



Single-Cell RNA Sequencing Reveals Cellular Heterogeneity and Differentiation Dynamics in Lung Adenocarcinoma Microenvironment

Fang Chen^{1,2#}, Zengwen Ma^{1,2#}, Lu Tang³, Weiwei Qian^{1,2*} and Xiaodong Du^{1,2*}

¹Department of Emergency Medicine, Sichuan University, Chengdu, Sichuan, 610041, China

²Department of Emergency Medicine, Shangjinnanfu Hospital, Sichuan University, Chengdu, Sichuan, 611730, China

³Department of Respiratory and Critical Care Medicine, Shangjinnanfu Hospital, West China Hospital, Sichuan University, Chengdu, Sichuan, 611730, China

KEYWORDS Cell Plasticity. Gene Expression Profiling. Immune Landscape. Marker Genes. Pseudotime. Transcriptome

ABSTRACT Lung adenocarcinoma contains diverse malignant, immune, and stromal cell populations that may influence tumour progression and therapeutic response. This study aimed to characterise cellular composition, transcriptional heterogeneity, and differentiation-related patterns in the lung adenocarcinoma microenvironment. The researchers performed an integrative single-cell RNA sequencing analysis using quality control, normalisation, principal component analysis, dimensionality reduction, clustering, cell annotation, marker gene identification, and pseudotemporal trajectory inference. Twelve major cell populations were identified, spanning immune, stromal, endothelial, epithelial, epithelial-like, tuft-like, and proliferating compartments. Principal component analysis suggested that cellular heterogeneity was mainly associated with epithelial-immune differences, extracellular matrix features, T-cell activation, and proliferation-related programs. Pseudotemporal analysis suggested branched cellular-state trajectories consistent with potential differentiation- or activation-related transitions. Marker genes, including *Sftpc*, *Sftpa1*, *Sftpb*, *Cxcl15*, *Mgp*, and *Slc34a2*, helped define distinct cell populations. These findings provide a single-cell transcriptomic overview of cellular diversity and possible differentiation dynamics in lung adenocarcinoma.

INTRODUCTION

Lung adenocarcinoma is the most common histological subtype of non-small cell lung cancer, accounting for approximately 40 percent of all lung cancer cases and remaining an important contributor to cancer-related mortality worldwide (Kratzer et al. 2024). Despite advances in targeted therapies and immunotherapies, many patients with advanced lung adenocarcinoma still experience disease progression, treatment resistance, or relapse, largely due to heterogeneity and complex interactions between malignant cells and the surrounding tumour microenvironment (Hu et al. 2023; Wang et al. 2023).

The tumour microenvironment in lung adenocarcinoma includes multiple cellular popula-

tions, including malignant epithelial cells, stromal fibroblasts, endothelial cells, and various immune cell infiltrates such as T cells, B cells, macrophages, neutrophils, monocytes, and myeloid cells (Bischoff et al. 2021; Sabit et al. 2025). This cellular diversity may contribute to tumour growth, immune escape, extracellular matrix remodelling, angiogenesis, and therapeutic response. However, traditional bulk RNA sequencing approaches provide only averaged gene expression profiles across mixed cell populations, which may obscure the distinct molecular signatures and functional states of individual cell types within the tumour (Liu et al. 2024; Tirosh and Suva 2024).

The emergence of single-cell RNA sequencing (scRNA-seq) technology provided a powerful approach for examining cellular composition and gene-expression heterogeneity at single-cell resolution, enabling the identification of rare cell populations, characterisation of cellular state transitions, and reconstruction of developmental trajectories within complex tissues (Kim et al. 2020; Renaut et al. 2024). Recent single-cell stud-

[#] These authors contributed to this work equally as the co-first author.

^{*}Address for correspondence:

Xiaodong Du

E-mail: 18980601189@163.com

Weiwei Qian

E-mail: qiangww21@163.com

ies in lung cancer and other tumour types have revealed previously unrecognised cellular diversity, identified novel therapeutic targets, and provided insights into mechanisms of drug resistance and immune evasion (Barkley et al. 2022; Kinker et al. 2020). In lung cancer specifically, single-cell analyses have begun to uncover the complexity of tumour-immune interactions, stromal cell contributions, and epithelial cell plasticity (Maynard et al. 2020). Nevertheless, these studies also indicate that the cellular organisation of lung adenocarcinoma is highly context dependent and may vary across patients, tumour regions, and analytical strategies.

Despite these advances, several aspects of cellular heterogeneity and differentiation-related changes in lung adenocarcinoma remain insufficiently characterised. Most previous studies have focused on specific cell types or limited patient cohorts, and the potential state-transition relationships between different cellular populations within the tumour microenvironment are poorly understood (Li et al. 2024; Wang et al. 2024). Furthermore, the molecular mechanisms underlying cellular state transitions and the identification of key regulatory genes driving tumour progression require further investigation using integrated single-cell analytical approaches. Recent single-cell studies further indicate that lung adenocarcinoma progression is accompanied by stage-dependent changes in tumour-cell states and microenvironmental composition, highlighting the need to interpret cellular heterogeneity within a dynamic disease context rather than as a collection of fixed cell populations (Qing et al. 2026).

Pseudotemporal trajectory analysis has emerged as a useful computational approach for inferring developmental relationships and differentiation dynamics from single-cell transcriptomic data (Cheng et al. 2023; Hou et al. 2023). By ordering cells along continuous trajectories based on gene expression similarities, this method can reveal possible cellular state transitions, identify branch points in differentiation programs, and uncover dynamic gene-expression changes associated with disease-related cellular states. However, pseudotemporal analysis is inferential and cannot directly prove true lineage relationships without experimental validation. Therefore, its results should be interpreted as hypotheses

regarding potential differentiation-related processes.

Objectives

The objective of this study was to characterise cellular heterogeneity and differentiation-related patterns in the lung adenocarcinoma microenvironment using single-cell transcriptomic analysis. Specifically, the researchers aimed to identify major cell populations, determine representative marker genes, describe transcriptional differences among cell compartments, and infer potential pseudotemporal trajectories associated with cellular state transitions.

METHODOLOGY

Data Source and Acquisition

This study used single-cell RNA sequencing data from the Gene Expression Omnibus (GEO) accession number GSM8633704, which is included in the GSE281964 dataset. The data consisted of lung tumor tissue from a genetically engineered mouse lung adenocarcinoma model (*Mus musculus*). Based on the original records, the mouse lung tumor tissue was dissociated into single-cell suspensions, and single-cell RNA sequencing was performed using a 10x Genomics Chromium platform. Processed gene expression matrices, feature files, and barcode files were downloaded from the GEO database for subsequent bioinformatics analysis.

Quality Control and Preprocessing

Quality control was performed using Seurat based on the number of detected genes per cell (nFeature_RNA), total unique molecular identifier counts per cell (nCount_RNA), and the proportion of mitochondrial gene expression (percent.mt). Cells were retained if they expressed between 200 and 6,000 genes, had total UMI counts between 500 and 50,000, and had mitochondrial gene expression below 20 percent. Genes detected in fewer than three cells were excluded from the analysis. Potential doublets were identified and removed using the Doublet-Finder package. The filtered expression matrix was subsequently used for normalisation, di-

mensionality reduction, clustering, cell-type annotation, and trajectory analyses.

