


# SNAI2 as a Prognostic Biomarker Based on Cancer-Associated Fibroblasts in Patients With Lung Adenocarcinoma

Clinical Medicine Insights: Oncology  
Volume 18: 1–10  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11795549241280506



Tian-Tian Li<sup>1\*</sup>, Qing-Gang Hao<sup>2\*</sup>, Zhao-Wei Teng<sup>3</sup>, Yuan Liu<sup>4</sup>,  
Jia-Fan Wu<sup>4</sup>, Jun Zhang<sup>5</sup> and Li-Rong Yang<sup>5</sup> 

<sup>1</sup>Department of pneumology, The Central Hospital of Wuhan, Wuhan, China. <sup>2</sup>Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming, China. <sup>3</sup>The Central Laboratory and Department of Orthopedic, The Second Affiliated Hospital of Kunming Medical University, Kunming, China. <sup>4</sup>Department of general surgery, Kunming Medical University, Kunming, China. <sup>5</sup>Department of Oncology, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, Chengdu, China.

## ABSTRACT

**BACKGROUND:** Lung adenocarcinoma (LUAD) is a common type of malignant tumor with therapeutic challenges. Cancer-associated fibroblasts (CAFs) promote LUAD growth and metastasis, regulate the tumor immune response, and influence tumor treatment responses and drug resistance. However, the molecular mechanisms through which CAFs control LUAD progression are largely unknown. In this study, we aimed to determine the correlations between CAF-related genes and overall survival (OS) in patients with LUAD.

**METHODS:** We acquired the gene expression data and clinical information of 522 patients with LUAD patients from The Cancer Genome Atlas (TCGA) and 442 patients with LUAD from the Gene Expression Omnibus (GEO) databases. CAF infiltration levels were assessed using the Microenvironment Cell Population (MCP) counter, the Estimating the Proportions of Immune and Cancer cells (EPIC) algorithm, and Tumor Immune Dysfunction and Exclusion (TIDE) scores. A CAF-related gene network was constructed using the Weighted gene co-expression network analysis (WGCNA). Based on the CAF-related genes, univariate Cox regression and Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression analyses were performed to identify prognostic genes. Gene expression levels within the prognostic model were validated using the Cancer Cell Line Encyclopedia (CCLE) databases and Western blotting.

**RESULTS:** Our results demonstrated that high CAF scores were associated with lower survival rates in patients with LUAD. Gene modules that were highly correlated with high CAF scores were closely associated with tissue characteristics and extracellular matrix structures in LUAD. In addition, correlations between CAF scores and responses to immunotherapy and chemotherapy were observed. Finally, we found that SNAI2 expression was higher in lung cancer tissues than in normal tissues.

**CONCLUSION:** Deepening our understanding of the influence of CAFs on tumor progression and treatment response at the molecular level can aid the development of more effective therapeutic strategies. This study provides important insights into the functional mechanisms of action of CAFs in LUAD and highlights their clinical implications.

**KEYWORDS:** Lung adenocarcinoma, cancer-associated fibroblasts, SNAI2, biomarker

**RECEIVED:** April 9, 2024. **ACCEPTED:** August 15, 2024.

**TYPE:** Original Research Article

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Kunming University of Science and Technology and the First People's Hospital of Yunnan Province Joint Special Project on Medical Research (grant number: KUST-KH2022002Z), the Joint Special Fund of Applied Fundamental Research of Kunming Medical University granted by the Science and Technology Office of Yunnan (grant number: 202301AY070001-299), the Yunnan Province Orthopedics and Sports Rehabilitation Clinical Medical Research Center Open Project (grant number: 2022YJZX-GK04), and the National Natural Science Foundation of China

(grant number: 82060167). The funding agencies had no role in the design, conduct, write-up, or publication of the study.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHORS:** Jun Zhang, Department of Oncology, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, 82 Qinglong Road, Chengdu 610031, Sichuan, China. Email: Zhangjun2021123@163.com

Li-Rong Yang, Department of Oncology, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, 82 Qinglong Road, Chengdu 610031, Sichuan, China. Email: yanglir321@163.com

## Introduction

Lung cancer is among the most common malignancies worldwide and is the leading cause of cancer-related deaths.<sup>1</sup> The incidence of lung cancer has declined in the last century; however, mortality rates remain high. Non-small-cell lung cancer (NSCLC) accounts for approximately 80% to 85% of all lung cancer cases.<sup>2</sup> Lung adenocarcinoma (LUAD) is the predominant subtype of NSCLC

worldwide and is a highly heterogeneous tumor.<sup>3</sup> Many patients with LUAD have similar pathologies, gene mutations, and disease stages; however, their prognoses differ considerably depending on the location and time of the gene mutation and the tumor microenvironment (TME).<sup>4,5</sup>

The TME negatively influences the efficacy of immunotherapy because it suppresses immune cell function and promotes cancer cell survival, local invasion, and metastatic dissemination.<sup>6-8</sup> The TME of LUAD is mainly composed of various highly heterogeneous cell types, including tumor cells,

\*Tian-Tian Li and Qing-Gang Hao contributed equally to this work and share first authorship.



fibroblasts, endothelial cells, immune cells, vasculature, and the extracellular matrix (ECM).<sup>8-10</sup> Fibroblasts found in tumor tissues, termed cancer-associated fibroblasts (CAFs), are one of the major types of stromal cells that influence homeostasis.<sup>11</sup> CAFs promote tumor progression by secreting factors that stimulate angiogenesis, invasion, and cancer cell proliferation.<sup>12</sup> Moreover, they enhance the formation of an immunosuppressive network during antitumor therapy by secreting growth factors and cytokines, paracrine signaling, and ECM remodeling.<sup>13-15</sup> Therefore, the use of CAFs as candidate targets for targeted molecular therapy is an appealing and emerging strategy.

Snail family transcriptional repressor 2 (SNAI2) is a classical epithelial-mesenchymal transition (EMT)-related transcriptional inhibitor with an N-terminal SNAG domain and a C-terminal DNA-binding domain.<sup>16</sup> SNAI2 plays an important role in the progression of several cancers. Reprogramming of mesenchymal fibroblasts by SNAI2 contributes to tumor fibrous tissue proliferation and promotes progression and invasive metastasis of ovarian cancer by regulating EMT through the lncRNA AC005224.4/miR-140-3p/SNAI2 axis.<sup>17,18</sup> FBXO28 can inhibit hepatocellular carcinoma invasion and metastasis by promoting protein kinase A (PKA)-dependent phosphorylation-mediated degradation of SNAI2.<sup>19</sup> SNAI2 can significantly influence the TME by reactivating the tumor stroma and generating an immunosuppressive microenvironment in prostate cancer.<sup>20</sup> In breast cancer, ASB13 inhibits metastasis by promoting SNAI2 degradation.<sup>21</sup> However, whether SNAI2 is a CAF-related gene requires further investigation.

Furthermore, the clinical relevance of and molecular mechanisms by which CAFs regulate tumor progression remain unclear. In this study, we aimed to construct a model to determine the correlations between CAF-related genes and overall survival (OS) in patients with LUAD. CAF infiltration scores were calculated using the Estimating the Proportions of Immune and Cancer cells (EPIC) method, Microenvironment Cell Population (MCP) counter, and Tumor Immune Dysfunction and Exclusion (TIDE) scoring to determine the relationship between the CAF scores and OS in patients with LUAD based on data from the publicly available Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. Weighted gene co-expression network analysis (WGCNA) was used to identify the hub modules that were most strongly correlated with CAF infiltration.

