Cancer-associated Fibroblasts Are Associated With Poor Prognosis in Solid Type of Lung Adenocarcinoma: Machine Learning Approach

Kyueng-Whan Min ( kyueng@gmail.com )
Hanyang University Guri Hospital, Hanyang University College of Medicine

Dong-Hoon Kim
Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine

Byoung Kwan Son
Uijeongbu Eulji Medical Center, Eulji University School of Medicine

Mi Jung Kwon
Hallym University Sacred Heart Hospital, Hallym University College of Medicine

Ji-Yong Moon
Hanyang University Guri Hospital, Hanyang University College of Medicine

Research Article

Keywords: Lung, Adenocarcinoma, Histology, Genes, Cancer-associated fibroblasts

Posted Date: May 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-540837/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on August 18th, 2021. See the published version at https://doi.org/10.1038/s41598-021-96344-1.
Abstract

Cancer-associated fibroblasts (CAFs) participate in critical processes in the tumor microenvironment, such as extracellular matrix remodeling, reciprocal signaling interactions with cancer cells and crosstalk with infiltrating inflammatory cells. However, the relationships between CAFs and survival are not well known in lung cancer. The aim of this study was to reveal the correlations of CAFs with survival rates, genetic alterations and immune activities. This study reviewed the histological features of 517 patients with lung adenocarcinoma from The Cancer Genome Atlas (TCGA) database. We performed gene set enrichment analysis (GSEA), network-based analysis and survival analysis based on CAFs in four histological types of lung adenocarcinoma: acinar, papillary, micropapillary and solid. We found four hallmark gene sets, the epithelial-mesenchymal transition, angiogenesis, hypoxia, and inflammatory response gene sets, that were associated with the presence of CAFs. CAFs were associated with tumor proliferation, elevated memory CD4+ T cells and high CD274 (encoding PD-L1) expression. In the pathway analyses, CAFs were related to blood vessel remodeling, matrix organization, negative regulation of apoptosis and transforming growth factor-b signaling. In the survival analysis of each histological type, CAFs were associated with poor prognosis in the solid type. These results may contribute to the development of therapeutic strategies against lung adenocarcinoma cases in which CAFs are present.

Highlights

- CAFs were associated with increased angiogenesis, tumor growth and matrix remodeling.
- CAFs were frequently observed in patients with solid lung adenocarcinoma.
- The presence of CAFs could be a poor prognostic factor in solid adenocarcinoma.

Introduction

Lung cancer is a major cancer and the most common cause of cancer death in the world. According to the National Comprehensive Cancer Network (NCCN) guidelines in oncology, early lung cancer requires surgical procedures, but advanced cancer is treated with systemic treatment. However, about half of the patients recur, usually within the first year after starting treatment.

Genetic mutation of cells can induce cancer development, but disease progression and treatment sensitivity are affected by nonmutant cells within the tumor microenvironment. One type of nonmutant cell within the dense collagenous stroma is fibroblast-like cells, so-called cancer-associated fibroblasts (CAFs). CAFs can drive cancer metastasis through remodeling of the extracellular matrix (ECM) and the production of growth factors and can affect angiogenesis, and these effects influence therapy response. Recently, there has been a growing appreciation of the ability of CAFs to modulate the immune system.

Several studies have reported that survival rates are associated with CAF histological features in different types of malignancy. Previous studies have demonstrated that CAFs and desmoplastic reaction are
predictive of poor prognosis in colorectal cancer\textsuperscript{9}. Another study suggested that adipocyte-derived fibroblasts are correlated with poor survival and desmoplastic reaction in breast cancer\textsuperscript{7}. However, another study reported that histological type, specifically the desmoplastic type, is an independent predictor of favorable prognosis in colorectal cancer\textsuperscript{10}.

Recently, molecular studies have utilized bioinformatic tools to find the mechanisms of CAFs. CAFs are a different cell population in terms of origin and pathobiological roles and are derived mainly from mesenchymal stromal cells that are resident in or recruited by the cancer\textsuperscript{11}. CAFs are located close to tumor cells and stromal components such as lymphocytes, neutrophils, plasma cells, endothelial cells and ECM\textsuperscript{12}. Fibroblasts include CAFs as well as myofibroblastic cells, quiescent fibroblastic cells and pericytic cells. The identification of fibroblasts within the cancer remains challenging due to the lack of specific biomarkers for known and still unclear subtypes\textsuperscript{12}.