Data Normalisation and Integration

Count data were log-normalised using the `NormalizeData` function in Seurat with a scale factor of 10,000. Highly variable genes were identified using the `FindVariableFeatures` function with the “vst” method, selecting the top 2,000 most variable genes for downstream analysis. For multi-sample integration, the researchers employed the Seurat integration workflow using canonical correlation analysis (CCA) to correct for batch effects while preserving biological variation. Data scaling was performed using the `ScaleData` function, regressing out unwanted sources of variation including total UMI counts, mitochondrial gene percentage, and cell cycle scores calculated using Seurat’s `CellCycleScoring` function with established S and G2/M phase marker genes.

Dimensionality Reduction and Principal Component Analysis

Principal component analysis (PCA) was performed on the scaled data using the top 2,000 highly variable genes. The optimal number of principal components was determined using multiple approaches, including elbow plot analysis of standard deviations, JackStraw procedure for statistical significance testing, and visual inspection of PC loadings. Based on these analyses, the first 20 principal components were selected for downstream analysis. To validate the significance of principal components, the researchers performed permutation testing and calculated p-values for each component. Components with p-values < 0.001 were considered statistically significant and retained for further analysis.

Clustering Analysis and Cell Type Identification

Unsupervised clustering was performed using the Seurat `FindNeighbors` and `FindClusters` functions. A shared nearest neighbour (SNN) graph was constructed using the selected principal components, followed by Louvain clustering with varying resolution parameters (0.1–1.0) to identify optimal cluster granularity. The final

clustering resolution was selected based on silhouette analysis and biological interpretability. For visualisation, uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbour embedding (t-SNE) were applied to the principal component space using default parameters. These low-dimensional embeddings were used to visualise cellular relationships and cluster assignments. Cell type annotation was performed through a multi-step approach including identification of cluster-specific marker genes using the `FindAllMarkers` function with Wilcoxon rank-sum test, comparison with established cell type signature databases, and manual curation based on known cell type-specific marker genes. Major cell types were identified including epithelial cells, immune cells (T cells, B cells, macrophages, neutrophils, myeloid cells), stromal cells, endothelial cells, and proliferating cells.

Trajectory Inference and Pseudotime Analysis

Pseudotemporal trajectory analysis was conducted using Monocle 3 to infer developmental relationships and differentiation dynamics among cellular populations. The Seurat object was converted to a `CellDataSet` object compatible with Monocle 3. Cells were re-clustered within Monocle 3 to define partitions and trajectories. Principal graph learning was performed using the ‘`learn_graph`’ function, and root nodes were selected based on biological knowledge and expression of early-stage marker genes. Pseudotime values were calculated using the ‘`order_cells`’ function, representing the relative position of each cell along the inferred trajectory. Multiple trajectory paths were identified and analysed separately for branching events. To identify genes with dynamic expression patterns along pseudotime, the researchers performed differential expression analysis using the ‘`graph_test`’ function in Monocle 3. Genes with q-value < 0.05 and significant association with pseudotime progression were considered dynamically expressed.

Differential Gene Expression and Marker Gene Analysis

Cluster-specific marker genes were identified using the `FindAllMarkers` function in Seurat

with parameters including log fold-change threshold ≥ 0.25 , minimum percentage of cells expressing the gene ≥ 0.1 , and adjusted p-value < 0.05 using Bonferroni correction. The Wilcoxon rank-sum test was used as the statistical method for differential expression analysis. For visualisation of gene expression patterns, the researchers generated feature plots, violin plots, and heatmaps using Seurat and custom R scripts. Expression levels were displayed on UMAP and t-SNE embeddings to illustrate spatial distribution patterns of key marker genes across cellular populations.

Statistical Analysis and Visualisation

All statistical analyses were performed in R (version 4.2.0) using Seurat (version 4.3.0), Monocle 3 (version 1.3.1), and other Bioconductor packages. Data visualisation was conducted using ggplot2, ComplexHeatmap, and custom plotting functions. Correlation analyses were performed using Pearson correlation coefficients, and statistical significance was assessed using appropriate tests as specified for each analysis. Results are presented as mean $\pm$ standard deviation unless otherwise specified. Statistical significance was defined as $p < 0.05$ for all analyses.

RESULTS

Single-cell RNA Sequencing Quality Control Analysis of Lung Adenocarcinoma

Quality-control analysis showed that the single-cell RNA sequencing dataset had an overall acceptable distribution of transcriptomic features. The number of detected genes per cell was mainly concentrated between approximately 1,000 and 5,000, with the highest density around 2,500 genes per cell, suggesting stable gene capture across most cells (Fig. 1A). Total RNA counts showed a right-skewed distribution, with most cells located in the lower-to-moderate count range and only a small proportion showing very high transcript counts (Fig. 1B). The proportion of mitochondrial gene expression was generally low, with most cells showing mitochondrial proportions below 15 percent, indicating that most retained cells were unlikely to

be severely damaged or dying (Fig. 1C). Correlation analysis further supported the quality-control pattern. RNA counts showed a weak negative correlation with mitochondrial gene proportion, whereas RNA counts were strongly positively correlated with the number of detected genes (Fig. 1D). These observations indicated that the dataset was suitable for subsequent normalisation, dimensionality reduction, clustering, and cell-type annotation.

Principal Component Analysis and Key Driver Genes in Lung Adenocarcinoma Single-cell RNA Sequencing Data

Principal component analysis revealed distinct transcriptional programmes across the first six principal components (Fig. 2). The first principal component (PC_1) is primarily driven by genes including S100a4, Cd3, and Cd9, demonstrating clear cellular population separation that likely reflects the fundamental distinction between epithelial and immune cell compartments. This primary axis of variation captures the most significant transcriptional differences within the dataset. PC_2 is dominated by extracellular matrix-related genes, particularly collagen family members (Col5a2, Col1a1, Col3a1) and matrix-associated genes such as Bgn and Prkg1. This component clearly identifies stromal cells or fibroblasts, highlighting the importance of the tumour microenvironment's structural components in defining cellular heterogeneity. The third through sixth principal components reveal increasingly specific cellular features and functional states. PC_3 encompasses diverse functional genes including Egfr7, Shank3, and Sox18, potentially representing distinct cellular subtypes or activation states. PC_4 is characterised by S100a14, gastrin-related genes (Gkn2), and cell junction proteins (Dsg2), suggesting associations with epithelial cell differentiation status. PC_5 demonstrates a clear enrichment of immune-related genes, particularly T-cell markers including Itk, Cd3d, and Ctla4, definitively identifying T-cell populations and their activation states. Finally, PC_6 is dominated by cell cycle-related genes such as Top2a, Cenpa2, Cenpe, and Tpx2, reflecting differences in cellular proliferative activity across the dataset.