## Material and Methods

### *Data of patients with LUAD*

The gene expression profiles of 522 patients with LUAD and corresponding patient clinical characteristics, including outcome, age, and survival time, were downloaded from TCGA (<https://gdc.nci.nih.gov/>). The batch effects were revised using the *sva* R package. The fragments per kilobase of transcript per million mapped reads (FPKM) format was used for the RNA-seq data from TCGA datasets. The normalized FPKM values were converted to

transcripts per million (TPM) and  $\log_2(\text{TPM} + 1)$  transformed. In addition, we included the normalized gene expression profiles of 442 patients with LUAD from the GSE72094 dataset from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), with complete information regarding OS-related transcriptomics, in the validation dataset.

### *Assessment of CAF scores*

CAF infiltration scores were evaluated using three methods—estimation of the marker gene expression-based MCP counter,<sup>22</sup> the EPIC algorithm,<sup>23</sup> and TIDE scores downloaded from the TIDE database (<http://tide.dfci.harvard.edu>). Tumor sample data obtained from TCGA and the GEO databases were divided into high- and low-CAF-score groups according to the cutoff value of the CAF score. Then, we analyzed the relationship between CAF scores and survival using the Survival R package, and samples were screened based on *P*-values < .05.

### *CAF co-expression network construction and hub gene selection*

Gene co-expression network analysis and the selection of hub genes that target CAF infiltration were performed using the WGCNA R package,<sup>24</sup> and the genes with the top 5000 median absolute deviations (MADs) in both the TCGA and GSE72094 datasets were chosen for WGCNA analysis. Thereafter, the appropriate soft-thresholding power  $\beta$  for adjacency computation was graphically confirmed. Subsequently, the adjacency matrix was clustered using a topological overlap matrix (TOM) and the corresponding dissimilarity matrix (1-TOM); a hierarchical clustering dendrogram was constructed, and genes with similar expressions were divided into different modules. The first principal component of each module expression was summarized by the module eigengene (ME). Pearson's correlations between the ME and clinical features were calculated, and the module with the strongest correlation was selected for subsequent analyses. Next, we evaluated module membership (MM) and gene significance (GS) for the traits of individual genes in the hub module, and hub genes with  $\text{GS} > 0.4$  and  $\text{MM} > 0.6$  were selected. Finally, the overlapping hub genes between the TCGA and GSE72094 datasets constituted the final set of hub genes.

### *Establishment and validation of the CAF-related prognostic signature*

CAF-related hub genes from TCGA were subjected to univariate Cox regression analysis to screen for prognosis-related hallmark genes using the Survival R package, with  $P < .05$ . To construct a prognostic correlation model, we performed Least Absolute Shrinkage and Selection Operator (LASSO)<sup>25</sup> Cox regression analysis. CAF-related prognostic genes from TCGA were used as the training set, and CAF-related hub genes from

the GEO dataset were used as the testing set. Patients in the two datasets were divided into high- and low-risk groups based on their median risk score. Risk scores were calculated using the following formula:

$$\text{Risk score} = \text{coef}(\text{gene 1}) \diamond \text{expr}(\text{gene 1}) + \text{coef}(\text{gene 2}) \diamond \text{expr}(\text{gene 2}) + \dots + \text{coef}(\text{gene n}) \diamond \text{expr}(\text{gene n}),$$

where  $\text{coef}(\text{gene } n)$  represents the coefficient of CAF-related genes correlated with survival and  $\text{expr}(\text{gene } n)$  represents the expression of CAF-related genes.

#### *Gene set enrichment analysis and single-sample gene set enrichment analysis of TCGA cohort*

To explore unique Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and hallmark gene sets in the high- and low-risk groups, gene set enrichment analysis (GSEA) was performed based on the hallmark gene sets, using the clusterProfiler package in R, and pathways with  $P < .05$  were considered to be significantly enriched. Single-sample GSEA (ssGSEA) was performed on several of the enriched pathways using the GSVA R package.

#### *Cancer cell line encyclopedia and HPA datasets*

To verify the expression levels of genes involved in the construction of prognostic models, we downloaded mRNA expression data from the Cancer Cell Line Encyclopedia (CCLE) database (<https://portals.broadinstitute.org/ccle>)<sup>26</sup> and compared the expression levels of these genes between fibroblasts and lung cell lines using the plyr R package and the Wilcoxon test. In addition, the protein expression of CAF-related genes in lung tissues was determined using immunohistochemical (IHC) images from the HPA online database (<https://www.proteinatlas.org/>).<sup>27</sup>

#### *Chemotherapeutic sensitivity predictions*

To predict chemosensitivity in the high- and low-risk groups, we downloaded drug sensitivity and data representation files from the largest publicly available pharmacogenomics database, Genomics of Drug Sensitivity in Cancer (GDSC; <https://www.Cancerrxgene.org/>),<sup>28</sup> and performed chemosensitivity analysis using the oncoPredict R package.

#### *Genetic somatic mutation analysis*

Genetic somatic mutation data for the LUAD samples were downloaded from TCGA. The maftools R package was used for the recognition and visualization of somatic variants in both the low- and high-risk groups. The tumor mutational burden (TMB), a sum of the nonsynonymous mutations per million bases in coding regions, has been proposed as a biomarker of immunotherapy efficacy.<sup>29</sup>

#### *Cell culture*

The human LUAD cell lines H1573, H1975, HCC827, A549, and SPC-A-1 were purchased from the Kunming Institute of Zoology. These cell lines were cultured in RPMI 1640 medium supplemented with 10% v/v fetal bovine serum (FBS), 10000 units of penicillin, and 10 mg/mL streptomycin, at 37°C in an incubator with 5% CO<sub>2</sub>. The human bronchial epithelial cells BESA-2b, obtained from iCell, China, were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% v/v FBS, 10000 units of penicillin, and 10 mg/mL of streptomycin at 37°C in an incubator with 5% CO<sub>2</sub>.

#### *Western blot analysis*

A total of  $1 \times 10^6$  cells were collected and lysed in radioimmunoprecipitation assay (RIPA) lysis buffer (Beyotime, Shanghai, China). The cell lysate was centrifuged to isolate proteins (30 or 35 µg), which were collected, resolved using Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with a 10% resolving gel, and electroblotted onto Polyvinylidene fluoride (PVDF) membranes (Millipore, Carrigtwohill, Germany). The membranes were blocked with 5% (w/v) dry skimmed milk in Tris buffered saline (TBS) and incubated overnight at 4°C with primary antibodies against SNAI2 (Proteintech, Cat#12129-1-AP, USA). The membranes were then washed with TBS containing 0.1% (v/v) Tween-20 and incubated with a horseradish peroxidase-conjugated secondary antibody (Proteintech, USA).

#### *Statistical analysis*

All statistical analyses were performed using the R software. The Wilcoxon test was used for pairwise comparisons. The Survival and Survminer R packages were used for OS analysis using the Kaplan–Meier curve with the log-rank test.  $P < .05$  was considered statistically significant.

