Int J Clin Exp Med 2021;14(9):2325-2332
www.ijcem.com /ISSN:1940-5901/IJCEM0132982

Original Article

Comprehensive analysis of cancer-associated fibroblasts based on single-cell RNA sequencing in lung adenocarcinoma patients

Han He[1], Jun Jiang[1], Yanfei Min[1], Baoguo Xie[2], Rong Huang[1]

[1]*Department of Radiotherapy for Thoracic and Abdominal Cancer, The First People's Hospital of Foshan, Foshan 528000, China;* [2]*Reproductive Medicine Center, The First Affiliated Hospital of Hainan Medical University, Haikou 570102, China*

Received March 10, 2021; Accepted June 15, 2021; Epub September 15, 2021; Published September 30, 2021

Abstract: Background: Cancer-associated fibroblasts (CAFs) in lung adenocarcinoma (LUAD) are important parts of the tumor microenvironment and are related to the prognosis. This study explores the biological characteristics of CAFs in LUAD tissue, and prognosis-related genes were screened. Methods: Using single-cell RNA sequencing data, we identified CAFs in LUAD tissue and analyzed them according to their RNA information. We evaluated their biological function, screened CAF-specific high-expression genes, and identified survival-related genes by combining the results with data from LUAD patients in The Cancer Genome Atlas (TCGA) database. All bioinformatics analyses were based on R-language software packages. Results: In total, 88,144 cells were collected from 22 tumor tissues and adjacent normal tissues of LUAD patients, and 68,881 cells were analyzed after quality control. Unsupervised clustering analysis was performed by graph clustering in Seurat, and 22 clusters were identified. Cluster 13 had 2140 cells that were identified as fibroblasts with genes *COL1A1*, *COL3A1*, *DCN,* and *LUM*. Of these, 840 cells were from tumor tissue, and 1300 cells were from adjacent normal tissue. *RPS19*, *PMEPA1,* and *CTHRC1* were highly expressed in CAFs, and high expression of these genes indicated poor 10-year survival rate. *SPARCL1, BTG2, ABCA8, CHRDL1, ADH18, GPX3,* and *SFTPC* had low expression in CAFs of tumor tissue, anfigd low expression was also associated with poor 10-year survival rate. Conclusion: Single-cell analysis revealed the heterogeneity and complexity of CAF functions, and a group of abnormally expressed genes may be related to LUAD patients' survival.

Keywords: Cancer-associated fibroblasts, lung adenocarcinoma, single-cell RNA sequencing, the cancer genome atlas

Introduction

Lung adenocarcinoma (LUAD) is the most common type of lung cancer, accounting for about 40% of all cases [1]. In recent years, many antitumor drugs have been developed, but due to tumor heterogeneity and complex tumor microenvironment (TME), most patients with lung cancer develop drug resistance, recurrence, and metastasis after treatment, which lead to treatment failure and a high mortality rate [2, 3]. As a major component of the TME, cancer-associated fibroblasts (CAFs) are generally considered as fibroblasts within tumor tissues that are activated by normal resident tissue fibroblasts or transdifferentiated from non-fibroblastic lineages, such as epithelial cells and adipocytes [4]. There is evidence that CAFs can promote tumor development through the establishment and remodeling of extracellular matrix, resulting in drug resistance and shortening survival time [5, 6].

Single-cell RNA sequence (scRNA-seq) is an optimized next-generation sequencing (NGS) technology that is used to examine the sequence information of individual cells, provide higher-resolution cell differences, and give better understanding of the function of individual cells in tissue [7]. ScRNA-seq has become an important tool in tumor research. Some studies based on scRNA-seq investigated the involvement of CAFs in the microenvironment of lung cancer. The results confirm that the microenvi-

ronment caused by adaptive changes in CAFs can lead to immune escape and cancer progression. The mechanism of tumor drug resistance has also been studied at the single-cell level [4, 8].

However, because of the cost, it is still difficult to use scRNA-seq to study clinical prognoses in a large population. Therefore, a combination of scRNA-seq and analysis of data from The Cancer Genome Atlas (TCGA) is also an reasonable research strategy [9]. In this study, we used LUAD transcriptome and clinical data from the TCGA project and combined it with scRNA-seq data from LUAD tissue. The goal was to explore the cellular biological characteristics of CAFs in LUAD and to identify genes related to the survival of patients with LUAD.

