RESEARCH





Integration of single-cell and bulk RNA sequencing to identify a distinct tumor stem cells and construct a novel prognostic signature for evaluating prognosis and immunotherapy in LUAD

Fengyun Zhao^{1*}, Mengting Chen², Tianjiao Wu³, Mingfang Ji¹ and Fugui Li^{1*}

Abstract

Background Cancer stem cells (CSCs) are crucial for lung adenocarcinoma (LUAD). This study investigates tumor stem cell gene signatures in LUAD using single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing (RNAseq), aiming to develop a prognostic tumor stem cell marker signature (TSCMS) model.

Methods LUAD scRNA-seq and RNA-seq data were analyzed. CytoTRACE software quantified the stemness score of tumor-derived epithelial cell clusters. Gene Set Variation Analysis (GSVA) identified potential biological functions in different clusters. The TSCMS model was constructed using Lasso-Cox regression, and its prognostic value was assessed through Kaplan–Meier, Cox regression, and receiver-operating characteristic (ROC) curve analyses. Immune infiltration was evaluated using the Cibersortx algorithm, and drug response prediction was performed using the pRRophetic package. TAF10 functional investigations in LUAD cells involved bioinformatics analysis, qRT-PCR, Western blot, immunohistochemistry, and assays for cell proliferation.

Results Seven distinct cell clusters were identified by CytoTRACE, with epithelial cell cluster 1 (Epi_C1) showing the highest stemness potential. The TSCMS model included 49 tumor stemness-related genes; high-risk patients exhibited lower immune and ESTIMATE scores and increased tumor purity. Significant differences in immune landscapes and chemotherapy sensitivity were observed between risk groups. TAF10 positively correlated with RNA expression-based stemness scores in various tumors, including LUAD. It was over-expressed in LUAD cell lines and clinical tumor tissues, with high expression linked to poor prognosis. Silencing TAF10 inhibited LUAD cell proliferation and tumor sphere formation.

Conclusions This study demonstrates the TSCMS model's prognostic value in LUAD, reveals insights into immune infiltration and therapeutic response, and identifies TAF10 as a potential therapeutic target.

Keywords Single-cell RNA, Gene signature, Prognostic model, LUAD, Tumor stem cell

*Correspondence: Fengyun Zhao 20193101013@stu.gzucm.edu.cn Fuquili leef2002@126.com Full list of author information is available at the end of the article



© The Author(s) 2025. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



Background

Lung adenocarcinoma (LUAD), a prevalent and challenging malignancy of the lungs, has exhibited a steady rise in its incidence in recent years [1-3]. Single-cell RNA sequencing (scRNA-seq) has emerged as a potent tool for delving deeper into the intricate landscape of this disease [4, 5]. Characterized by rapid technological advancements, single-cell technologies have garnered substantial attention and application across diverse solid and hematologic malignancies, such as utilizing scRNAseq to unveil the landscape of infiltrating T cells in liver cancer [6], and investigating the clonal evolution of circulating tumor cells within peripheral blood [7]. In LUAD, by furnishing high-resolution gene expression profiles, single-cell sequencing has endowed researchers with an unprecedented ability to decipher the intricate heterogeneity and underlying molecular mechanisms driving the pathogenesis [8–10].

Leveraging the capabilities of single-cell sequencing, researchers are navigating complex gene expression signatures, functional attributes, and intricate cellular interactions within distinct subpopulations in LUAD [10]. Of particular significance within the landscape of LUAD research is the burgeoning interest in tumor stem cells (TSCs). TSCs are characterized by their intrinsic capacity

for self-renewal, differentiation into diverse cellular lineages, and the pivotal role they play in the initiation and propagation of tumors [11, 12]. They exhibit distinctive biological attributes, including resistance to apoptosis, chemotherapeutic drugs, and radiation therapy, as well as the propensity for long-distance metastasis [12]. These characteristics render TSCs a highly attractive therapeutic target and a critical determinant of tumor aggressiveness and prognosis [13, 14]. In the context of LUAD, several TSCs specific markers have been identified, such as CD44, CD133 (PROM1), and aldehyde dehydrogenase 1 (ALDH1), which are upregulated in TSC populations [15–17]. These markers have been correlated with adverse clinical outcomes and resistance to therapeutic interventions. Preclinical investigations have demonstrated promise in targeting these markers as a therapeutic strategy aimed at overcoming drug resistance and mitigating the aggressive behavior of LUAD [18]. As such, understanding the functional role of TSCs in LUAD has significant implications for the development of targeted therapies aimed at improving patient prognosis.

In pursuit of refining prognostication, numerous studies have embarked on the development of prognostic models [19–21]. The publicly accessible databases, including TCGA and GEO repositories, provide

abundant LUAD samples and associated clinical data, thereby facilitating the construction and rigorous validation of these prognostic models [22, 23]. In this study, we employed single-cell sequencing to investigate tumor stem cells in LUAD, identify key gene signatures, and develop a prognostic risk model. Our analysis revealed a tumor stem cell marker signature (TSCMS) that could predict prognosis and guide therapeutic decisions, including immune checkpoint blockade efficacy. We further identified TATA-box binding protein associated factor 10 (TAF10) as a critical oncogenic gene linked to stemness and poor prognosis in LUAD, suggesting its potential as a therapeutic target. These findings highlight the potential of the TSCMS model to improve prognostication and personalized treatment strategies for LUAD patients.

Materials and methods

Data collection

Single-cell sequencing data (scRNA-seq) and bulk RNA sequencing data were obtained online from the GEO and TCGA databases. The single-cell data originated from GEO (GSE131907, 11 normal and 11 LUAD), while the bulk RNA data were downloaded from TCGA (TCGA-LUAD cohort, n=500) and GEO (GSE26939, n=115, and GSE72094, n=398). The LUAD immunotherapy cohort 'IMvigor210CoreBiologies' was sourced from previously published research [24]. Data for drug IC50 predictions were acquired from a statistical study [25]. All these datasets were sourced from public databases or shared by others.

Preprocessing of ScRNA-seq data

Employing the R package Seurat, we imported the unprocessed expression matrix [26]. Subsequently, we performed filtering to include single-cell data originating from both LUAD and normal tissues. Cells exhibiting mitochondrial gene content exceeding 30% and those manifesting expression of more than 10,000 genes were excluded from the analysis. For normalization, we applied the SCTransform function, which mitigates technical noise and ensures uniform scaling across cells. Subsequently, the RunPCA function was applied with the parameter npcs=50, and the RunUMAP function used parameters reduction="pca" and dim=1:30. The FindNeighbors function was employed with parameters reduction = "pca" and dims = 1:30. Leveraging these neighborhood relationships, clustering was performed with the FindClusters function, wherein a resolution parameter of 0.1 was chosen to delineate 16 distinct cell clusters.

Annotation of cellular subpopulations

After obtaining the 16 clusters, we proceeded to annotate these clusters with cell types based on the expression of specific marker genes [27]. Immunological cells were identified using a spectrum of markers, including PTPRC, and various subclasses such as B cells (CD79A and MS4A1), plasma cells (IGLC2 and IGHM), T cells (CD3D and CD3E), monocytes (CD14 and S100A8), NK cells (NKG7 and GNLY), mast cells (CPA3 and KIT), and macrophages (CD68 and MARCO). Additionally, nonimmune cell types were characterized, including epithelial cells (EPCAM and KRT8), endothelial cells (PECAM1 and VWF), and fibroblasts (COL1A1 and DCN). This analysis ultimately yielded the identification of 10 major cell types within the dataset.

