scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A single-cell RNA-seq dataset describing macrophages in NSCLC tumor and peritumor tissues

Aitian Li[1,4], Huishang Wang[1,4], Lei Zhang[2,4], Qitai Zhao[1], Yang Yang[2] ✉, Yi Zhang[1,3] ✉ & Li Yang[1,3] ✉

Examining tumor-associated macrophages in the immune microenvironment of non-small cell lung cancer (NSCLC) is essential for gaining an understanding of the genesis and development of NSCLC as well as for identifying key clinical therapeutic targets. Although previous studies have reported the diverse phenotypes and functions of macrophages in tumor tissues, thereby highlighting their significant role in the tumor microenvironment, the characteristic differences and correlations between tumor and peritumor tissue-derived macrophages that are necessary for an understanding of NSCLC progression remain unclear. Based on single-cell RNA sequencing, we generated a comprehensive dataset of transcriptomes from NSCLC tumor and peritumor tissues, thereby facilitating comprehensive analysis and providing significant insights. In summary, our dataset will serve as a valuable transcriptomic resource for further studies investigating NSCLC development.

Background & Summary

During the past decade, significant advances have been made in the treatment of non-small cell lung cancer (NSCLC), leading to a rapid transition to the era of immunotherapy[1,2]. Current cancer research has revealed that the tumor microenvironment (TME), particularly the tumor immune microenvironment (TIME), plays a pivotal role in the development, therapeutic resistance, and prognosis of NSCLC[3–5]. The intricate communication and interplay between immune and tumor cells as well as among different types of immune cells constitutes a complex and potent network within the TIME[6–8]. Regarding the TIME in NSCLC, tumor-associated macrophages (TAMs) have garnered considerable attention as one of the most abundant components of the immune cell population[9,10]. Particularly noteworthy is the plasticity of these macrophages, the activities of which can undergo modification in response to both tumor cells and the surrounding TME to facilitate tumor invasion and drug resistance. Conversely, such interactions can promote an alteration in the polarization state and thereby contribute to effective anti-tumor therapy[11,12]. In previous studies, we have employed RNA-seq to validate the phenotypic function of sorted $CD163^{+}$ and $CD163^{-}$ macrophages in NSCLC, and we analyzed their metabolic pathways. However, a drawback of this approach is its low cell yield and the necessity for substantial amounts of tissue samples for sequencing purposes that undoubtedly hamper the research process.

Based on their origin, TAMs can be categorized into tissue-resident and monocyte-derived subtypes[13,14]. Traditionally, macrophage functions have been dichotomized with respect to M1 (antitumor) or M2 (protumor) polarized states[15]. However, with advances in macrophage research, the roles of TAMs are now acknowledged to be multifaceted and are dynamically modulated in response to diverse microenvironmental conditions within tumors[16]. Consequently, single-cell sequencing has emerged as a valuable approach for delving deeper into the precise roles of TAMs in NSCLC and for assessing their potential therapeutic applications. Since the introduction of single-cell sequencing, this technique has been widely used in NSCLC studies. For example, Merad *et al.* used single-cell RNA sequencing to identify distinct sub-populations of macrophages within the TME of NSCLC. Based on molecular and functional characterization, they discovered that tissue-resident macrophages facilitate epithelial–mesenchymal transition in tumor cells and induce regulatory T cell responses[17].

[1]Biotherapy Center and Cancer Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China. [2]Thoracic Surgery Department, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China. [3]School of Life Sciences, Zhengzhou University, Zhengzhou, Henan, China. [4]These authors contributed equally: Aitian Li, Huishang Wang, Lei Zhang. ✉e-mail: fccyangy1@zzu.edu.cn; yizhang@zzu.edu.cn; fccyangl1@zzu.edu.cn

Patient	Sex	Age	Diagnosis	AJCC stage	Sample
Patient 1	Female	52	Lung adenocarcinoma	IIIA	Ca 1
					N 1
Patient 2	Female	70	Lung adenocarcinoma	IB	Ca 2
					N 2
Patient 3	Female	78	Lung adenocarcinoma	IA2	Ca 3
					N 3
Patient 4	Female	49	Lung adenocarcinoma	IIIA	Ca 4
					N 4
Patient 5	Female	57	Lung adenocarcinoma	IA3	Ca 5
					N 5

Table 1. Sample information.

Additionally, using single-cell RNA sequencing, Zhang *et al*. examined the changes in the TME in patients with NSCLC before and after neoadjuvant immunotherapy and observed that TAMs were remodeled to a neutral rather than an anti-tumor phenotype after treatment[18]. However, these aforementioned studies focused exclusively on sequencing macrophages within tumor tissues without performing the necessary comparisons between tumor and adjacent non-tumor tissues.

In this study, we performed single-cell RNA sequencing to investigate the alterations in macrophages during their migration within the tumor and adjacent peritumor tissues from five patients with NSCLC (Table 1). Sequencing data from all the samples were uniformly processed to eliminate potential batch effects. After initially performing quality control and dimensionality reduction, we labeled and clustered cell populations based on their gene expression profiles. Furthermore, we performed subgroup annotation of macrophage clusters and selected classical monocytes and macrophage clusters for subsequent analysis and comparison. Pseudo-time analysis revealed the potential lineage differentiation of macrophage subpopulations. Our dataset provides detailed coverage of the gene expression profiles of individual macrophages within NSCLC tumor and peritumor tissues, thereby providing new insights into the spatiotemporal interactions of macrophages.