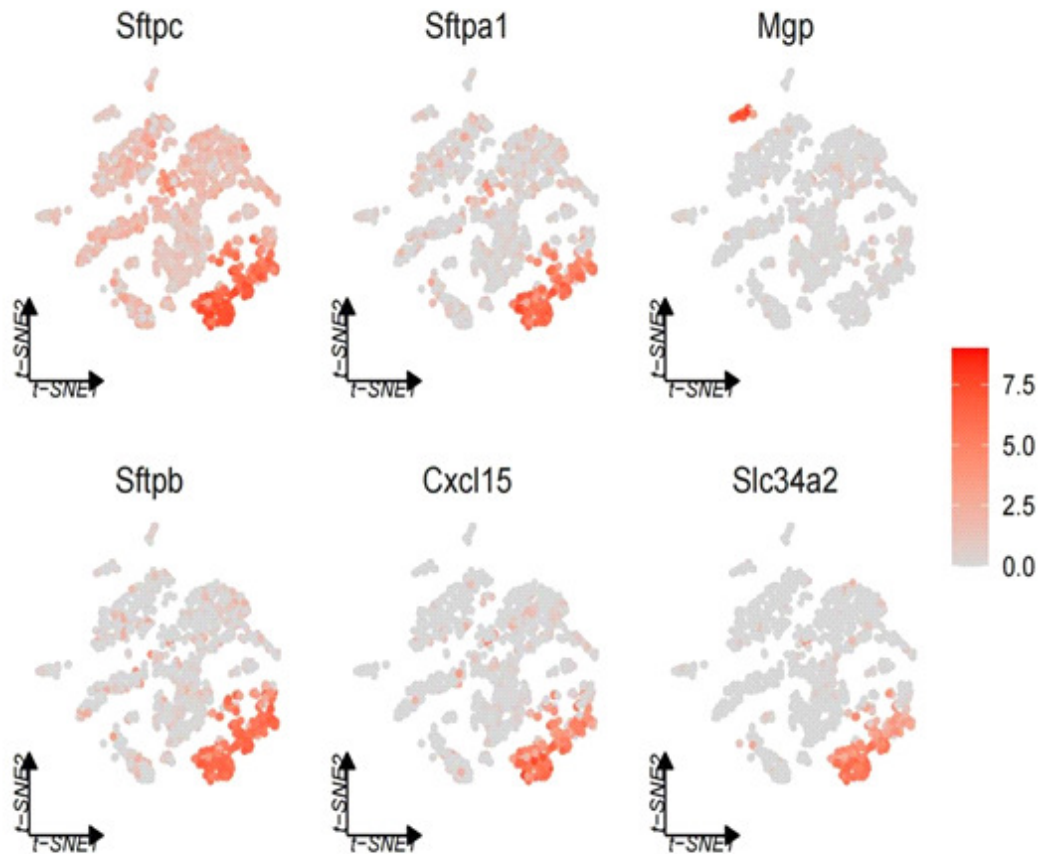


Fig. 1. Quality control metrics for single-cell RNA sequencing data from lung adenocarcinoma samples (A) Distribution of genes detected per cell (nFeature_RNA). (B) Distribution of total RNA counts per cell (nCount_RNA). (C) Distribution of mitochondrial gene expression percentage (percent.mt). (D) Correlation analysis between quality metrics: RNA counts vs. mitochondrial percentage (left, $r=-0.17$) and RNA counts vs. detected genes (right, $r=0.92$). All cells are coloured by tumour identity

Single-cell RNA Sequencing Data Analysis and Cell Type Identification in Lung Adenocarcinoma

Dimensionality reduction and unsupervised clustering resolved transcriptionally distinct cell populations in the lung adenocarcinoma single-cell dataset. Principal component significance testing showed that multiple leading principal components captured non-random transcriptional variation, supporting their use for downstream dimensionality reduction and clustering (Fig.

3A). The standard deviation plot showed a progressive decline across principal components, with the largest reduction occurring among the early components, indicating that the leading components retained the major structure of the dataset (Fig. 3B).

Using these principal components, unsupervised clustering identified 14 transcriptionally distinct clusters. Both uniform manifold approximation and projection and t-distributed stochastic neighbour embedding showed clear separa-

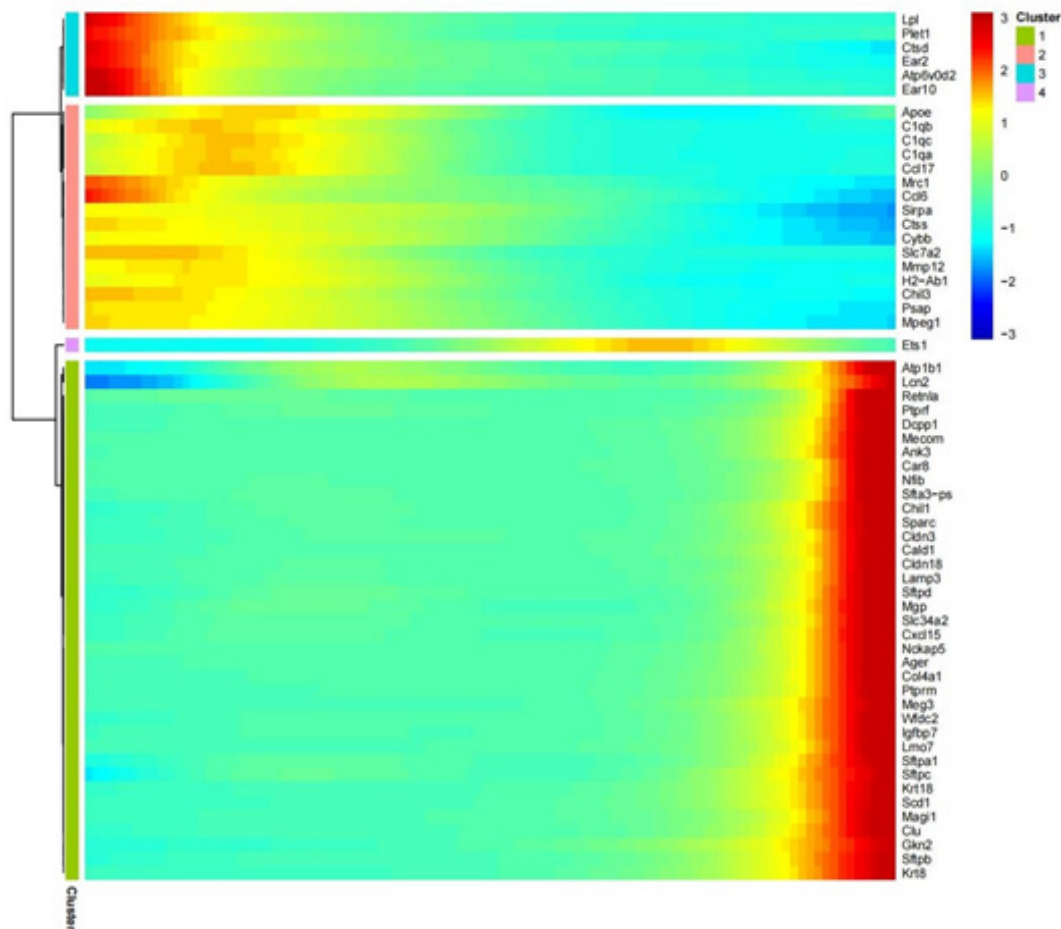


Fig.2. Principal component analysis reveals cellular heterogeneity in lung adenocarcinoma

Heatmaps showing expression patterns of key driver genes across the first six principal components (PC_1 to PC_6). Each component identifies distinct cellular populations: PC_1 (epithelial vs. immune cells), PC_2 (stromal/fibroblast cells), PC_3 (cellular subtypes), PC_4 (epithelial differentiation), PC_5 (T-cell populations), and PC_6 (proliferating cells). Yellow indicates high expression, purple indicates low expression

tion of several clusters, indicating robust transcriptional differences among cell populations (Figs. 3C and 3D). Cell-type annotation based on canonical marker genes further assigned these clusters to major cellular compartments of the lung adenocarcinoma microenvironment. These included T cells, B cells, macrophages, monocytes, neutrophils, myeloid cells, stromal cells, endothelial cells, epithelial cells, epithelial-like cells, tuft-like cells, and proliferating cells (Figs. 3E and 3F). Overall, the clustering and annotation results showed that the lung adeno-

carcinoma microenvironment contained diverse immune, stromal, endothelial, epithelial, and proliferative cell populations, providing the basis for subsequent marker-gene and trajectory analyses.