Differential gene analysis

The identification of highly expressed genes in scRNAseq cells was performed using the Seurat package's Find-AllMarkers function with parameters set as only.pos = T and logfc.threshold=0.25, while keeping other parameters as default. Differential gene analysis for the epithelial cell cluster in scRNA-seq was presented in Table S1, and results were visualized using the R package EnhancedVolcano. For bulk RNA-seq differential analysis, the DESeq2 package was utilized with default parameters. Differential analysis was conducted by grouping samples into high and low-risk categories based on the median, and the results of differentially expressed genes between the high-risk and low-risk groups are available in Table S7.

Prediction of tumor epithelial cell stemness

CytoTRACE utilizes gene expression and an intrinsic stemness gene set to predict cell stemness at the singlecell level [28]. To identify the clusters of tumor epithelial cells with the highest stemness or lowest differentiation, we employed the CytoTRACE pipeline from the R package. The results of stemness-related genes (cor > 0.3) can be found in Table S3.

Gene functional enrichment analysis

The enrichment analysis of seven types of tumor tissuederived epithelial cells in scRNA-seq was conducted using the R package GSVA [29]. Initially, 50 tumor Hallmark gene sets were obtained using the R package msigdbr. The GSVA function was applied with the parameter method = "ssgsea" to perform enrichment analysis on the expression matrices of the seven epithelial cell types. The ssgsea enrichment scores can be found in Table S2. For the GSEA enrichment analysis of the TSCMS model, the R package fgsea was used with default parameters. Differential genes between high and low-risk groups based on the TCGA training set were ranked according to their FoldChange. The enrichment results of KEGG pathways from GSEA can be found in Table S10.

Construction and validation of the prognostic risk model TSCMS

Intersecting the stemness-related genes with the highly expressed genes within the tumor epithelial cell cluster Epi_C1, we conducted a univariate Cox regression analysis to ascertain the prognostic significance of these overlapping genes in relation to overall survival among LUAD patients sourced from the TCGA dataset. Genes yielding a p-value of less than 0.05 were designated as prognostic candidates. Subsequently, we subjected the identified prognostic genes to a least absolute shrinkage and selection operator (LASSO) Cox proportional hazards regression, leveraging the "glmnet" package [30]. Employing ten-fold cross-validation, we curated a gene list featuring nonzero coefficients, culminating from an optimal model feature selection process (Table S5). The resultant risk model was meticulously formulated by a linear summation of the products of genes and their corresponding risk coefficients. Patient stratification into low-risk or highrisk groups was based on a median threshold (Table S6). To methodically validate the prognostic efficacy of the TSCMS model, we computed the area under the curve (AUC) utilizing the "timeROC" package [31]. Survival analysis, grounded in the Kaplan-Meier methodology, was adeptly undertaken. Further statistical assessment of differences was facilitated through the application of the log-rank test, seamlessly integrated within the R package "survminer" [32]. Notably, the predictive robustness of the model was subject to rigorous validation via survival analysis and AUC computation across two distinct GEO datasets.

Immune cell infiltration analysis

Immune cell infiltration analysis was conducted by using R Packages CIBERSORT and ESTIMATE in TCGA-LUAD Patients [33]. The infiltration scores for 22 distinct immune cell types were computed using CIBERSORT (Table S11). Based on the median risk score, patients were divided into two groups, and differences in immune cell infiltration across the 22 types were compared between these groups. Furthermore, the ESTIMATE package was utilized to calculate overall immune scores, stromal scores, ESTIMATE scores, and tumor purity (Table S12). Following the division into two groups based on the median risk score, inter-group differences were assessed.

Prediction of immunotherapy response

The IMvigor210 cohort is an immunotherapy-focused dataset for bladder cancer (BLCA), encompassing gene expression matrices, patient clinical information [24], and records of immunotherapy responses. Patients were stratified into two groups based on the median cutoff of their risk scores. Comparative analysis was performed to assess differences in the expression of immune checkpoint markers between the two groups, as well as disparities in patients' immunotherapy responses.

Drug response prediction

We conducted drug response prediction using the pRRophetic package [25]. The gene expression profiles of high- and low-risk groups were employed to estimate the IC50 values for various commonly used clinical or preclinical anti-tumor drugs. By leveraging statistical methods, we identified drugs with significantly distinct IC50 values between these risk groups (Table S8 and S9).

Gene expression and bioinformatics analysis of TAF10 from public database

The expression and RNA expression-based stemness score (RNAss) data for TAF10 in various tumor types in TCGA database were obtained from the SangerBox database (http://SangerBox.com/Tool) [34]. For the prognosis analysis of TAF10 in LUAD, Kaplan-Meier (KM) survival curves for disease-free survival (DFS) and overall survival (OS) were generated using the GEPIA2 platform (http://gepia2.cancer-pku.cn/) [35]. To perform Gene Set Enrichment Analysis (GSEA) based on TAF10 expression levels, tumor samples from the LUAD cohort were initially selected; patients were categorized into TAF10-High and TAF10-Low groups according to the median expression level of TAF10. GSEA was subsequently conducted for KEGG, GOBP, and Hallmark gene sets; all visualizations were generated using the ggplot2 package in R (version 4.4.1).

Cell culture

Human LUAD cell lines (A549, PC9, H1975) and human normal bronchial epithelial cells (16HBE) were purchased from the American Type Culture Collection (ATCC, RRID: CVCL_0023 for A549, CVCL_B260 for PC9, CVCL_1511 for H1975, and CVCL_0021 for 16HBE). The cell lines were authenticated by STR profiling and karyotyping upon initial receipt, and were tested negative for mycoplasma using a PCR-based detection method. All cell lines were maintained in either RPMI-1640 medium or DMEM (Thermo Fisher Scientific, MA, USA) medium supplemented with 10% fetal bovine serum. Cells were cultured at 37 °C in a humidified atmosphere with 5% CO2.

Plasmids and cell transfections

Short hairpin RNA (shRNA) sequences targeting TAF10 were cloned into psiF-copGFP vectors (System Biosciences, Mountain View, CA). The shRNA sequences for TAF10 were 5'-CCAGAAATTCATCTCAGATAT-3', and the sequence for the negative control (shCtl) was 5'-GGTGTGCAGTTGGAATGTA-3'. For plasmidbased transfection, 2 µg of each plasmid (psiF-copGFPshTAF10 or psiF-copGFP-shCtl) were used to transfect HEK-293 T cells, with a plasmid ratio of 1:1 for pMD2.G and psPAX2. The plasmids pMD2.G and psPAX2 were obtained from Addgene (plasmid #12,259 and #12,260, respectively). Lentivirus was harvested 48 h post-transfection. LUAD cell lines were transduced with the virus in the presence of 8 μ g/mL polybrene (Sigma-Aldrich, Cat. # S-2667) and subsequently selected with 2 µg/mL puromycin (Thermo Fisher Scientific, Cat. #A11138-03) for 7 days to establish stable knockdown cell lines.

Cell proliferation and clone formation assay

Cell proliferation was assessed using the Cell Counting Kit-8 (CCK-8) assay (Dojindo, Tokyo, Japan) according to the manufacturer's instructions. For each condition, three biological replicates were performed. Absorbance was measured at 450 nm using a microplate reader (ELX808, BioTek, USA). The cell proliferation rate was calculated relative to the control group. The results are presented as the mean±standard deviation (SD) from three independent experiments.

