0.38	6e-18	-0.13	0.31
	CX3CR1	0.45	8.8e-26	0.43	6e-04
Checkpoints	CD14	0.27	1.9e-09	0.12	0.36
	PDCD1	0.25	1.8e-08	0.17	0.2
	CD274	0.27	1.2e-09	0.24	7.1e-06
	CTLA4	0.32	3.7e-13	0.37	0.0038
	LAG3	0.15	0.0014	0.12	0.38
	TIGIT	0.37	1.5e-17	0.3	0.022

GEPIA2, Gene Expression Profiling Interactive Analysis 2; LUAD, lung adenocarcinoma; Th1, T helper 1 cells; Th2, T helper 2 cells; Tfh, follicular helper T cells; Th17, T helper cell 17.

that the immune microenvironment is frequently associated with lung cancer outcome. T cells, especially cytotoxic and memory T cells, have been consistently shown to have positive effects on prognosis (28, 32–34). B cells, dendritic cells, and eosinophils have also been reported to be associated with a favorable prognosis (32, 33, 35). In contrast, high proportions of Tregs and macrophages M0 have been associated with a poor prognosis among LUAD patients (28, 36). Thorsson *et al.* identified six immune subtypes that span cancer tissue types and molecular subtypes, and the inflammatory subtype had the best prognosis (29). In this study, *COL6A6* expression was highest in the inflammatory immune subtype, indicating a better prognosis in LUAD; the positive impact of *COL6A6* expression prognosis among LUAD patients was consistent with the potential roles of

diverse immune cell types. This further confirms the association between *COL6A6* expression and survival of LUAD patients.

Intriguingly, the GSEA results suggested that *COL6A6* was associated with differentiation and activation of several lymphocytes including B and T cells. Consistently, functional analysis of the *COL6A6*-associated immunomodulators indicated that regulation of T cell activation, leukocyte cell–cell adhesion, and T cell receptor signaling pathway were related to *COL6A6*-mediated immune events. *COL6A6* participates in the cell adhesion process, and many other cell adhesion molecules, especially integrins, are known to play key roles in the regulation of immune cell infiltration into the TME (37). Integrins have various effects on T cell infiltration, including T cell recruitment into tumors, priming, and effector function (37, 38). Multiple integrin pairs, including