Pseudotemporal Trajectory Analysis Reveals Cellular Differentiation Dynamics in Lung Adenocarcinoma

Pseudotemporal trajectory analysis was performed to explore potential cellular state transitions within the lung adenocarcinoma microen-

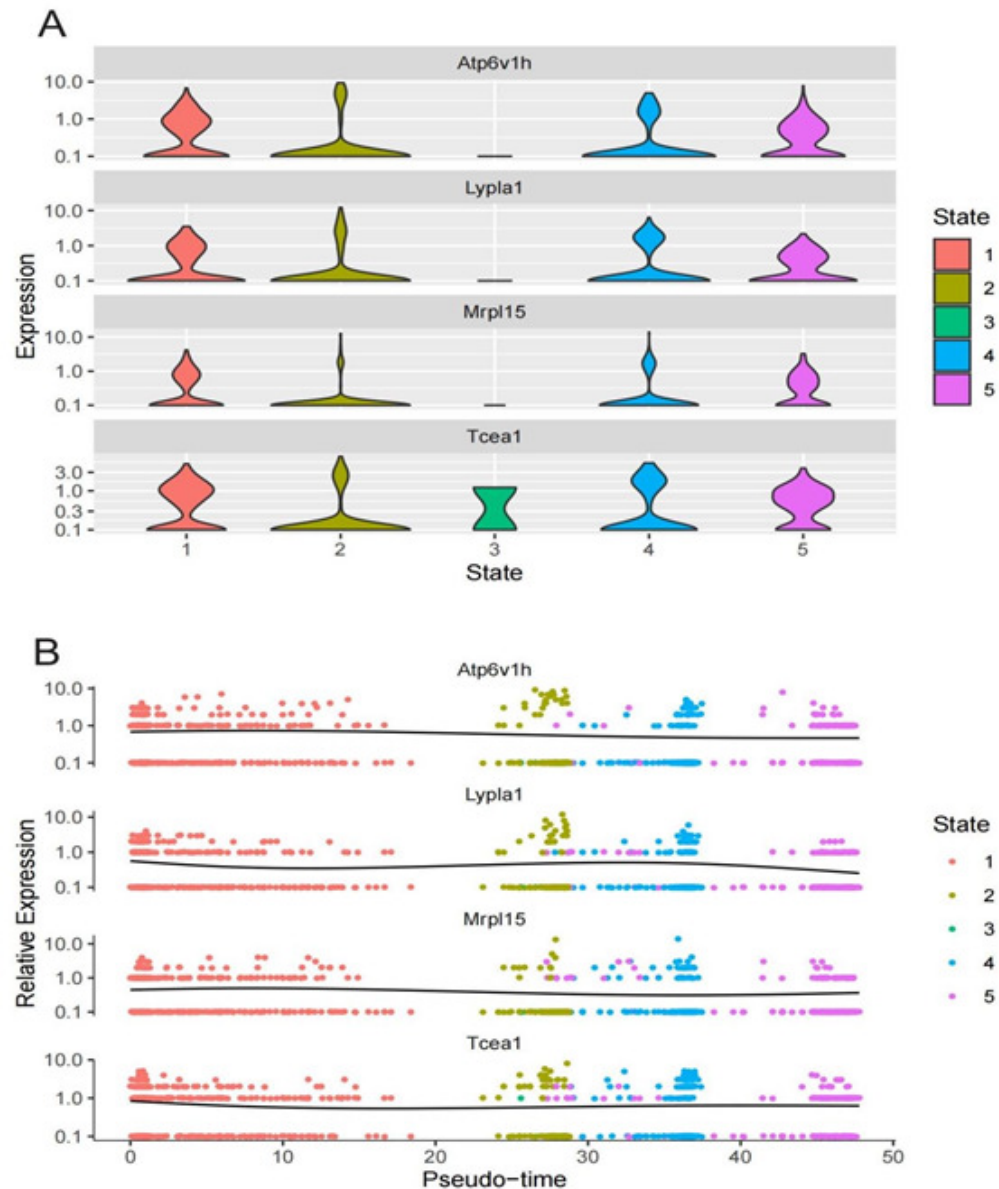


Fig. 3. Single-cell RNA sequencing workflow and cell type identification in lung adenocarcinoma (A) PCA validation showing p-values for the first 15 principal components. (B) Standard deviation plot across principal components showing the elbow pattern. (C-D) UMAP and t-SNE dimensionality reduction plots coloured by computational clusters (0-13). (E-F) The same UMAP and t-SNE plots with biological cell type annotations, identifying 12 major cell populations including immune cells (T cells, B cells, neutrophils, myeloid cells, macrophages, monocytes), structural cells (stromal, endothelial), epithelial cells, and proliferating cells

vironment. Cells were arranged along a branched trajectory, suggesting that the dataset contained multiple transcriptional states rather than a single linear progression. When cells were coloured by annotated cell type, epithelial, stromal, endothelial, immune, and proliferating cell populations showed non-random distributions along different trajectory branches, indicating that these populations occupied distinct transcriptional positions in the inferred cellular-state space (Fig. 4A, left). Pseudotime ordering further showed a gradual transition across the trajectory, with cells distributed from early to later

pseudotime positions along the major branches (Fig. 4A, right).

When the trajectory was coloured by computational state, five states were identified and showed clear spatial separation along the branching structure. This pattern suggested that different transcriptional programs may underlie the observed cellular-state heterogeneity in the dataset (Fig. 4B). The alternative trajectory layout provided a simplified view of the branching relationships among annotated cell populations and further supported the presence of distinct cell-state groups within the tumour microenvi-

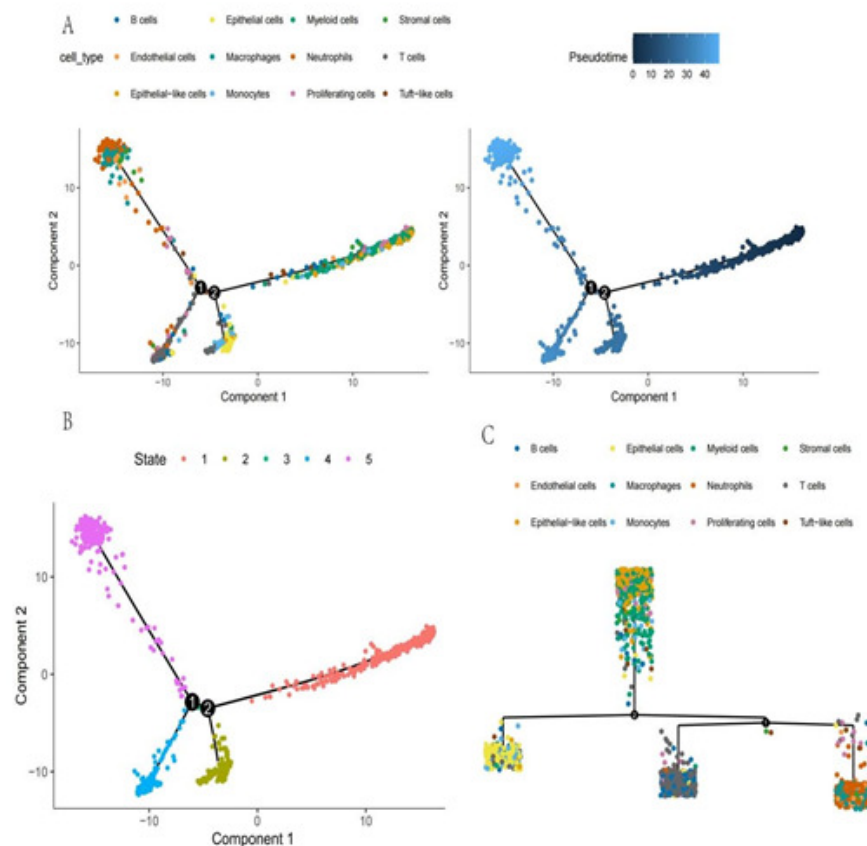


Fig. 4. Pseudotemporal trajectory analysis of lung adenocarcinoma cellular differentiation
(A) Trajectory reconstruction coloured by cell types (left) and pseudotime progression (right, dark to light blue indicates early to late timepoints). (B) Same trajectory coloured by computational states (States 1-5) showing distinct developmental stages. (C) Alternative linearised trajectory visualisation maintaining cell type annotations. Black lines indicate inferred differentiation paths connecting different cellular populations

ronment (Fig. 4C). Overall, the pseudotemporal analysis suggested possible differentiation-related or activation-related transitions among cellular populations, but these inferred trajectories should be interpreted as computational hypotheses rather than direct evidence of lineage progression.