For the colony formation assay, cells were seeded in 6-well plates at a density of 500 cells/well, with three technical replicates per condition. Cells were incubated at 37 °C, and the medium was refreshed every 3 days until colonies formed. After 14 days, cells were fixed with paraformaldehyde and stained with crystal violet (Sigma-Aldrich, St. Louis, MO). Colonies with more than 50 cells per colony were counted under a microscope. The number of colonies was quantified and compared between experimental groups, with results presented as the mean \pm SD from three independent experiments.

Sphere formation assay

LUAD cell lines were stably transfected with either shCtl or shTAF10 and seeded at 500 cells per well in 24-well ultra-low attachment plates (Corning, USA). Cells were cultured in DMEM/F12 serum-free medium (Gibco, Cat. No. 11320–033) supplemented with 2% B27 (Gibco, Cat. No. 17504–044), 20 ng/mL basic fibroblast growth factor (bFGF, PeproTech, Cat. No. 100-18B), 20 ng/mL epidermal growth factor (EGF, PeproTech, Cat. No. 37000015), 5 μ g/mL insulin (Sigma-Aldrich, Cat. No. 19278), and 0.4% BSA (Sigma-Aldrich, Cat. No. A1933-1G). After 12 days, spheroid formation was assessed by counting the number of spheres with a diameter > 50 μ m under a light microscope. The sphere formation efficiency was calculated as the ratio of the number of spheres to the total number of cells plated. Results were presented as the mean ± SD from three independent experiments.

Quantitative real-time PCR (qRT-PCR)

First-strand cDNA was synthesized using the GenStar A212-05 kit according to the standard protocol. qPCR was performed using the SYBR Green Supermix and CFX96 real-time PCR detection system. Each experiment was performed in triplicate, and the mRNA expression of genes was analyzed using the $2^{-\Delta\Delta Ct}$ method. The following primers were used: 5'-ATTGATGCCATA CTCGCTGAG-3' and 5'- GAAGTGAAGCCCGTA GTGTCC-3' for TAF10, 5'-CTCAAGGTGCTGATG GAGAAGG-3' and 5'- GAACTCACTGAAGTCCAC CTGG-3' for S100P, 5'-ACTCCTTGGTCCAGCTCA TGCA -3' and 5'- ATTCTCCAGCCGCCACAGTACA -3' for PAFAH1B3, 5'-AGAAGGCATAGTTGCTCT GCGC -3' and 5'- CAAGCAGTCAGGACTTAGGTCG -3' for CCT6A, 5'-CCTGCAAAAGCAGTGGACCATG -3' and 5'- CTCCTACCAGTGGCTGAGCATA -3' for DCBLD2, 5'-TCGTGCGTGACATTAAGGAG-3' and 5'-ATGCCAGGGTACATGGTGGT-3' for β-actin. Gene expression data were normalized to β -actin, and results are presented as mean ± SD from three independent biological replicates.

Western blot

Cells were lysed with RIPA buffer supplemented with protease inhibitor and boiled at 95 °C for 5 min. Equal amounts of protein (10 µg) were added to sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gel electrophoresis and transferred to a polyvinylidene difluoride membrane. The membrane was blocked with 5% nonfat dry milk for 1 h at room temperature and then incubated with primary antibodies overnight at 4 °C. Antibodies for TAF10 (Novus Cat# NBP1-80,706, RRID: AB 11006462) were purchased from NOVUS Biologicals, and GAPDH antibody (Sigma-Aldrich Cat# SAB5600208, RRID: AB_2920926) was purchased from Sigma as a loading control. On the following day, membranes were incubated with HRP-conjugated anti-rabbit or anti-mouse secondary antibodies (Santa Cruz Biotechnology, Dallas, TX) for 1 h at room temperature. Immunoreactive proteins were visualized using the SuperSignal West Dura Chemiluminescent Substrate (Thermo Fisher Scientific). Protein bands were quantified using ImageJ software (NIH, Bethesda, MD) and normalized to GAPDH. Quantitative data are presented as the mean±SD from three independent experiments.

Immunohistochemistry

The protocol and all procedures involving human samples in this study were reviewed and approved by the Institutional Review Board (IRB) of Zhongshan City People's Hospital (approval number: 2024-116). All paraffinembedded tissues of patients in this study were obtained with informed patient consent (n = 5 pairs of adjacent and tumorous tissues, totaling 10 tissues). For immunohistochemistry staining, deparaffinized and rehydrated sections were boiled in Na-citrate buffer (10 mM, pH 6.0) for 30 min for antigen retrieval. The sections were incubated with primary antibodies and developed using the Ultra Vision Detection System. Images were captured using an Olympus IX51 microscope and processed using cellSens Dimension software. The H-Score (Histochemistry score) is calculated as \sum (pi×i), where i represents the intensity grading of positive cells: 0 for negative (no staining); 1 for weak positive (pale yellow); 2 for moderate positive (brown-yellow); and 3 for strong positive (brown). Here, pi represents the percentage of cells at each respective intensity level. The formula can be expressed as: H-Score = (percentage of weak positive cells \times 1) + (percentage of moderate positive cells $\times 2$) + (percentage of strong positive cells \times 3). The resulting H-Score ranges from 0 to 300. A higher H-Score indicates a stronger overall positivity, reflecting both the intensity and the proportion of positive cells.

Statistical analysis

In appropriate scenarios, we employed either the Student's t-test or the Wilcoxon rank-sum test to assess the significance of differences between groups. The selection of the test depended on the distribution of the data and the assumption of normality. For survival analysis, the Log-Rank test was utilized to determine the significance of survival differences between different groups or conditions. P value less than 0.05 was considered statistically significant. Statistical significance levels were denoted as follows: * for p < 0.05, ** for p < 0.01, *** for p < 0.001, and **** for p < 0.001.

Results

Workflow and cell population landscape in LUAD

To explore the potential functions of LUAD tumor stem cells, we collected bulk RNA-seq data from TCGA-LUAD and GEO datasets (GSE26939 and GSE72094), as well as single-cell RNA-seq data from the GEO dataset (GSE131907). Using scRNA-seq, we predicted the stemness score of tumor epithelial cells. Then, we constructed a LUAD prognostic model based on tumor stemness genes and further validated its predictive ability (Fig. 1A). First, we conducted quality control on all cells, applying filters with a minimum cell count of 3, a minimum feature count of 200, and mitochondrial gene content of less than 30% (Fig. S1). Next, we annotated a total of 22 samples (11 normal and 11 LUAD) from the single-cell dataset, which comprised 88,144 individual cells distributed across 16 clusters (Fig. 1B, C, and Fig. S2). Based on the expression of cell markers within clusters, we identified ten major cell populations (Fig. 1B). Compared to normal lung tissue, LUAD exhibited reduced infiltration of NK cells and macrophages. However, LUAD patients demonstrated heterogeneity, with different LUAD samples showing varying proportions of epithelial cells (Fig. 1D). We used common cell markers such as EPCAM for epithelial cells, PECAM1 for endothelial cells, PTPRC for immune cells, and COL1A1 for fibroblasts to define each cell type (Fig. 1E). Therefore, LUAD exhibits substantial tumor heterogeneity, with varying compositions of tumors and their microenvironments among different patients.

