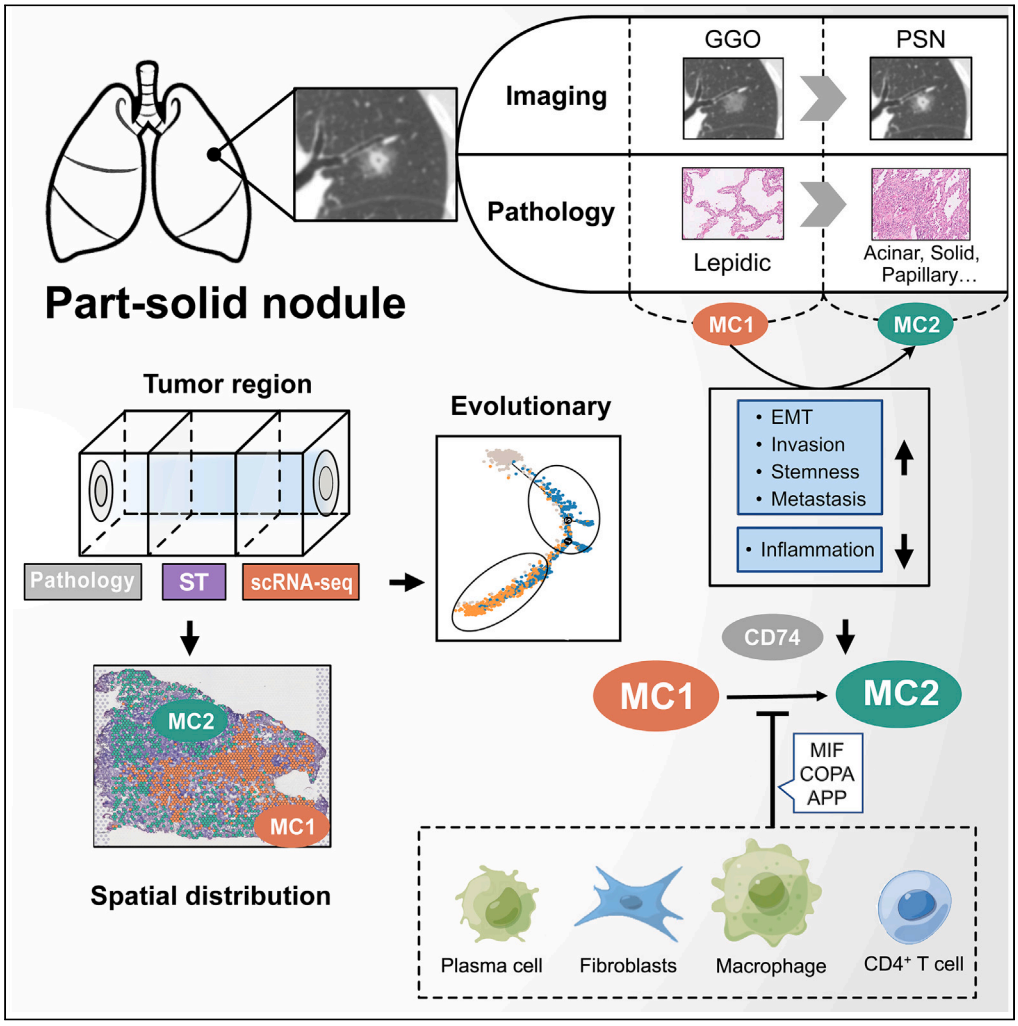
iScience

CellPress
OPEN ACCESS

Article

Spatial downregulation of CD74 signatures may drive invasive component development in part-solid lung adenocarcinoma



Jia-Tao Zhang, Juan Zhang, Song-Rong Wang, ..., Qing Zhou, Yi-Long Wu, Wen-Zhao Zhong

zhongwenzhao@gdph.org.cn

Highlights

Linear evolution pattern of malignant cells is observed in part-solid lung cancer

Early presenting malignant cells correspond to the preinvasive component

Downregulation of CD74 may drive immune escape in part-solid lung cancer

Zhang et al., iScience 26, 107699
October 20, 2023 © 2023
https://doi.org/10.1016/j.isci.2023.107699

iScience

CellPress
OPEN ACCESS

Article

Spatial downregulation of CD74 signatures may drive invasive component development in part-solid lung adenocarcinoma

Jia-Tao Zhang,[1] Juan Zhang,[2] Song-Rong Wang,[1] Li-Xu Yan,[1] Jing Qin,[3] Kai Yin,[1] Xiang-Peng Chu,[1] Meng-Min Wang,[1] Hui-Zhao Hong,[1] Zhi-Yi Lv,[1] Song Dong,[1] Ben-Yuan Jiang,[1] Xu-Chao Zhang,[1] Xiang Liu,[2] Qing Zhou,[1] Yi-Long Wu,[1] and Wen-Zhao Zhong[1,4,*]

SUMMARY

Pulmonary nodules with part-solid imaging features manifest during the progression from preinvasive to invasive lung adenocarcinoma. To define the spatial composition and evolutionary trajectories of early-stage lung adenocarcinoma, we combined spatial transcriptomics (ST) and pathological annotations from 20 part-solid nodules (PSNs), four of which were matched with single-cell RNA sequencing. Two malignant cell populations (MC1 and MC2) were identified, and a linear evolutionary relationship was observed. Compared to MC2, the pre-existing malignant MC1 exhibited a lower metastatic signature, corresponding to the preinvasive component (lepidic) on pathology and the ground glass component on PSN imaging. Higher immune infiltration was observed among MC1 regions in ST profiles, and further analysis revealed that macrophages may be involved in this process through the CD74 axis. This work provides deeper insights into the evolutionary process and spatial immune cell composition behind PSNs and highlights the mechanisms of immune escape behind this adenocarcinoma trajectory.

INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide.[1] Since low-dose computed tomography has been proven to be an effective screening tool,[2,3] suspicious lung nodules are increasingly being diagnosed. Among these, the incidence of part-solid nodules (PSNs) appears especially prominent.[4–6] The development of lung PSNs is considered a stepwise process of ground-glass nodules emerging as a solid component, which is also a strong indication of tumors with invasive pathological components, including minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC).[7,8] Because of this classical evolving disease process, PSNs are an ideal tumor evolution model for early-stage lung adenocarcinoma (LUAD).

Single-cell RNA sequencing (scRNA-seq) has changed our fundamental understanding of tumor heterogeneity.[9,10] Continuous efforts have been undertaken to reveal the intrinsic and extrinsic factors involved in the evolutionary process behind LUAD, and changes in certain signaling pathways and tumor microenvironmental components have been observed. However, the association between these transcriptomic features and pathological or imaging changes has not been well established. In other words, we still lack a clear understanding of whether these transcriptomic events trigger binary-classified evolutionary progressions in PSNs, such as non-invasive to invasive components observed through pathology testing or pure ground-glass opacity to solid components detected on imaging. Recent advances in spatial transcriptomic (ST) technologies have enabled the comprehensive characterization of gene expression patterns using spatial tissue information. In cases of LUAD, it might be possible to correlate information from tumor evolution, spatial patterns, imaging, and pathological features in a frame-by-frame manner. Here, to achieve that, we applied ST and scRNA-seq in the context of a pathological annotation framework.