Methods

An overview of the scRNA-seq analysis of NSCLC tissues is provided in Fig. 1a. The overall process included acquiring tissue samples from patients with NSCLC, preparing single-cell suspensions, and SeekGene sample processing.

Ethical approval. The study was approved by the Scientific Research and Clinical Trial Ethics Committee of the First Affiliated Hospital of Zhengzhou University (No. 2019-KY-256). We first screened newly diagnosed patients with NSCLC who were treated in the Thoracic Surgery Department of the First Affiliated Hospital of Zhengzhou University between January and April 2023. We randomly selected patients who met the following criteria: 1. did not receive any drug treatment; 2. tumors were classified as AJCC stage IIIA or earlier and met the requirements for radical resection of lung cancer. Patients whose tumor sizes met the requirements for clinicopathological diagnosis and experimental research sampling were included. We provided detailed information regarding the study to the patients and their guardians, including its purpose, procedures, risks, benefits, and participants' rights. They were fully aware of their freedom to decide whether to participate in the study and make informed decisions. Subsequently, the participants and their guardians signed an informed consent form confirming their understanding of all aspects of the study and their consent to participate. Subsequently, we sampled the resected tumor and peritumor tissues and assigned identifiers to these tissues. Throughout the experimental processes of sample handling, sequencing, data analysis, and data presentation, the samples were named using serial numbers to ensure that they were anonymized.

NSCLC tissue isolation and cell preparation. Fresh samples of NSCLC tumors and adjacent normal tissues were collected from patients at the First Affiliated Hospital of Zhengzhou University. All samples were obtained from treatment-naive patients who underwent surgical resection. Adjacent normal tissues were obtained from regions at least 2 cm away from the tumor border.

The freshly excised specimens were placed in ice-cold tissue storage buffer (Miltenyi) and transferred to the laboratory within 2 h of collection. Subsequently, tissue samples were washed two times in cold RPMI 1640 to remove visible blood clots, after which they were mechanically broken into small fragments using a sterile scalpel and further dissociated into single-cell suspensions using tissue dissociation reagent A (Seekgene, K01301-30) at 37 °C for 1 h. The digested cells were then filtered through a 70 μm cell filter, and following the addition of 5 mL of RPMI 1640, the resulting cell suspension was centrifuged at 300 × *g* for 5 min at 4 °C. Thereafter, the erythrocytes were removed using red blood cell lysis buffer (Solarbio, R1010) for 3 min, and the remaining cells were centrifuged and resuspended as described above. Cell numbers and viability were determined using a fluorescence cell analyzer (Countstar® Rigel S2) with AO/PI reagent, and then debris and dead cells were removed (Miltenyi, 130-109-398/130-090-101). Finally, if cell viability exceeded 80%, the cells were resuspended at 1×10^6 cells/mL in 1 × PBS and 0.04% bovine serum albumin for single-cell RNA-seq library construction and sequencing.

Single-cell RNA-seq library construction and sequencing. Single-cell RNA-seq libraries were prepared using a SeekOne Digital Droplet Single Cell 5′ library preparation kit (SeekGene, K00501). Briefly, cell

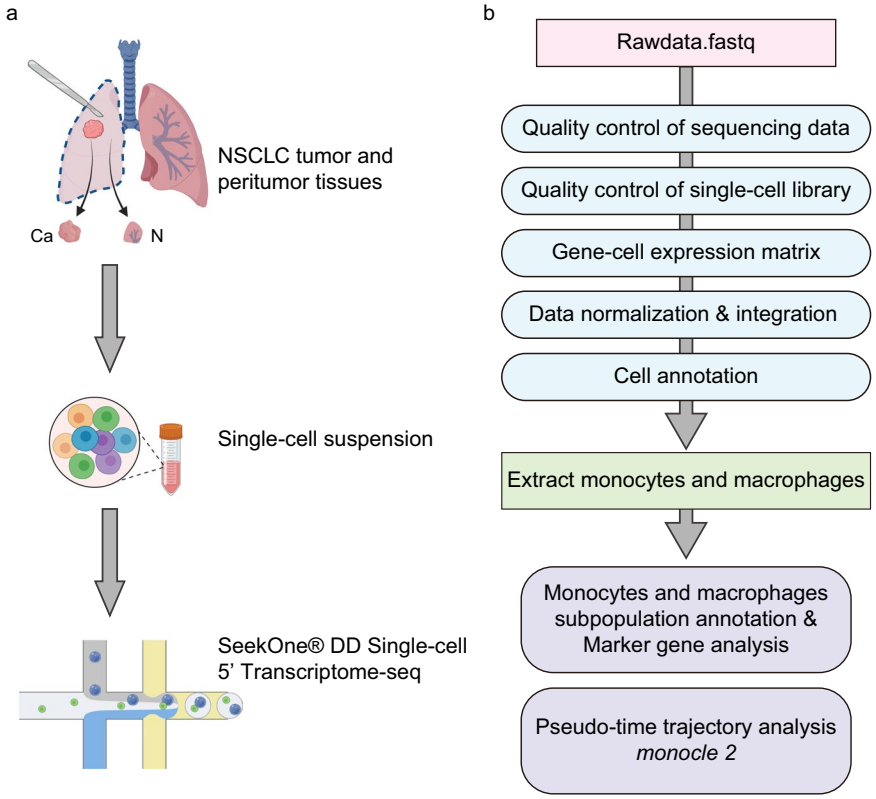

Fig. 1 Outline of the sample handling and data analysis workflow. (**a**) An overview of the process for conducting single-cell RNA sequencing using cancerous and adjacent non-cancerous tissues from five patients with non-small cell lung cancer, including tissue sample collection, single-cell suspension preparation, and on-machine testing. (**b**) A flow chart showing the analysis of single-cell RNA sequencing data.