Gene Expression Dynamics During Pseudotemporal Progression in Lung Adenocarcinoma

The expression patterns of selected trajectory-associated genes were further examined across the inferred cellular states and pseudotime.

Violin plots showed that *Atp6v1h*, *Lypla1*, *Mrpl15*, and *Tcea1* displayed state-dependent expression distributions across the five computational states identified in the trajectory analysis (Fig. 5A). *Atp6v1h* and *Lypla1* were mainly detected in States 1, 2, 4 and 5, with limited expression in State 3. *Mrpl15* showed a similar but relatively lower expression pattern across most states. In contrast, *Tcea1* was detectable across all five states and showed a more evident expression distribution in State 3, suggesting that this state may have a distinct transcriptional profile compared with the other trajectory states (Fig. 5A).

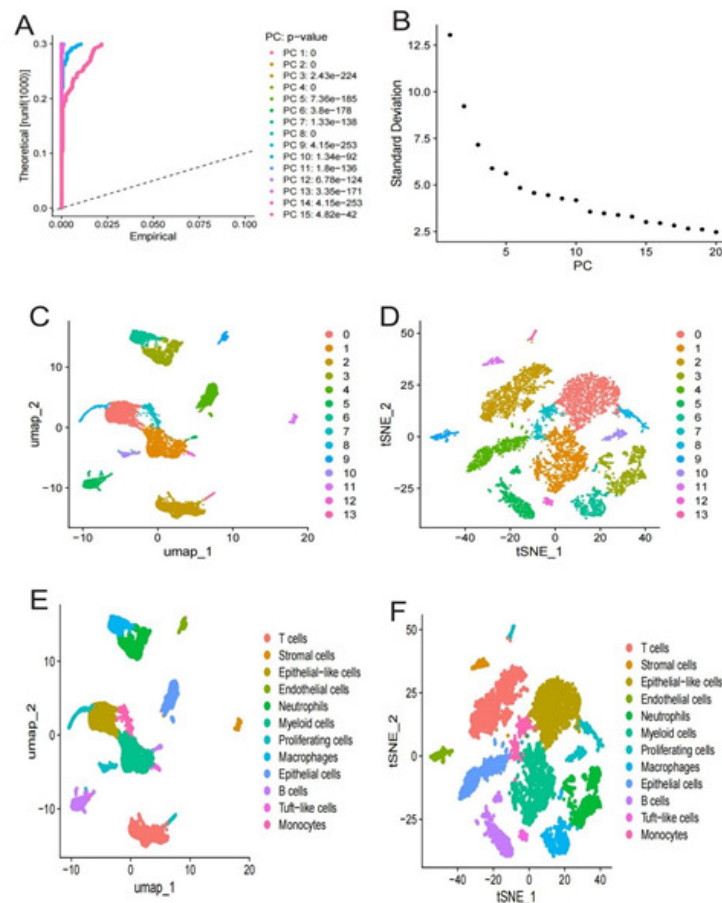


Fig. 5. Gene expression dynamics across pseudotemporal states in lung adenocarcinoma
(A) Violin plots showing expression distribution of four selected genes (*Atp6v1h*, *Lypla1*, *Mrpl15*, *Tcea1*) across computational states 1-5. (B) Expression of the same genes plotted against pseudotime progression, with cells coloured by state and fitted with trend lines (black curves) showing temporal expression dynamics during cellular differentiation

along the inferred cellular trajectory. The heatmap showed that selected genes were grouped into four major expression modules according to their variation across the ordered cells (Fig. 6). One module showed higher expression at the early part of the trajectory, including genes such as *Lpl*, *Plet1*, *Ctsd*, *Ear2*, *Atp6v0d2* and *Ear10*. A second module contained immune and macrophage-associated genes, including *Apoe*, *C1qa*, *C1qb*, *C1qc*, *Mrc1*, *Ccl6*, *Sirpa*, *Cybb*, *Mmp12*, *H2-Ab1*, *Chil3*, and *Mpeg1*, suggesting that immune-related transcriptional programs contributed to the inferred cellular-state changes. In contrast, a larger gene module showed increased expression toward the later part of the trajectory and included epithelial or lung-associated genes such as *Sftpa1*, *Sftpb*, *Sftpc*, *Sftpd*, *Slc34a2*, *Wfdc2*, *Krt8*, *Krt18*, *Cldn18*, *Lamp3*, *Ager*, and *Lcn2*. These patterns indicated that the inferred pseudotemporal process was accompanied by coordinated changes in immune-associated and epithelial-associated transcriptional programs. Overall, the heatmap supported the presence of

Pseudotime-associated gene-expression analysis revealed dynamic transcriptional changes

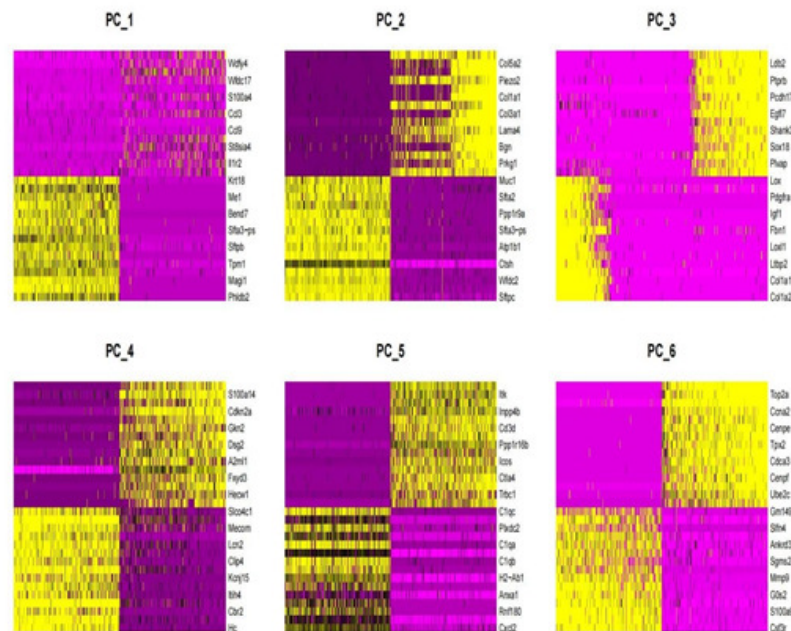


Fig. 6. Cluster-specific marker gene expression heatmap in lung adenocarcinoma
Heatmap showing expression of marker genes across different computational clusters. Each row represents a cluster (coloured bars on left), each column represents a gene. Expression values are scaled from -3 (blue, low) to +3 (red, high). Red blocks indicate cluster-specific highly expressed genes used for cell type identification

distinct gene-expression modules along the trajectory, although these inferred temporal patterns require further experimental validation.