In recent years, big data analytics and next-generation sequencing (NGS) have allowed the analyses of genetic biomarkers, the quantification of the several types of tumor-infiltrating lymphoid cells and the molecular pathway network-based integration of multiomics data\textsuperscript{13–15}. Considering the multiple gene-environment relationships of lung cancer, the clinicopathological application of gene expression data is difficult. For these reasons, we believe that analyses based on gene expression data should focus on identifying a simple, robust, and druggable biomarkers based on high-throughput experimental tools and bioinformatics to achieve accessible therapeutic strategies. The Cancer Genome Atlas (TCGA) has a big database, including digital pathologic slides, clinicopathological information, RNA sequencing, mutation, copy number variable and methylation data\textsuperscript{13}. Moreover, the histological features reported in the TCGA database provide data on the presence of CAFs and the tumor microenvironment in lung cancer.

This study aimed to determine whether the presence of CAFs contributes to the clinical outcomes of lung cancer and to evaluate the prognostic value of CAFs\textsuperscript{16}. We further aimed to find the gene sets related to CAFs based on gene set enrichment analysis (GSEA)\textsuperscript{14} and molecular pathway network analyses\textsuperscript{17,18}. The relationships between lymphoid cells and CAFs were analyzed\textsuperscript{15}. Using the Catalog of Somatic Mutations in Cancer (COSMIC) and Genomics of Drug Sensitivity in Cancer (GDSC) databases, we performed drug sensitivity screening in lung cancer cells according to fibroblast activation protein-a (FAP) expression (as a marker associated with CAFs)\textsuperscript{19,20}.

**Materials And Methods**

**Patient selection**

We obtained a total of 1,053 non-small-cell lung carcinoma (NSCLC) cases comprising 566 lung adenocarcinomas and 487 squamous cell carcinomas with known mRNA expression and mutation data from the TCGA database\textsuperscript{13}. The analysis was performed on 517 cases containing both virtual histopathological slides and clinical data (from a total of 566 lung adenocarcinoma samples).
Cancer-associated fibroblasts

In this review, cells with both immature fibroblastic proliferation (proportion: > 10%) at the tumor invasive front and high FAP gene expression were defined as CAFs (Fig. 1A). To determine the optimal cutoff value for FAP expression, we generated receiver operating characteristic (ROC) curves comparing sensitivity versus 1–specificity. The cutoff value calculated by the ROC curve was used to evaluate the relationship between survival and FAP expression. On the basis of the ROC curve, FAP expression was classified as low (mRNA level ≤ 562.9965) or high (mRNA level > 562.9965). Of the 517 cases, CAFs were present in 101 cases (19.5%).

Gene set enrichment analysis and pathway-based network analysis based on TCGA data

To detect significant gene sets, GSEA (version 4.1.0) was performed with 31,117 gene sets in the Molecular Signatures Database (MSigDB version 7.2) from the Broad Institute at MIT. Specific gene sets (50 hallmark gene sets) were tested to determine which were associated with CAFs. For this analysis, 1,000 permutations were utilized to calculate the p values, and the permutation type was set to phenotype. Significant gene sets were those with the following characteristics: false discovery rate < 0.001; family wise-error rate ≤ 0.001; and p < 0.001.

We analyzed tumor-infiltrating lymphocytes using deep learning-based lymphocyte classification with convolutional neural networks in whole-slide image analysis and identified immune subtypes using CIBERSORT and Kallisto software and algorithms.

Pathway-based network analysis was based on the identified genes linked to CAFs using Cytoscape (version 3.8.1) network visualization software. To visualize the biological relevance of the histological subtypes and their relevant elements on the basis of GSEA, we performed functional enrichment analyses using ClueGO, an application within Cytoscape software.