numbers equivalent to a target cell recovery of 10,000 cells per sample after sequencing were added to the sample wells in a SeekOne DD Chip S3 together with a reverse transcription reagent. Subsequently, in Chip S3, barcoded hydrogel beads and partitioning oil were added to the respective wells, and having generated emulsified droplets, we performed reverse transcription (42 °C for 90 min and 85 °C for 5 min). The obtained cDNA was purified and PCR-amplified. The amplified cDNA products were cleaned, fragmented, end-repaired, A-tailed, and ligated to sequencing adaptor. Indexed PCR was performed using 12 cycles to amplify cDNA representing 5' expression genes that contained cell barcodes and unique molecular identifiers (UMI). The indexed sequencing libraries were cleaned using SPRI beads and quantified (KAPA Biosystems, KK4824). Paired-end reads (150 bp) were generated using the Illumina NovaSeq 6000 platform.

Bioinformatic analysis. Single-cell data analysis was performed using SeekSoul® Tools, which can identify cell barcode labels, perform quantitative alignment, and generate downstream analysis results in the form of a cell expression matrix, and we used this software for subsequent cell clustering and differential analysis (http://seeksoul.seekgene.com/en/v1.2.1/index.html). The analysis pipeline is provided in Fig. 1b. Below, we describe the individual steps of the bioinformatic analysis pipeline used in this study.

Quality control of sequencing data. Initially, FASTP was used to process raw sequencing data[19]. Sequencing splices and low-quality fragments were removed by trimming. Subsequent analysis was based on clean data, and Fastp software was used for raw data trimming. The data quality control results are presented in Table 2.

Quality control of single-cell libraries. The quality control software SeekSoul® Tools was used to control the clean reads, after which the sequencing data were aligned to the reference genome (GRCh38) for gene expression quantification and to generate a preliminary cell–gene expression matrix. Sequencing data labeled with barcodes and UMIs were aligned to the reference genome using the STAR software. Statistical analysis of expression profiles was performed using feature Counts software, in which reads with the same barcode and UMI were used to represent replicate sequencing data of the same molecule and were merged to obtain a preliminary expression matrix. Seurat (version 4.0.0) was used to filter low-quality cells, and the UMI threshold was set to 300[20]. In addition, cells with fewer than 200 genes or more than 5,000 genes were removed, and cells affected by mitochondrial genes were deleted using MAD-variance normal[21]. The remaining cells were used for subsequent analyses. The results of the library quality control and filtration are shown in Table 3.

Sample	Clean reads	Clean ratio	Q30 bases in RNA read	Valid barcodes	Reads mapped confidently to genome
Ca 1	374,984,591	94.94%	84.34%	91.28%	59.95%
N 1	356,785,437	94.9%	87.52%	90.28%	71.48%
Ca 2	396,606,273	94.21%	92.33%	87.69%	72.19%
N 2	374,592,556	94.14%	92.61%	86.76%	74.16%
Ca 3	263,157,024	92.36%	85.5%	90.72%	78.32%
N 3	262,939,734	92.88%	85.65%	90.69%	80.05%
Ca 4	341,585,030	92.65%	89.12%	89.09%	68.77%
N 4	344,516,041	90.98%	88.88%	87.92%	74.25%
Ca 5	346,697,369	93.81%	93.95%	95.06%	93.19%
N 5	404,330,114	94.46%	94%	96.02%	92.76%

Table 2. Detailed quality control of FASTQ files.

Sample	Estimated number of cells	Mean reads per cell	Median genes per cell	Median UMI per cell	Fraction reads in cells	Sequencing saturation
Ca 1	6,375	58,821	2,665	6,813	89.74%	67.49%
N 1	9,575	37,262	1,778	4,072	82.74%	62.79%
Ca 2	7,101	55,852	1,671	3,545	74.49%	72.26%
N 2	6,376	58,750	1,629	3,053	74.74%	76.17%
Ca 3	8,109	32,452	2,085	4,760	68.78%	52.32%
N 3	6,485	40,545	1,690	3,356	76.86%	66.90%
Ca 4	6,115	55,860	1,802	3,873	71.88%	64.79%
N 4	5,730	60,124	1,402	2,708	64.49%	74.83%
Ca 5	6,280	55,206	1,477	3,119	63.85%	85.10%
N 5	6,753	59,874	1,384	2,716	65.04%	86.84%

Table 3. Sequencing statistics for generated libraries.

Data normalization and integration. After filtering low-quality cells, the samples were pooled, and the batch effect was eliminated using the CCA method. All samples were grouped using FindIntegrationAnchors and IntegrateData, and the UMI numbers were normalized using ScaleData. Subsequently, nonlinear dimension reduction was performed using RunPCA and RunUMAP. Cells were clustered using FindClusters at a resolution of 0.8 and 30 dims.