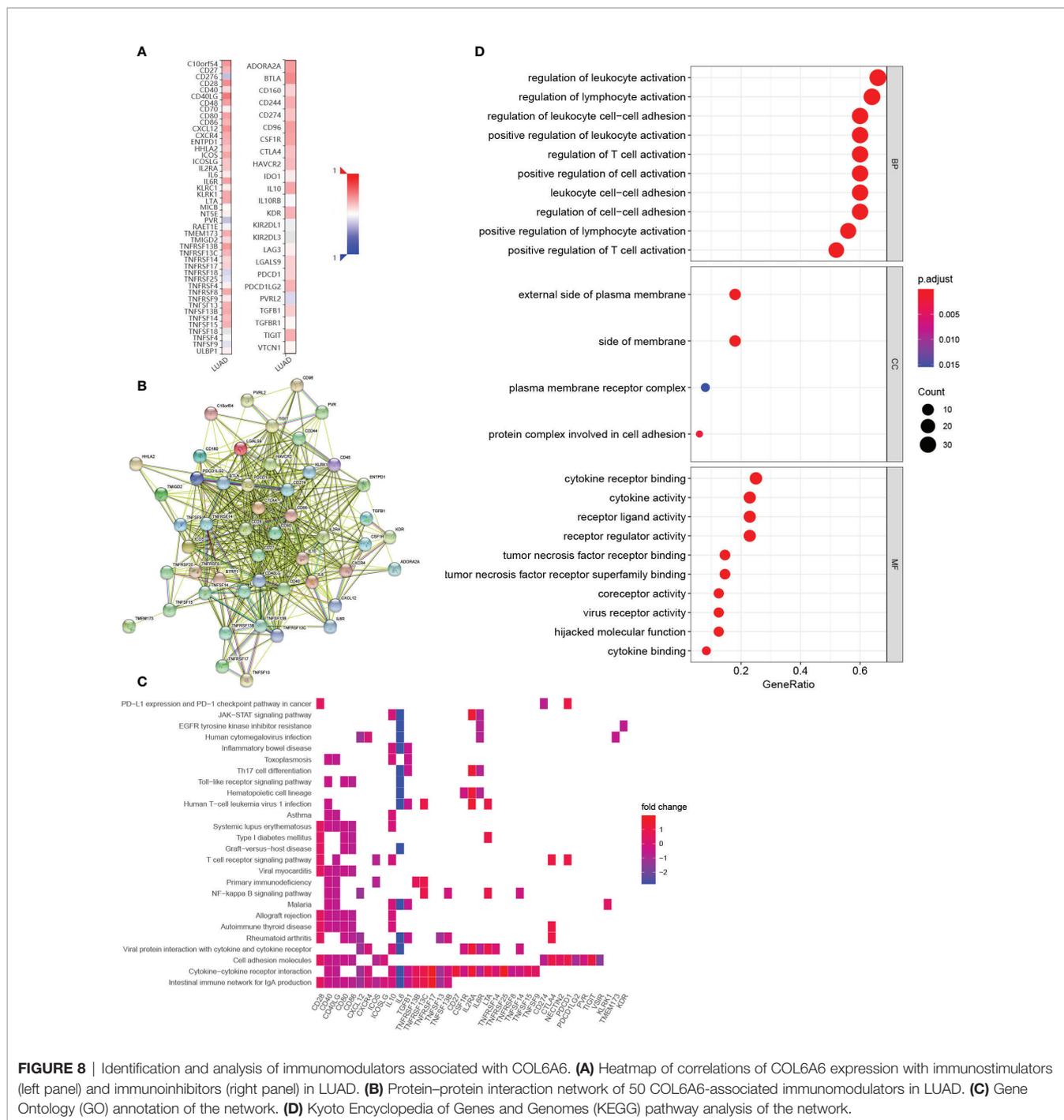


FIGURE 8 | Identification and analysis of immunomodulators associated with COL6A6. **(A)** Heatmap of correlations of COL6A6 expression with immunostimulators (left panel) and immunoinhibitors (right panel) in LUAD. **(B)** Protein-protein interaction network of 50 COL6A6-associated immunomodulators in LUAD. **(C)** Gene Ontology (GO) annotation of the network. **(D)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the network.

TABLE 4 | Function of the genes in the prognostic signature.

Gene symbol	Full name	Function	Risk coefficient
CD160	CD160	Associated with peripheral blood NK cells and CD8 T lymphocytes with cytolytic effector activity	-0.6985597
CD40LG	CD40 ligand	Regulation of B cell function	-0.4634196
CD86	CD86	regulation of T cell activation	-0.5033796
IL10	Interleukin 10	Pleiotropic effects in immunoregulation and inflammation	-0.3230688
PVR	Poliovirus receptor	Cell adhesion and regulation of immune response	0.2712843
TNFSF13B	TNF superfamily, member 13b	Regulation of B cells proliferation and differentiation	0.8373417
TNFSF14	TNF superfamily member 14	Stimulating the activation of lymphoid cells and proliferation of T cells	0.3225482