Machine Learning algorithm for validation

We integrated CAFs with clinical risk factors (T stage, N stage, age, sex, smoking history) to composite prognostic models for survival prediction by applying machine learning (ML) algorithms in 517 cases (randomization: train set, 70%; validation set, 30%). A learning algorithm was independently applied to select and combine multiple covariates from gradient boosting machines (GBM) based on multivariate Gaussian models. In this step, “forward” search method, which initiates on a prototype set and selects a feature if and only if the addition of the feature could increase the performance of the prognostic model, is adopted to select optimal features sequentially. Hyperparameters of the ML algorithms, such as learning rate in GBM were optimized for each combination of selected covariates and learning algorithm by grid search cross-validation through a predefined range. We searched across 81 models with varying learning rates and tree depth. The final optimal models were trained based on the selected covariates and the optimized hyperparameters. To explore the performances of the GBM method, the receiver operator characteristic (ROC) curve was used.
Data extraction from the GDSC and COSMIC databases

Drug screening was performed using datasets from the GDSC and COSMIC databases, which are large-scale cancer cell line and drug response databases containing data from 987 cancer cell lines and 365 compounds, respectively. The response of 8 lung cancer cell lines to 316 drugs was measured and reported as the natural log half-maximal inhibitory concentration (LN IC50) value. A drug was defined as a sensitive drug when the LN IC50 value was lower in lung cell lines with high FAP expression than in those with low FAP expression.

Statistical analysis

Student’s t-test was used to evaluate the differences or relationships among continuous parameters. Disease-free survival (DFS) and disease-specific survival (DSS) were compared using the log rank test. Multivariate analysis was performed to identify independent prognostic markers for DFS and DSS using a Cox multistep regression model. All data were analyzed using R packages. A two-tailed p value < 0.05 was considered statistically significant.

Results

Gene set enrichment analysis of cancer-associated fibroblasts in lung adenocarcinoma

FAP was highly expressed in tumors compared with normal tissues (p < 0.001) (Fig. 1B). We performed GSEA to identify various gene sets associated with CAFs. In the analyses of hallmark gene sets, we found four gene sets (the epithelial-mesenchymal transition, angiogenesis, hypoxia and inflammatory response gene sets) that were associated with lung adenocarcinoma (Fig. 2A).

On the basis of the GSEA results, we analyzed the association between CAFs and each gene set-related marker. Vimentin, a biomarker related to epithelial-mesenchymal transition, was highly expressed in the presence of CAFs (p < 0.001). Vascular endothelial growth factor-A (VEGF-A), as a marker related to angiogenesis, was increased in the presence of CAFs (p = 0.022). Hypoxia-inducible factor 1 subunit alpha (HIF1a), which is linked to hypoxia, was elevated in the presence of CAFs (p < 0.001). The lymphocyte infiltration signature score, which is associated with the inflammatory response, showed a tendency to increase in the presence of CAFs, but it was not statistically significant (p = 0.209) (Fig. 2B).

Immune profiles, proliferation and biomarkers of cancer-associated fibroblasts

In the analyses of CAFs, we referred to the immune cell profiles, tumor cell proliferation parameters and biomarkers used in a study by Thorsson et al. and in silico cytometry.24

In comparing the immune cell fractions between samples with and without CAFs, memory B cells were decreased in samples with CAFs (p < 0.001), while activated memory CD4+ T cells were increased in samples with CAFs (p = 0.002). CD274 (encoding PD-L1, programmed death-ligand 1) expression was more elevated in samples with CAFs than in those without CAFs (p < 0.001). CD8+ T cells showed a
tendency to be decreased in samples with CAFs, but this trend was not statistically significant ($p = 0.602$) (Fig. 3A).

The presence of CAFs was associated with higher proliferation and matrix metallopeptidase-11 (MMP-11), interleukin-6 (IL-6) and transforming growth factor-β (TGF-β) levels than the absence of CAFs ($p < 0.001$, $< 0.001$, $= 0.015$ and $< 0.001$, respectively) (Fig. 3B).

**Pathway-based network analysis of cancer-associated fibroblast-related genes and gene sets**

We performed pathway-based network analysis using the genes and gene sets associated with CAFs. The CAFs were linked to 10 functionally enriched Gene Ontology (GO) terms and pathways: 1) blood vessel remodeling; 2) lung fibrosis; 3) regulation of tissue remodeling; 4) extracellular matrix organization; 5) cell-matrix adhesion; 6) fibrillar collagen trimer; 7) negative regulation of extrinsic apoptosis signaling pathway; 8) collagen fibril organization; 9) morphogenesis of an epithelial sheet; and 10) TGF-β receptor signaling (Fig. 4).

**Survival analyses of cancer-associated fibroblasts and machine learning**

In the TCGA data, the distributions of the five predominant histological types were as follows: 7 lepidic type (1.4%), 115 acinar type (22.2%), 107 papillary type (20.7%), 65 micropapillary type (12.6%) and 223 solid type (43.1%). There was an absence of CAFs in the lepidic cases; thus, they were not included in the survival analyses.