Cell annotation. FindAllMarkers was used to compare each cluster with all other clusters to identify cluster-specific marker genes. Marker genes that retained were those expressed in at least 10% of the cells and expressed under a minimum log-fold change threshold of 0.25. Accordingly, cluster-related differential gene expression was considered significant if the adjusted p-value was less than 0.05 and avg_log2FC was greater than 0.25. Cell-type-specific signature marker genes were used to annotate the clusters.

Monocyte and macrophage sub-population annotation and marker gene analysis. Numerous studies have documented the heterogeneity of mononuclear macrophages, with Zhang's team categorizing them into six subgroups based on single-cell sequencing of esophageal cancer tissue[22]. To annotate the monocyte and macrophage subpopulations in our dataset, we leveraged the gene-barcode matrix. Based on 25 dimensions and a resolution of 0.5, clusters were identified in the macrophages, and cell type assignment was performed based on marker genes reported in previous studies[22].

Pseudo-time trajectory analysis. Monocyte and macrophage fates and pseudo-time trajectories were reconstructed using the Monocle2 package (version 2.3.6)[23]. The importCDS function in Monocle was used to convert the original count in the Seurat object to a CellDataSet, and the genes used for pseudo-time ordering were selected using the dispersionTable function. The DDRTree method was utilized for dimension reduction and cell ordering along the pseudo-time trajectory. No root state was specified, and ordering was performed in an unsupervised manner. Branch analysis was performed using the BEAM function. Significantly altered genes at the branch point were clustered using plot_genes_branched_pseudo-time functions according to the distinct patterns of the changes in gene expression.

Data Records

The sequencing data generated in this study have been deposited in the Genome Sequence Archive at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA006716) and are accessible through BioProject PRJCA023562 (GSA-human: HRA006716)[24].