Prediction of tumor epithelial stem cells

Further exploration of tumor stem cells involved the selection of 7252 tumor-derived epithelial cells for calculating stemness scores using the CytoTRACE software (Fig. 2A). After applying dimensionality reduction and clustering techniques, 7 distinct cell clusters were identified (Fig. 2B). Comparing the CytoTRACE-predicted stemness scores across these 7 tumor epithelial cell clusters revealed that Epi_C1 exhibited the highest stemness potential (Fig. 2C). Subsequent differential gene analysis of Epi_C1 highlighted elevated expression of genes such as CDKN2A, TMSB10, SOO2A, PTGS2, and SNCG (Fig. 2D and Table S1). Additionally, the Hallmark GSVA enrichment analysis demonstrated that Epi_C1 displayed higher enrichment scores in pathways associated with hypoxia, EMT, Kras signaling, MYC signaling, as well as E2F targets and G2M checkpoint, which are closely linked to cell cycle regulation, compared to other epithelial cell clusters (Fig. 2E and Table S2). Thus, the Epi_C1 cluster is likely to represent a subpopulation of stem-like epithelial cells within LUAD tumors.

Construction and validation of the prognostic model TSCMS

To investigate the impact of stem-like tumor epithelial cells on LUAD patients, we intersected 1068 highly expressed genes in Epi_C1 with 2509 CytoTRACEcomputed genes showing correlation (cor > 0.3), resulting in 964 genes (Fig. 3A and Table S1, 3). These genes



Fig. 1 Landscape of cell type in LUAD and normal tissues. A Workflow of this study. B UMAP plot of major nine cell types of LUAD. C UMAP plot of sites. Different cell types and sites are grouped by different colors. D The proportion of different cell types within each sample. E Expression of representative genes for different cell types. Bubble size reflects expression proportion, while the color gradient from blue to red signifies higher expression levels

were utilized for univariate Cox regression analysis to predict their association with survival in LUAD patients. We used the LUAD mRNA count expression matrix and corresponding clinical information from the TCGA database as the training set. Out of the 964 genes, 92 genes with p-value < 0.05 (Fig. S3) were further subjected to Lasso regression and multiple-factor Cox regression with tenfold cross-validation, ultimately selecting 49 genes with non-zero coefficients as features to construct tumor stem cell marker signature (TSCMS) prognostic risk model (Fig. 3B and Table S4, 5). The risk score was calculated based on the cumulative expression values of the genes multiplied by their corresponding coefficients, and the TCGA training set samples were divided into



Fig. 2 Identification and functional analysis of tumor stem cells. **A**, **B** UMAP plot of 7 distinct tumor epithelial cell types with CytoTRACE stemness scores (**A**) and cell clusters (**B**). **C** Tumor stemness scores of 7 epithelial cell clusters using CytoTRACE. **D** Volcano plot of differentially expressed genes in Epi_C1. **E** Hallmark enrichment analysis of 7 epithelial cell clusters. The intensity of enrichment increases from blue to red

high and low-risk groups using the median risk score (Fig. 3C and Table S6). In the training set, the model significantly stratified patients' survival (p < 0.0001), with area under the curve (AUC) values of 0.818, 0.851, and

0.871 for 1-year, 3-year, and 5-year survival, respectively (Fig. 3D, G). Furthermore, we validated the TSCMS model using two independent external LUAD datasets, GSE26939 and GSE72094. The model demonstrated



Fig. 3 Construction and validation of the prognostic model TSCMS. A Overlapping CytoTRACE predicted stemness-associated genes and marker genes of Epi_C1. B Each independent variable's trajectory and distribution for the lambda. C Expression of 49 TSCMS genes in TCGA-LUAD cohort. D–F Kaplan–Meier plot of prognostic survival for TCGA (D), validation sets GSE26939 (E) and GSE72094 (F). G–I ROC curves for TCGA (G) test set, validation sets GSE26939 (H) and GSE72094 (I). Red for 1-year, blue for 3-year, and black for 5-year survival rates

robust prognostic stratification ability in GSE26939 (p=0.012) and GSE72094 (p=0.00015) (Fig. 3E, F), with corresponding AUC values of 0.707, 0.637, and 0.595 for

1-year, 3-year, and 5-year survival in GSE26939 (Fig. 3H), and 0.702, 0.667, and 0.751 in GSE72094 (Fig. 3I). In conclusion, the newly developed prognostic risk model based

on stem-like tumor epithelial cells exhibits an excellent predictive capacity for the prognosis of LUAD patients.

The association between TSCMS and immune cell infiltration in the TME

As immune cells play a pivotal role in tumor immunity and promotion, we explored the relationship between TSCMS and immune cell infiltration within LUAD patients. Leveraging the cibersortx program, we investigated the infiltration of 22 distinct immune cell types. Notably, the high-risk group demonstrated diminished levels of B cell naive, CD4⁺ T cell memory resting, monocytes, and mast cells when compared with the low-risk group (Fig. 4A). Conversely, macrophage M0 infiltration exhibited heightened levels within the high-risk group (Fig. 4A). Moreover, employing the ESTIMATE program, we calculated infiltration scores for both highand low-risk groups, the high-risk group revealed markedly reduced immune scores (Fig. 4B), lower ESTIMATE scores (Fig. 4D), and heightened tumor purity (Fig. 4E). Remarkably, no pronounced disparities emerged in stromal scores (Fig. 4C). After partitioning TCGA-LUAD samples into two distinct groups according to the median risk score, we conducted an analysis of differential gene expression (Table S7), followed by GSEA enrichment analysis ranked by fold change. Specifically, the highrisk group exhibited enrichments in pivotal pathways such as cell cycle regulation, DNA repair, and P53 signaling, while the low-risk group showcased enrichments in chemokine signaling, chemokine receptor interactions, and T cell receptor signaling (Fig. 4F). These results imply a positive correlation between the TSCMS risk score and tumor cell proliferation, and a negative correlation with immune functionality. The diminished predictive prognosis of TSCMS might be attributed to its association with reduced immune infiltration capacity.

TSCMS could predict immunotherapy benefits in LUAD patients

Building upon the pivotal role of TSCMS in immune cell infiltration, we further explored its predictive influence on immune checkpoint blockade and immunotherapy response. Firstly, within the IMvigor210 cohort, we analyzed immune checkpoint expression including PD1, PD-L1, and CTLA4. Notably, there were no significant differences in PD1 and CTLA4 expression between high- and low-risk groups (Fig. 5A, D), while PD-L1 expression was higher in the high-risk group (Fig. 5B). Evaluating the response to anti-PD-L1 therapy, the risk score was notably lower in the R (complete response/partial response; CR/PR) group compared to the NR (stable disease/progressive disease; SD/PD) group (Fig. 5C). In terms of treatment response, the low-risk group showed a nearly 9% higher proportion of CR/PR compared to the high-risk group (Fig. 5E). Moreover, the stratification of patient prognosis by TSCMS within this cohort exhibited statistically significant implications (Fig. 5F). In conclusion, these findings suggest that patients with a lower risk score may benefit more from anti-PD-L1 therapy, indicating TSCMS as a potentially helpful biomarker for anti-PD-L1 treatment.

TSCMS-based prediction of anti-tumor drug efficacy

In addition to immunotherapy, chemotherapy remains a pivotal approach in the battle against tumors. Thus, we computed the IC50 sensitivities of commonly used clinical or preclinical anti-tumor drugs between highand low-risk TSCMS groups. Among the findings, IC50 values for 61 drugs were observed to be lower in the high-risk group in comparison to the low-risk group (Table S8). Furthermore, for 8 drugs, the IC50 values in the low-risk group were significantly lower than those in the high-risk group (Table S9). Through prioritizing results based on significance, we revealed the top 6 drugs with better sensitivity in the high-risk group (Fig. 6A) and the top 3 drugs with better sensitivity in the low-risk group (Fig. 6B) in terms of IC50 outcomes. These results hold the potential to offer invaluable guidance for personalized treatment strategies in LUAD patients.