Methods

Preprocessing for transcriptome and clinical data

A total of 594 transcriptional group samples of LUAD tumors and adjacent normal tissue (TCGA-LUAD) were downloaded from TCGA data (https://portal.gdc.cancer.gov/), including 513 primary tumor samples. The Kaplan-Meier survival of 513 LUAD samples was analyzed by R survival software. Patients were divided into a high expression group and a low expression group according to the median of gene expression in the sample. The overall survival was compared between them. The cohort of TCGA samples was used to perform a survival analysis using the survival package in R. The survminer package in R was then used to construct a Kaplan-Meier curve.

ScRNA-seq dataset

The scRNA-seq dataset was downloaded from the NCBI Gene Expression Omnibus data (accession number: GSE131907). There were 22 samples, including 11 samples of lung cancer and 11 samples of adjacent normal tissues. All patients were newly diagnosed LUAD patients without drug or radiotherapy treatment, and the adjacent lung tissue was separated from the cancerous area by at least 5 cm. All data in this study are based on previously published studies, so ethical approval or patient consent was not required.

ScRNA-seq data processing and statistical analysis

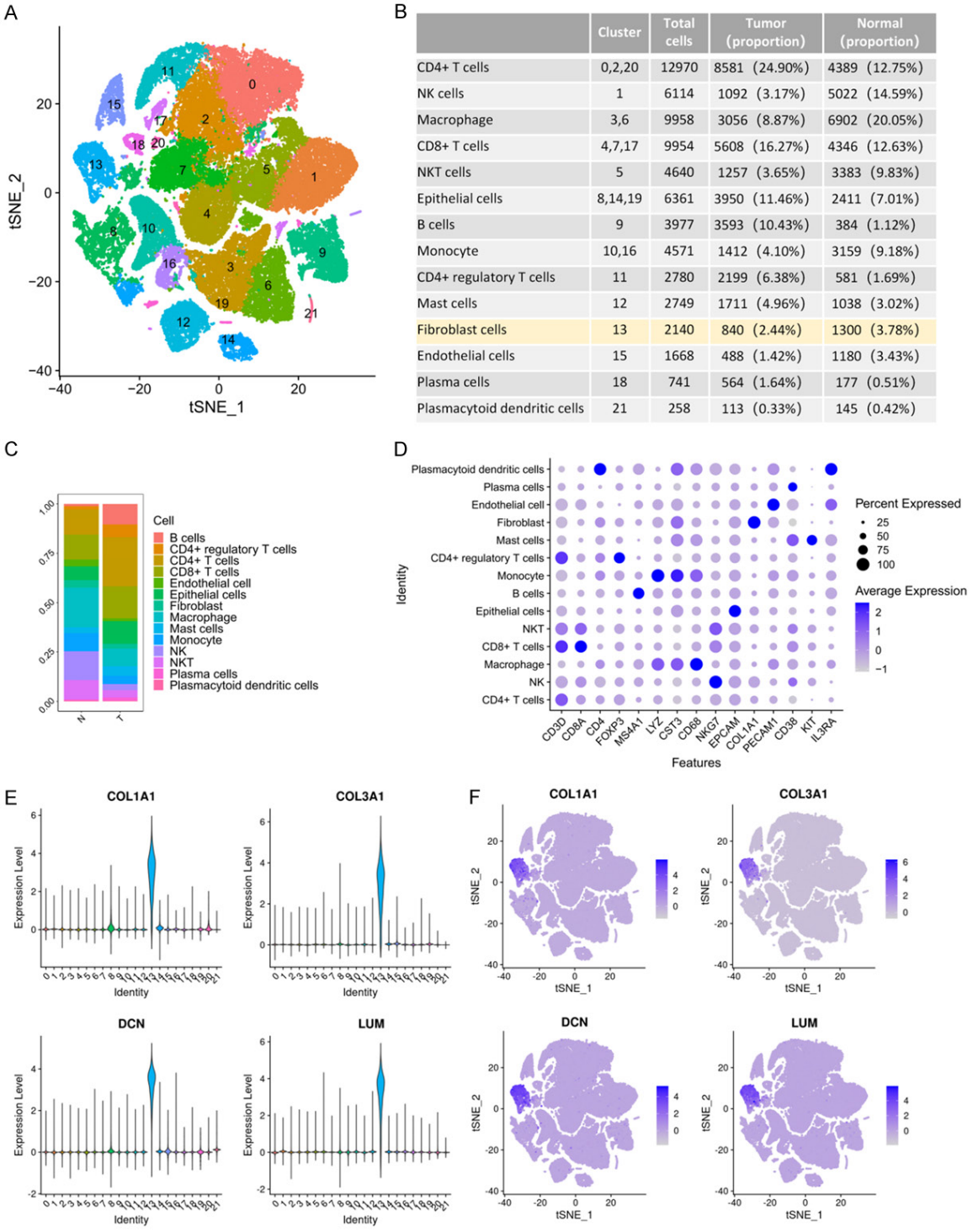
The FASTQ files were transformed into Cell Ranger-specific FASTQ files and processed individually using a modified version of Cell Ranger count pipeline with the GRCh38.p5 human genome as the reference genome. Aligned reads were then filtered for valid cell barcodes and UMIs to generate gene-cell matrices for downstream analysis. Genes were filtered out if they were expressed in fewer than five cells. Cells with > 200 genes and < 10% mitochondrial genes were further processed.