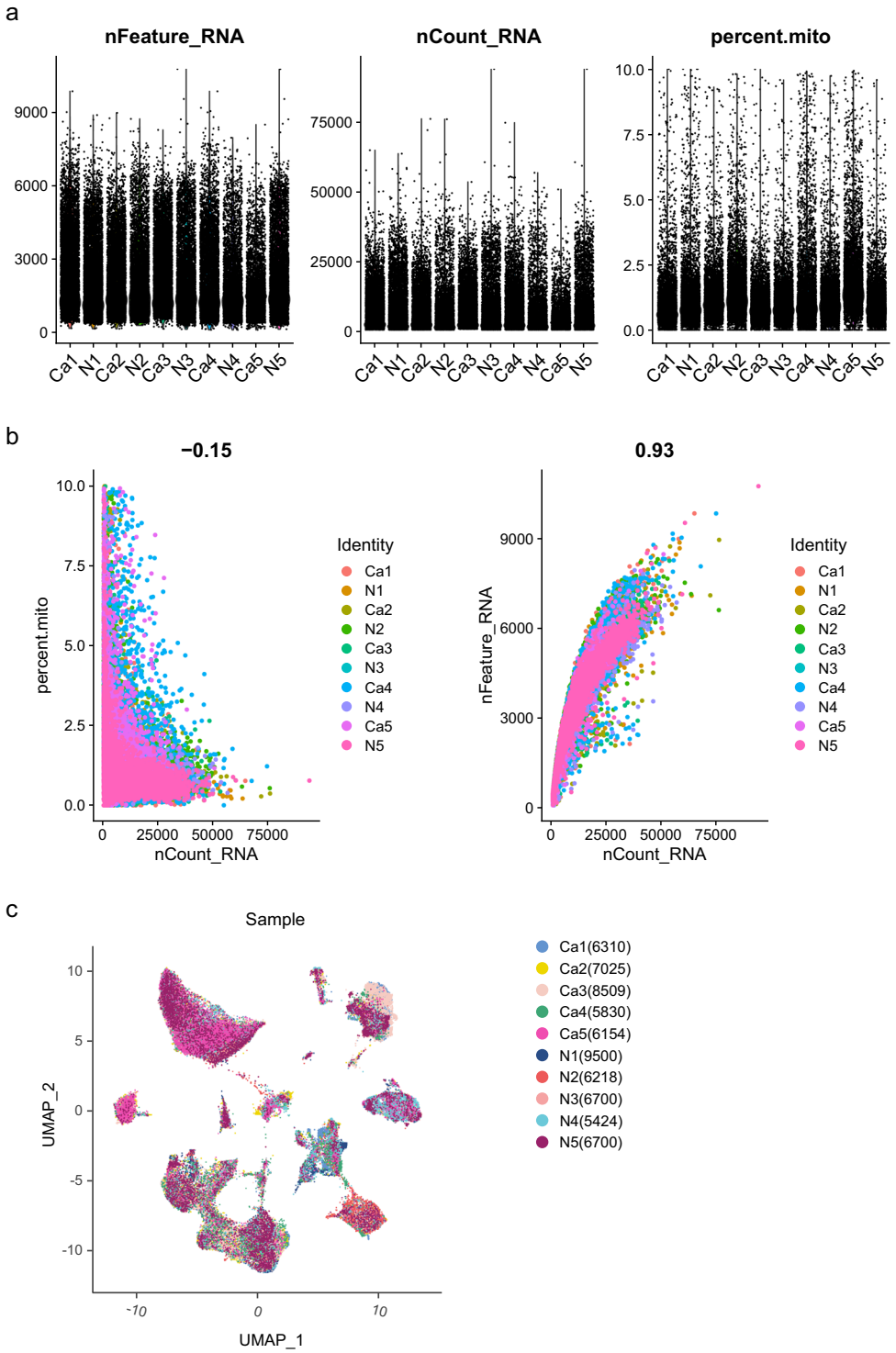

Fig. 2 Quality control of single-cell RNA sequencing data from cancerous and adjacent non-cancerous tissues from patients with non-small cell lung cancer. (**a**) A scatterplot illustrating the number of genes, unique molecular identifiers (UMIs), and percentage of mitochondrial genes in each cell of 10 samples. (**b**) The relationship between the percentage of mitochondrial genes and the mRNA reads together with the relationship between the amount of mRNA and the mRNA reads. (**c**) We detected a batch effect between 10 different samples.

Technical Validation

Fresh cancer and para-cancer tissue specimens were collected from five female patients undergoing surgery for NSCLC and processed to obtain single-cell suspensions. We utilized Seurat for quality control, during which the counts of genes, UMIs, and the proportion of mitochondrial genes in each cell were computed (Fig. 2a). We

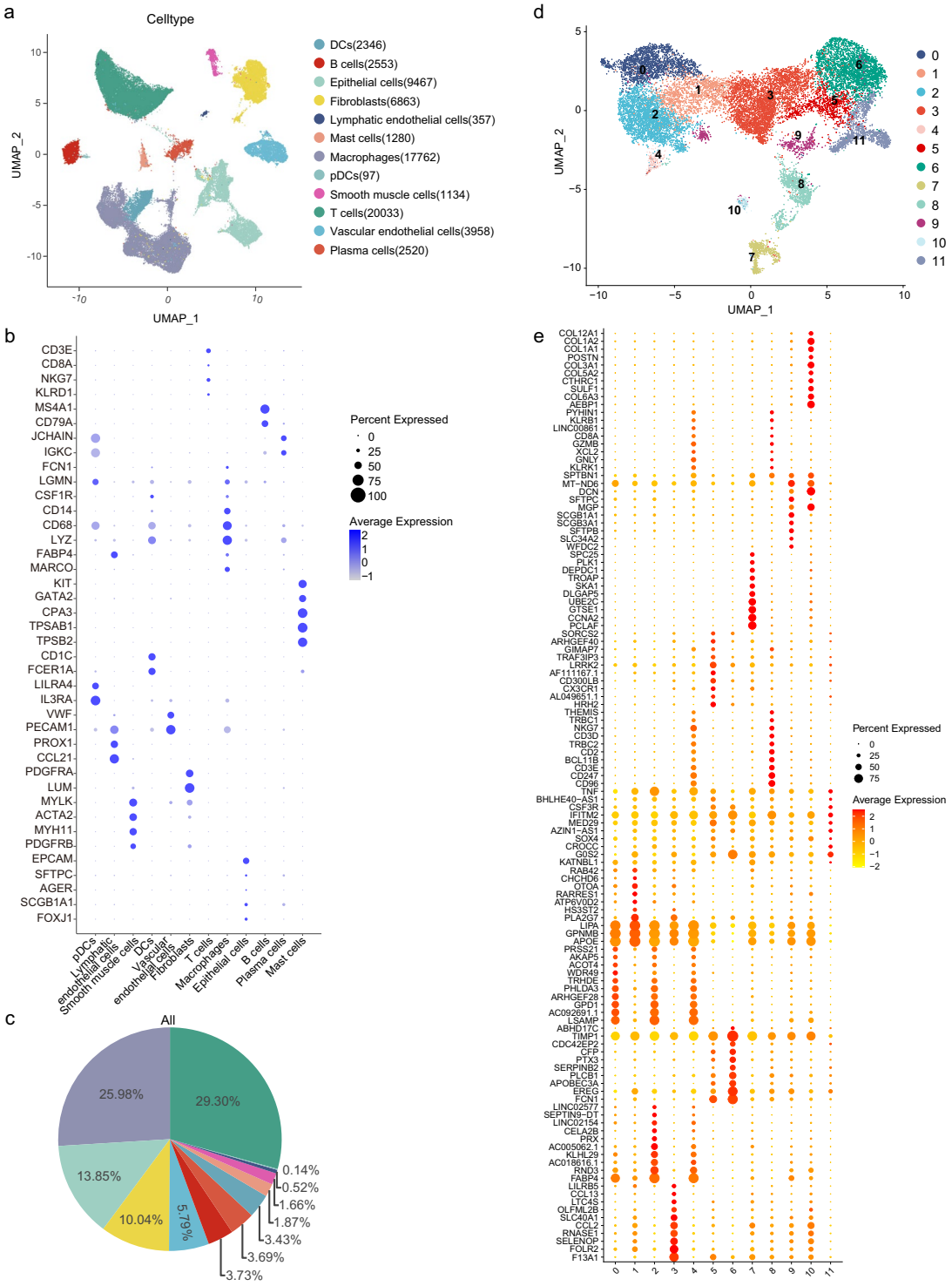

Fig. 3 Macrophage sub-population annotation and marker gene analysis. (**a**) A Uniform Manifold Approximation and Projection (UMAP) plot of colored-based cell types analyzed in non-small cell lung cancer tumor and peritumor tissues (n = 5). (**b**) A dot plot showing the marker genes of each cluster. (**c**) A pie chart showing the proportions of each cell type. (**d**) A UMAP plot of 12 macrophage sub-populations. (**e**) A dot plot showing the top 10 differentially expressed genes in macrophage sub-populations.

analyzed and graphically presented the correlations between the proportion of mitochondrial genes and mRNA reads, as well as the associations between mRNA quantity and mRNA reads (Fig. 2b). Furthermore, we investigated the batch effect in diverse cancer and para-cancerous tissue samples (Fig. 2c).