Expression Patterns of Cell Type-Specific Marker Genes on the t-SNE Embedding

Representative marker-gene expression was further visualised on the t-distributed stochastic neighbour embedding map to support cell-type annotation and characterise transcriptional heterogeneity in the lung adenocarcinoma single-cell dataset. Surfactant-associated genes, including *Sftpc*, *Sftpa1* and *Sftpb*, showed enriched expression in a distinct cell cluster located in the lower-right region of the embedding, supporting the annotation of an alveolar epithelial-like population (Fig. 7).

Slc34a2 showed a partially overlapping expression pattern with these surfactant genes, consistent with its association with lung epithelial identity and ion-transport-related epithelial function. In contrast, *Mgp* was mainly expressed in a separate small cluster, suggesting a transcriptionally distinct mesenchymal or stromal-associated population. *Cxcl15* showed a more restricted and lower-intensity expression pattern, indicating potential inflammatory or chemokine-related activity in a subset of cells. Overall, these marker-gene distributions supported the separation of epithelial, stromal-associated, and inflammation-related cell states and provided additional molecular evidence for the cell annotations derived from unsupervised clustering (Fig. 7).

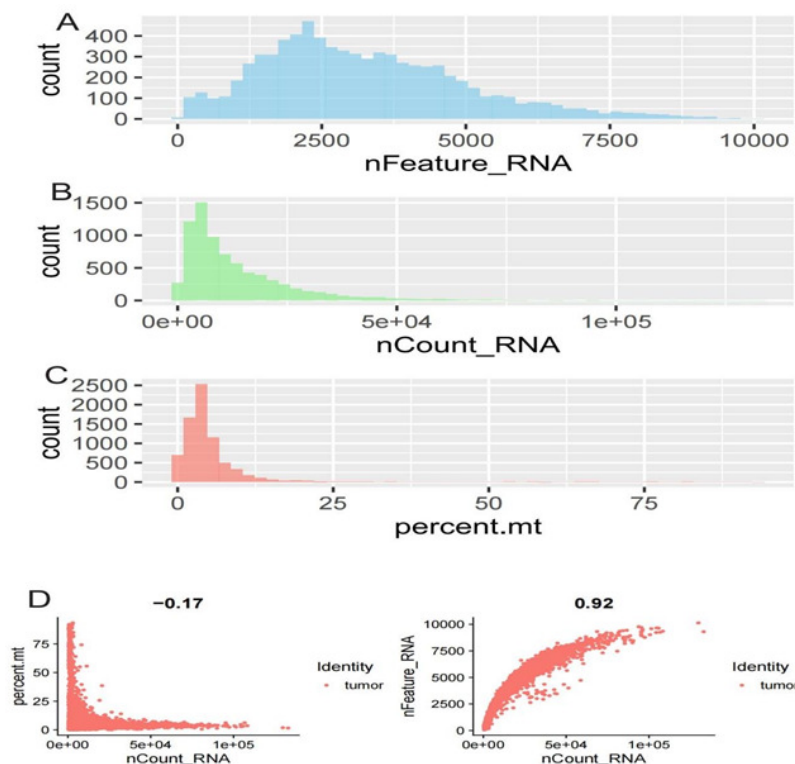


Fig. 7. Expression Patterns of Marker Genes on the t-SNE Embedding in Lung Adenocarcinoma Single-Cell Data

t-SNE plots showing expression intensity of six marker genes: surfactant proteins (*Sftpc*, *Sftpa1*, *Sftpb*) identifying epithelial cells, matrix protein (*Mgp*) marking stromal cells, chemokine (*Cxcl15*) indicating immune signaling, and transporter (*Slc34a2*) representing specialised epithelial cells. Expression levels range from 0 (gray) to 7.5 (red)

DISCUSSION

This study used single-cell transcriptomic analysis to characterise cellular heterogeneity and potential differentiation-related patterns in the lung adenocarcinoma microenvironment. The main findings were that the dataset contained diverse epithelial, immune, stromal, endothelial, and proliferating cell populations, that major transcriptional variation was associated with epithelial-immune differences, extracellular matrix-related programs, immune activation, and cell-cycle activity, and that pseudotemporal analysis suggested branched cellular-state patterns accompanied by state-associated gene-expression changes. These observations are consistent with the current understanding that lung adenocarcinoma is not a uniform epithelial malignancy, but a heterogeneous tumour ecosystem shaped by interactions among malignant cells, immune cells, stromal cells, and vascular components (Bischoff et al. 2021; Bray et al. 2024; He et al. 2021; Thai et al. 2021; Wu et al. 2021).

The identification of multiple immune, epithelial, stromal, endothelial, and proliferating compartments supports previous single-cell studies showing that lung adenocarcinoma contains substantial intercellular and intratumoral heterogeneity (Bischoff et al. 2021; He et al. 2021; Wu et al. 2021). Wu et al. (2021) reported extensive tumour and microenvironmental heterogeneity in advanced non-small cell lung cancer, including diverse malignant epithelial states and immune-cell subsets. Similarly, He et al. (2021) showed that early-stage epidermal growth factor receptor-mutant lung adenocarcinomas contained heterogeneous tumour and immune-cell populations, while Bischoff et al. (2021) demonstrated distinct tumour microenvironmental patterns across lung adenocarcinoma specimens. In this context, the researchers' clustering and cell-type annotation results are broadly consistent with published single-cell evidence. However, the present findings should be interpreted as a transcriptomic classification of cellular states rather than a definitive histological or functional validation of each cell population.

The principal component and dimensionality-reduction analyses suggested that the major sources of variation in the dataset were related to epithelial identity, immune-cell composition,

stromal and extracellular matrix features, inflammatory programs, and proliferative activity. This pattern is biologically plausible because lung adenocarcinoma heterogeneity is shaped not only by malignant epithelial programs, but also by immune suppression, stromal remodelling, and treatment-associated or cell-cycle-associated transcriptional states (Hu et al. 2023; Maynard et al. 2020; Yang et al. 2022). Recent work has emphasised that recurring cancer-cell states and non-malignant microenvironmental programs can be shared across tumours and may contribute to phenotypic diversity, immune escape, and therapeutic resistance (Kinker et al. 2020; Tirosh and Suva 2024). Therefore, the principal component results are most appropriately viewed as a technical and biological summary of major transcriptional axes in the dataset, rather than as independent evidence of discrete biological mechanisms.

The marker-gene analyses further supported the annotation of epithelial-like and stromal-associated populations. Surfactant-associated genes, including *Sftpc*, *Sftpa1*, and *Sftpb*, together with *Slc34a2*, were enriched in a subset of cells with epithelial features. This pattern is compatible with alveolar epithelial or alveolar-like transcriptional identity, which has been repeatedly observed in lung adenocarcinoma and lung epithelial single-cell studies (Bischoff et al. 2021; He et al. 2021; Wu et al. 2021). In contrast, *Mgp* showed a more restricted expression pattern in a separate subset, suggesting a stromal or mesenchymal-associated transcriptional program. *Cxcl15* showed a restricted and lower-intensity expression pattern, which may indicate inflammatory or chemokine-related activity in a subset of cells. These marker-gene patterns provide molecular support for the clustering results, but they should not be overinterpreted as diagnostic biomarkers or therapeutic targets without independent validation using protein-level assays, spatial localisation, or functional experiments.