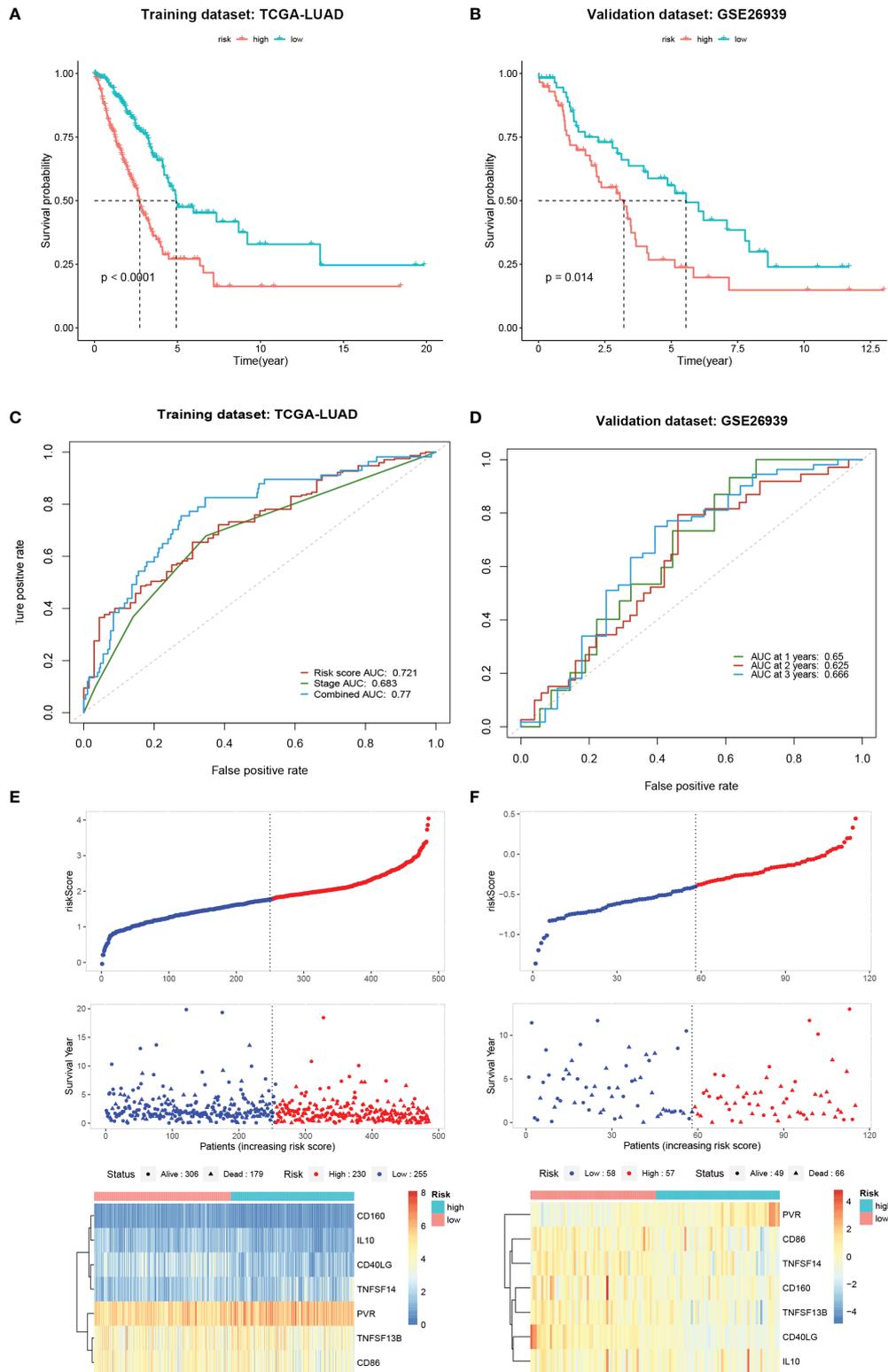


FIGURE 9 | Establishment and validation of a seven-gene signature. **(A, B)** Kaplan-Meier curves of overall survival for the high and low-risk groups in the training and validation datasets. **(C, D)** Time-dependent receiver operating characteristic curves for the risk score in the training and validation datasets. **(E, F)** Distribution of risk scores, survival status and gene expression profiles for the training and validation datasets.