All de-duplicated UMIs were processed for quality control, preprocessing, dimensionality reduction, and differential expression analysis. The results of unsupervised clustering were visualized by using the tSNE project with R (Seurat package version 2.2). Genes were analyzed for KEGG and GO pathway enrichment and differential expression of hallmark gene sets by gene set enrichment analysis (GSEA). The normalized enrichment score (NES) was used to account for differences in the gene set sizes. The thresholds for significantly enriched pathways were logFDR < 0.5 and |NES| > 1.0. Significant differences were determined by Kaplan-Meier analysis and a t-test. $P < 0.05$ was considered statistically significant.

Results

Characteristics of cancer-associated fibroblasts in patients with LUAD at the single-cell level

A total of 88,144 cells were collected from 22 tumor tissues and adjacent normal tissues from *LUAD* patients, and 68,881 cells were analyzed after qualified quality control. Unsupervised clustering analysis was performed by a graph clustering method in Seurat, and 22 clusters (clusters 0-21) were identified, including T cells, NK cells, monocytes/macrophages, B cells, plasma cells, and dendritic cells. Of all the cell types, immune cells accounted for the largest proportion, especially CD4+ cells, macrophages, and NK cells, which were more abundant, while plasma cells and dendritic cells were less abundant. Cluster 13 with 2140 cells in total was identified as fibroblasts with the



	Cluster	Total cells	Tumor (proportion)	Normal (proportion)
CD4+ T cells	0,2,20	12970	8581 (24.90%)	4389 (12.75%)
NK cells	1	6114	1092 (3.17%)	5022 (14.59%)
Macrophage	3,6	9958	3056 (8.87%)	6902 (20.05%)
CD8+ T cells	4,7,17	9954	5608 (16.27%)	4346 (12.63%)
NKT cells	5	4640	1257 (3.65%)	3383 (9.83%)
Epithelial cells	8,14,19	6361	3950 (11.46%)	2411 (7.01%)
B cells	9	3977	3593 (10.43%)	384 (1.12%)
Monocyte	10,16	4571	1412 (4.10%)	3159 (9.18%)
CD4+ regulatory T cells	11	2780	2199 (6.38%)	581 (1.69%)
Mast cells	12	2749	1711 (4.96%)	1038 (3.02%)
Fibroblast cells	13	2140	840 (2.44%)	1300 (3.78%)
Endothelial cells	15	1668	488 (1.42%)	1180 (3.43%)
Plasma cells	18	741	564 (1.64%)	177 (0.51%)
Plasmacytoid dendritic cells	21	258	113 (0.33%)	145 (0.42%)

Figure 1. CAFs in LUAD patients at the single cell level. (A) t-SNE plot of all classified cells. (B) Cell types and their proportion in LUAD tumor tissues and adjacent normal tissues. (C) The proportion of various types of cells in tumor tissues (T) and adjacent normal tissues (C, D) Specific highly expressed genes in various cells. (E, F) The characteristic gene map of CAFs.

genes *COL1A1, COL3A1, DCN,* and *LUM* (**Figure 1**). Among them, there were 840 cells from cancer tissue and 1300 cells from adjacent normal tissue.