TAF10 plays oncogenic role in LUAD

To investigate the role of genes incorporated into the TSCM prognostic risk model in LUAD, we assessed the mRNA expression of the top five genes in LUAD cell lines. Our results demonstrated that these genes were significantly upregulated in LUAD cell lines compared to human normal bronchial epithelial cells (16HBE), with TAF10 exhibiting the highest expression levels (Fig. 7A). Given its strong association with the prognosis of LUAD patients among the 49 genes analyzed, TAF10 was selected for further investigation. Using data from TCGA and GTEx databases, we found that the mRNA expression of TAF10 was elevated in various tumors, including LUAD, compared to corresponding normal tissues (Fig. 7B). Importantly, analysis of stemness features indicated a positive correlation between TAF10 expression and RNA expression-based stemness scores (RNAss) across several tumor types, including LUAD (R=0.325, p = 0.009) (Fig. 7C). Additionally, high TAF10 expression was correlated with poor prognosis in LUAD patients (Fig. 7D, E). Western blotting results further confirmed the high protein expression of TAF10 in LUAD cells (Fig. 7F). Furthermore, TAF10 expression was significantly higher in tumor tissues compared to adjacent normal tissues (Fig. 7G). To further investigate the role of TAF10 in LUAD, we used a loss-of-function approach



Fig. 4 Immune infiltration and functional analysis of TSCMS. A Fraction scores of 22 immune cell infiltration using CIBERSORTx software. B-E Box plots of immune scores (B), stromal scores (C), ESTIMATE scores (D), and tumor purity (E) for TSCMS high- and low-risk groups using ESTIMATE software. F Enhanced GSEA plot for TSCMS gene set enrichment analysis



Fig. 5 Prediction of immunotherapy efficacy using TSCMS in the IMvigor210 cohort. A Box plot of PD1 expression in high-risk and low-risk groups. B Box plot of PD-L1 expression in high-risk and low-risk groups. C Box plot of TSCMS scores in the anti-PD-L1 treatment group. D Box plot of CTLA4 expression in high-risk and low-risk groups. E Bar chart showing treatment response proportions in high-risk and low-risk groups. F Kaplan–Meier plot of TSCMS in the IMvigor210 Cohort



Fig. 6 Comparison of anti-tumor drug sensitivity between high-risk and low-risk groups. A Bortezomib, Pazopanib, AKT inhibitor VIII, AZD6482, CGP.082996, and CEP-701 demonstrated enhanced drug sensitivity in the high-risk group. B CCT007093, GDC.0449, and Lapatinib exhibited superior drug sensitivity in the low-risk group. Statistics based on Wilcoxon test

to evaluate the impact of TAF10 on LUAD cells (Fig. 7H). We found that TAF10 silencing significantly reduced colony formation (Fig. 7I) and suppressed cell proliferation in LUAD cell lines (Fig. 7J). Notably, TAF10-knockdown LUAD cells formed fewer and smaller tumor spheres than those transduced with the negative control (shCtl) (Fig. 7K). These results indicate that upregulated TAF10 expression promotes stemness and cell proliferation in LUAD, highlighting TAF10 as an ideal gene for further mechanistic studies in LUAD.

To investigate the signaling pathways associated with TAF10, we performed GSEA based on its expression levels. We found that high TAF10 expression is linked to key tumor-related pathways, including the MAPK signaling pathway, Notch signaling, and pathways involved in the cell cycle and DNA replication (Fig. 7L). Notably, high TAF10 expression is also associated with non-small cell lung cancer (NSCLC) (Fig. 7L). From a biological perspective, high TAF10 expression positively regulates the cell cycle and protein translation, potentially influencing the proliferation and differentiation of tumor stem cells (Fig. 7M). Furthermore, hallmark of Cancer pathway analysis revealed significant correlations between high TAF10 expression and critical pathways related to cell proliferation, including the classic p53 pathway, tumor DNA repair mechanisms, and the G2/M checkpoint, as well as E2F targets (Fig. 7N). These findings suggest that TAF10 may promote tumor stemness by regulating key pathways involved in cell proliferation, the cell cycle, and DNA repair, thus contributing to the maintenance and differentiation of tumor stem cells.

Discussion

In the rapidly evolving field of biomedical research, advanced optimization and feature selection techniques have emerged as transformative elements. Numerous studies have introduced innovative optimization algorithms that not only enhance the performance of diagnostic models but also reshape medical research, enabling more accurate disease diagnoses and a deeper understanding of biological mechanisms [36-39]. Previously, significant efforts focused on developing prognostic models for LUAD by utilizing tumor-associated cancer-associated fibroblasts (CAFs) and immune cells [40–43]. In contrast, the present study centers on tumor stem cells in LUAD. By integrating stemness-associated genes and scRNA-seq datasets, we have successfully identified distinct epithelial cell clusters within the tumor microenvironment that display stemness characteristics. This endeavor led us to the development of a novel prognostic risk model termed TSCMS. Notably, this model relies not only on clinical parameters but also provides enhanced precision in predicting patient survival outcomes, thereby serving as a robust aid for informed clinical decision-making.

The TSCMS was constructed from a set of 49 key genes, among which TAF10, S100P, PAFAH1B3, CCT6A, DCBLD2, CCDC85B, PSMD11, TFAP2A, TM4SF1, and DRG1 hold prominent coefficients. Functionally, TAF10, TFAP2A, PAFAH1B3, and DRG1 play crucial roles in regulating cell proliferation and differentiation in tumors [44–47]. S100P and TM4SF1 are mainly responsible for affecting cell-cell interaction and migration [48, 49]. CCDC85B and PSMD11 are involved in protein degradation [50, 51], and DCBLD2 promotes angiogenesis for tumor growth [52]. CCT6A facilitates lung adenocarcinoma progression and glycolysis via STAT1/HK2 axis [53]. Notably, the TSCMS exhibits robust predictive capacity, highlighting its potential significance in prognostic evaluation.

In the TCGA training dataset, the TSCMS reveals a substantial median survival difference of more than 5 years between high-risk and low-risk patients. Similarly, across two independent external validation datasets, the TSCMS indicates a median survival difference of 3 years between the high and low-risk patient groups. Furthermore, when evaluating the accuracy of TSCMS for predicting survival rates at 1 year, 3 years,

(See figure on next page.)