RESULTS

Characterization of PSN-resident cell types by scRNA-seq

We collected 20 fresh PSN samples that were pathologically confirmed as MIA (n = 3, 15.0%) or IAC (n = 17, 85%, Figure 1A; Table S1; Figure S1). Based on imaging features, the consolidation-to-tumor ratio (CTR) of the PSNs was calculated according to our previous three-dimensional computerized methods,[11] with −470 HU as the cutoff value to distinguish the solid component from ground-glass opacity.[12] Two of

[1]Guangdong Lung Cancer Institute, Guangdong Provincial People's Hospital, (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China
[2]Echo Biotech Co, Ltd, Beijing, China
[3]School of Pharmaceutical Sciences (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong, China
[4]Lead contact
*Correspondence: zhongwenzhao@gdph.org.cn
https://doi.org/10.1016/j.isci.2023.107699

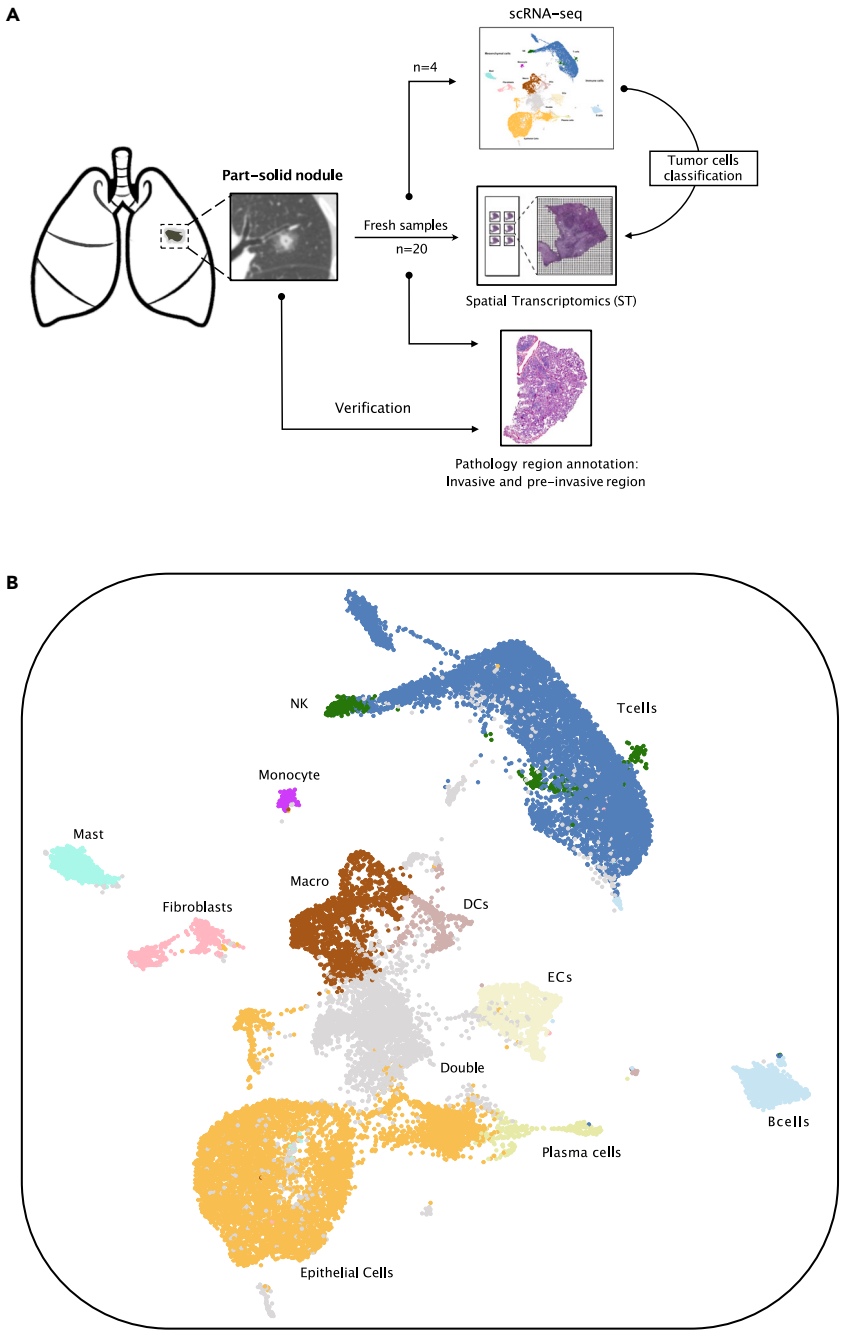


Figure 1. Study schematic and resident cell types characterization by scRNA-seq

(A) Overview of study design.

(B) UMAP plot of 26,710 single cells colored according to the 12 major cell types. scRNA-seq, single-cell RNA sequencing; UMAP, uniform manifold approximation and projection; PSN, part-solid nodules.

these samples had CTRs of 0–25%, seven had CTRs of 25–50%, eight had CTRs of 50–75%, and three had CTRs >75%. All samples were subjected to next-generation sequencing of 1,021 cancer-associated genes. *EGFR* L858R (n = 7, 35%), *EGFR* 19del (n = 4, 20%), and *KRAS* (n = 2, 10%) were among the top driver-gene mutations.

Four PSN samples with matched scRNA-seq were analyzed to obtain a comprehensive single-cell solution atlas. Unsupervised clustering of 26,710 QC-passed cells revealed cell clusters representing 11 major cellular lineages based on the Seurat package and their marker gene expression, including epithelial cells, immune cell types (T cells, B cells, plasma cells, nature killer cells, dendritic cells, macrophages,