CAF gene enrichment analysis at single-cell level

We found that there was a great difference between the expression of genes in tumor-derived CAFs and that in adjacent normal tissues, including 32 genes with high expression (top 10: *BGN, PTMA, RPS19, IGLC2, CTHRC1, NAP1L1, IGHA1, COL8A1, ACTN1, APOE*) and 112 genes with low expression (top 10: *DCN, CFD, IGFBP6, GSN, PLAC9, FBLN1, OGN, CST3, WISP2,* and *ADH1B*). In this study, hallmark collection for GSEA was used to assess the distributions of gene sets in CAFs at single-cell levels, and the analysis showed that the genes with positive NES (NES > 1.0, P < 0.05) were enriched in MYC targets v1, oxidative phosphorylation, epithelial mesenchymal transition, DNA repair, and allograft rejection (**Figure 2B**). Genes with negative NES (NES < -1.0, P < 0.05, top 10) were enriched in TNF-α signaling via NF-κB, apoptosis, UV response up, xenobiotic metabolism, UV response dn, hypoxia, peroxisome, coagulation, reactive oxygen species pathway, and fatty acid metabolism (**Figure 2C**).

The results of the Gene Ontology (GO) annotation and enriched bubble map analysis of the differential genes indicated that the differential genes mainly involved biological processes (BP), molecular function (MF), and cellular components (CC). Only a part of the analysis results are presented, including SRP-dependent cotranslational protein targeting to membrane, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, embryo implantation, cytosolic ribosome, collagen-containing extracellular matrix, cytosolic part, and focal adhesion (**Figure 2D**).

Survival-related genes in fibroblasts in LUAD patients

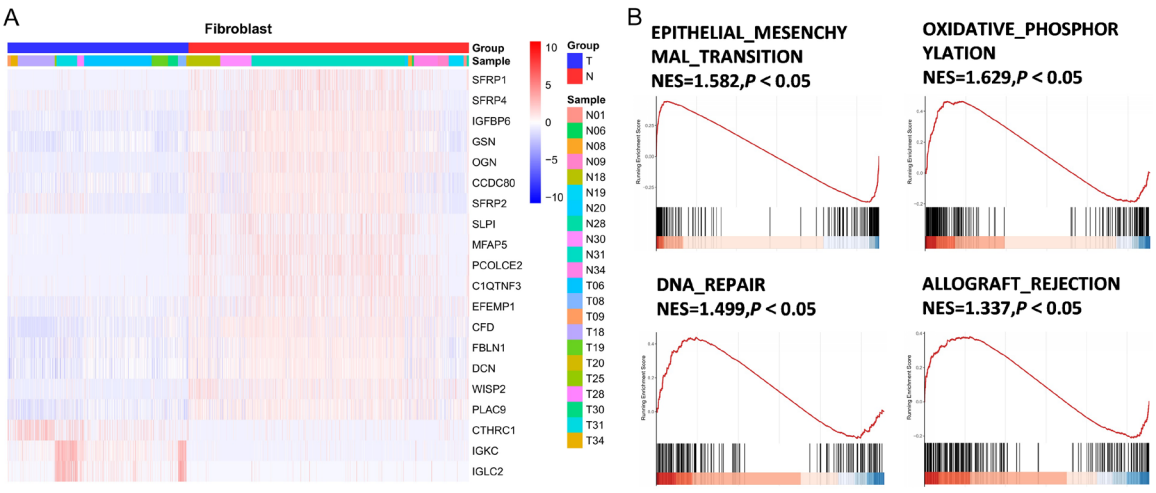
Our aim was to screen genes associated with LUAD survival. Cancer cells and the tumor microenvironment form a network, and CAFs are closely related to the tumor microenvironment. CAF activation is also associated with tumor prognosis in LUAD patients. In this study, we compared para-carcinoma tissues and screened the differentially expressed genes in CAFs. Combining the results with the TCGA database, we screened for genes associated with survival rate, which identified 10 highly expressed genes associated with survival. Among the differentially expressed genes, *RPS19, PMEPA1,* and *CTHRC1* were highly expressed in CAFs of tumor tissue, and patients with high expression had a poor 10-year survival rate. *SPARCL1, BTG2, ABCA8, CHRDL1, ADH18, GPX3,* and *SFTPC* havd low expression in CAFs of tumor tissue, and low expression was also associated with poor 10-year survival rate (**Figure 3D**).