Fig. 7 TAF10 plays oncogenic role in LUAD. **A** mRNA expression levels of the corresponding gene in human normal bronchial epithelial cells (16HBE) and LUAD cell lines. **B** TAF10 mRNA expression levels in various tumors and matched normal tissues from the TCGA and GTEx databases, analyzed using SangerBox platform. **C** The stemness features (RNA expression-based stemness scores) analyses of TAF10 across different types of tumors in the TCGA database, analyzed by SangerBox platform. (**D-E**) Disease-free survival (**D**) and overall survival (**E**) analyses of TAF10 in LUAD samples from the TCGA database, performed using the GEPIA2 platform. (**F**) Protein expression levels of TAF10 in 16HBE and LUAD cell lines. **G** Representative IHC analysis of TAF10 expression in paired adjacent and tumorous tissues from LUAD patients (n = 5 pairs, 10 tissues in total). Black scale bar: 50 μm; red scale bar: 20 μm. **H** TAF10 knockdown in LUAD cells was confirmed by Western blot analysis. **I** LUAD cell lines were stably transfected with either shCtl or shTAF10 for 24, 48, and 72 h, and cell viability was measured using a CCK-8 assay. (**J**) The effect of TAF10 knockdown on colony formation in LUAD cells (shCtl). Scale bar, 100 μm. **L**–**N** GSEA plot of KEGG (**L**), GOBP (**M**), and Hallmark pathways (**N**), grouped by TAF10 expression into TAF10-high and TAF10-low subgroups. NES represents the normalized enrichment score, and FDR represents the adjusted p-value



Fig. 7 (See legend on previous page.)

and 5 years, our results consistently surpass an average threshold of 0.7 across the three cohorts. Importantly, in comparison to risk models proposed by Ren et al.

centered on CAFs [40] and Zhang et al. focusing on T-cell markers [42], our TSCMS displays superior accuracy in predicting patient survival. This underscores the robust predictive power of TSCMS for patient survival.

Immune cells play a pivotal role within the tumor microenvironment, exerting influence over tumor development and therapeutic responses. Employing methodologies such as CIBERSORT and ESTIMATE, we observed distinctions in the distribution of B cells, monocytes, mast cells, and macrophages between the high-risk and low-risk groups. Sarvaria et al. have highlighted the crucial role of B cells in promoting inflammation and carcinogenesis [54]. M0 macrophages exhibited significantly higher infiltration in the high-risk group, a trend consistent with the findings of Huang et al., who reported M0 macrophages promoting malignant growth in glioma [55]. Mast cells promote angiogenesis by releasing classical pro-angiogenic factors, and support tumor invasion by releasing matrix metalloproteinases [56]. We found that the proportion of activated mast cells was notably higher in the high-risk group compared to the low-risk group. Additionally, the low-risk group exhibited higher enrichment of immune-related pathways, including T cell receptor signaling and chemokine-chemokine receptor signaling pathways. In the tumor microenvironment, tumor infiltration of T cells was driven by chemokines [57], and T cells integrate chemokine signals to enhance antitumor responses in peripheral tissues [58]. Those results suggest a potential association between tumor stemness and immune infiltration, thereby providing valuable leads for further exploration into immune-based therapies.

Immunotherapy and drug treatment are effective strategies in combating cancer. TSCMS has demonstrated robust predictive capability within the IMvigor210 immunotherapy cohort. In the high-risk group, PD-L1 expression levels are significantly elevated compared to low-risk patients. Additionally, patients who respond favorably to anti-PD-L1 treatment exhibit lower risk scores. Moreover, we successfully predicted the response of different risk groups to various anti-cancer drugs. These findings provide substantial support for personalized treatment and drug selection, holding the potential to make a positive impact in clinical practice.

Among the 49 key genes, TAF10 holds the highest prognostic correlation coefficient. Numerous studies have demonstrated that TAF10 plays an oncogenic role within a wide variety of tumors, including transcription, the cell cycle, and apoptosis [59, 60]. For example, in gastric cancer cells, high expression of TAF10 plays an important role in maintaining tumor cell survival [61]. We found that TAF10 is overexpressed in LUAD cell lines and tumor tissues, and that elevated TAF10 expression is associated with poor prognosis in LUAD. Silencing TAF10 inhibited tumor sphere formation in LUAD cells, with TAF10-knockdown cells forming significantly fewer and smaller spheres compared to control cells. These results suggest that TAF10 may play a critical role in regulating tumor stemness.

In terms of signaling pathways, our initial and exploratory analysis revealed that high TAF10 expression is associated with several key tumor-related pathways, including the MAPK signaling pathway, Notch signaling, and pathways involved in the cell cycle and DNA replication. These pathways are fundamental for tumor cell proliferation and differentiation, indicating that TAF10 may promote LUAD progression by modulating these critical pathways. Furthermore, high TAF10 expression was linked to NSCLC, further supporting its potential role in tumorigenesis. From a biological perspective, TAF10 appears to regulate the cell cycle and translation, impacting tumor stem cell proliferation and differentiation. Hallmark of Cancer analysis also revealed significant correlations between high TAF10 expression and key pathways involved in cell proliferation, such as the p53 pathway, DNA repair mechanisms, and the G2/M checkpoint. Taken together, these findings suggest that TAF10 may promote tumor stemness and progression through its regulation of essential pathways, although further studies are needed to fully elucidate its mechanisms in LUAD.

However, our study has certain limitations. Firstly, while the TSCMS model demonstrates robust predictive capacity, its clinical utility requires independent validation in larger, prospective cohorts of LUAD patients. The sample size in this study may limit the generalizability of the findings, and future research should address this to confirm the robustness of the model. Secondly, although we identify key pathways associated with TAF10 and tumor stemness, direct experimental validation is lacking, and further mechanistic studies are needed to explore how TAF10 regulates tumor progression and stemness at the molecular level, including in vivo studies to better understand its role in LUAD. Furthermore, although the predictive capacity of TSCMS in immunotherapy cohorts is promising, additional clinical trials and mechanistic studies are needed to fully assess its effectiveness in predicting treatment outcomes, particularly in the context of immunotherapy. These validations will be crucial for strengthening both the clinical applicability and biological understanding of our findings.

In terms of clinical implications, targeting TAF10 with specific inhibitors or RNA interference could be explored as potential therapeutic strategies to reduce LUAD metastasis. Additionally, integrating the TSCMS model into clinical diagnostic protocols could aid in the early identification of high-risk patients, enabling timely and personalized treatment interventions. Moreover, incorporating the TSCMS model could facilitate the development of more effective, tailored therapies that address the unique molecular profiles of LUAD patients. Collaborative efforts with clinical institutions for real-world testing will be essential for translating these findings into clinical practice.

Lastly, while we have outlined potential therapeutic implications, further studies are required to refine these strategies and explore the most effective approaches for targeting TAF10 in LUAD. The use of advanced optimization and feature selection methods could also improve the analytical rigor of future studies, enhancing the accuracy and reliability of molecular interactions identified in cancer progression.

Conclusion

In conclusion, our research highlights the prognostic value of the TSCMS model in evaluating the clinical outcomes of LUAD patients, provides important insights into immune cell infiltration and therapeutic response, and suggests that TAF10 may serve as a potential therapeutic target for LUAD. However, further studies are needed to validate its clinical relevance and therapeutic potential.

Abbreviations

| LUAD CSCs TSCMS OS TME | Lung adenocarcinoma Cancer stem cells Tumor stem cell marker signature Overall survival Tumor microenvironment |
|------------------------------------|--|
| KM | Kaplan-Meier |
| GSVA | Gene set variation analysis |
| DEGs | Differentially expressed genes |
| LogFC | LogFoldChange |
| scRNA-seq | Single-cell RNA sequencing |
| TCGA | The cancer genome atlas |
| GEO | Gene expression omnibus |
| PCA | Principal components analysis |
| KEGG | Kyoto encyclopedia of genes and genomes |
| GSEA | Gene set enrichment analysis |
| ROC | Receiver operating characterastic |
| AUC | Area under the curve |
| RNAss | RNA expression-based stemness score |
| CCK-8 | Cell counting kit-8 |
| qRT _ PCR | Quantitative rea-time PCR |
| LASSO | Least absolute shrinkage and selection operator |
| shRNA | Short hairpin RNA |

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12967-025-06243-6.

Additional file 1.

Additional file 2.