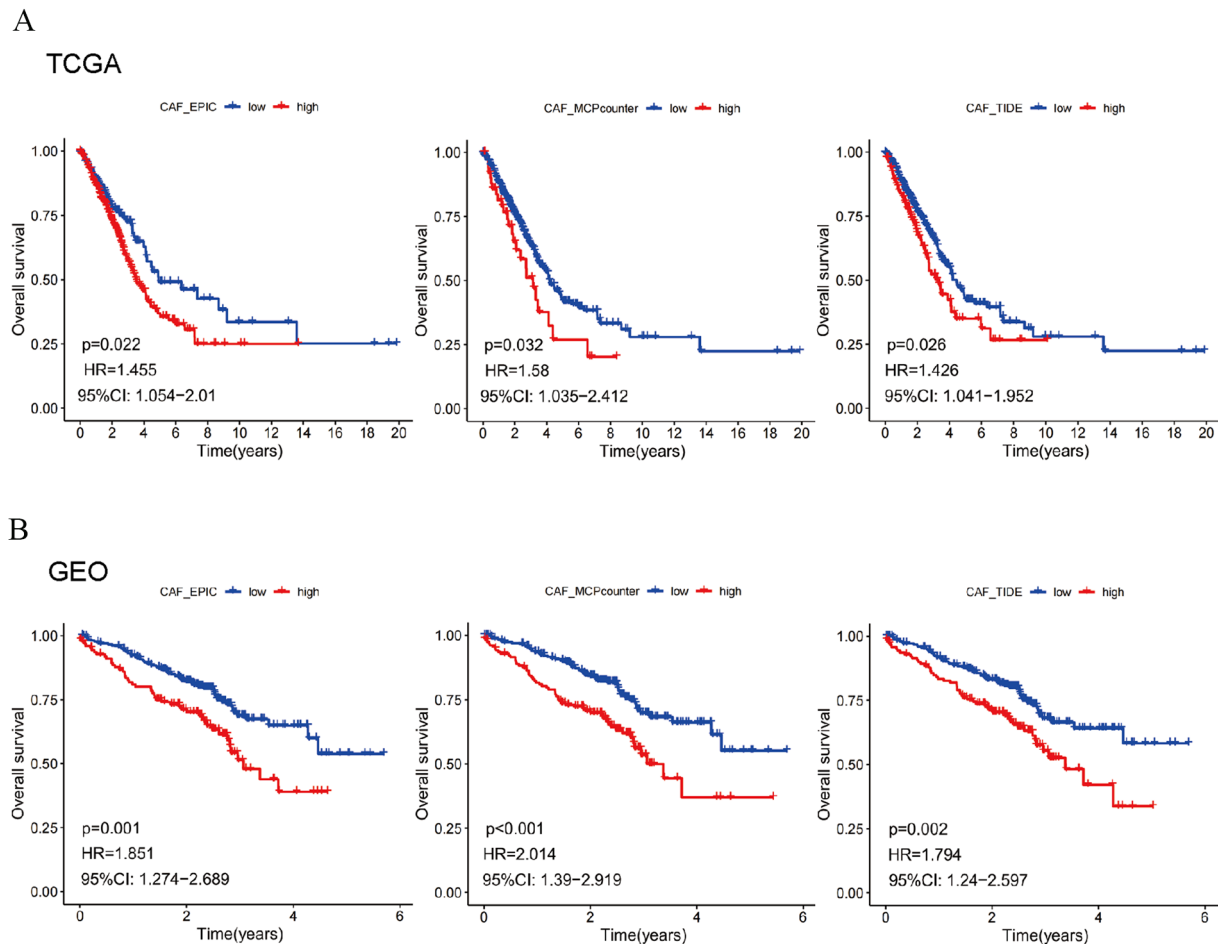
## **Results**

#### *Survival analysis of CAF scores in patients with LUAD*

We calculated CAF infiltration scores using the EPIC, MCP counter, and TIDE methods. OS analysis using the Kaplan–Meier curve with the log-rank test demonstrated that the high-CAF-score group exhibited worse OS than the low-CAF-score group in both the GSE72094 (Figure 1A) and TCGA (Figure 1B) cohorts of patients with LUAD. This highlights the importance of further research on CAF in LUAD.

#### *Construction of CAF-related modules and selection of hub genes*

To construct a co-expression network of CAF scores, we performed WGCNA in the GSE72094 and TCGA cohorts. Regarding the TCGA cohort, the CAF scores had the strongest



**Figure 1.** Survival analysis of patients with lung adenocarcinoma (LUAD) using cancer-associated fibroblast (CAF) scores. (A) Patients with LUAD and increased CAF infiltration had worse overall survival (OS), as revealed by Kaplan–Meier analysis of The Cancer Genome Atlas (TCGA) data. (B) Patients with LUAD and increased CAF infiltration had worse OS, as revealed by Kaplan–Meier analysis of the GSE72094 data.

\* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

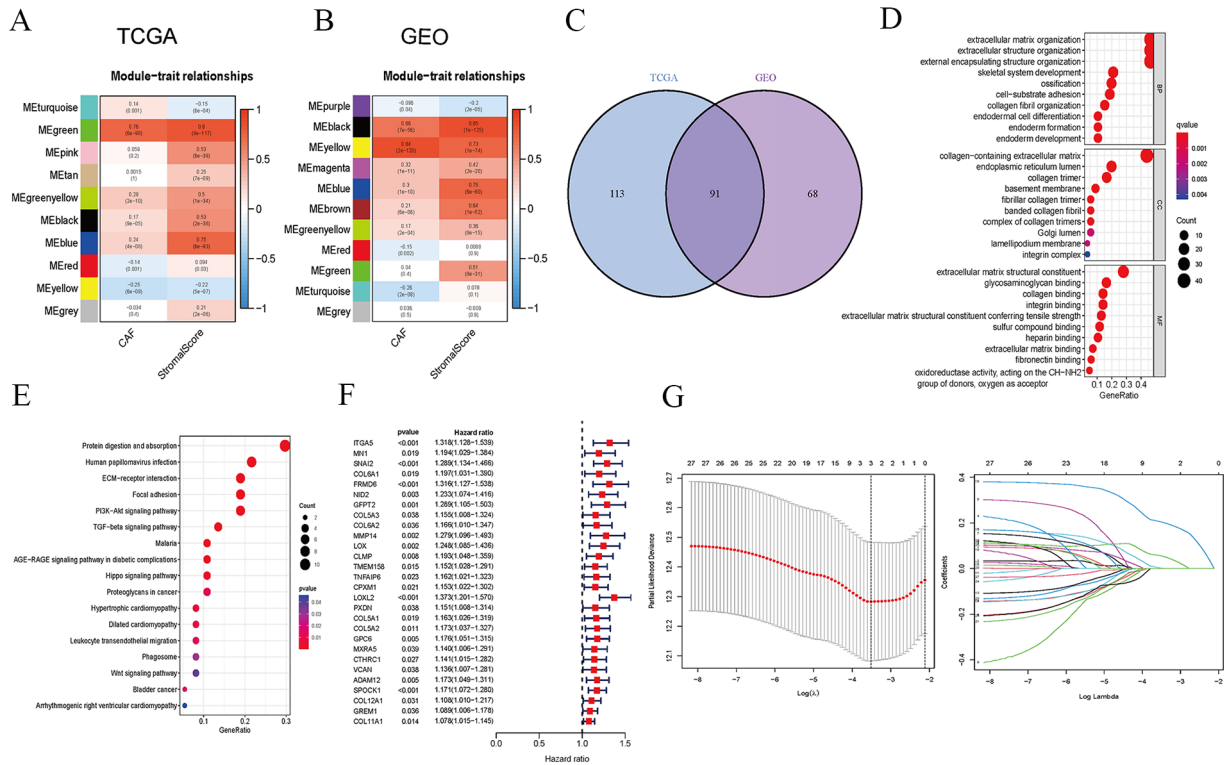
positive correlation with the green module (Figure 2A,  $R = 0.76$ ,  $P < .001$ ). For the GSE72094 cohort, the CAF scores had the strongest positive correlation with the yellow module (Figure 2B,  $R = 0.84$ ,  $P < .001$ ). Accordingly, we further explored the genes in these two modules and screened 91 genes of the 204 genes in the green module of TCGA dataset and 159 genes in the yellow module of GSE72094 as hub genes, which were visualized using a Venn diagram (Figure 2C).