Discussion

The tumor microenvironment is composed of CAFs, tumor-associated macrophages, immune cells, vascular endothelial cells, and various interstitial components [10]. Previous studies have shown that CAFs play an important role in cancer progression, and their presence is related to poor prognosis of patients with LUAD [11]. CAFs can easily cause tumor cells to invade blood vessels and distant metastasis by remodeling the structure of extracellular matrix [12]. Importantly, CAFs can also mediate the interaction between tumor cells and stromal cells by secreting a variety of growth factors, cytokines, chemokines, and extracellular vesicles, thus promoting the progression of tumor malignancy [13, 14].

ScRNA-seq can provide more accurate information on molecular mechanisms, overcome the shortcomings of bulk RNA sequencing, and play a great role in the study of the tumor microenvironment [15, 16]. The present study was based on scRNA-seq, comparison with adjacent normal tissues, and GSEA analysis to identify CAFs enriched in TNF-α signaling via NF-κB, apoptosis, xenobiotic metabolism, etc. These signaling pathways are mostly classical tumor-related biological pathways. We also performed GO enrichment analysis for the differential genes, and the results were similar to those reported in the literature.

In this study, CAFs were identified from tissue with marker genes *COL1A1, COL3A1, DCN*, and *LUM* [17, 18]. Studies have confirmed that the high expression of *COL1A1* and COL3A1 is related to the role of fibroblasts in the transformed tumor microenvironment. This includes the resulting abnormal extracellular matrix or fibrous connective tissue, which is one of the mechanisms of tumor drug resistance or metabolic remodeling [19, 20]. DCN is a small proteoglycan that is involved in the regulation of



C

Down HALLMARK Description	Enrichment Score	NES	P Value	P Adjust
TNFA_SIGNALING_VIA_NFKB	-0.637	-2.204	0.002	0.008
APOPTOSIS	-0.557	-1.885	0.002	0.008
UV_RESPONSE_UP	-0.541	-1.833	0.002	0.008
XENOBIOTIC_METABOLISM	-0.505	-1.743	0.002	0.008
UV_RESPONSE_DN	-0.526	-1.743	0.002	0.008
HYPOXIA	-0.496	-1.715	0.002	0.008
PEROXISOME	-0.529	-1.684	0.002	0.008
COAGULATION	-0.509	-1.680	0.002	0.008
REACTIVE_OXYGEN_SPECIES_PATHWAY	-0.593	-1.668	0.006	0.013
FATTY_ACID_METABOLISM	-0.482	-1.630	0.002	0.008