The pseudotemporal trajectory analysis suggested that cellular states in the dataset were arranged along branched transcriptional trajectories rather than a single linear path. This finding is consistent with the concept that tumour and microenvironmental cells may occupy continuous and plastic cell states rather than fixed categories (Marine et al. 2020; Pérez-González et

al. 2023; Tirosh and Suva 2024). Therapy-induced single-cell studies in lung cancer have shown that malignant and microenvironmental compartments can undergo marked transcriptional adaptation during treatment (Maynard et al. 2020), and broader cancer studies have linked cellular plasticity and non-genetic state transitions to tumour progression, metastasis, and therapeutic resistance (Marine et al. 2020; Pérez-González et al. 2023). Therefore, the branched patterns observed in this study may reflect potential differentiation-related or activation-related transitions within the tumour microenvironment. Nevertheless, pseudotime analysis is computational and inferential. It orders cells according to transcriptional similarity and does not directly prove true lineage relationships or temporal causality (Cheng et al. 2023; Roux de Bézieux et al. 2024; Saelens et al. 2019). The inferred branches should therefore be considered hypothesis-generating and require further validation by lineage tracing, longitudinal sampling, spatial transcriptomics, or experimental perturbation.

The state-associated expression patterns of *Atp6v1h*, *Lypla1*, *Mrpl15*, and *Tceal* also support the presence of transcriptional differences across inferred trajectory states. However, these genes should be discussed cautiously. Their differential expression across pseudotime suggests association with cellular-state variation, but does not establish that they are key regulators of differentiation. In single-cell transcriptomic studies, state-associated genes may reflect differences in metabolism, transcriptional activity, cell stress, cell cycle, or cell-type composition rather than direct causal regulation (Cheng et al. 2023; Luecken and Theis 2019; Roux de Bézieux et al. 2024; Saelens et al. 2019). Therefore, these genes may serve as candidates for further investigation, but their biological roles in lung adenocarcinoma progression should be tested in independent datasets and experimental models.

The immune-related findings are also relevant to current studies of lung cancer immunobiology. Single-cell studies have shown that epidermal growth factor receptor-mutant lung adenocarcinoma can display an immunosuppressive tumour microenvironment with altered T-cell and myeloid-cell states (Yang et al. 2022). Other studies have demonstrated that neoadju-

vant immunotherapy can reshape immune and stromal compartments in non-small cell lung cancer, and that these changes may differ between responders and non-responders (Hu et al. 2023). In the present study, the identification of T cells, B cells, macrophages, monocytes, neutrophils, and myeloid cells indicates a complex immune landscape. However, because this analysis did not include paired clinical response data, immune receptor sequencing, cytokine assays, or functional immune profiling, the findings cannot directly predict immunotherapy response. Instead, they provide a cellular framework for generating hypotheses about immune-cell involvement in lung adenocarcinoma biology. Recent single-cell and spatial studies have further clarified how immune populations may interact within lung tumours. Macrophage–T-cell crosstalk has been identified as an important component of the immunosuppressive microenvironment in non-small cell lung cancer, including spatially organised interactions involving SPP1–CD44, NECTIN2–TIGIT, and HLA-E–CD8B (Zhu et al. 2025). A myeloid-associated immune-defence programme enriched at tumour borders has also been reported in primary and brain-metastatic lung adenocarcinoma (Meng et al. 2025). Together, these findings suggest that immune-cell states in lung cancer should be interpreted not only according to their transcriptional identities but also in the context of intercellular communication and spatial organisation.

Recent spatial and single-cell studies have further emphasised that transcriptional identity alone is insufficient to fully understand the lung tumour microenvironment (De Zuani et al. 2024; Sorin et al. 2023; Takano et al. 2024). Sorin et al. (2023) showed that single-cell phenotypes in lung adenocarcinoma are organised within spatial immune landscapes. De Zuani et al. (2024) combined single-cell and spatial transcriptomics to analyse non-small cell lung cancer and demonstrated the value of integrating cellular identity with tissue context. Takano et al. (2024) further showed that spatially resolved gene-expression profiling can identify key microenvironmental changes during lung adenocarcinoma development. Compared with these studies, the present analysis provides transcriptomic evidence for cellular heterogeneity and inferred state transitions, but it does not directly resolve

spatial relationships among tumour, immune, and stromal populations. Future work integrating single-cell RNA sequencing with spatial transcriptomics, multiplex immunofluorescence, or imaging mass cytometry would help determine whether the inferred cellular states occupy biologically meaningful tissue niches.

The findings also have methodological implications. The quality-control results indicated that the dataset was suitable for downstream analysis, and the use of normalisation, dimensionality reduction, clustering, marker-gene visualisation, and pseudotemporal inference followed common single-cell analytical principles (Luecken and Theis 2019; Saelens et al. 2019). However, single-cell analysis is sensitive to sample quality, dissociation effects, doublets, batch effects, clustering resolution, and marker-gene selection (Luecken and Theis 2019). These technical factors can influence cell-type annotation and trajectory structure. Therefore, future studies should report preprocessing parameters transparently, validate cell annotations using multiple marker sets or reference atlases, and test whether the identified cellular states are reproducible across independent cohorts.

Overall, this study supports the view that lung adenocarcinoma contains a heterogeneous and dynamically organised cellular microenvironment. The results are consistent with recent single-cell and spatial studies showing that lung tumours comprise diverse epithelial, stromal, endothelial, immune, and proliferating cell states. The main value of the present analysis lies in providing an integrated transcriptomic overview of cellular composition, marker-gene expression, and inferred pseudotemporal patterns. At the same time, the conclusions should remain cautious. The results describe associations and computationally inferred cellular-state relationships, not direct causal mechanisms. Further validation in larger datasets, spatially resolved tissue sections, and functional experimental systems will be necessary to determine whether the identified cell populations and trajectory-associated genes contribute directly to lung adenocarcinoma progression or therapeutic response.

CONCLUSION

This comprehensive single-cell RNA sequencing analysis has provided unprecedented

insights into the cellular architecture and dynamic processes underlying lung adenocarcinoma pathogenesis. Through the identification of 12 distinct cellular populations, characterisation of pseudotemporal differentiation trajectories, and discovery of key molecular signatures, the study advances fundamental understanding of tumour microenvironment complexity and cellular plasticity in lung cancer.

RECOMMENDATIONS

Further studies using larger independent cohorts, spatial transcriptomics, protein-level validation, and functional experiments are recommended to confirm the identified cellular populations and trajectory-associated genes. Future work should also evaluate whether these cellular states are related to lung adenocarcinoma progression, treatment resistance, and therapeutic response.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations of the study warrant consideration. First, the cross-sectional nature of single-cell RNA sequencing provides a snapshot of cellular states rather than direct temporal dynamics. Future studies incorporating time-series sampling or lineage tracing approaches could provide more definitive insights into cellular differentiation trajectories. Second, while the analysis identified diverse cellular populations, functional validation of identified cell states and their roles in tumour progression requires additional experimental approaches, including in vitro and in vivo functional studies. The spatial organisation of identified cellular populations within the tumour tissue represents another important dimension that was not directly addressed in the analysis. Integration of single-cell RNA sequencing with spatial transcriptomics approaches could provide crucial insights into how cellular interactions and spatial organisation influence tumour behaviour. Additionally, extending this analysis to larger patient cohorts with detailed clinical annotation would enable more robust associations between cellular composition and clinical outcomes.

DECLARATIONS

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Conceptualisation: Fang Chen

Investigation: Fang Chen, Zengwen Ma, Lu Tang, Xiaodong Du, and Weiwei Qian.