Then, we performed gene ontology (GO) and KEGG analyses of the 91 hub genes. ECM and extracellular structure organization in the biological process (BP) category, collagen-containing ECM in the cellular component (CC) category, and ECM structural constituents in the molecular function (MF) category were the main significantly enriched GO terms. These are key components of ECM processes in the TME (Figure 2D). The enriched KEGG pathways were mainly involved in desmoplastic and proliferation processes, such as focal adhesion, ECM–receptor interaction, the PI3K–Akt signaling pathway, and the TGF- $\beta$  signaling pathway (Figure 2E). These enrichment terms are associated with the prognosis of patients with LUAD.

#### Prognostic risk model based on CAF-related genes

We screened 28 OS-related genes (*ITGA5*, *MN1*, *SNAI2*, *COL6A1*, *FRMD6*, *NID2*, *GFPT2*, *COL5A3*, *COL6A2*, *MMP14*, *LOX*, *CLMP*, *TMEM158*, *TNEAIP6*, *CPXM1*, *LOXL2*, *PXDN*, *COL5A1*, *COL5A2*, *GPC6*, *MXRA5*, *CTHRC1*, *VCAN*, *ADAM12*, *SPOCK1*, *COL12A1*, *GREM1*, and *COL11A1*) with  $P < .05$  from the 91 common hub genes to establish a prognostic risk model based on CAF-related genes in patients with LUAD using LASSO-penalized Cox analysis (Figure 2F). Initially, 522 samples from TCGA were used as the training group, and 442 samples from GSE72094 were used as the testing group; a cross-validation method was used to optimize the prognostic model (Figure 2G). Subsequently, patients with LUAD were divided into low- and high-risk groups according to the median cutoff values. Kaplan–Meier curves revealed that the OS of patients with LUAD in the low-risk group was longer than that of the patients in the high-risk group in both the TCGA and GSE72094 datasets (Figure 3A and B).





**Figure 2.** Cancer-associated fibroblast (CAF)-related modules constructed using weighted gene co-expression network analysis (WGCNA). (A, B) Module-trait relationships showing correlations between each gene module eigengene and phenotype in The Cancer Genome Atlas [TCGA] (A) and GSE72094 cohorts (B). (C) Venn diagram presenting overlapping TCGA and GSE72094 module genes. (D) Gene ontology (GO) analysis of 91 genes and the enriched biological process (BP), cellular component (CC), and molecular function (MF) terms. (E) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the enriched pathways of the 91 genes. (F) Forest diagram showing overall survival (OS)-associated genes, as revealed by univariate Cox analysis of TCGA data. (G) Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression was conducted to construct the prognostic model. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

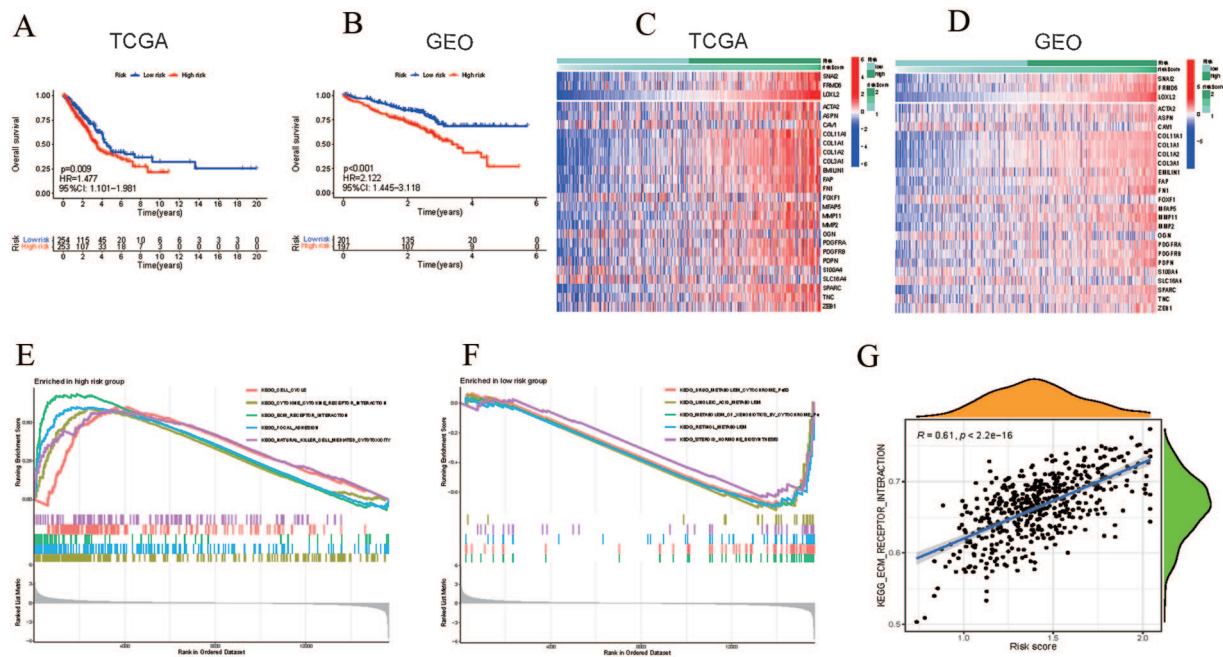
Moreover, three CAF-related genes (*SNAI2*, *FRMD6*, and *LOXL2*) were highly expressed in the high-risk group in both the TCGA and GSE72094 datasets, which is consistent with the results obtained using various CAF markers (Figure 3C and D). GSEA was performed on the TCGA dataset between the high- and low-risk groups, and the essential KEGG signaling pathways were involved in the cell cycle, cytokine-cytokine receptor interactions, and ECM-receptor interactions in the high-risk group (Figure 3E). Genes in the low-risk group were mainly enriched in drug metabolism, cytochrome P450, linoleic acid metabolism, and the metabolism of xenobiotics by cytochromes (Figure 3F). The ssGSEA results also showed that CAF risk scores were positively correlated with ECM-receptor interaction (Figure 3G).