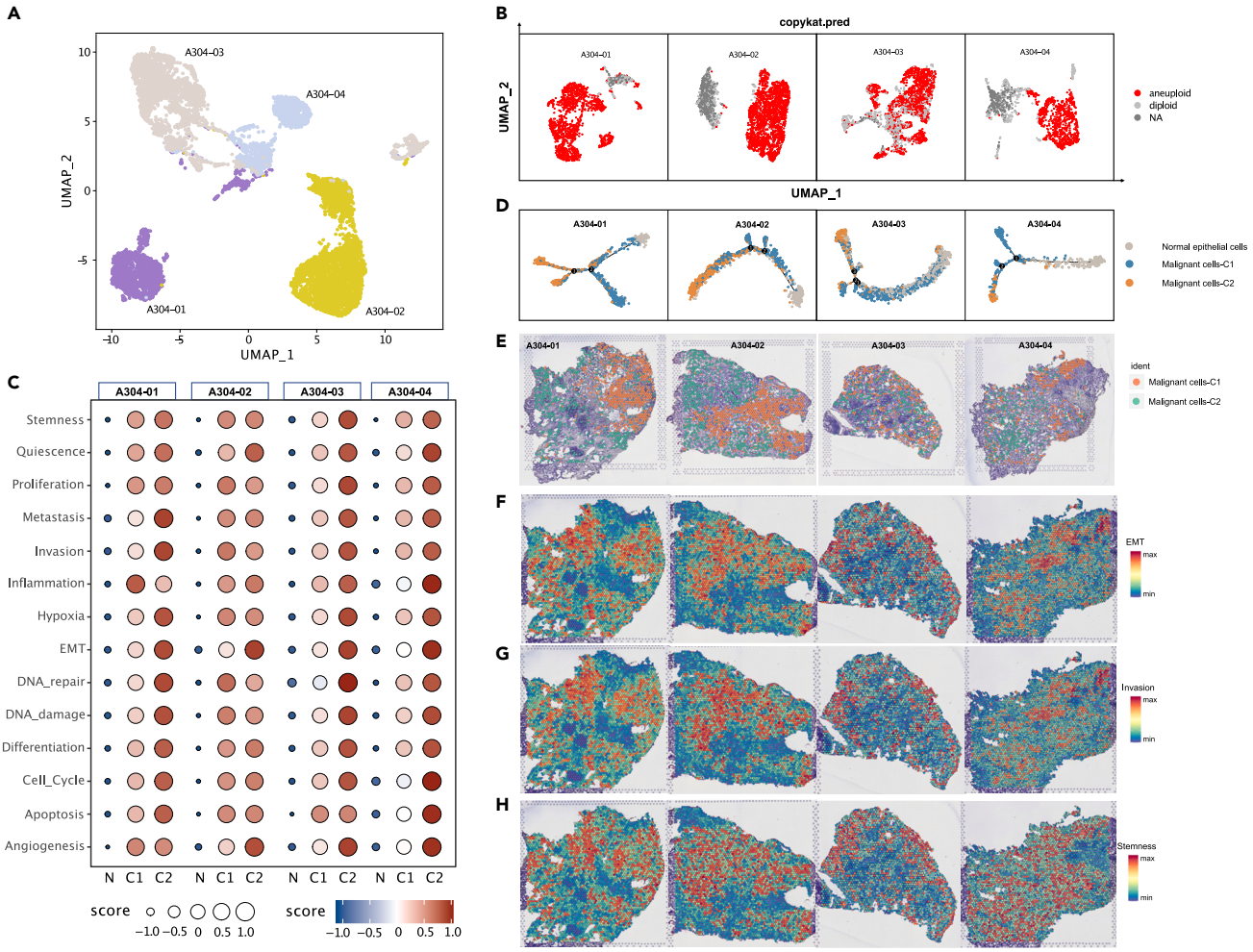


Figure 2. Identification, clustering, and spatial distribution of malignant cells

(A) UMAP plot of 8492 epithelial cells colored according to the PSN sample identity.

(B) UMAP embedding of scRNA-seq data with annotation of the inferred CopyKAT diploid and aneuploid copy number profiles.

(C) GSVA of 14 functional states from the CancerSEA dataset among normal epithelial cells, MC1 and MC2.

(D) Illustration of pseudotime trajectory process from normal epithelial cells to malignant cells inferred by Monocle2.

(E) ST spot overlay of cell type predictions for MC1 and MC2 via FindTransferAnchors in the Seurat package.

(F–H) The spatial visualization of the expression level of EMT, invasion, and stemness signature from CancerSEA dataset.PSN, part-solid nodule; UMAP, uniform manifold approximation and projection; scRNA-seq, single-cell RNA sequencing; GSVA, gene set variation analysis; ST, spatial transcriptomics; EMT, epithelial-mesenchymal transition.

monocytes, and mast cells), and stromal cell types (fibroblasts and endothelial cells) (Figure 1B; Figures S2A and S2B). The relative abundances of each major cell type were similar in the four patients (Figure S2C), and epithelial and T cells were the two dominant resident cell types in the tumor microenvironment of the PSN samples (Figure S2D).

Identification of two major subpopulations of malignant cells

As expected, epithelial cells showed extremely high inter-tumor heterogeneity (Figure 2A), similar to the findings of previous studies.[13–15] Therefore, we attempted to identify malignant cells in the epithelial cell population through CopyKAT inference in a case-by-case manner. We obtained1355 (A304-01), 2071 (A304-02), 1542 (A304-03), and 716 (A304-04) malignant epithelial cells from the epithelial cell types (Figure 2B). The malignant cell clusters inferred from the CopyKAT patterns were consistent with the differentiation states predicted by CytoTRACE (Figures S3A–S3H).

Next, we used CopyKAT to characterize the subclonal structures of the malignant epithelial cells on a case-by-case basis, to control for inter-tumoral heterogeneity (Figures S3I–S3L). Two major subclones (designated as malignant cells-C1 and -C2) were identified by two distinct lineages in a neighbor-joining tree in these samples. Malignant cells-C2 (MC2) comprised 54.91%, 45.92%, 38.33%, and 24.44% of

the malignant cell population in A304–01, A304–02, A304-03, and A304-04, respectively. Interestingly, the clustering of aneuploid copy number profiles identified clonal amplifications in chromosome 5p and clonal deletions in chromosome 13p that were shared across the four samples (Figure S3B). We then used single-cell gene set variation analysis (GSVA) to assess the differences in biological features between malignant cells-C1 (MC1) and MC2. Notably, the MC2 of these four patients exhibited a pro-metastatic pattern, including cell cycle, epithelial-to-mesenchymal transition (EMT), metastasis, invasion, and other cancer signatures. Interestingly, only the inflammation signature was upregulated in MC1, suggesting a hot immune microenvironment surrounding these cells (Figure 2C).

We then performed a pseudotime analysis for all epithelial cells based on Monocle2 (Figure 2D). As expected, non-malignant epithelial cells gathered at one end. The MC2 subpopulation was mostly clustered at the other end. The MC1 subpopulation presented dispersal patterns and was found to be in an intermediate transition transcriptional state between the normal epithelial cells and MC2. Thus, we speculated that MC2 originated from a subgroup of MC1 in the later stages of tumor evolution. Furthermore, the MSigDB Hallmark GSVA analysis revealed stepwise exacerbating tumor malignancy among normal epithelial cells, in MC1 and MC2 (Figure S4). These results suggest similar evolutionary routes in these tumors, despite the high degree of transcriptional inter-tumor heterogeneity, similar evolutionary routes were observed in these tumors, which is in accordance with the stepwise process of PSN-like LUAD.

Spatial distribution of malignant cells related to evolutionary occurrence time

Although single-cell transcriptional profiling provides a powerful tool to characterize the identity and molecular makeup of malignant cells, information regarding the spatial distribution of MC1 and MC2 is lost. We therefore used the matched ST data to capture more comprehensive spatial information for various cell types. The spatial spot locations of MC1 and MC2 were precisely visualized via FindTransferAnchors on a case-by-case basis (Figure 2E), and two distribution patterns were observed. One pattern displayed a mutually exclusive distribution: MC1 and MC2 were relatively concentrated in different regions (e.g., A304-01, 02). The other pattern displayed a more "mixed" distribution (e.g., A304-03, 04). From the viewpoint of evolutionary order, compared with A304-01 and 02, the MC2 of A304-03 and 04 were more likely to focus on the end of evolutionary routes, which suggests that the occurrence time of MC2 was later in these two patients and earlier in A304-01 and 02 (Figure 2D). We may further speculate that the earlier MC2 appeared, the easier it was for it to give rise to a "spatial subclonal tribe", which was then displayed as a mutually-exclusive spatial distribution. Conversely, the later MC2 appeared, the more the display tended toward a mixed spatial distribution.

To characterize the spatial aggregation of malignant cells, we mapped representative biological functions to the corresponding ST sections. The activity maps roughly exhibited spatial patterns that closely mirrored the spatial distributions of MC1 and MC2 (Figures 2F–2H). The signatures of EMT, invasion, and stemness tended to be upregulated in the MC2 regions, which was corroborated by their pro-metastatic natures.

Malignant cells-C1 corresponds to the pathological lepidic subtype

Next, we attempted to determine the relationship between MC1/C2 and the pathological subtype of LUAD. A previous report by Tavernari et al.[16] elaborates in detail on diverse transcriptional signatures among LUAD histologic subtypes (lepidic, acinar, papillary, and solid). As shown in Figure 3A, a strong correlation was observed between MC1 and the lepidic subtype signature (all $p < 0.001$, Figure 3A). Thus, we performed further pathological annotations of the MC1 and MC2 mapping regions. As expected, the MC1-aggregated regions were enriched for the lepidic component, while the MC2-aggregated regions were enriched for the invasive adenocarcinoma component (Figures 3B and 3C). These data suggest that MC1 may correspond to the lepidic subtype.