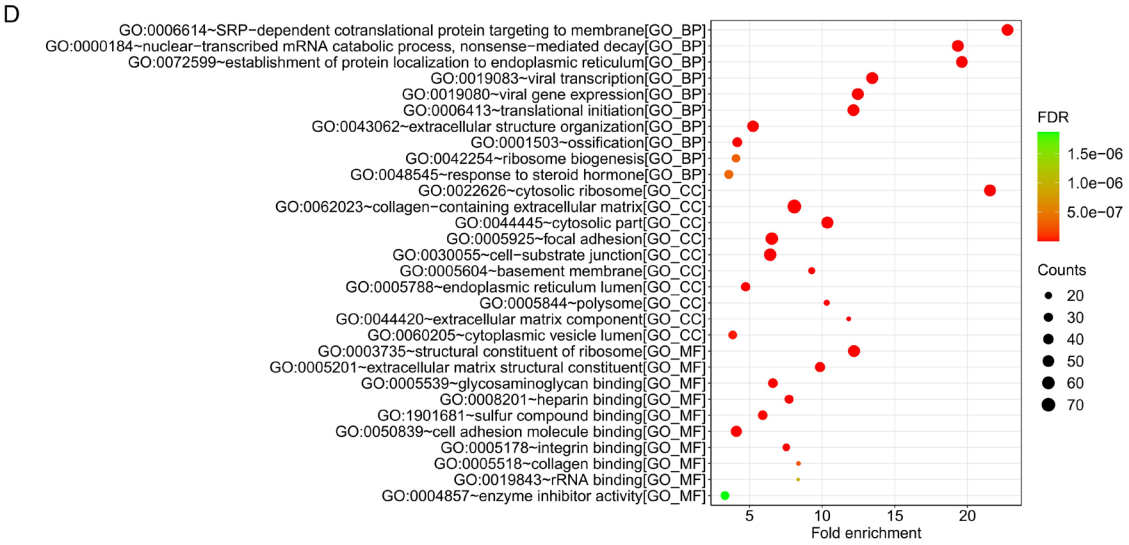


Figure 2. GSEA analysis of CAFs and fibroblasts from adjacent normal tissues. A. Heat map of differential genes between CAFs and fibroblasts from adjacent normal tissues. B. The enriched hallmark gene sets with normalized-enrichment scores (NES) > 1.0 and P < 0.05. C. The reduced hallmark gene sets with NES < -1.5 and P < 0.05. D. GO enrichment analysis for the differential genes.

organ wound healing [21]. It has been found that DCN, as the biological ligand of EGFR, regulates the migration of fibroblasts by down-regulating EGFR, which plays an important role in