Author contributions

F-Y.Z. designed the study, collected samples, performed the experiments, analyzed the data, and wrote the manuscript. M.-T.C. and T.-J.W. wrote the initial draft of the manuscript, performed the experiments. M.-F.J. participated in RNA-seq data analysis and data interpretation. F.-G.L. designed and supervised the study, provided reagents, provided his expertise in the analysis of the data, and edited the manuscript. All authors critically reviewed, revised and approved the final version of the manuscript.

Funding

This study was supported by the grants from Zhongshan Social Welfare Science and Technology Research Project (No. 2022B3001).

Availability of data and materials

Databases analyzed for this study are available in online repositories. Detailed information can be found in the article.

Declarations

Ethics approval and consent to participate

The protocol and all procedures involving human samples in this study were reviewed and approved by the Institutional Review Board (IRB) of Zhongshan City People's Hospital (approval number: 2024–116). Informed patient consent was obtained from all individuals whose paraffin-embedded tissues were used in this study. All research procedures were conducted in compliance with relevant ethical guidelines and local laws.

Consent for publication

Not appliable.

Competing interests

The authors declare no conflict of interest.

Author details

¹Cancer Research Institute of Zhongshan City, Zhongshan City People's Hospital, Zhongshan 528403, Guangdong, China. ²South China Normal University, Guangzhou 510630, Guangdong, China. ³Guangdong Medical University, Zhanjiang 523000, Guangdong, China.

Received: 18 December 2024 Accepted: 11 February 2025 Published online: 22 February 2025

References

- 1. Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. Cell Death Dis. 2018;9:117.
- Wang Y, Ding Y, Liu S, Wang C, Zhang E, Chen C, Zhu M, Zhang J, Zhu C, Ji M, et al. Integrative splicing-quantitative-trait-locus analysis reveals risk loci for non-small-cell lung cancer. Am J Hum Genet. 2023;110:1574–89.
- Xu Y, Li J, Zhu K, Zeng Y, Chen J, Dong X, Zhang S, Xu S, Wu G. FIBP interacts with transcription factor STAT3 to induce EME1 expression and drive radioresistance in lung adenocarcinoma. Int J Biol Sci. 2023;19:3816–29.
- Suva ML, Tirosh I. Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. Mol Cell. 2019;75:7–12.
- Lei Y, Tang R, Xu J, Wang W, Zhang B, Liu J, Yu X, Shi S. Applications of single-cell sequencing in cancer research: progress and perspectives. J Hematol Oncol. 2021;14:91.
- Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, Kang B, Hu R, Huang JY, Zhang Q, et al. Landscape of infiltrating t cells in liver cancer revealed by single-cell sequencing. Cell. 2017;169:1342–56.
- Shi M, Dong X, Huo L, Wei X, Wang F, Sun K. The potential roles and advantages of single cell sequencing in the diagnosis and treatment of hematological malignancies. Adv Exp Med Biol. 2018;1068:119–33.
- Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, Kang B, Liu Z, Jin L, Xing R, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;24:978–85.
- Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, Lee JI, Suh YL, Ku BM, Eum HH, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun. 2020;11:2285.
- Zhang L, Zhang Y, Wang C, Yang Y, Ni Y, Wang Z, Song T, Yao M, Liu Z, Chao N, et al. Integrated single-cell RNA sequencing analysis reveals distinct cellular and transcriptional modules associated with survival in lung cancer. Signal Transduct Target Ther. 2022;7:9.

- Wang X, Chen Y, Wang X, Tian H, Wang Y, Jin J, Shan Z, Liu Y, Cai Z, Tong X, et al. Stem cell factor SOX2 confers ferroptosis resistance in lung cancer via upregulation of SLC7A11. Cancer Res. 2021;81:5217–29.
- Guo L, Mohanty A, Singhal S, Srivastava S, Nam A, Warden C, Ramisetty S, Yuan YC, Cho H, Wu X, et al. Targeting ITGB4/SOX2-driven lung cancer stem cells using proteasome inhibitors. iScience. 2023;26:107302.
- Masciale V, Banchelli F, Grisendi G, Samarelli AV, Raineri G, Rossi T, Zanoni M, Cortesi M, Bandini S, Ulivi P, et al. The molecular features of lung cancer stem cells in dedifferentiation process-driven epigenetic alterations. J Biol Chem. 2024;300: 107994.
- 14. Pan Z, Liu H, Chen J. Lung cancer stem-like cells and drug resistance. Zhongguo Fei Ai Za Zhi. 2022;25:111–7.
- Bertolini G, Roz L, Perego P, Tortoreto M, Fontanella E, Gatti L, Pratesi G, Fabbri A, Andriani F, Tinelli S, et al. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. Proc Natl Acad Sci U S A. 2009;106:16281–6.
- Leung EL, Fiscus RR, Tung JW, Tin VP, Cheng LC, Sihoe AD, Fink LM, Ma Y, Wong MP. Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties. PLoS ONE. 2010;5: e14062.
- Jiang F, Qiu Q, Khanna A, Todd NW, Deepak J, Xing L, Wang H, Liu Z, Su Y, Stass SA, Katz RL. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. Mol Cancer Res. 2009;7:330–8.
- Raei M, Bagheri M, Aghaabdollahian S, Ghorbani M, Sadeghi A. Ionizing radiation promotes epithelial-mesenchymal transition phenotype and stem cell marker in the lung adenocarcinoma. in vitro and bioinformatic studiesc. Cell J. 2022;24:522–30.
- Zhang W, Zhao L, Zheng T, Fan L, Wang K, Li G. Comprehensive multiomics integration uncovers mitochondrial gene signatures for prognosis and personalized therapy in lung adenocarcinoma. J Transl Med. 2024;22:952.
- Cai Q, He B, Zhang P, Zhao Z, Peng X, Zhang Y, Xie H, Wang X. Exploration of predictive and prognostic alternative splicing signatures in lung adenocarcinoma using machine learning methods. J Transl Med. 2020;18:463.
- Gan J, Huang M, Wang W, Fu G, Hu M, Zhong H, Ye X, Cao Q. Novel genome-wide DNA methylation profiling reveals distinct epigenetic landscape, prognostic model and cellular composition of early-stage lung adenocarcinoma. J Transl Med. 2024;22:428.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41:D991–5.
- Blum A, Wang P. Zenklusen JC: snapshot: TCGA-analyzed tumors. Cell. 2018;173:530.
- Necchi A, Joseph RW, Loriot Y, Hoffman-Censits J, Perez-Gracia JL, Petrylak DP, Derleth CL, Tayama D, Zhu Q, Ding B, et al. Atezolizumab in platinumtreated locally advanced or metastatic urothelial carcinoma: postprogression outcomes from the phase II IMvigor210 study. Ann Oncol. 2017;28:3044–50.
- Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS ONE. 2014;9: e107468.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.
- Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, et al. Cell Marker 20: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 2023;51:D870–6.
- Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. Science. 2020;367:405–11.
- 29. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform. 2013;14:7.
- Liang L, Yu J, Li J, Li N, Liu J, Xiu L, Zeng J, Wang T, Wu L. Integration of scRNA-Seq and bulk RNA-seq to analyse the heterogeneity of ovarian cancer immune cells and establish a molecular risk model. Front Oncol. 2021;11: 711020.
- 31. Wang P, Zhang T, Wang X, Xiao H, Li H, Zhou LL, Yang T, Wei B, Zhu Z, Zhou L, et al. Aberrant human ClpP activation disturbs mitochondrial

proteome homeostasis to suppress pancreatic ductal adenocarcinoma. Cell Chem Biol. 2022;29:1396–408.