Figure 2C shows that the inflammatory signature was upregulated in MC1. Thus, we further examined whether MC1 harbored a hot tumor immune microenvironment (TIME). We used ssGSEA to estimate the infiltrating immune cell composition of the 20 PSN samples and two distinct clusters corresponding to a hot and cold TIME (Figure 3D). As expected, samples with a hot TIME exhibited significantly higher expressions of lepidic patterns and lower expressions of acinar patterns (Figure 3E). We then explored the relevance of the PSN imaging features in this context. It is widely acknowledged that CTR imaging is a strong indicator of the pathological invasiveness of PSN.[17–19] Figure 3F shows that there was a trend toward lower CTR values ($p = 0.099$) for PSNs with a hot TIME. Moreover, CTR positively correlated with the acinar pattern (Rho = 0.64, $p = 0.002$) and negatively correlated with the lepidic pattern (Rho = −0.41, $p = 0.073$, Figure 3G). In addition, compared to acinar expression and CTR, the lepidic expression had a strong positive correlation with most immune cell populations (Figure 3H). Overall, these findings suggest that MC1 correspond to the pathological lepidic subtype and the pure group glass component on imaging and that they tend to have a hot TIME.

No shared molecular regulation mechanisms

We next attempted to determine whether shared molecular mechanisms are involved in the evolutionary routes from MC1 to MC2. We first investigated the transcriptional changes associated with transitional states from normal epithelial cells to MC1 and MC2 in the four samples discussed above. As shown in Figures 4A–4D, several obvious different transcriptional changes occurred during the progression from MC1 to MC2. For example, in A304-04 (Figure 4D), MC1 was predominantly enriched in the intermediate phase, characterized by the upregulated expression of *BMP2*, *KDM6B*, and *MAP2K3*. Pathway analysis indicated that signaling pathways were involved in TNFA signaling, apoptosis, and hypoxia, suggesting that MC1 was under evolutionary pressure. Conversely, MC2 was predominantly enriched in the late phase, characterized by high EMT activity, which corresponds to its pro-metastatic feature. Subsequently, we attempted to find the intersection of upregulated genes involved in the MC1 and MC2 phases of these four samples. However, no obvious candidate genes were identified for this intersection (Figures 4E and 4F). Similarly, single-cell regulatory network inference and clustering (SCENIC) analysis did not reveal any shared

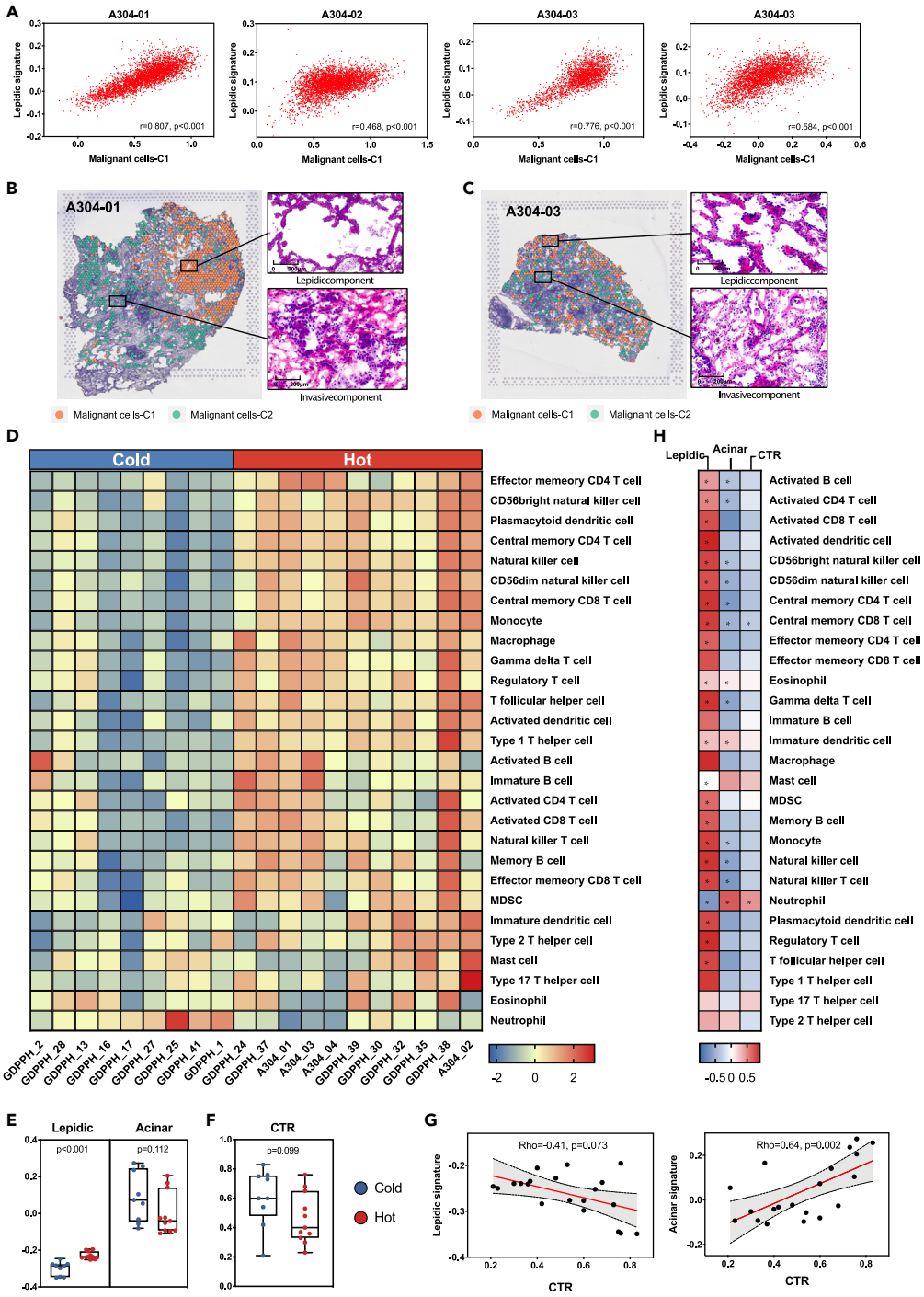


Figure 3. Correlation analysis of MC1/2, pathological annotation, and imaging features

(A) Correlation analysis of MC1 and the lepidic signature.

(B and C) Pathological annotation of MC1- and MC2- aggregated regions of A304-01 and –03 samples.

(D) Clustering heatmap of the estimated immune cell infiltrates by ssGSEA.

(E) Bar plots showing the differences in lepidic and acinar expression between the hot and cold TIME groups.

(F) Bar plots showing the differences in CTR between the hot and cold TIME groups.

(G) Correlation analysis of CTR and the lepidic or acinar signature.

(H) Heatmap of Pearson correlation values for the lepidic signature, acinar signature, CTR, and each immune cell type, estimated based on ssGSEA score. MC1, malignant cells-C1; MC2, malignant cells-C2; PSN, part-solid nodule; CTR, consolidation-to-tumor ratio; TIME, tumor immune microenvironment; ssGSEA, single-sample gene set enrichment analysis.