After quality control, 68,370 tissue cells were subjected to further analysis. Subsequently, we successfully identified 12 distinct cell clusters, with individual clusters containing between 97 and 20,033 cells (Fig. 3a). Cell

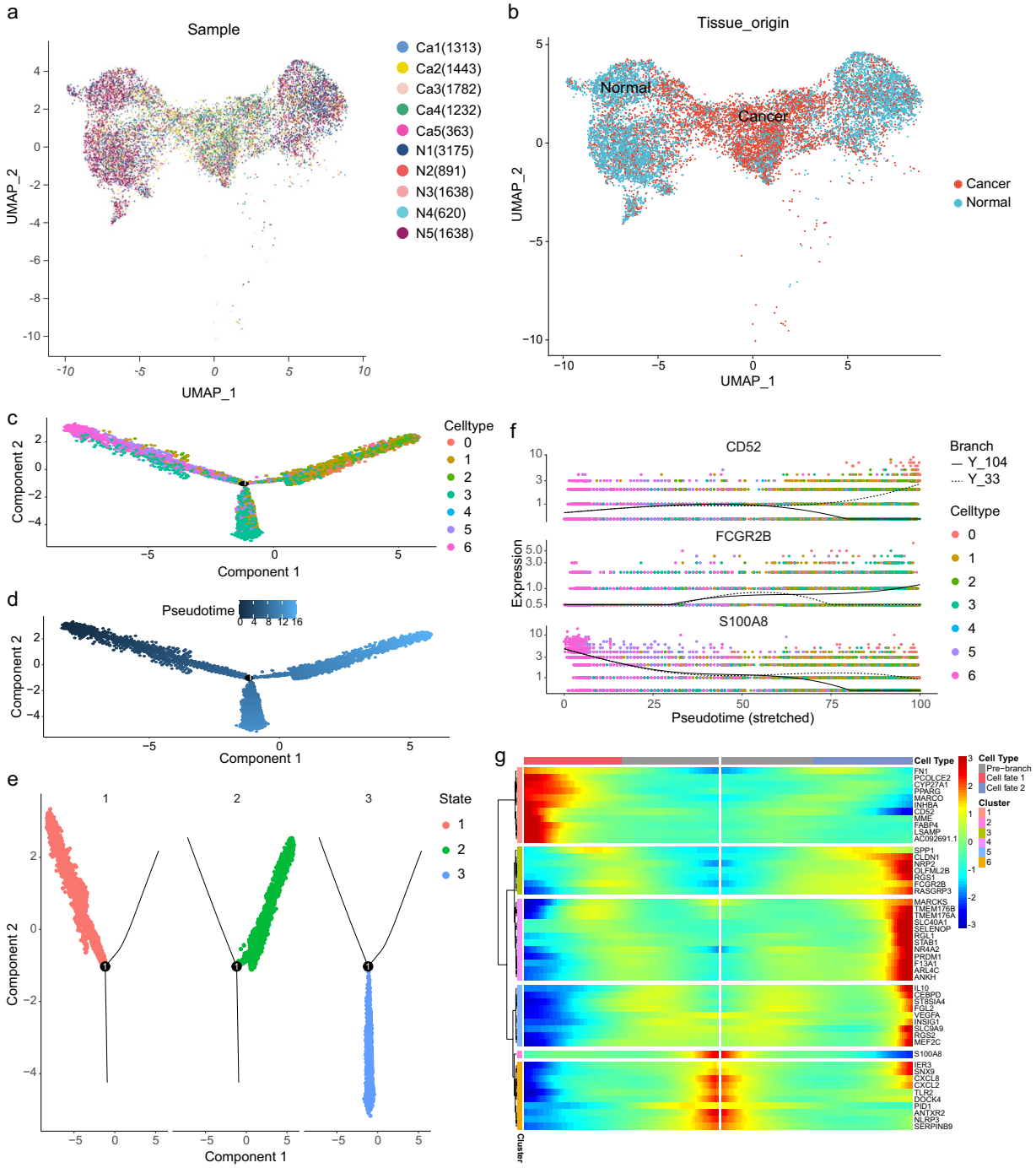

Fig. 4 Reconstruction of the developmental trajectories of macrophages. (**a**) A Uniform Manifold Approximation and Projection (UMAP) plot showing each cell from all tissues. (**b**) A UMAP plot showing each cell from tumor and peritumor tissues. (**c**) Monocle2-generated pseudo-temporal trajectory of seven clusters. (**d**) Pseudo-time colored in a gradient from dark to light blue, with dark blue signifying the start of pseudo-time. (**e**) Using Monocle2, the pseudo-time trajectory was divided into three different states. (**f**) The top three genes influencing macrophage fate are represented as line plots displayed as the expression level over pseudo-time, as determined using Monocle2. (**g**) A heat map presenting the clustering of the top 50 genes that influenced cell fate. These 50 genes were divided into six clusters, showing genes at the initial, transitory, and final stage of the developmental trajectory.