- Wang L, Wang D, Yang L, Zeng X, Zhang Q, Liu G, Pan Y. Cuproptosis related genes associated with Jab1 shapes tumor microenvironment and pharmacological profile in nasopharyngeal carcinoma. Front Immunol. 2022;13: 989286.
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol. 2018;1711:243–59.
- Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell. 2018;173:338–54.
- Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556–60.
- 36. A novel multi class disease detection of chest x-ray images using deep learning with pre trained transfer learning models for medical imaging applications
- 37. Gharaibeh H, Alawad NA, Nasayreh A, Al Mamlook RE, Makhadmeh SN, Bashkami A, Al-Na'Amneh Q, Abualigah L, Ezugwu AE. A novel binary modified beluga whale optimization algorithm using ring crossover and probabilistic state mutation for enhanced bladder cancer diagnosis. Inform Med Unlocked. 2024;50: 101581.
- Got A, Zouache D, Moussaoui A, Abualigah L, Alsayat A. Improved manta ray foraging optimizer-based svm for feature selection problems: a medical case study. J Bionic Eng. 2024;21:409–25.
- Bashkami A, Nasayreh A, Makhadmeh SN, Gharaibeh H, Alzahrani AI, Alwadain A, Heming J, Ezugwu AE, Abualigah L. A review of artificial intelligence methods in bladder cancer: segmentation, classification, and detection. Artif Intell Rev. 2024. https://doi.org/10.1007/ s10462-024-10953-6.
- Ren Q, Zhang P, Lin H, Feng Y, Chi H, Zhang X, Xia Z, Cai H, Yu Y. A novel signature predicts prognosis and immunotherapy in lung adenocarcinoma based on cancer-associated fibroblasts. Front Immunol. 2023;14:1201573.
- Zhang H, Wang Y, Wang K, Ding Y, Li X, Zhao S, Jia X, Sun D. Prognostic analysis of lung adenocarcinoma based on cancer-associated fibroblasts genes using scRNA-sequencing. Aging (Albany NY). 2023;15:6774–97.
- 42. Zhang J, Liu X, Huang Z, Wu C, Zhang F, Han A, Stalin A, Lu S, Guo S, Huang J, et al. T cell-related prognostic risk model and tumor immune environment modulation in lung adenocarcinoma based on single-cell and bulk RNA sequencing. Comput Biol Med. 2023;152: 106460.
- Zhang P, Liu J, Pei S, Wu D, Xie J, Liu J, Li J. Mast cell marker gene signature: prognosis and immunotherapy response prediction in lung adenocarcinoma through integrated scRNA-seq and bulk RNA-seq. Front Immunol. 2023;14:1189520.
- 44. Liu H, Wang L, Shi X, Yin L, Zhai W, Gao S, Chen Y, Zhang T. Calcium saccharate/DUSP6 suppresses renal cell carcinoma glycolytic metabolism and boosts sunitinib efficacy via the ERK-AKT pathway. Biochem Pharmacol. 2024;224: 116247.
- Xiong Y, Feng Y, Zhao J, Lei J, Qiao T, Zhou Y, Lu Q, Jiang T, Jia L, Han Y. TFAP2A potentiates lung adenocarcinoma metastasis by a novel miR-16 family/TFAP2A/PSG9/TGF-beta signaling pathway. Cell Death Dis. 2021;12:352.
- Lu L, Lv Y, Dong J, Hu S, Peng R. DRG1 is a potential oncogene in lung adenocarcinoma and promotes tumor progression via spindle checkpoint signaling regulation. Oncotarget. 2016;7:72795–806.
- Fan J, Yang Y, Qian J, Zhang X, Ji J, Zhang L, Li S, Yuan F. Aberrant expression of PAFAH1B3 affects proliferation and apoptosis in osteosarcoma. Front Oncol. 2021;11: 664478.
- Gao L, Bai Y, Zhou J, Liang C, Dong Y, Han T, Liu Y, Guo J, Wu J, Hu D. S100P facilitates LUAD progression via PKA/c-Jun-mediated tumor-associated macrophage recruitment and polarization. Cell Signal. 2024;120: 111179.
- Tang Q, Chen J, Di Z, Yuan W, Zhou Z, Liu Z, Han S, Liu Y, Ying G, Shu X, Di M. TM4SF1 promotes EMT and cancer stemness via the Wnt/ beta-catenin/SOX2 pathway in colorectal cancer. J Exp Clin Cancer Res. 2020;39:232.
- Iwai A, Hijikata M, Hishiki T, Isono O, Chiba T, Shimotohno K. Coiledcoil domain containing 85B suppresses the beta-catenin activity in a p53-dependent manner. Oncogene. 2008;27:1520–6.

- Sun L, Liu Z, Wu Z, Wu Z, Qiu B, Liu S, Hu J, Yin X. PSMD11 promotes the proliferation of hepatocellular carcinoma by regulating the ubiquitination degradation of CDK4. Cell Signal. 2024;121: 111279.
- Xie P, Yuan F, Huang M, Zhang W, Zhou H, Li X, Liu Z. DCBLD2 affects the development of colorectal cancer via EMT and angiogenesis and modulates 5-FU drug resistance. Front Cell Dev Biol. 2021;9: 669285.
- Yu S, Yu T, Wang Y, Sun A, Liu J, Lu K. CCT6A facilitates lung adenocarcinoma progression and glycolysis via STAT1/HK2 axis. J Transl Med. 2024;22:460.
- 54. Sarvaria A, Madrigal JA, Saudemont A. B cell regulation in cancer and anti-tumor immunity. Cell Mol Immunol. 2017;14:662–74.
- Huang L, Wang Z, Chang Y, Wang K, Kang X, Huang R, Zhang Y, Chen J, Zeng F, Wu F, et al. EFEMP2 indicates assembly of M0 macrophage and more malignant phenotypes of glioma. Aging (Albany NY). 2020;12:8397–412.
- 56. Komi DEA, Redegeld FA. Role of mast cells in shaping the tumor microenvironment. Clin Rev Allergy Immunol. 2020;58:313–25.
- Abdulrahman Z, Santegoets SJ, Sturm G, Charoentong P, Ijsselsteijn ME, Somarakis A, Hollt T, Finotello F, Trajanoski Z, van Egmond SL, et al. Tumorspecific T cells support chemokine-driven spatial organization of intratumoral immune microaggregates needed for long survival. J Immunother Cancer. 2022. https://doi.org/10.1136/jitc-2021-004346.
- 58. Groom JR. Regulators of T-cell fate: Integration of cell migration, differentiation and function. Immunol Rev. 2019;289:101–14.
- Iturbide A, Pascual-Reguant L, Fargas L, Cebria JP, Alsina B, Garcia De Herreros A, Peiro S. LOXL2 oxidizes methylated TAF10 and controls TFIID-dependent genes during neural progenitor differentiation. Mol Cell. 2015;58:755–66.
- Xiong Y, Wang L, Xu S, Fu B, Che Y, Zaky MY, Tian R, Yao R, Guo D, Sha Z, et al. Small molecule Z363 co-regulates TAF10 and MYC via the E3 ligase TRIP12 to suppress tumour growth. Clin Transl Med. 2023;13: e1153.
- Zhao X, Lu J, Wu W, Li J. METTL14 inhibits the malignant processes of gastric cancer cells by promoting N6-methyladenosine (m6A) methylation of TAF10. Heliyon. 2024;10: e32014.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.