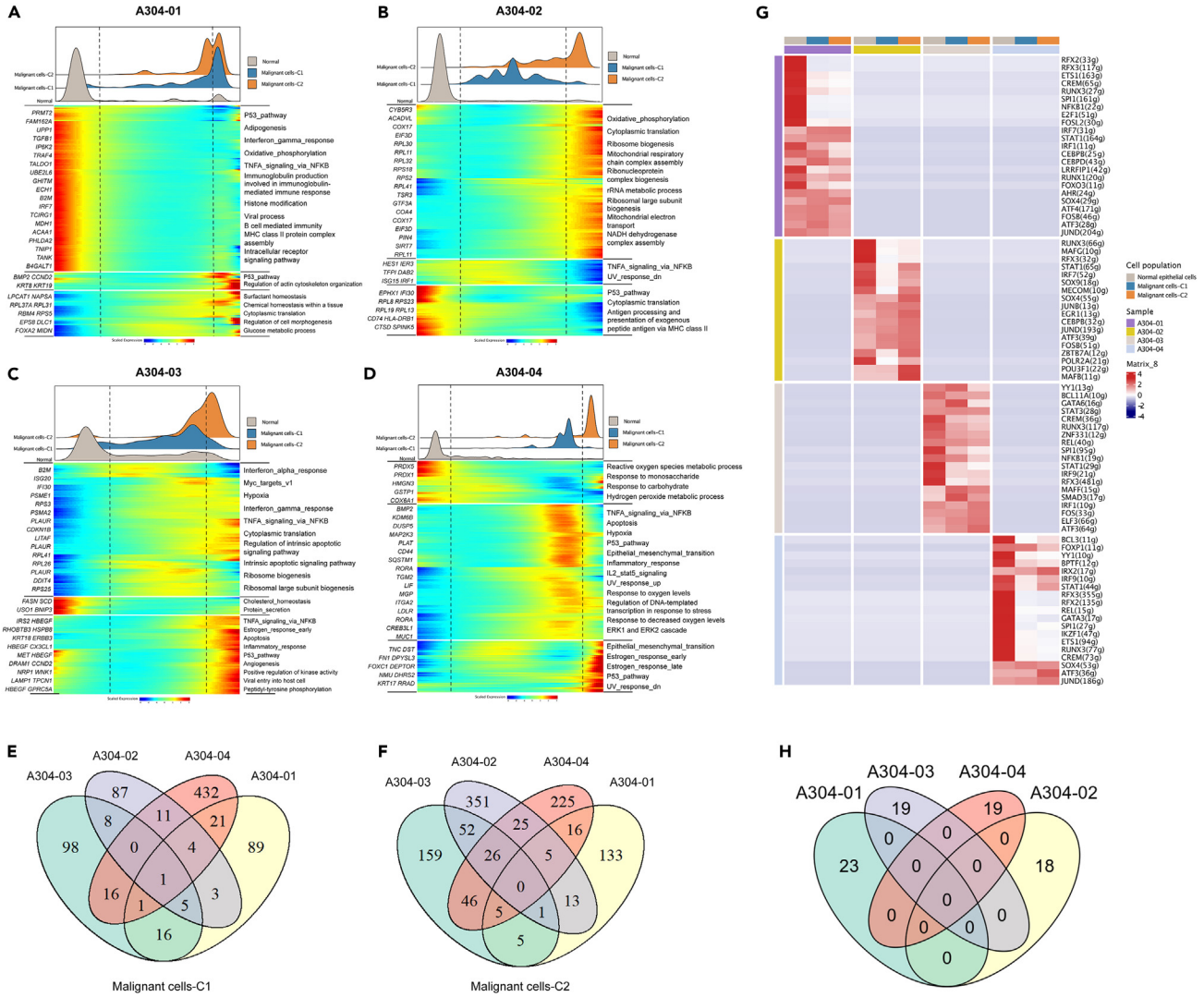


Figure 4. Analysis of the trigger for malignant cell transition

(A–D) Heatmap showing the dynamic changes in gene expression along the pseudotime (lower panel). The distribution of normal epithelial cells, MC1, and MC2, along with the pseudotime (upper panel).

(E) The intersection of upregulated genes involved in the MC1 phase.

(F) The intersection of upregulated genes involved in the MC2 phase.

(G) Heatmap of the area under the curve scores of expression regulation by transcription factors, estimated by SCENIC analysis.

(H) The intersection of upregulated transcription factors involved in MC2. SCENIC, single-cell regulatory network inference and clustering.

gene regulatory networks for these four samples (Figures 4G and 4H). Thus, we were unable to speculate on the shared molecular events that might drive MC1 to evolve into MC2.

Macrophages may be involved in PSN tumor evolution through the CD74 axis

We next investigated which immune cell population was involved in the evolution from MC1 to MC2. We applied the "STdeconvolve" approach to recover the cell-type transcriptional profiles and their proportional representation in each simulated pixel of these 20 ST datasets. The most commonly curated immune cell types included macrophages and plasma cells (Figure 5A; Figure S5). Moreover, $CD4^+$ T cells, plasma cells, macrophages, and fibroblasts were significantly enriched in the MC1 regions, whereas B cells and dendritic cells were enriched in the MC2 ones (all p values <0.001, Figure 5B). We also found that macrophage infiltration, assessed using the ssGSEA score, positively correlated with the lepidic signature (Rho = 0.79, $p < 0.001$, Figure 5C) and negatively correlated with the acinar signature (Rho = −0.5, $p = 0.024$, Figure 5D). Thus, based on the above results, we speculated that macrophages may be associated with the evolution process from MC1 to MC2.

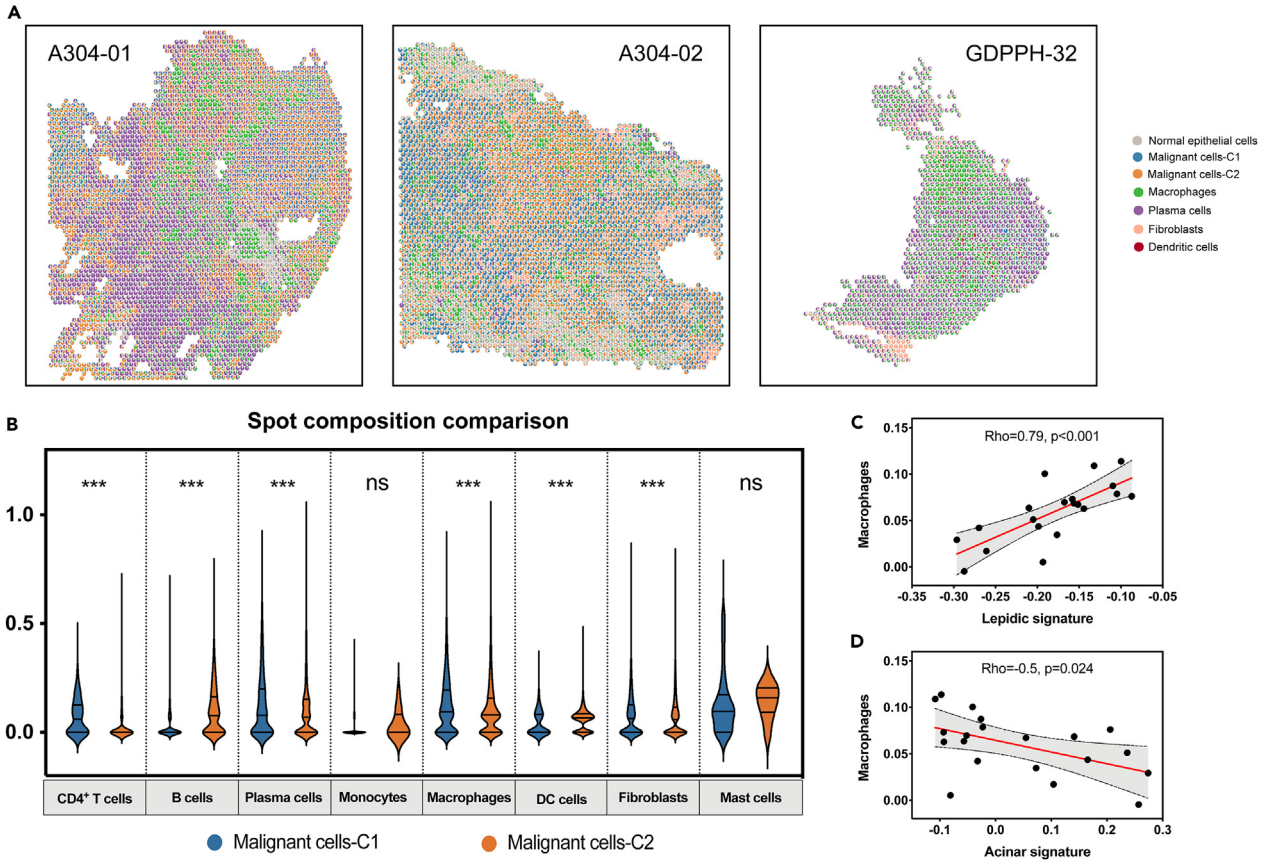


Figure 5. Mapping cell subpopulations across the ST sections to identify immune microenvironments

(A) Spatial scatter pie plot representing proportions of cell types in three representative ST sections.