clustering was visualized using UMAP (Fig. 3a). Based on the selected marker genes, we categorized cells into 12 clusters, including B lymphocytes, dendritic cells, epithelial cells, fibroblasts, lymphatic endothelial cells, mast cells, macrophages, plasmacytoid dendritic cells, plasma cells, smooth muscle cells, T lymphocytes, and vascular endothelial cells[25–28] (Fig. 3a,b). Macrophages accounted for 25.98% of the total cell population (Fig. 3c).

Subsequently, dimensionality reduction was performed on the macrophage clusters, and these cells were further subdivided into 12 distinct subpopulations based on characteristic gene expression (Fig. 3d). After identifying the top 10 differentially expressed genes in each subgroup, we established that subgroups 0 to 6 comprised classical monocyte and macrophage populations, including five macrophage subsets (cluster 0–4) and two monocyte subsets (cluster 5 and cluster 6)[29–31] (Fig. 3e). Consequently, we conducted an in-depth comparison and analysis of seven subgroups (Fig. 4a). UMAP plots of the tumor and peritumor tissues revealed that the monocytes and macrophages in these two tissues tended to be derived from different clusters (Fig. 4b). Macrophage clusters 1 and 3 were enriched in tumor tissues, highly expressed APOE and FOLR2, and were generally defined as immunosuppressive macrophages[25,32]. In contrast, there were two inflammatory monocyte clusters expressing FCN1 and three macrophage clusters (0, 2, and 4) characterized by the expression of FABP4, a gene involved in lipid metabolism, in the peritumoral tissues (Figs. 3d,e and 4b)[33,34]. Subsequently, we used Monocle2 to conduct a pseudotemporal analysis of the cells within these subgroups, which revealed their fate and pseudo-time trajectories (Fig. 4c–e). We concurrently identified and characterized the top three genes with the most significant influence on the differentiation trajectory (Fig. 4f), and obtained more comprehensive gene information (Fig. 4g). Pseudotime analysis revealed a differentiation trajectory from inflammatory monocytes into immunosuppressive macrophages, thus indicating peritumoral monocyte pools may be recruited to the tumor milieu to replenish the pro-tumoral macrophage polarization and exacerbate an immunosuppressive microenvironment.

In conclusion, we present a comprehensive single-cell transcriptional map of tumor and peritumor tissue cells in patients with NSCLC. Additionally, we investigated the differentiation of macrophages in tissues from a spatiotemporal perspective. These single-cell sequencing data will facilitate further examination of the intricate regulatory network within the tumor-associated microenvironment in NSCLC and the identification of key targets.

Code availability

All single-cell RNA-Seq analyses were performed using FastQC v 0.23.1 (https://github.com/OpenGene/fastp), SeekSoul® Tools v1.0.0 (http://seeksoul.seekgene.com/zh/v1.0.0/index.html), Seurat v 4.0.0 (https://satijalab.org/seurat/), and Monocle2 package v 2.3.6 (http://cole-trapnell-lab.github.io/monocle-release/docs/).

Received: 8 April 2024; Accepted: 5 September 2024;
Published online: 01 October 2024