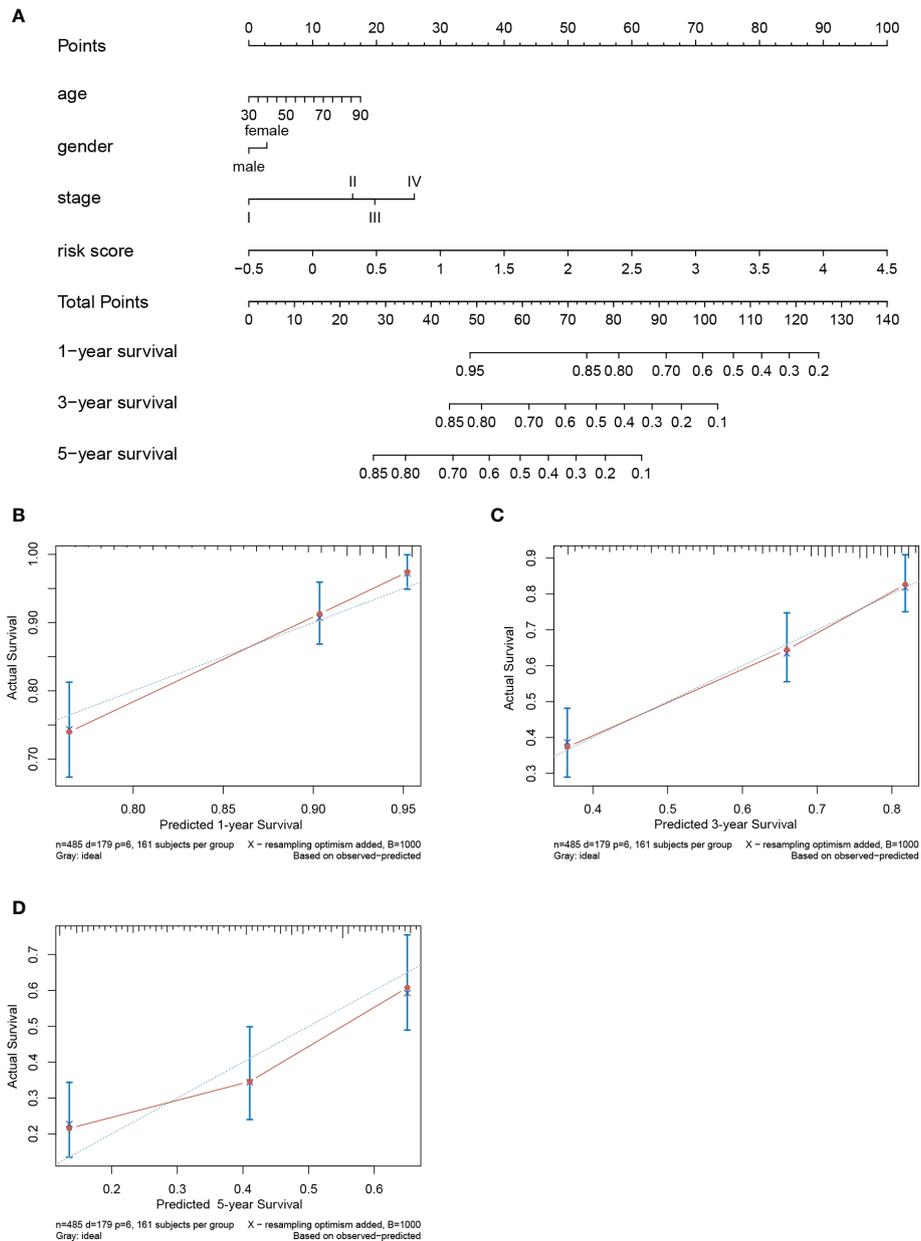


FIGURE 10 | Nomogram to predict prognostic probabilities in the The Cancer Genome Atlas-lung adenocarcinoma (TCGA-LUAD) dataset. **(A)** Nomogram for predicting 1-, 3-, and 5-year overall survival (OS) of LUAD patients. **(B–D)** Calibration curves of 1-, 3-, and 5-year OS of LUAD patients. The Y axis represents the actual OS while the X axis represents nomogram predicted OS.