(B) Cell type proportion comparison within each spot between MC1 and MC2 spots; ns represents $p \geq 0.05$, and *** represents $p < 0.001$.

(C) Correlation analysis of the macrophage score and lepidic signature.

(D) Correlation analysis of the macrophage score and acinar signature.

Cell-cell communication mediated by ligand-receptor complexes is critical for coordinating tumor evolution. We used CellPhoneDB to identify which ligand-receptor pairs displayed significant interactions between MC1 and MC2 (Figures 6A and 6B). Unexpectedly, we found very strong signals of CD74 interactions with multiple tumor microenvironment cell types in MC1 rather MC2, macrophage migration inhibitory factor (MIF)-CD74 was dominant in these communications (Figure 6C). ST co-localization showed that MIF-CD74 was highly expressed in MC1 regions (Figure 6D). We found that CD74 expression in these samples was significantly negatively correlated with CTR (Rho = −0.69, $p < 0.001$, Figure 6E) and the acinar signature (Rho = −0.64, $p = 0.002$, Figure 6F) but positively correlated with the lepidic signature (Rho = 0.73, $p < 0.001$, Figure 6G). Thus, we performed immunohistochemistry staining for CD74 in the PSN samples of GDPPH-30 (see Figure 6H), and found that CD74 expression in the preinvasive tumor region was remarkably higher than that in the adjacent tissue and also slightly higher than that in the invasive region (Figure 6I). Moreover, LUAD patients with high CD74 expression levels had better prognoses based on the data from the TCGA and CHOICE datasets (Figure 6J). Overall, these findings suggest that the activation of CD74 signaling may impede the progression of MC1, thereby serving a protective role in this context.

DISCUSSION

In this study, we focused on the evolutionary processes of malignant cells and spatial distribution characteristics of early-stage LUAD samples manifested as PSN. We highlight two interesting findings of this study. First, two distinct evolutionary states and spatial distributions of malignant epithelial cells were identified (MC1 and MC2), which conceptually correspond to the binary-classified phenotype for PSNs: preinvasive (lepidic subtype) to invasive components on pathology and pure ground glass opacity to solid components on imaging, corresponding to the evolution from MC1 to MC2, respectively. Second, owing to the large heterogeneity between individuals, the molecular mechanisms driving MC1 to MC2 could not been clearly defined. However, considering the wide differences in immune cell proportions and their spatial distributions between MC1 and MC2, immune evasion of these tumor cells is undoubtedly the major step. Macrophages may impede this process through CD74 signaling.

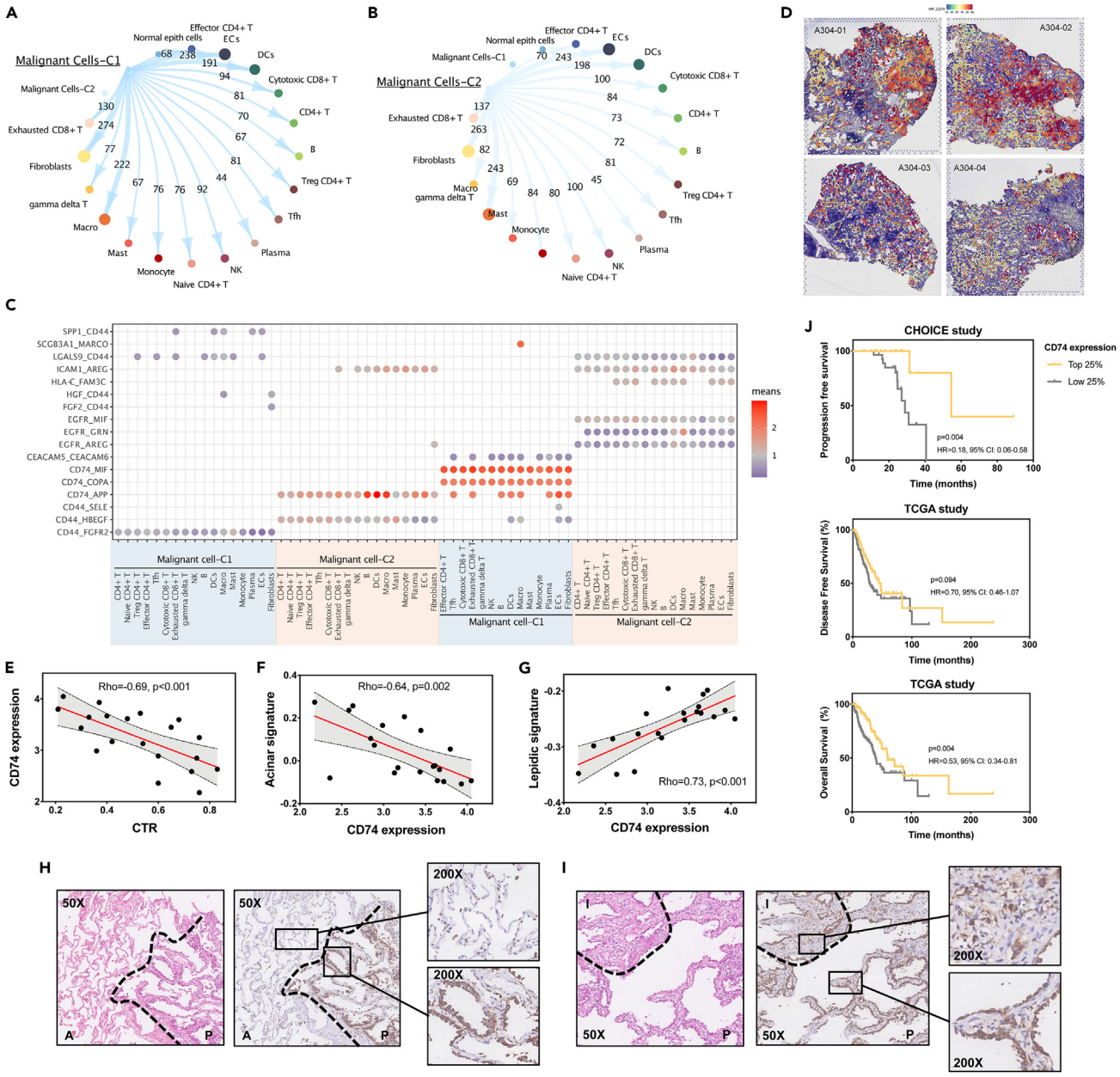


Figure 6. Cell-cell communications with malignant cells

(A and B) Ligand-receptor pairs of distinct cell types and MC1 or MC2.

(C) Dot plot showing ligand-receptor pair analysis of the interactions between distinct cell types and MC1 or MC2.

(D) Co-location of MIF-CD74 ligand-receptor pair on ST sections.

(E) Correlation analysis of CD74 expression and CTR.

(F) Correlation analysis of CD74 expression and the lepidic or acinar signature.

(G–I) Representative images of CD74 staining in two PSN samples. "A" represents adjacent normal tissue, "P" represents the preinvasive region, and "I" represents an invasive region.