### Comparison of the TMB and the efficacy of immunotherapy and chemotherapy between the high- and low-risk groups

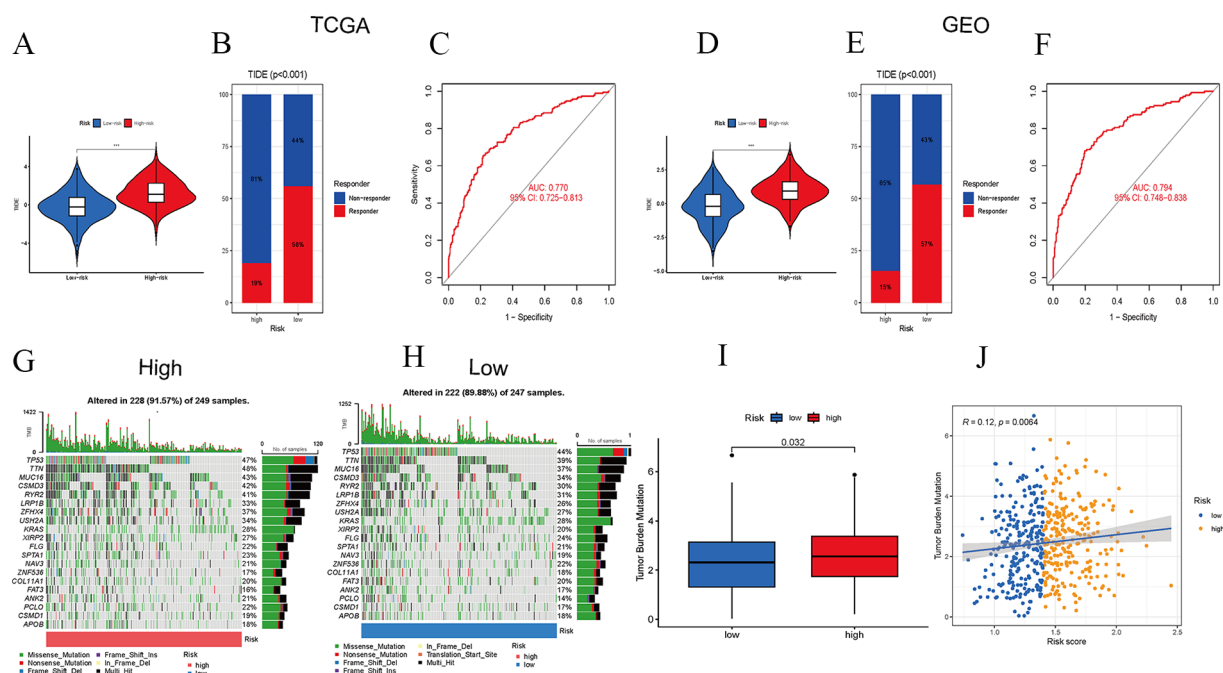
Immunotherapy has become a powerful clinical treatment strategy for patients with cancer.<sup>30</sup> Using the TIDE method, we investigated the potential clinical efficacy of immunotherapy in patients with LUAD. We found that the low-risk group

displayed significantly lower TIDE scores than the high-risk group in both the TCGA and GSE72094 datasets (Figure 4A and D), which implied that patients in the low-risk group had lower potential for immune evasion and were more likely to benefit from immune checkpoint inhibitor (ICI) therapy than those in the high-risk group. In the TCGA dataset, the low-risk group had similar proportions of responders (56%) and non-responders (44%); however, the proportion of non-responders was significantly higher than that of responders (81% vs 19%) in the high-risk group (Figure 4B). Similar results were obtained for the GSE72094 cohort (Figure 4E). These findings suggested that the low-risk group exhibited higher sensitivity to immunotherapy. Moreover, the Area under the curve (AUC) values were 0.77 and 0.794 in the TCGA (Figure 4C) and GSE72094 cohorts, respectively (Figure 4F), indicating the excellent sensitivity and specificity of our CAF prognostic model for immunotherapy response prediction.

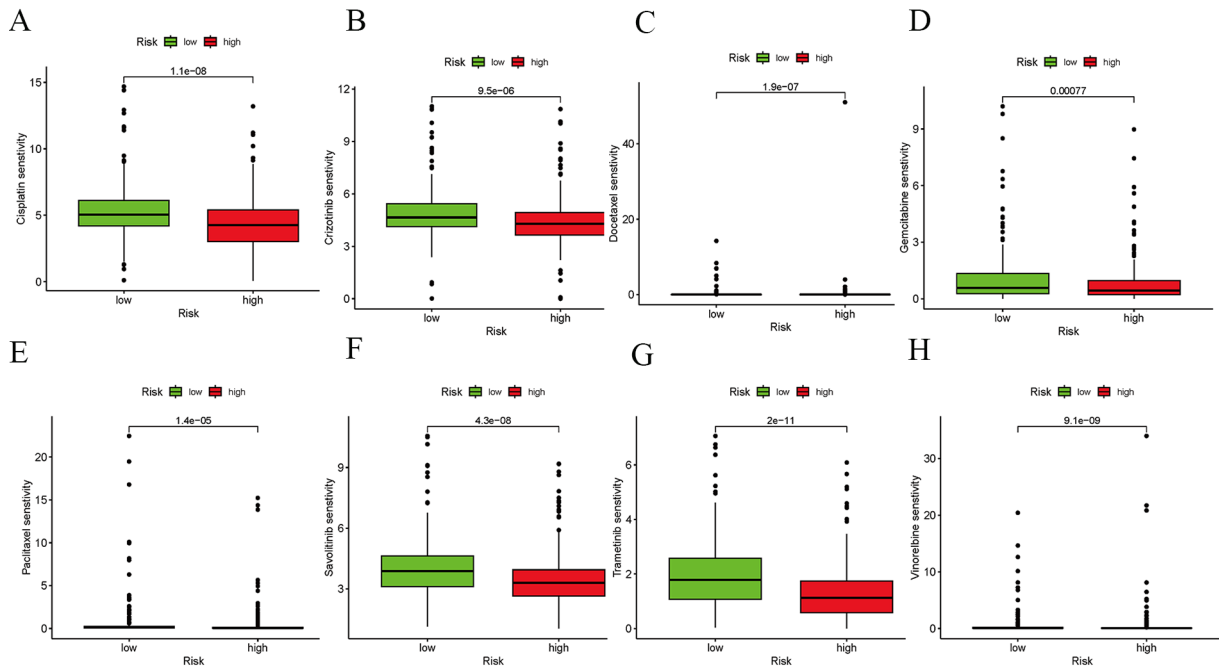
The top 20 genes with the highest mutational frequencies were identified in the low- and high-risk subgroups. Surprisingly, the genes (*TP53*, *TTN*, *MUC16*, *CSMD3*, *RYR2*, *LRP1B*, *ZFH4*, *USH2A*, *KRAS*, *XIRP2*, *FLG*, *SPTA1*, *NAV3*, *ZNF536*, *COL11A1*, *FAT3*, *ANK2*, *PCLO*, *CSMD1*, and *APOB*; Figure 4G) and frequencies of mutations were highly



**Figure 3.** Differences in pathways between the two groups. (A, B) Kaplan–Meier survival curves for patients with lung adenocarcinoma (LUAD) in the high- and low-risk groups from The Cancer Genome Atlas (TCGA) and GSE72094 cohorts. (C, D) Heat map revealing the expression patterns of the three identified cancer-associated fibroblast (CAF) genes and the CAF risk scores in TCGA and GSE72094. (E, F) GSEA of KEGG gene sets in the high- and low-risk groups. (G) Single sample gene set enrichment analysis (ssGSEA) results and the positive correlation between CAF risk scores and extracellular matrix (ECM)–receptor interaction-related genes in the TCGA cohort. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .



**Figure 4.** (A–F) Tumor Immune Dysfunction and Exclusion (TIDE) immunotherapy prediction analyses. (A, D) TIDE cancer-associated fibroblast (CAF) risk scores were higher in the high-risk groups than those in the low-risk groups in the GSE72094 and The Cancer Genome Atlas (TCGA) cohorts. (B, E) Proportions of responders and non-responders in the high- and low-risk groups in the GSE72094 and TCGA cohorts. (C, F) Receiver operating characteristic (ROC) curves of the risk signatures in the GSE72094 and TCGA cohorts. (G, H) Oncoplots showing the top 20 mutational genes in the high- and low-CAF-risk groups of the TCGA cohort. (I) Boxplot displaying the tumor mutation burden (TMB) values of the high- and low-risk groups. (J) CAF risk scores were significantly positively correlated with TMB values. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .



**Figure 5.** (A-H) Boxplot showing a comparison of the sensitivities of the high- and low-risk groups to several chemotherapy drugs. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

similar between the two groups. In addition, we found that the high-risk group had higher TMB values than the low-risk group (Figure 4H), and the CAF risk score was significantly positively correlated with the TMB value (Figure 4I).

Chemotherapy is the standard approach for treating LUAD. Low-risk patients with LUAD exhibited increased sensitivity to cisplatin, crizotinib, docetaxel, gemcitabine, savolitinib, trametinib, and paclitaxel. In contrast, the high-risk group was estimated to be more sensitive to vinorelbine in the TCGA cohort (Figure 5).

#### Validation of CAF-related genes

Through analysis of the CCLE database, we found that the mRNA expression of *SNAI2*, *FRMD6*, and *LOXL2* was high in fibroblast cell lines (Figure 6A). In addition, IHC data confirmed that *SNAI2* expression was high in both cancerous and normal lung tissues (Figure 6B). However, we found that *SNAI2* was not expressed in CAF cells (Supplemental Figure 1). The Western blot results showed that *SNAI2* expression was higher in lung cancer cells (H1975, HCC827, A549, SPC-A-1, and H1573) than that in Beas-2b pulmonary epithelial cells (Figure 6B). Moreover, we found that the percentage of cells in the S phase significantly increased, while the proportion of those in the G2/M phase and G1 phase significantly decreased after knocking down *SNAI2* in LUAD cells, indicating that the cell cycle of LUAD cells was halted in the S phase. Thus, these results suggest that knocking down *SNAI2* may significantly inhibit the proliferation of LUAD cells via cell cycle arrest in the S phase (Supplemental Figure 2). These results suggest that *SNAI2* is a CAF-specific marker.

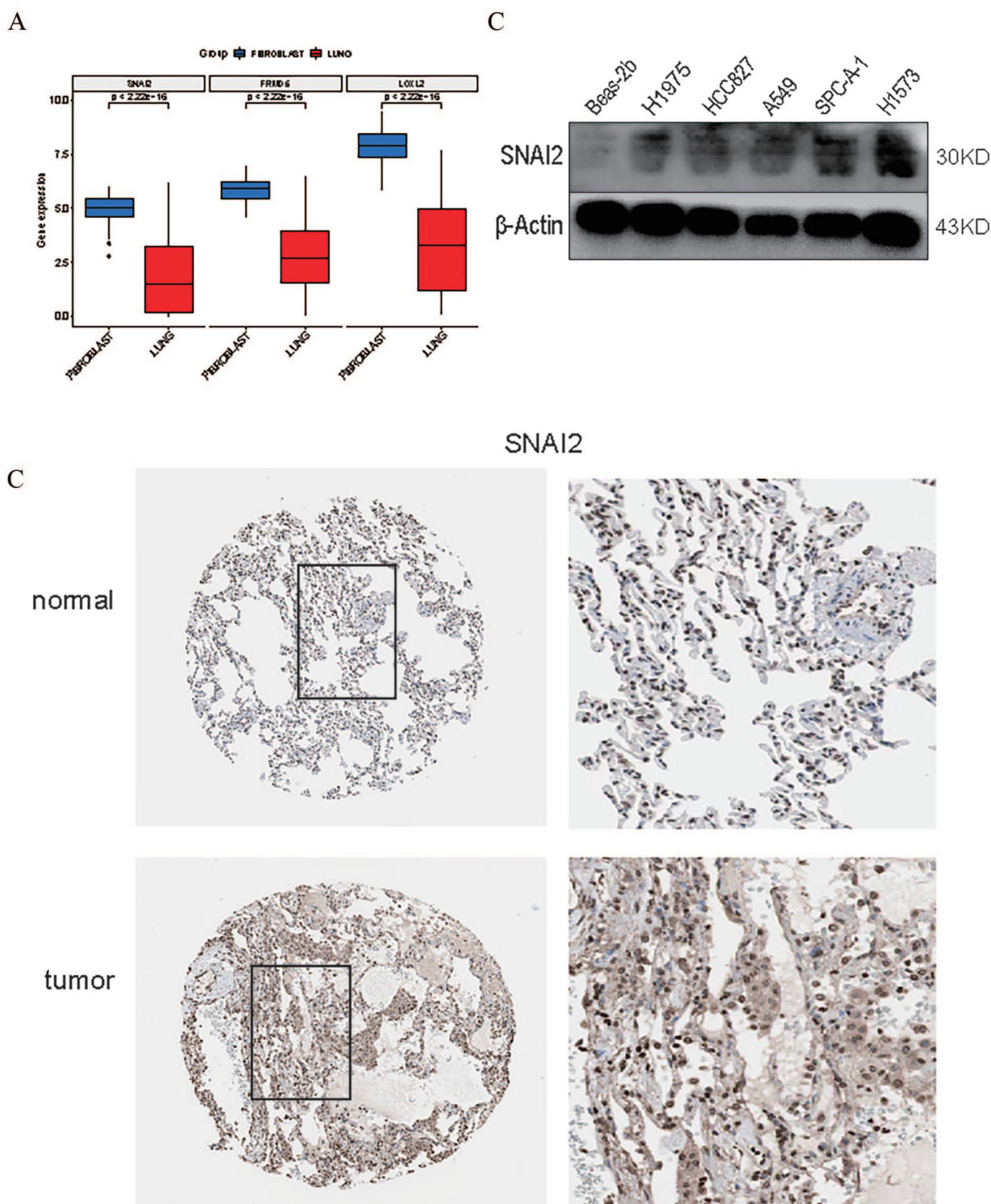
#### Discussion

CAFs have long been considered to play a crucial role in the development and prognosis of LUAD. Our results revealed that CAF infiltration was associated with poor prognosis in patients with LUAD in two independent cohorts (GSE72094 and TCGA). Patients with high CAF scores had lower survival rates than those with low CAF scores, which highlights the significant role of CAF in the TME.