$\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 10\beta 1$, and $\alpha v\beta 3$ integrin, have been identified as potential binding partners for COL6 (39, 40). The PI3K-Akt pathway is another crucial pathway involved in tumor immunity. As a major pathway in tumorigenesis, its role in tumor immunity has also been reported. Okkenhaug demonstrated that the pathway contributes to the development and differentiation of B cells and T cell subsets (41). PI3K inhibition in dendritic cells enhances type I immune responses. Moreover, inhibition of the PI3K pathway enhances CD8+ T cell infiltration into tumor tissue (42). PI3K-AKT-mTOR inhibition can regulate immunosuppressive cytokine

secretions, effect Treg infiltrations into tumor tissues and promote the development of memory T cells (43). The PI3K pathway may also regulate the expression of PD-L1 (44). Furthermore, COL6A6 was reported to inhibit the PI3K-Akt pathway; Long *et al.* found that COL6A6 suppressed the growth and metastasis of pituitary adenocarcinoma by blocking the PI3K-Akt pathway (11). Together, these findings show that COL6A6 may play a pivotal role in the tumor immune microenvironment of LUAD.

Gene signatures have been widely used as prognostic predictors in various cancers (45, 46). Zhuang *et al.* analyzed M1 macrophage-

related genes and identified a four-gene signature for predicting thyroid cancer prognosis (45). Li et al. analyzed the gene expression profiles of frozen tumor tissue samples from 19 public NSCLC cohorts and identified a prognostic immune-related gene signature involving 25 gene pairs (46). Their immune signature achieved moderate prognostic accuracy (C-index=0.64). In this study, we established a seven-gene signature for LUAD based on COL6A6-associated immunomodulators. The risk score derived from the immune gene signature was an independent predictor of LUAD prognosis. The signature had high accuracy in both the training dataset and validation dataset. Combined with clinical features, we further built a prognostic nomogram for personalized prediction, which was found to have a C-index of 0.723. Our findings may provide clinicians with a convenient and accurate way to evaluate the prognosis of patients with LUAD after surgery.

However, there are several limitations in this study. First, our study was mainly based on online databases. Nonetheless, integrating bioinformatic analyses, such as meta-analysis add strength to this study. We will further verify the expression and function of COL6A6 as well as its correlation with immune cell infiltration in the laboratory. Second, we were unable to analyze the relationship between COL6A6 expression and immunotherapy response, as these data in the TCGA and other databases were insufficient. Lastly, the prognostic value of the seven-gene signature should be further validated in an in-house patient population.

In summary, COL6A6 is downregulated in LUAD tissues, while increased COL6A6 expression predicts a favorable prognosis. COL6A6 expression is associated with the infiltration of various immune cells, such as B cells, T cells, and dendritic cells, in LUAD. COL6A6 may play a role in the tumor immune microenvironment of LUAD. The seven-gene signature derived from COL6A6-associated immunomodulators is an independent predictor of OS in LUAD. The prognostic nomogram showed good performance in predicting the OS of LUAD and might thus be beneficial for individualized treatment and medical decision making.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: TCGA database (<https://portal.gdc.cancer.gov/>) and GEO database (<http://www.ncbi.nlm.nih.gov/geo/>).

REFERENCES

- Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. *Adv Exp Med Biol* (2016) 893:1–19. doi: 10.1007/978-3-319-24223-1_1
- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* (2008) 83(5):584–94. doi: 10.4065/83.5.584
- Polanski J, Jankowska-Polanska B, Rosinczuk J, Chabowski M, Szymanska-Chabowska A. Quality of life of patients with lung cancer. *Onco Targets Ther* (2016) 9:1023–8. doi: 10.2147/OTT.S100685
- Zhang H, Shen J, Yi L, Zhang W, Luo P, Zhang J. Efficacy and Safety of Ipilimumab plus Chemotherapy for Advanced Lung Cancer: A Systematic Review and Meta-Analysis. *J Cancer* (2018) 9(23):4556–67. doi: 10.7150/jca.27368
- Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* (2013) 39(1):1–10. doi: 10.1016/j.immuni.2013.07.012
- Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM, Vincent BG. Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. *J Natl Cancer Inst* (2016) 108(11):djw144. doi: 10.1093/jnci/djw144
- Fitzgerald J, Rich C, Zhou FH, Hansen U. Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). *J Biol Chem* (2008) 283(29):20170–80. doi: 10.1074/jbc.M710139200
- Groulx JF, Gagne D, Benoit YD, Martel D, Basora N, Beaulieu JF. Collagen VI is a basement membrane component that regulates epithelial cell-fibronectin interactions. *Matrix Biol* (2011) 30(3):195–206. doi: 10.1016/j.matbio.2011.03.002
- Naba A, Pearce OMT, Del Rosario A, Ma D, Ding H, Rajeev V, et al. Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *J Proteome Res* (2017) 16(8):3083–91. doi: 10.1021/acs.jproteome.7b00191
- Aktary Z, Alaei M, Pasdar M. Beyond cell-cell adhesion: Plakoglobin and the regulation of tumorigenesis and metastasis. *Oncotarget* (2017) 8(19):32270–91. doi: 10.18632/oncotarget.15650