The presence of CAFs was associated with unfavorable DFS and DSS in the acinar type ($p = 0.039$ and 0.067, respectively), but the relationship between CAFs and DSS was not statistically significant. The presence of CAFs was significantly related to shorter DFS and DSS than the absence of CAFs in the papillary type ($p < 0.001$ and 0.019, respectively). The presence of CAFs correlated with worse DFS and DSS in the micropapillary type ($p = 0.47$ and 0.069, respectively), but the correlations were not statistically significant. In the solid type, the presence of CAFs was significantly associated with shorter DFS and DSS than the absence of CAFs ($p = 0.007$ and 0.002, respectively) (Fig. 5A and B). After adjustment for confounders including T stage, N stage, age, sex and smoking history, the presence of CAFs was associated with worse DFS and DSS than the absence of CAFs ($p < 0.001$ and = 0.008, respectively). There was a relationship between shorter DSS and the presence of CAFs in only solid type samples ($p = 0.003$) (Table 1).

We compared the performance of the two GBM models in predicting survival rates (Model 1; T stage, N stage, age, sex, smoking history versus Model 2; CAFs, T stage, N stage, age, sex, smoking history). ROC curves were performed (area under the curve: Model 1, 0.642; Model 2, 0.677) (Fig. 5C). We found that the GBM algorithm performed the best while the addition of CAFs to the prediction model improved the prognostic performance. With cross-validation estimates, 15 decision trees were utilized sequentially while the maximum depth of each decision tree was optimized at 1, corresponding to one-way interactions, and the learning rate was optimized at 0.018.
Drug screening of lung cancer cell lines

Mitomycin-C, a drug targeting DNA replication, and IOX2, which inhibits PHD2 activation, effectively suppressed the growth of lung cancer cells with high FAP expression. Dabrafenib, an anti-v-raf murine sarcoma viral oncogene homolog B (BRAF)-targeting drug, showed greater inhibition of the proliferation of cells with high FAP expression than mitomycin-C and IOX2 [dabrafenib: $r = -0.848$, $p = 0.008$ (Pearson's correlation) and 0.006 (Student's t-test); mitomycin-C: $r = -0.782$, $p = 0.038$ (Pearson's correlation) and 0.026 (Student's t-test); IOX2: $r = -0.727$, $p = 0.041$ (Pearson's correlation) and 0.041 (Student's t-test)] (Fig. 5D).

Discussion

This study showed survival differences between patients with and without CAFs and analyzed genetic/molecular alterations in patients with lung adenocarcinoma. In previous studies, genetic/molecular signatures related to CAFs have been shown to correlate with prognosis in colorectal, ovarian and breast cancer\textsuperscript{27–29}. Our results revealed that the presence of CAFs was associated with a shorter survival rate than the absence of CAFs in lung adenocarcinoma, especially the solid type. In this study, the machine learning model analysis which includes CAFs increased the accuracy of predicting the survival rate. A study by Marcela et al. reported that CAFs were related to increased survival in patients with diffuse large B cell lymphoma\textsuperscript{30}. Moreover, another study of colorectal carcinoma demonstrated that the presence of desmoplasia and CAFs was associated with better survival than the absence of desmoplasia\textsuperscript{10}. Thus, there is controversy regarding the association between CAFs and the survival of patients with cancer. It is thought that CAFs in the tumor microenvironment are phenotypically heterogeneous and may exhibit both a protumorigenic and antitumorigenic phenotypes\textsuperscript{31}. We analyzed hallmark gene sets related to CAFs in lung adenocarcinoma. A total of four gene sets associated with the presence of CAFs were identified: the epithelial-mesenchymal transition (EMT), angiogenesis, hypoxia, and inflammatory response gene sets. Subsequently, we determined the correlations of representative biomarkers associated with these gene sets with to the presence or absence of CAFs. First, vimentin is an EMT biomarker and is involved in cell migration, motility and adhesion and associated with metastasis\textsuperscript{32}. Second, VEGF-A is an angiogenesis biomarker and induces high microvascular density and permeability and promotes tumor expansion\textsuperscript{33}. Third, HIF1a is a hypoxia biomarker and is associated with the upregulation of glycolytic genes related to oxygen deprivation with increased cancer metabolism\textsuperscript{34}. Fourth, the lymphocyte infiltration signature score, which is an inflammatory response marker, is related to prognosis and host-tumor immune interactions in different types of malignancies\textsuperscript{35}. Some representative markers, such as vimentin, VEGF-A and HIF1a, were elevated in the presence of CAFs compared to the absence of CAFs. There was no significant difference in the lymphocyte infiltration signature score between samples with and without CAFs. These results suggest that the presence or absence of CAFs has minimal effect on the host-tumor immune response in lung adenocarcinoma.
In addition, we analyzed the changes in immune cell subtypes according to the presence or absence of CAFs by considering the characteristics of different immune cells. Memory B cells were decreased in the presence of CAFs, but activated memory CD4+ T cells were increased in the presence of CAFs. CD274 was elevated in the presence of CAFs. There was no significant difference in CD8+ T cells between samples with or without CAFs. A study by Costa et al. demonstrated that FAP-high fibroblasts, such as CAFs, are correlated with T_{reg} cell-mediated immunosuppression and poor outcome in breast cancer. Our results showed that, in lung adenocarcinoma, CAFs have little effect on the immunomodulation associated with CD8+ T cells in the tumor microenvironment.