CAF-related genes in multiple cancer types have been an important focus of research to identify critical molecular pathways in CAF subtypes that may be associated with clinical outcomes, disease progression, and immunotherapy resistance.<sup>31-34</sup> Recently, there has been a decline in the incidence of lung cancer owing to improved cancer screening and treatment in clinical practice.<sup>35</sup> This implies that targeted cancer control, interventions, early detection, and treatment can help reduce cancer-related mortality.<sup>36</sup> Some studies have demonstrated several roles of CAFs in LUAD; however, the molecular markers of CAF need to be elucidated to better classify tumor subtypes and facilitate the establishment of CAF-specific targeted therapies. Furthermore, the Heterogeneous nuclear ribonucleoprotein K (HNRNPK)/Chloride channel 3 (CLCN3) axis promotes the progression of LUAD through CAF-tumor interactions,<sup>37</sup> and CAF activation can facilitate the invasion of breast cancer cells by CAFs.<sup>38</sup>

Cancer progression, recurrence, and metastasis are strongly associated with the TME, of which the ECM is a key component.<sup>39</sup> The ECM regulates cell differentiation, proliferation, migration, death, and survival and plays a crucial role in maintaining cellular homeostasis.<sup>40</sup> Disruption of the ECM composition and structure is linked to carcinogenesis and cancer





**Figure 6.** Analysis of the expression of cancer-associated fibroblast (CAF)-related genes. (A) mRNA expression levels of the three CAF-related genes in CAFs and lung adenocarcinoma (LUAD) cell lines. (B) Protein expression of SNAI2 in LUAD tissues from the Human Protein Atlas database. (C) Western blotting demonstrated the protein expression of SNAI2 in LUAD cell lines. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ .

progression.<sup>41</sup> By constructing a co-expression network of CAF-related genes and screening for hub genes, we identified gene modules that were highly correlated with CAF scores, as well as functionally enriched GO and KEGG pathways. Specifically, GO analysis revealed that CAF-related genes were mainly associated with the ECM structure and tissues, which

is consistent with the findings of previous research, indicating that CAF-related genes play important roles in ECM remodeling, tumor development, and metastasis.<sup>38</sup> In addition, KEGG analysis revealed that the enriched pathways of CAF-related genes, such as ECM-receptor interactions, were related to cancer development. The mechanical remodeling of the



ECM by CAF is crucial for tumor cell migration and invasion.<sup>42</sup> Changes in these enrichment pathways have been implicated in the development and metastasis of LUAD, which further supports the association between CAF and LUAD. A total of 28 genes that were significantly associated with the survival of patients with LUAD were identified using univariate Cox regression analysis, and a prognostic risk model based on CAF-related genes was established using the LASSO analysis. The training and testing groups in the TCGA and GSE72094 cohorts were divided into low- and high-risk subgroups based on the median. Kaplan–Meier curves showed that the OS of the low-risk group was better than that of the high-risk group, confirming that CAF-related genes may serve as predictive biomarkers for LUAD. Furthermore, our results showed that three CAF-related genes (*SNAI2*, *FRMD6*, and *LOXL2*) were highly expressed in the high-risk group, which is consistent with the panel of CAF markers collected from the TCGA and GSE72094 datasets. Additional experiments confirmed the high mRNA expression of *SNAI2*, *FRMD6*, and *LOXL2* in fibroblast cell lines in the CCLE database. In addition, Western blotting validated the higher expression of *SNAI2* in lung cancer cell lines compared with that in normal tissues. Moreover, we found that knocking down *SNAI2* can significantly inhibit the proliferation of LUAD cells.

*SNAI2* (*SLUG*) is a C<sub>2</sub>H<sub>2</sub> zinc-finger transcriptional repressor belonging to a three-member family of SNAIL proteins (*SNAIL*, *SNAI2*, and *SMUC*).<sup>43</sup> *SNAI2* is known for its role in orchestrating EMT, during which tumor cells lose their epithelial characteristics and acquire a fibroblast-like cell phenotype and enhanced migratory abilities.<sup>18</sup> Recent studies have revealed additional roles of *SNAI2* in cancer progression, including in the activation of tumor-initiating cells, and cell cycle regulation, invasion, and metastasis.<sup>44–46</sup> In addition, *SNAI2* is a key transcription factor in the regulation of squamous cell differentiation, which is a hallmark histopathological feature of lung squamous cell carcinoma.<sup>31</sup> Our findings suggest that *SNAI2* may serve as a specific marker for CAFs.

CAF s can regulate the formation of an immunosuppressive network to weaken antitumor immunity.<sup>47,48</sup> Our study also explored the differences between the low- and high-CAF groups in terms of TMB and response to immunotherapy and chemotherapy. The results showed that in both the TCGA and GSE72094 cohorts, the low-risk group had significantly lower TIDE scores than the high-risk group, suggesting that patients in the low-risk group had a lower potential for immune escape and were more likely to benefit from ICI therapy than those in the high-risk group. Furthermore, our findings revealed a significant positive correlation between CAF risk scores and TMB values. Regarding chemotherapy, the low-risk group exhibited higher sensitivity to drugs such as cisplatin, crizotinib, docetaxel, gemcitabine, sunitinib, and paclitaxel, whereas the high-risk group exhibited higher sensitivity to etoposide.

Thus, further exploration of correlations between CAF-related genes and OS in patients with LUAD is needed, including more basic experiments and clinical samples, which we will address in our future studies.

Furthermore, there may have been sample bias in the public database used. In addition, confounding factors, such as patient clinical characteristics, may have affected the results. Future studies should focus on investigating the function of *SNAI2* in LUAD and the underlying molecular mechanisms, as well as designing clinical trials to study the potential of CAFs as a therapeutic target.

## Conclusions

Overall, our findings enhance our understanding of the role of CAFs in LUAD progression and provide a foundation for future research and clinical applications. We will further study the potential of the CAF marker *SNAI2* in clinical trials in the future. *SNAI2* has potential as a CAF therapeutic target and could be combined with existing treatment options, such as chemotherapy and immunotherapy, to reduce the development of drug resistance. In addition, the CAF marker *SNAI2* has possible applications in prognosis assessment and treatment response monitoring of patients with LUAD to help clinicians better manage the disease and optimize treatment plans.

## Author Contributions

L-RY and JZ: Conceptualization, Writing—review & editing. Z-WT: Data curation, Funding acquisition. Q-GH, J-FW, and YL: Formal analysis, Validation, Software. T-TL: Methodology, Resources, Project administration. T-TL and Q-GH: Writing—original draft.

## Data Availability Statement

The datasets generated during the current study are available in the [TCGA] repository, [https://portal.gdc.cancer.gov/] and in the [GEO] repository, [https://www.ncbi.nlm.nih.gov/geo/].

## ORCID iD

Li-Rong Yang  <https://orcid.org/0000-0002-9939-4649>

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72:7–33.
2. Huang X, Xiao S, Zhu X, et al. MiR-196b-5p-mediated downregulation of FAS promotes NSCLC progression by activating IL6-STAT3 signaling. *Cell Death Dis.* 2020;11:785.
3. Lamort AS, Kaiser JC, Pepe MAA, et al. Prognostic phenotypes of early-stage lung adenocarcinoma. *Eur Respir J.* 2022;60:2101674.
4. Kalinke L, Janes SM. Two phenotypes that predict prognosis in lung adenocarcinoma. *Eur Respir J.* 2022;60:2200569.
5. Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer.* 2019;19:495–509.
6. Marin-Acevedo JA, Kimbrough EO, Lou Y. Next generation of immune checkpoint inhibitors and beyond. *J Hematol Oncol.* 2021;14:45.