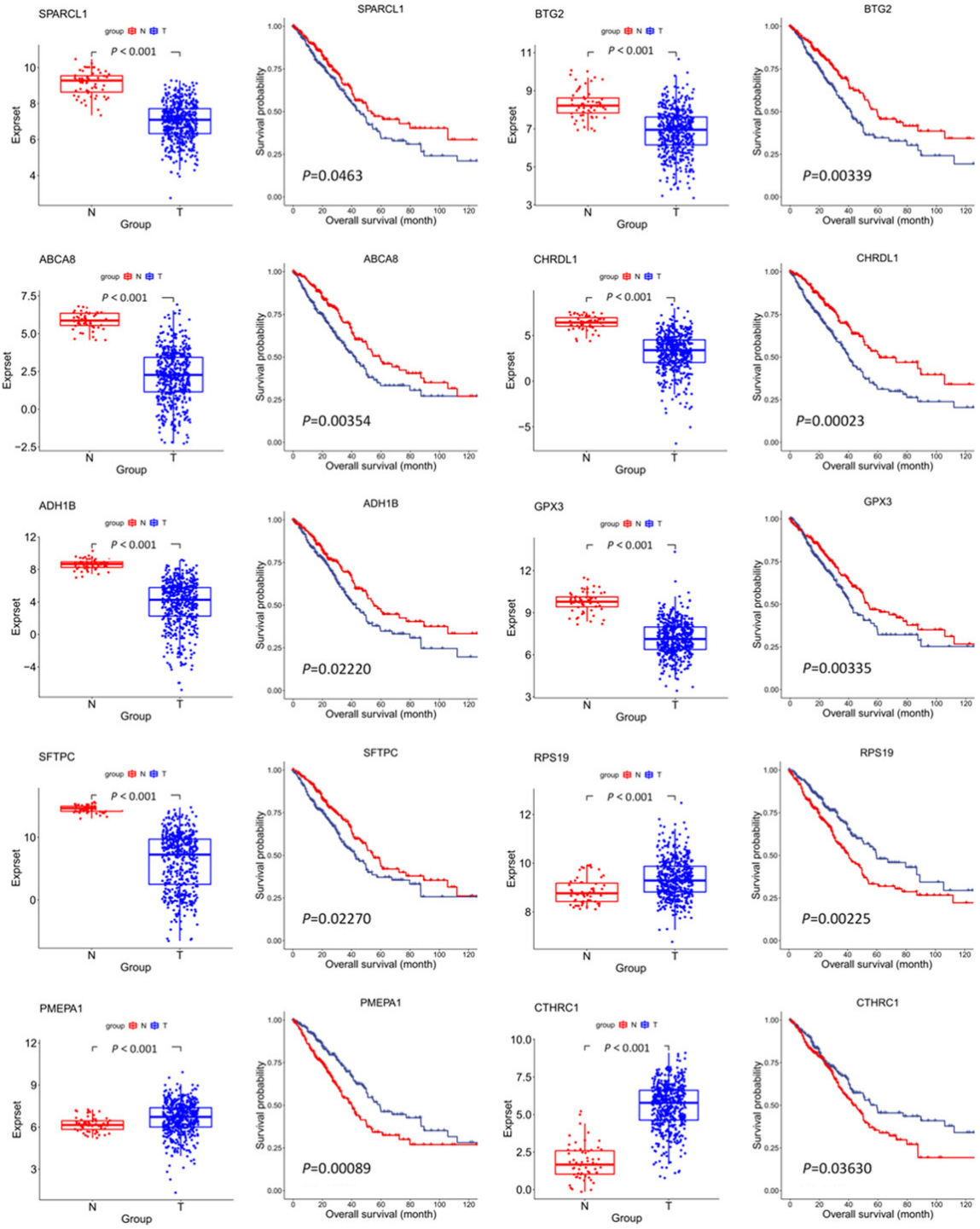


Figure 3. Survival-related genes in fibroblasts in patients with LUAD.

promoting wound healing [22]. Therefore, DCN is also used as a marker of fibroblast activation [23].

CAFs are closely related to tumor survival, but there are few reports on lung cancer survival based on CAFs at single-cell levels. In this study, we screened the genes with high expression and low expression in CAFs and then combined the results with the genes related to the survival of LUAD patients in TCGA data. We screened 10 genes that may be related to the

survival of LUAD, among which the high expression of *RPS19, PMEPA1*, and *CTHRC1* indicates poor prognosis, while those with low expression of *SPARCL1, BTG2, ABCA8, CHRDL1, ADH18, GPX3*, and *SFTPC* have poor prognosis.

RPS19, PMEPA1, and *CTHRC1* have also been reported in a number of studies to be related to the survival of cancers such as liver cancer and lung cancer, but most of these studies are based on tumor cells or bulk tissue sequencing, and these genes are rarely located in a cell subgroup.

Conclusions

Based on scRNA-seq data and TCGA data, this study analyzed the heterogeneity of CAFs and screened a group of genes related to the survival of LUAD patients. The results could provide clues for further study of molecular mechanism, and drug targets. As a follow-up study, the molecular mechanisms could be examined, as well as the histopathology and clinical prognosis, which will provide more scientific value.

Disclosure of conflict of interest

None.

Address correspondence to: Rong Huang, Department of Radiotherapy for Thoracic and Abdominal Cancer, The First People's Hospital of Foshan, North Lingnan Avenue No. 81, Chancheng District, Foshan 528000, Guangdong Province, China. E-mail: dr_huangrong@yeah.net; Baoguo Xie, Reproductive Medicine Center, The First Affiliated Hospital of Hainan Medical University, Longhua Road No. 31, Haikou 570102, Hainan Province, China. E-mail: dr_Baoguo@yeah.net

References

[1] Riihimaki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J and Hemminki K. Metastatic sites and survival in lung cancer. Lung Cancer 2014; 86: 78-84.

[2] Wu DJ, Wang DC, Cheng YF, Qian MJ, Zhang MM, Shen Q and Wang XD. Roles of tumor heterogeneity in the development of drug resistance: a call for precision therapy. Semin Cancer Biol 2017; 42: 13-19.

[3] Ramos P and Bentires-Alj M. Mechanism-based cancer therapy: resistance to therapy, therapy for resistance. Oncogene 2015; 34: 3617-3626.

[4] Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T, Hynes RO, Jain RK, Janowitz T, Jorgensen C, Kimmelman AC, Kolonin MG, Maki RG, Powers RS, Pure E, Ramirez DC, Scherz-Shouval R, Sherman MH, Stewart S, Tlsty TD, Tuveson DA, Watt FM, Weaver V, Weeraratna AT and Werb Z. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer 2020; 20: 174-186.