CAFs can induce increased levels of growth factors, matrix remodeling and increased levels of numerous cytokines related to immunomodulation. In our results, the proliferation index was increased in the presence of CAFs. TGF-b and IL-6 are related to tumor growth and/or immunosuppression and were increased in the presence of CAFs. A representative marker of ECM and cancer invasion, MMP-11, was elevated in the presence of CAFs. The pathway-based network analysis showed biological functions related to CAFs, such as blood vessel remodeling, extracellular matrix organization, negative regulation of the extrinsic apoptotic signaling pathway and TGF-b receptor signaling.

We investigated 316 anticancer drugs to which 8 breast cancer cell lines with high FAP expression were responsive by using pharmacogenomic screens from the GDSC database. We identified three anticancer drugs that effectively reduced the growth of lung cancer cells with high FAP expression. Among them, dabrafenib, which showed a high negative correlation between high FAP expression and a promising LN IC50 value, was identified as the most sensitive anti-CAF drug for the lung cancer cell lines assessed. The combination of dabrafenib, as a BRAF kinase inhibitor, and trametinib, an inhibitor of mitogen-activated extracellular signal-regulated kinase (MEK), is currently being investigated in advanced malignant tumors. A previous study demonstrated that trametinib produced a potential tumor response against protumorigenic CAFs in neuroblastoma. In the GDSC analysis, trametinib showed a tendency to inhibit lung cancer cells with high FAP expression, but the trend was not statistically significant [(r = -0.666, p = 0.103 (Pearson's correlation) and 0.423 (Student's t-test)]. Given the lack of vivo studies, dabrafenib-based clinical trials in lung cancer patients with high FAP expression are needed in the future.

This study has several limitations that should be acknowledged. First, because this is a cross-sectional study and the in silico analyses with TCGA did not show sustained relationships over time, it is difficult to reach a definitive conclusion. Second, experimental data allowing for novel biological insights into CAFs were not obtained in our study. Further in vitro and/or in vivo studies may be necessary to clarify the molecular mechanisms of CAFs in solid lung adenocarcinoma. Third, CAF function may be highly heterogeneous in solid lung adenocarcinoma patients, as many components of signaling pathways are affected by disease status, microenvironment, and immunity. Fourth, the difficulty in identifying CAFs results largely from the lack of unique markers. In our study, CAFs were defined by a combination of the histological features of fibroblasts and high FAP gene expression.
This study demonstrated that CAFs are associated with increased tumor cell growth, angiogenesis, and ECM remodeling, effects that produce an unfavorable prognosis in patients with solid lung adenocarcinoma. CAFs were found to be associated with enhanced recruitment of activated memory CD4+ T cells with high CD274 expression. The presence of CAFs was related to decreased numbers of CD8+ T cells, but the relationship was not statistically significant. Patients with CAFs with high CD274 expression without elevated CD8+ T cells might develop resistance to anti-PD-L1 therapies. Our workflow results regarding CAFs will contribute to designing future clinical and experimental studies for patients with solid lung adenocarcinoma.