Writing - Original Draft preparation: Fang Chen

Writing - Review and Editing: Xiaodong Du, Weiwei Qian

All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

ARTIFICIAL INTELLIGENCE (AI) DISCLOSURE

No artificial intelligence (AI) tools, including large language models or machine learning software, were used in the preparation, analysis, or presentation of this manuscript.

ACKNOWLEDGEMENT

Not applicable.

AVAILABILITY OF DATA AND INFORMATION

Data and information will be available on reasonable request.

ETHICAL APPROVAL AND CONSENT FOR PARTICIPATION

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

REFERENCES

- Barkley D, Moncada R, Pour M, et al. 2022. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat Genet*, 54: 1192-1201. 10.1038/s41588-022-01141-9
- Bergen V, Lange M, Peidli S, et al. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*, 38: 1408-1414. 10.1038/s41587-020-0591-3
- Bischoff P, Trinks A, Obermayer B, et al. 2021. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene*, 40: 6748-6758. 10.1038/s41388-021-02054-3
- Bray F, Laversanne M, Sung H, et al. 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 74: 229-263. 10.3322/caac.21834
- Cheng J, Smyth GK, Chen Y 2023. Unraveling the timeline of gene expression: A pseudotemporal trajectory analysis of single-cell RNA sequencing data. *F1000Res*, 12: 684. 10.12688/f1000research.134078.2
- De Zuani M, Xue H, Park JS, et al. 2024. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat Commun*, 15: 4388. 10.1038/s41467-024-48700-8
- He D, Wang D, Lu P, et al. 2021. Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene*, 40: 355-368. 10.1038/s41388-020-01528-0
- Hou W, Ji Z, Chen Z, et al. 2023. A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. *Nat Commun*, 14: 7286. 10.1038/s41467-023-42841-y
- Hu J, Zhang, L, Xia, H, et al. 2023. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome Med*, 15: 14. 10.1186/s13073-023-01164-9
- Kim N, Kim HK, Lee K, et al. 2020. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun*, 11: 2285. 10.1038/s41467-020-16164-1
- Kinker GS, Greenwald AC, Tal R, et al. 2020. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet*, 52: 1208-1218. 10.1038/s41588-020-00726-6
- Kratzer TB, Bandi P, Freedman ND, et al. 2024. Lung cancer statistics, 2023. *Cancer*, 130: 1330-1348. 10.1002/cncr.35128
- La Manno G, Soldatov R, Zeisel A, et al. 2018. RNA velocity of single cells. *Nature*, 560: 494-498. 10.1038/s41586-018-0414-6
- Li Y, Jiang M, Aye L, et al. 2024. UPP1 promotes lung adenocarcinoma progression through the induction of an immunosuppressive microenvironment. *Nat Commun*, 15: 1200. 10.1038/s41467-024-45340-w
- Liu W, Puri A, Fu D, et al. 2024. Dissecting the tumor microenvironment in response to immune checkpoint inhibitors via single-cell and spatial transcriptomics. *Clin Exp Metastasis*, 41: 313-332. 10.1007/s10585-023-10246-2
- Luecken MD, Theis FJ 2019. Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol Syst Biol*, 15: e8746. 10.15252/msb.20188746

- Marine JC, Dawson SJ, Dawson MA 2020. Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer*, 20: 743-756. 10.1038/s41568-020-00302-4
- Maynard A, McCoach CE, Rotow JK, et al. 2020. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell*, 182: 1232-1251.e1222. 10.1016/j.cell.2020.07.017
- Meng F, Sun Y, Li J, et al. 2025. Single-cell and spatial transcriptome landscapes reveal the spatial immune defense pattern shared by primary and brain metastatic lung adenocarcinoma. *J Thorac Dis*, 17: 9927-9942. 10.21037/jtd-2025-1224
- Pérez-González A, Bévant K, Blanpain C 2023. Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat Cancer*, 4: 1063-1082. 10.1038/s43018-023-00595-y
- Qing B, Li X, He X, et al. 2026. Characterizing stage-specific cellular dynamics and microenvironmental remodeling in lung adenocarcinoma by single-cell RNA sequencing. *Adv Sci (Weinh)*, 13: e10847. 10.1002/advs.202510847
- Rao A, Barkley D, França GS, et al. 2021. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596: 211-220. 10.1038/s41586-021-03634-9
- Renaut S, Saavedra Armero V, Boudreau DK, et al. 2024. Single-cell and single-nucleus RNA-sequencing from paired normal-adenocarcinoma lung samples provide both common and discordant biological insights. *PLoS Genet*, 20: e1011301. 10.1371/journal.pgen.1011301
- Roux de Bézieux H, Van den Berge K, Street K, et al. 2024. Trajectory inference across multiple conditions with condiments. *Nat Commun*, 15: 833. 10.1038/s41467-024-44823-0
- Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al. 2020. The Human Tumor Atlas Network: Charting tumor transitions across space and time at single-cell resolution. *Cell*, 181: 236-249. 10.1016/j.cell.2020.03.053
- Sabit H, Arneth B, Pawlik TM, et al. 2025. Leveraging single-cell multi-omics to decode tumor microenvironment diversity and therapeutic resistance. *Pharmaceuticals (Basel)*, 18. 10.3390/ph18010075
- Saelens W, Cannoodt R, Todorov H, et al. 2019. A comparison of single-cell trajectory inference methods. *Nat Biotechnol*, 37: 547-554. 10.1038/s41587-019-0071-9
- Sorin M, Rezanejad M, Karimi E, et al. 2023. Single-cell spatial landscapes of the lung tumour immune microenvironment. *Nature*, 614: 548-554. 10.1038/s41586-022-05672-3
- Takano Y, Suzuki J, Nomura K, et al. 2024. Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat Commun*, 15: 10637. 10.1038/s41467-024-54671-7
- Thai AA, Solomon BJ, Sequist LV, et al. 2021. Lung cancer. *Lancet*, 398: 535-554. 10.1016/s0140-6736(21)00312-3
- Tirosh I, Suva ML 2024. Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors. *Cancer Cell*, 42: 1497-1506. 10.1016/j.ccell.2024.08.005
- Wang Q, Shao X, Zhang Y, et al. 2023. Role of tumor microenvironment in cancer progression and therapeutic strategy. *Cancer Med*, 12: 11149-11165. 10.1002/cam4.5698
- Wang Y, Luo B, Wang Z, et al. 2024. Advancements in Single-cell RNA Sequencing Technology in the Study of the Tumor Microenvironment in Lung Cancer. *Zhongguo Fei Ai Za Zhi*, 27: 441-450. 10.3779/j.issn.1009-3419.2024.101.15
- Weinstein JN, Collisson EA, Mills GB, et al. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, 45: 1113-1120. 10.1038/ng.2764
- Williams CG, Lee HJ, Asatsuma T, et al. 2022. An introduction to spatial transcriptomics for biomedical research. *Genome Med*, 14: 68. 10.1186/s13073-022-01075-1
- Wu F, Fan J, He Y, et al. 2021. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun*, 12: 2540. 10.1038/s41467-021-22801-0
- Yang L, He YT, Dong S, et al. 2022. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer*, 10. 10.1136/jitc-2021-003534
- Zhu P, Liu Z, Husain H, et al. 2025. Single-cell and spatial analysis reveals macrophage-T cell crosstalk in non-small cell lung cancer immunosuppression. *Transl Lung Cancer Res*, 14: 4002-4020. 10.21037/tlcr-2025-912

Paper received for publication in February, 2026
Paper accepted for publication in August, 2026