(J) Kaplan–Meier survival curves of CD74 expression grouped by top 25% and bottom 25% in LUAD patients (from TCGA data and CHOICE data). MIF, macrophage migration inhibitory factor; ST, spatial transcriptomics; CTR, consolidation-to-tumor ratio; LUAD, lung adenocarcinoma.

PSNs have a strong clinical association with invasiveness acquisition in early-stage LUAD, for example, from atypical adenomatous hyperplasia (AAH) and adenocarcinoma *in situ* (AIS) to MIA and IAC.[7,8,20,21] PSNs have, therefore, been recognized as well-characterized evolutionary tumor models. Comparing individual subjects with ground-glass nodules to those with solid nodules is a classical line of thought.[22–26] With the development of scRNA-seq technologies, a number of studies have further explored the tumor microenvironment constitution of

PSNs at single-cell resolution.[13,15,27,28] In 2021, Xing et al.[13] compared the scRNA-seq data of 16 PSN tumor samples: six adjacent normal lung tissues and nine primary LUAD tissues with lymph node metastasis. Cytotoxic natural killer/T cells were found to be dominant in PSN samples, and some metabolic pathways were highly expressed in malignant PSN cells, including oxidative phosphorylation and arginine and proline metabolism. In a later study, Wang et al.[15] performed scRNA-seq on three AAH, five AIS, nine MIA, 17 IAC, and 18 adjacent normal lung tissues. Interestingly, the authors found that a group of cells displayed features that made them resemble stem-like cells, and that the energy metabolism and ribosome synthesis were progressively upregulated during the early progression of LUAD. More recently, Li et al.[28] manually subjected the two regions (ground glass and solid) from 12 PSNs to scRNA-seq and found that the cytotoxic function and macrophage pro-inflammatory function of T/Nature killer cell subsets were suppressed in solid components. However, given the high individual heterogeneity contained in the scRNA-seq data, these studies did not elaborate on the malignant cells and could thus not account for the binary-classified phenotype for PSNs; this was the question we sought to answer in this study.

Conclusions

In conclusion, we used scRNA-seq and ST technology to explore the intratumor heterogeneity and spatial distribution of PSNs. Two distinct malignant cell populations were identified in our samples. Our scRNA-seq and ST data suggest that malignant cells arise first on the evolutionary route corresponding to the lepidic component on pathology and the ground glass component on imaging. Based on the spatial distribution of immune cells, these two malignant cells harbor a wide degree of immune cell infiltration. Macrophages may impede tumor evolution through CD74 signaling in PSNs. Taken together, our findings contribute to our understanding of the evolutionary patterns of PSNs.

Limitations of the study

The major limitation of this study lies in the 16 ST datasets without matched scRNA-seq data, which significantly affected the analysis of the ST profiles. Despite the ST deconvolve method compensating for this gap, the problem of imprecision still exists. Another limitation is the insufficient resolution of the 10x Chromium ST technology itself. Each pixel contains several or several tens of cells, which can cause interference. Our analyses focused on malignant cell clustering but not on the comparison between tumors and adjacent regions, which may have further amplified the impact of this methodological flaw. Moreover, our predictions of the evolutionary route from MC1 to MC2 were mainly derived from the scRNA-seq data and need further experimental verification.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Patient enrollment and sample collection
  - scRNA sequencing
  - scRNA-seq data preprocessing, integration, and cluster annotation
  - Identifying malignant cells among normal lung epithelial cells
  - Gene set variation analysis
  - Pseudotime trajectory analysis and SCENIC analysis
  - Cell-cell interaction network analysis
  - Spatial transcriptomics
  - ST data processing
  - ST spot scoring and spatial distribution analysis of malignant epithelial cells
  - Gene signature score evaluation
  - Spatial cell-cell interactions
  - STdeconvolve
  - Immunohistochemistry and hematoxylin and eosin staining
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107699.

ACKNOWLEDGMENTS

We are grateful to the patients and their families for their involvement in this study. This work was supported by the Guangdong Provincial People's Hospital Young Talent Project (Grant No. GDPPHYTP201902 to W.Z.Z.), the High-Level Hospital Construction Project (Grant No. DFJH201801 to W.Z.Z.), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019B1515130002 to W.Z.Z.), the Key Lab System Project of Guangdong Science and Technology Department, Guangdong Provincial Key Lab of Translational Medicine in Lung Cancer (Grant No. 2017B030314120 to Y.L.W.), and the Guangdong Provincial People's Hospital Scientific Research Funds for Leading Medical Talents in Guangdong Province (Grant No. KJ012019426 to Y.L.W.). The funding sources had no role in the preparation of this manuscript.

AUTHOR CONTRIBUTIONS

J-T.Z.: Formal analysis, writing of the original draft. J.Z.: Formal analysis, writing of the original draft. J.Q.: Formal analysis. S-R.W.: Resources, methodology, and data curation. K.Y.: Resources and Methodology. L-X.Y.: Formal analysis. X-P.C.: Methodology. M-M.W.: Resources and methodology. H-Z.H.: Resources and methodology. Z-Y.L.: Resources and methodology. Z-H.C.: Data curation. J.S.: Data curation. R-Q.L.: Data curation. S.D.: Investigation. B.Y.J.: Data curation. X-C.Z.: Methodology. X.L.: Formal analysis and supervision. Y-L.W.: Supervision, funding acquisition, writing, review, and editing. W-Z.Z.: Supervision, funding acquisition, writing, review, and editing.

DECLARATION OF INTERESTS

W-Z.Z. declares honoraria from AstraZeneca, BMS, MSD, Roche, and Innovent outside the submitted work. Y-L.W. declares advisory services for AstraZeneca, Boehringer Ingelheim, Novartis, and Takeda; speaker fees from AstraZeneca, Beigene, Boehringer Ingelheim, BMS, Eli Lilly, MSD, Pfizer, Roche, and Sanofi; and grants from AstraZeneca, Boehringer Ingelheim, BMS, Hengrui, and Roche outside the submitted work. The authors declare no potential conflicts of interest.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: January 30, 2023
Revised: May 23, 2023
Accepted: August 18, 2023
Published: August 21, 2023