**Abbreviations**

CAFs, cancer-associated fibroblasts; ECM, extracellular matrix; TCGA, the Cancer Genome Atlas; GSEA, gene set enrichment analysis; GDSC, genomics of drug sensitivity in cancer; COSMIC, catalog of somatic mutations in cancer; FAP, fibroblast activation protein-a; GBM, Gradient Boosting Machine; NSCLC, non-small-cell lung carcinoma; ROC, receiver operating characteristic; ML, Machine Learning; MSigDB, molecular signatures database; LN IC50, natural log half-maximal inhibitory concentration; DFS, disease-free survival; DSS, disease-specific survival; VEGF-A, vascular endothelial growth factor-A; HIF1a, hypoxia-inducible factor 1 subunit alpha; MMP-11, matrix metallopeptidase-11; IL-6, interleukin-6; TGF-β, transforming growth factor-β; GO, gene ontology; BRAF, v-raf murine sarcoma viral oncogene homolog B

**Declarations**

**Data availability**

The authors declare that all data supporting the findings of this study are available within the article.

**Acknowledgements**

The results <published or shown> here are in whole or part based upon data generated by the TCGA Research Network: [https://www.cancer.gov/tcga](https://www.cancer.gov/tcga).

**Author contributions**


**Competing interests**

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to K.-W.M.
References


Tables

Table 1. Univariate and multivariate analyses of disease-free survival and disease-specific survival based on cancer-associated fibroblasts

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Disease-free survival</th>
<th></th>
<th></th>
<th>Disease-specific survival</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>P value</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Acinar type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>2.125</td>
<td>1.021</td>
<td>0.039</td>
<td>2.202</td>
<td>0.926</td>
</tr>
<tr>
<td>Multivariate*</td>
<td>1.712</td>
<td>0.775</td>
<td>0.184</td>
<td>1.842</td>
<td>0.698</td>
</tr>
<tr>
<td>Papillary type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>5.460</td>
<td>2.031</td>
<td>&lt;0.001</td>
<td>4.061</td>
<td>1.142</td>
</tr>
<tr>
<td>Multivariate*</td>
<td>13.630</td>
<td>4.112</td>
<td>&lt;0.001</td>
<td>3.585</td>
<td>0.706</td>
</tr>
<tr>
<td>Micropapillary type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>1.724</td>
<td>0.387</td>
<td>0.47</td>
<td>3.975</td>
<td>0.796</td>
</tr>
<tr>
<td>Multivariate*</td>
<td>0.535</td>
<td>0.051</td>
<td>0.603</td>
<td>3.844</td>
<td>0.367</td>
</tr>
<tr>
<td>Solid type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>1.852</td>
<td>1.173</td>
<td>0.007</td>
<td>2.161</td>
<td>1.296</td>
</tr>
<tr>
<td>Multivariate*</td>
<td>1.980</td>
<td>1.199</td>
<td>0.008</td>
<td>2.339</td>
<td>1.328</td>
</tr>
</tbody>
</table>

*Adjusted for T stage, N stage, age, sex and smoking history
Figure 1

please see the manuscript file for the full caption

Figure 2

please see the manuscript file for the full caption
Figure 3

please see the manuscript file for the full caption
Figure 4

Grouping of networks based on functionally enriched Gene Ontology (GO) terms and pathways associated with the presence of cancer-associated fibroblasts.
Survival analyses of the four types based on the absence or presence of cancer-associated fibroblasts. 
(A) Disease-free survival: acinar, papillary, micropapillary and solid types ($p = 0.039$, $< 0.001$, $= 0.47$ and $0.007$, respectively). 
(B) Disease-specific survival: acinar, papillary, micropapillary and solid types ($p = 0.067$, $0.019$, $0.069$ and $0.002$, respectively). 
(C) We supervised machine-learning models for prognosis prediction using gradient boosting machine (GBM). Covariates were included in the confounding factors.
[Model 1 (left); T stage, N stage, age, sex, smoking (pack years) versus Model 2 (right); Cancer associated fibroblasts (CAFs), T stage, N stage, age, sex, smoking (pack years)] and their relative importance using overall survival. Receiver operator characteristic curve for GBM was used based on a multivariate Gaussian model. (D) Pearson's correlations (top) and boxplots (bottom) showing the natural log half-maximal inhibitory concentration (LN IC50) values of dabrafenib, mitomycin-C and IOX2, the most sensitive anticancer drugs against lung cancer cell lines with high FAP (fibroblast activation protein-α) expression.