[5] Busch S, Acar A, Magnusson Y, Gregersson P, Ryden L and Landberg G. TGF-beta receptor type-2 expression in cancer-associated fibroblasts regulates breast cancer cell growth and survival and is a prognostic marker in premenopausal breast cancer. Oncogene 2015; 34: 27-38.

[6] Awaji M and Singh RK. Cancer-associated fibroblasts' functional heterogeneity in pancreatic ductal adenocarcinoma. Cancers (Basel) 2019; 11: 290.

[7] Haghverdi L, Lun ATL, Morgan MD and Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018; 36: 421-427.

[8] Bartoschek M, Oskolkov N, Bocci M, Lovrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringner M, Bergh J, Bjorklund A and Pietras K. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun 2018; 9: 1-13.

[9] Ma KY, Schonnesen AA, Brock A, Van Den Berg C, Eckhardt SG, Liu Z and Jiang N. Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes. JCI Insight 2019; 4: e121387.

[10] Liang W, Huang X, Carlos CJJ and Lu X. Research progress of tumor microenvironment and tumor-associated macrophages. Clin Transl Oncol 2020; 22: 2141-2152.

[11] Nakasone S, Mimaki S, Ichikawa T, Aokage K, Miyoshi T, Sugano M, Kojima M, Fujii S, Kuwata T, Ochiai A, Tsuboi M, Goto K, Tsuchihara K and Ishii G. Podoplanin-positive cancer-associated fibroblast recruitment within cancer stroma is associated with a higher number of single nucleotide variants in cancer cells in lung adenocarcinoma. J Cancer Res Clin Oncol 2018; 144: 893-900.

[12] Eble JA and Niland S. The extracellular matrix in tumor progression and metastasis. Clin Exp Metastasis 2019; 36: 171-198.

[13] Shiao SL, Chu GC and Chung LW. Regulation of prostate cancer progression by the tumor microenvironment. Cancer Lett 2016; 380: 340-348.

[14] Erdogan B and Webb DJ. Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. Biochem Soc Trans 2017; 45: 229-236.
[15] Gonzalez-Silva L, Quevedo L and Varela I. Tumor functional heterogeneity unraveled by scRNA-seq technologies. Trends Cancer 2020; 6: 13-19.
[16] Ren X, Kang B and Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. Genome Biol 2018; 19: 1-14.
[17] Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, Kong SL, Chua C, Hon LK, Tan WS, Wong M, Choi PJ, Wee LJK, Hillmer AM, Tan IB, Robson P and Prabhakar S. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 2017; 49: 708-718.
[18] Li Z, Liu Z, Shao Z, Li C, Li Y, Liu Q, Zhang Y, Tan B and Liu Y. Identifying multiple collagen gene family members as potential gastric cancer biomarkers using integrated bioinformatics analysis. PeerJ 2020; 8: e9123.
[19] Bianchi-Frias D, Vakar-Lopez F, Coleman IM, Plymate SR, Reed MJ and Nelson PS. The effects of aging on the molecular and cellular composition of the prostate microenvironment. PLoS One 2010; 5: e12501.
[20] Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwe H, Pircher A, Van den Eynde K, Weynand B, Verbeken E, De Leyn P, Liston A, Vansteenkiste J, Carmeliet P, Aerts S and Thienpont B. Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med 2018; 24: 1277-1289.
[21] Pang X, Dong N and Zheng Z. Small leucine-rich proteoglycans in skin wound healing. Front Pharmacol 2020; 10: 1649.
[22] Mohan RR, Tripathi R, Sharma A, Sinha PR, Giuliano EA, Hesemann NP and Chaurasia SS. Decorin antagonizes corneal fibroblast migration via caveolae-mediated endocytosis of epidermal growth factor receptor. Exp Eye Res 2019; 180: 200-207.
[23] Nurmik M, Ullmann P, Rodriguez F, Haan S and Letellier E. In search of definitions: cancer-associated fibroblasts and their markers. Int J Cancer 2020; 146: 895-905.