References

1. Thai, A. A., Solomon, B. J., Sequist, L. V., Gainor, J. F. & Heist, R. S. Lung cancer. *Lancet* **398**, 535–554, https://doi.org/10.1016/S0140-6736(21)00312-3 (2021).
2. Reck, M., Remon, J. & Hellmann, M. D. First-Line Immunotherapy for Non-Small-Cell Lung Cancer. *J Clin Oncol* **40**, 586–597, https://doi.org/10.1200/JCO.21.01497 (2022).
3. Pitt, J. M. *et al*. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann Oncol* **27**, 1482–1492, https://doi.org/10.1093/annonc/mdw168 (2016).
4. Hinshaw, D. C. & Shevde, L. A. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res* **79**, 4557–4566, https://doi.org/10.1158/0008-5472.CAN-18-3962 (2019).
5. Remark, R. *et al*. The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome. *Am J Respir Crit Care Med* **191**, 377–390, https://doi.org/10.1164/rccm.201409-1671PP (2015).
6. Chen, D., Zhang, X., Li, Z. & Zhu, B. Metabolic regulatory crosstalk between tumor microenvironment and tumor-associated macrophages. *Theranostics* **11**, 1016–1030, https://doi.org/10.7150/thno.51777 (2021).
7. Mao, X. *et al*. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer* **20**, 131, https://doi.org/10.1186/s12943-021-01428-1 (2021).
8. Vitale, I., Manic, G., Coussens, L. M., Kroemer, G. & Galluzzi, L. Macrophages and Metabolism in the Tumor Microenvironment. *Cell Metab* **30**, 36–50, https://doi.org/10.1016/j.cmet.2019.06.001 (2019).
9. Mantovani, A., Allavena, P., Marchesi, F. & Garlanda, C. Macrophages as tools and targets in cancer therapy. *Nat Rev Drug Discov* **21**, 799–820, https://doi.org/10.1038/s41573-022-00520-5 (2022).
10. Zheng, X. *et al*. Spatial Density and Distribution of Tumor-Associated Macrophages Predict Survival in Non-Small Cell Lung Carcinoma. *Cancer Res* **80**, 4414–4425, https://doi.org/10.1158/0008-5472.CAN-20-0069 (2020).
11. Mantovani, A., Sozzani, S., Locati, M., Allavena, P. & Sica, A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol* **23**, 549–555, https://doi.org/10.1016/s1471-4906(02)02302-5 (2002).
12. Cao, M. *et al*. Ginseng-derived nanoparticles alter macrophage polarization to inhibit melanoma growth. *J Immunother Cancer* **7**, 326, https://doi.org/10.1186/s40425-019-0817-4 (2019).
13. Epelman, S., Lavine, K. J. & Randolph, G. J. Origin and functions of tissue macrophages. *Immunity* **41**, 21–35, https://doi.org/10.1016/j.immuni.2014.06.013 (2014).
14. Ginhoux, F. & Guilliams, M. Tissue-Resident Macrophage Ontogeny and Homeostasis. *Immunity* **44**, 439–449, https://doi.org/10.1016/j.immuni.2016.02.024 (2016).
15. Locati, M., Curtale, G. & Mantovani, A. Diversity, Mechanisms, and Significance of Macrophage Plasticity. *Annu Rev Pathol* **15**, 123–147, https://doi.org/10.1146/annurev-pathmechdis-012418-012718 (2020).
16. Bai, R. *et al*. The hypoxia-driven crosstalk between tumor and tumor-associated macrophages: mechanisms and clinical treatment strategies. *Mol Cancer* **21**, 177, https://doi.org/10.1186/s12943-022-01645-2 (2022).
17. Casanova-Acebes, M. *et al*. Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584, https://doi.org/10.1038/s41586-021-03651-8 (2021).
18. Hu, J. *et al*. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome Med* **15**, 14, https://doi.org/10.1186/s13073-023-01164-9 (2023).
19. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
20. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420, https://doi.org/10.1038/nbt.4096 (2018).
21. Pijuan-Sala, B. *et al*. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495, https://doi.org/10.1038/s41586-019-0933-9 (2019).

22. Cheng, S. *et al*. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809 e723, https://doi.org/10.1016/j.cell.2021.01.010 (2021).
23. Trapnell, C. *et al*. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386, https://doi.org/10.1038/nbt.2859 (2014).
24. National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, https://ngdc.cncb.ac.cn/gsa-human/browse/HRA006716 (2024).
25. Kim, N. *et al*. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun* **11**, 2285, https://doi.org/10.1038/s41467-020-16164-1 (2020).
26. Wang, C. *et al*. The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. *Signal Transduct Target Ther* **7**, 289, https://doi.org/10.1038/s41392-022-01130-8 (2022).
27. Wu, F. *et al*. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun* **12**, 2540, https://doi.org/10.1038/s41467-021-22801-0 (2021).
28. Li, Q. *et al*. Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq. *Genome Med* **14**, 87, https://doi.org/10.1186/s13073-022-01089-9 (2022).
29. Evren, E. *et al*. Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259–275 e257, https://doi.org/10.1016/j.immuni.2020.12.003 (2021).
30. van Vlerken-Ysla, L., Tyurina, Y. Y., Kagan, V. E. & Gabrilovich, D. I. Functional states of myeloid cells in cancer. *Cancer Cell* **41**, 490–504, https://doi.org/10.1016/j.ccell.2023.02.009 (2023).
31. Mulder, K. *et al*. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883–1900 e1885, https://doi.org/10.1016/j.immuni.2021.07.007 (2021).
32. Xiang, C. *et al*. Single-cell profiling reveals the trajectory of FOLR2-expressing tumor-associated macrophages to regulatory T cells in the progression of lung adenocarcinoma. *Cell Death Dis* **14**, 493, https://doi.org/10.1038/s41419-023-06021-6 (2023).
33. Wang, J., Zhu, N., Su, X., Gao, Y. & Yang, R. Novel tumor-associated macrophage populations and subpopulations by single cell RNA sequencing. *Front Immunol* **14**, 1264774, https://doi.org/10.3389/fimmu.2023.1264774 (2023).
34. Liu, Z. *et al*. Lipid-associated macrophages in the tumor-adipose microenvironment facilitate breast cancer progression. *Oncoimmunology* **11**, 2085432, https://doi.org/10.1080/2162402X.2022.2085432 (2022).

Acknowledgements

We thank Zhichao Li for providing constructive feedback regarding this manuscript. This work was supported by grants from the State's Key Project of Research and Development Plan (grant number 2021YFE0110600), the National Natural Science Foundation of China (grant numbers 82072578 and 82350121), and Funding for Scientific Research and Innovation Team of The First Affiliated Hospital of Zhengzhou University (grant number ZYCXTD2023013).

Author contributions

L.Y. designed the study and reviewed and revised the manuscript. A.L. performed the experiments, analyzed the data, and wrote the manuscript. H.W. performed the experiments, analyzed the data, and wrote the manuscript. L.Z. collected the samples. Q.Z. analyzed the data. Y.Y. collected the samples. Y.Z. reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-03885-x.

Correspondence and requests for materials should be addressed to Y.Y., Y.Z. or L.Y.