AUTHOR CONTRIBUTIONS

FY, MQ, XL, and YM: study concept and design. YM, HG, HC, and JL: acquisition and analysis of the data. YM and MQ: drafting and revising of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Nature Science Foundation of China (81702256 to MQ and 81802824 to XL), Natural Science Foundation of Beijing (7182169 to MQ), and Peking University People's Hospital Scientific Research Development Funds (RDH 2020-10 to MQ).

ACKNOWLEDGMENTS

We thank Charlesworth Author Services for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.633420/full#supplementary-material>

Supplementary Table 1 | Clinical characteristics data of dataset GSE26939.

Supplementary Table 2 | Spearman correlation analysis between expression of COL6A6 and Immunomodulators in lung adenocarcinoma from TISIDB database.

Supplementary Figure 1 | COL6A6 protein concentration in plasma detected by mass spectrometry.

Supplementary Figure 2 | Cox regression analyses of gene signatures and risk score. **(A)** The hazard ratios of genes integrated into the prognostic signature. **(B)** Multivariate Cox regression analysis of the risk score in LUAD regarding OS.

Supplementary Figure 3 | COL6A6 expression was distinctly higher in the lung than in other tissues. **(A)** COL6A6 expression profile in different tissues according to HPA database. **(B)** COL6A6 expression profiles across all tumor samples and paired normal tissues according to GEPIA2.