7. Chen YP, Zhang Y, Lv JW, et al. Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. *Theranostics*. 2017;7:3585-3594.
8. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol*. 2020;30:R921-R925.
9. Liu Y, Cao X. Characteristics and significance of the pre-metastatic niche. *Cancer Cell*. 2016;30:668-681.
10. Zhao Y, Guo S, Deng J, et al. VEGF/VEGFR-targeted therapy and immunotherapy in non-small cell lung cancer: targeting the tumor microenvironment. *Int J Biol Sci*. 2022;18:3845-3858.
11. Erez N, Truitt M, Olson P, Arron ST, Hanahan D. Cancer-associated fibroblasts are activated in incipient neoplasia to orchestrate tumor-promoting inflammation in an NF-kappaB-dependent manner. *Cancer Cell*. 2010;17:135-147.
12. Eyden B. The myofibroblast: phenotypic characterization as a prerequisite to understanding its functions in translational medicine. *J Cell Mol Med*. 2008;12:22-37.
13. Monteran L, Erez N. The dark side of fibroblasts: cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front Immunol*. 2019;10:1835.
14. Sahai E, Astsaturov I, Cukierman E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*. 2020;20:174-186.
15. Kim HJ, Yang K, Kim K, et al. Reprogramming of cancer-associated fibroblasts by apoptotic cancer cells inhibits lung metastasis via Notch1-WISP-1 signaling. *Cell Mol Immunol*. 2022;19:1373-1391.
16. Serrano-Gomez SJ, Maziveyi M, Alahari SK. Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer*. 2016;15:18.
17. Yang Z, Yang X, Xu S, et al. Reprogramming of stromal fibroblasts by SNAI2 contributes to tumor desmoplasia and ovarian cancer progression. *Mol Cancer*. 2017;16:163.
18. Xiong T, Wang Y, Zhang Y, Yuan J, Zhu C, Jiang W. lncRNA AC005224.4/miR-140-3p/SNAI2 regulating axis facilitates the invasion and metastasis of ovarian cancer through epithelial-mesenchymal transition. *Chin Med J (Engl)*. 2023;136:1098-1110.
19. Qiao X, Lin J, Shen J, et al. FBXO28 suppresses liver cancer invasion and metastasis by promoting PKA-dependent SNAI2 degradation. *Oncogene*. 2023;42:2878-2891.
20. Mazzu YZ, Liao Y, Nandakumar S, et al. Dynamic expression of SNAI2 in prostate cancer predicts tumor progression and drug sensitivity. *Mol Oncol*. 2022;16:2451-2469.
21. Du S, Li H, Sun X, et al. MicroRNA-124 inhibits cell proliferation and migration by regulating SNAI2 in breast cancer. *Oncol Rep*. 2016;36:3259-3266.
22. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17:218.
23. Richards KE, Zeleniak AE, Fishel ML, Wu J, Littlepage LE, Hill R. Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. *Oncogene*. 2017;36:1770-1778.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
25. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients*. 2020;12:2652.
26. Ghandi M, Huang FW, Jané-Valbuena J, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature*. 2019;569:503-508.
27. Ouyang W, Winsnes CF, Hjeltnes M, et al. Analysis of the human protein atlas image classification competition. *Nat Methods*. 2019;16:1254-1261.
28. Yang W, Soares J, Greninger P, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 2013;41:D955-D961.
29. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med*. 2017;377:2500-2501.
30. Riley RS, June CH, Langer R, Mitchell MJ. Delivery technologies for cancer immunotherapy. *Nat Rev Drug Discov*. 2019;18:175-196.
31. Zheng H, Liu H, Li H, Dou W, Wang X. Weighted gene co-expression network analysis identifies a cancer-associated fibroblast signature for predicting prognosis and therapeutic responses in gastric cancer. *Front Mol Biosci*. 2021;8:744677.
32. Wang H, Li N, Liu Q, et al. Antiandrogen treatment induces stromal cell reprogramming to promote castration resistance in prostate cancer. *Cancer Cell*. 2023;41:1345-1362.e1349.
33. Li C, Chen T, Liu J, et al. FGF19-induced inflammatory CAF promoted neutrophil extracellular trap formation in the liver metastasis of colorectal cancer. *Adv Sci (Weinh)*. 2023;10:e2302613.
34. Zheng Y, Wang X, Yang X, Xing N. Single-cell RNA sequencing reveals the cellular and molecular characteristics of high-grade and metastatic bladder cancer. *Cell Oncol (Dordr)*. 2023;46:1415-1427.
35. Yang Q, Gong H, Liu J, Ye M, Zou W, Li H. A 13-gene signature to predict the prognosis and immunotherapy responses of lung squamous cell carcinoma. *Sci Rep*. 2022;12:13646.
36. Yabroff KR, Bradley CJ, Mariotto AB, Brown ML, Feuer EJ. Estimates and projections of value of life lost from cancer deaths in the United States. *J Natl Cancer Inst*. 2008;100:1755-1762.
37. Li Y, Yang Y, Ma Q, et al. HNRNP/K/CLCN3 axis facilitates the progression of LUAD through CAF-tumor interaction. *Int J Biol Sci*. 2022;18:6084-6101.
38. Tang X, Hou Y, Yang G, et al. Stromal miR-200s contribute to breast cancer cell invasion through CAF activation and ECM remodeling. *Cell Death Differ*. 2016;23:132-145.
39. Stevens LE, Cheung WKC, Adua SJ, et al. Extracellular matrix receptor expression in subtypes of lung adenocarcinoma potentiates outgrowth of micrometastases. *Cancer Res*. 2017;77:1905-1917.
40. Li C, Qiu S, Liu X, et al. Extracellular matrix-derived mechanical force governs breast cancer cell stemness and quiescence transition through integrin-DDR signaling. *Signal Transduct Target Ther*. 2023;8:247.
41. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. *Adv Drug Deliv Rev*. 2016;97:4-27.
42. Erdogan B, Webb DJ. Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. *Biochem Soc Trans*. 2017;45:229-236.
43. Nieto MA. The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol*. 2002;3:155-166.
44. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature*. 2005;436:518-524.
45. Mittal MK, Singh K, Misra S, Chaudhuri G. SLUG-induced elevation of D1 cyclin in breast cancer cells through the inhibition of its ubiquitination. *J Biol Chem*. 2011;286:469-479.
46. Uygur B, Abramo K, Leikina E, Vary C, Liaw L, Wu WS. SLUG is a direct transcriptional repressor of PTEN tumor suppressor. *Prostate*. 2015;75:907-916.
47. Zhu GQ, Tang Z, Huang R, et al. CD36(+) cancer-associated fibroblasts provide immunosuppressive microenvironment for hepatocellular carcinoma via secretion of macrophage migration inhibitory factor. *Cell Discov*. 2023;9:25.
48. Van Ginderachter JA. The heat is on: 20-HETE instructs an immunosuppressive phenotype in cancer-associated fibroblasts. *Cancer Res*. 2022;82:3882-3883.