11. Long R, Liu Z, Li J, Yu H. COL6A6 interacted with P4HA3 to suppress the growth and metastasis of pituitary adenoma via blocking PI3K-Akt pathway. *Aging (Albany NY)* (2019) 11(20):8845–59. doi: 10.18632/aging.102300
12. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* (2016) 44(8):e71. doi: 10.1093/nar/gkv1507
13. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* (2019) 47(W1):W556–W60. doi: 10.1093/nar/gkz430
14. Cai L, Lin S, Girard L, Zhou Y, Yang L, Ci B, et al. LCE: an open web portal to explore gene expression and clinical associations in lung cancer. *Oncogene* (2019) 38(14):2551–64. doi: 10.1038/s41388-018-0588-2
15. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6(269):pl1. doi: 10.1126/scisignal.2004088
16. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* (2015) 347(6220):1260419. doi: 10.1126/science.1260419
17. Hruz T, Laule O, Szabo G, Wessendorf F, Bleuler S, Oertle L, et al. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinf* (2008) 2008:420747. doi: 10.1155/2008/420747
18. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47(D1):D607–D13. doi: 10.1093/nar/gky1131
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
22. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* (2017) 77(21):e108–e10. doi: 10.1158/0008-5472.CAN-17-0307
23. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* (2019) 37(7):773–82. doi: 10.1038/s41587-019-0114-2
24. Ru BB, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* (2019) 35(20):4200–2. doi: 10.1093/bioinformatics/btz210
25. Choi I, Wells BJ, Yu CH, Kattan MW. An empirical approach to model selection through validation for censored survival data. *J BioMed Inform* (2011) 44(4):595–606. doi: 10.1016/j.jbi.2011.02.005
26. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* (2013) 32(30):5381–97. doi: 10.1002/sim.5958
27. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* (2008) 26(8):1364–70. doi: 10.1200/Jco.2007.12.9791
28. Bremnes RM, Busund LT, Kilvaer TL, Andersen S, Richardsen E, Paulsen EE, et al. The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer. *J Thorac Oncol* (2016) 11(6):789–800. doi: 10.1016/j.jtho.2016.01.015
29. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023
30. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* (2014) 13(2):397–406. doi: 10.1074/mcp.M113.035600
31. Duff MO, Olson S, Wei X, Garrett SC, Osman A, Bolisetty M, et al. Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* (2015) 521(7552):376–9. doi: 10.1038/nature14475
32. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* (2012) 12(4):298–306. doi: 10.1038/nrc3245
33. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors. *Adv Anat Pathol* (2017) 24(6):311–35. doi: 10.1097/PAP.0000000000000161
34. Remark R, Becker C, Gomez JE, Damotte D, Dieu-Nosjean MC, Sautes-Fridman C, et al. The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome. *Am J Respir Crit Care Med* (2015) 191(4):377–90. doi: 10.1164/rccm.201409-1671PP
35. Simon SCS, Utikal J, Umansky V. Opposing roles of eosinophils in cancer. *Cancer Immun Immunother* (2019) 68(5):823–33. doi: 10.1007/s00262-018-2255-4
36. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. *BioMed Pharmacother* (2017) 95:55–61. doi: 10.1016/j.biopha.2017.08.003
37. Harjunpaa H, Llort Asens M, Guenther C, Fagerholm SC. Cell Adhesion Molecules and Their Roles and Regulation in the Immune and Tumor Microenvironment. *Front Immunol* (2019) 10:1078. doi: 10.3389/fimmu.2019.01078
38. Wu TC, Xu K, Banchereau R, Marches F, Yu CI, Martinek J, et al. Reprogramming tumor-infiltrating dendritic cells for CD103+ CD8+ mucosal T-cell differentiation and breast cancer rejection. *Cancer Immunol Res* (2014) 2(5):487–500. doi: 10.1158/2326-6066.CIR-13-0217
39. Mereness JA, Bhattacharya S, Wang Q, Ren Y, Pryhuber GS, Mariani TJ. Type VI collagen promotes lung epithelial cell spreading and wound-closure. *PLoS One* (2018) 13(12):e0209095. doi: 10.1371/journal.pone.0209095
40. Tulla M, Penttinen OT, Viitasalo T, Kapyla J, Impola U, Nykivist P, et al. Selective binding of collagen subtypes by integrin alpha 1I, alpha 2I, and alpha 10I domains. *J Biol Chem* (2001) 276(51):48206–12. doi: 10.1074/jbc.M104058200
41. Okkenhaug K. Signaling by the phosphoinositide 3-kinase family in immune cells. *Annu Rev Immunol* (2013) 31:675–704. doi: 10.1146/annurev-immunol-032712-095946
42. Peng W, Chen JQ, Liu C, Malu S, Creasy C, Tetzlaff MT, et al. Loss of PTEN Promotes Resistance to T Cell-Mediated Immunotherapy. *Cancer Discov* (2016) 6(2):202–16. doi: 10.1158/2159-8290.CD-15-0283
43. O'Donnell JS, Massi D, Teng MWL, Mandalia M. PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux. *Semin Cancer Biol* (2018) 48:91–103. doi: 10.1016/j.semcancer.2017.04.015
44. Lastwika KJ, Wilson W3rd, Li QK, Norris J, Xu H, Ghazarian SR, et al. Control of PD-L1 Expression by Oncogenic Activation of the AKT-mTOR Pathway in Non-Small Cell Lung Cancer. *Cancer Res* (2016) 76(2):227–38. doi: 10.1158/0008-5472.CAN-14-3362
45. Zhuang G, Zeng Y, Tang Q, He Q, Luo G. Identifying M1 Macrophage-Related Genes Through a Co-expression Network to Construct a Four-Gene Risk-Scoring Model for Predicting Thyroid Cancer Prognosis. *Front Genet* (2020) 11:591079. doi: 10.3389/fgene.2020.591079
46. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. *JAMA Oncol* (2017) 3(11):1529–37. doi: 10.1001/jamaoncol.2017.1609

Conflict 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.

Copyright © 2021 Ma, Qiu, Guo, Chen, Li, Li and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.