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J. Clin. *71*, 209–249.
2. Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M., and Sicks, J.D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. *365*, 395–409.
3. de Koning, H.J., van der Aalst, C.M., de Jong, P.A., Scholten, E.T., Nackaerts, K., Heuvelmans, M.A., Lammers, J.W.J., Weenink, C., Yousaf-Khan, U., Horeweg, N., et al. (2020). Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N. Engl. J. Med. *382*, 503–513.
4. Yankelevitz, D.F., Yip, R., Smith, J.P., Liang, M., Liu, Y., Xu, D.M., Salvatore, M.M., Wolf, A.S., Flores, R.M., and Henschke, C.I.; International Early Lung Cancer Action Program Investigators Group (2015). CT Screening for Lung Cancer: Nonsolid Nodules in Baseline and Annual Repeat Rounds. Radiology *277*, 555–564.
5. Scholten, E.T., de Jong, P.A., de Hoop, B., van Klaveren, R., van Amelsvoort-van de Vorst, S., Oudkerk, M., Vliegenthart, R., de Koning, H.J., van der Aalst, C.M., Vernhout, R.M., et al. (2015). Towards a close computed tomography monitoring approach for screen detected subsolid pulmonary nodules? Eur. Respir. J. *45*, 765–773.
6. Henschke, C.I., Yip, R., Smith, J.P., Wolf, A.S., Flores, R.M., Liang, M., Salvatore, M.M., Liu, Y., Xu, D.M., and Yankelevitz, D.F.; International Early Lung Cancer Action Program Investigators (2016). CT Screening for Lung Cancer: Part-Solid Nodules in Baseline and Annual Repeat Rounds. AJR Am. J. Roentgenol. *207*, 1176–1184.
7. Chae, H.D., Park, C.M., Park, S.J., Lee, S.M., Kim, K.G., and Goo, J.M. (2014). Computerized texture analysis of persistent part-solid ground-glass nodules: Differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. Radiology *273*, 285–293.
8. Kakinuma, R., Noguchi, M., Ashizawa, K., Kuriyama, K., Maeshima, A.M., Koizumi, N., Kondo, T., Matsuguma, H., Nitta, N., Ohmatsu, H., et al. (2016). Natural History of Pulmonary Subsolid Nodules: A Prospective Multicenter Study. J. Thorac. Oncol. *11*, 1012–1028.
9. Baslan, T., and Hicks, J. (2017). Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat. Rev. Cancer *17*, 557–569.
10. Ren, X., Zhang, L., Zhang, Y., Li, Z., Siemers, N., and Zhang, Z. (2021). Insights Gained from Single-Cell Analysis of Immune Cells in the Tumor Microenvironment. Annu. Rev. Immunol. *39*, 583–609.
11. Qiu, Z.B., Zhang, C., Chu, X.P., Cai, F.Y., Yang, X.N., Wu, Y.L., and Zhong, W.Z. (2022). Quantifying invasiveness of clinical stage IA lung adenocarcinoma with computed tomography texture features. J. Thorac. Cardiovasc. Surg. *163*, 805–815.e3.
12. Travis, W.D., Asamura, H., Bankier, A.A., Beasley, M.B., Detterbeck, F., Flieder, D.B., Goo, J.M., MacMahon, H., Naidich, D., Nicholson, A.G., et al. (2016). The IASLC Lung Cancer Staging Project: Proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid Tumors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer. J. Thorac. Oncol. *11*, 1204–1223.
13. Xing, X., Yang, F., Huang, Q., Guo, H., Li, J., Qiu, M., Bai, F., and Wang, J. (2021). Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. Sci. Adv. *7*, eabd9738.
14. He, Y., Yu, F., Tian, Y., Hu, Q., Wang, B., Wang, L., Hu, Y., Tao, Y., Chen, X., and Peng, M. (2022). Single-Cell RNA Sequencing Unravels Distinct Tumor Microenvironment of Different Components of Lung Adenocarcinoma Featured as Mixed Ground-Glass Opacity. Front. Immunol. *13*, 903513.
15. Wang, Z., Li, Z., Zhou, K., Wang, C., Jiang, L., Zhang, L., Yang, Y., Luo, W., Qiao, W., Wang, G., et al. (2021). Deciphering cell lineage specification of human lung adenocarcinoma

with single-cell RNA sequencing. Nat. Commun. *12*, 6500.
16. Tavernari, D., Battistello, E., Dheilly, E., Petruzzella, A.S., Mina, M., Sordet-Dessimoz, J., Peters, S., Krueger, T., Gfeller, D., Riggi, N., et al. (2021). Nongenetic Evolution Drives Lung Adenocarcinoma Spatial Heterogeneity and Progression. Cancer Discov. *11*, 1490–1507.
17. Mazzone, P.J., and Lam, L. (2022). Evaluating the Patient With a Pulmonary Nodule: A Review. JAMA *327*, 264–273.
18. Lee, K.H., Goo, J.M., Park, S.J., Wi, J.Y., Chung, D.H., Go, H., Park, H.S., Park, C.M., and Lee, S.M. (2014). Correlation between the size of the solid component on thin-section CT and the invasive component on pathology in small lung adenocarcinomas manifesting as ground-glass nodules. J. Thorac. Oncol. *9*, 74–82.
19. Ye, T., Deng, L., Xiang, J., Zhang, Y., Hu, H., Sun, Y., Li, Y., Shen, L., Wang, S., Xie, L., and Chen, H. (2018). Predictors of Pathologic Tumor Invasion and Prognosis for Ground Glass Opacity Featured Lung Adenocarcinoma. Ann. Thorac. Surg. *106*, 1682–1690.
20. Hattori, A., Matsunaga, T., Takamochi, K., Oh, S., and Suzuki, K. (2017). Importance of Ground Glass Opacity Component in Clinical Stage IA Radiologic Invasive Lung Cancer. Ann. Thorac. Surg. *104*, 313–320.
21. McWilliams, A., Tammemagi, M.C., Mayo, J.R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., et al. (2013). Probability of cancer in pulmonary nodules detected on first screening CT. N. Engl. J. Med. *369*, 910–919.
22. Kobayashi, Y., Mitsudomi, T., Sakao, Y., and Yatabe, Y. (2015). Genetic features of pulmonary adenocarcinoma presenting with ground-glass nodules: the differences between nodules with and without growth. Ann. Oncol. *26*, 156–161.
23. Li, Y., Li, X., Li, H., Zhao, Y., Liu, Z., Sun, K., Zhu, X., Qi, Q., An, B., Shen, D., et al. (2020). Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. Eur. Respir. J. *55*, 1901409.
24. Zhang, C., Zhang, J., Xu, F.P., Wang, Y.G., Xie, Z., Su, J., Dong, S., Nie, Q., Shao, Y., Zhou, Q., et al. (2019). Genomic Landscape and Immune Microenvironment Features of Preinvasive and Early Invasive Lung Adenocarcinoma. J. Thorac. Oncol. *14*, 1912–1923.
25. Chen, K., Bai, J., Reuben, A., Zhao, H., Kang, G., Zhang, C., Qi, Q., Xu, Y., Hubert, S., Chang, L., et al. (2021). Multiomics Analysis Reveals Distinct Immunogenomic Features of Lung Cancer with Ground-Glass Opacity. Am. J. Respir. Crit. Care Med. *204*, 1180–1192.
26. Hu, X., Fujimoto, J., Ying, L., Fukuoka, J., Ashizawa, K., Sun, W., Reuben, A., Chow, C.W., McGranahan, N., Chen, R., et al. (2019). Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. Nat. Commun. *10*, 2978.
27. Lu, T., Yang, X., Shi, Y., Zhao, M., Bi, G., Liang, J., Chen, Z., Huang, Y., Jiang, W., Lin, Z., et al. (2020). Single-cell transcriptome atlas of lung adenocarcinoma featured with ground glass nodules. Cell Discov. *6*, 69.
28. Li, Y., Li, X., Chen, H., Sun, K., Li, H., Zhou, Y., Wang, J., Bai, F., and Yang, F. (2022). Single-cell RNA sequencing reveals the multi-cellular ecosystem in different radiological components of pulmonary part-solid nodules. Clin. Transl. Med. *12*, e723.
29. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Core Team).
30. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.
31. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.
32. Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat. Biotechnol. *39*, 599–608.
33. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411.
34. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7.
35. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.
36. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141.
37. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.
38. Garcia-Alonso, L., Handfield, L.F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nat. Genet. *53*, 1698–1711.
39. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088.
40. Pham, D., Tan, X., Xu, J., Grice, L., Lam, P.Y., Raghubar, A., Vukovic, J., Ruitenberg, M., and Nguyen, Q. (2020). stLearn: Integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. Preprint at bioRxiv125658. https://doi.org/10.1101/2020.05.31.125658.
41. Miller, B.F., Huang, F., Atta, L., Sahoo, A., and Fan, J. (2022). Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. Nat. Commun. *13*, 2339.
42. Zhang, Y., Yao, Y., Xu, Y., Li, L., Gong, Y., Zhang, K., Zhang, M., Guan, Y., Chang, L., Xia, X., et al. (2021). Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. Nat. Commun. *12*, 11.
43. Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. *18*, 248–262.
44. Zhang, X.C., Wang, J., Shao, G.G., Wang, Q., Qu, X., Wang, B., Moy, C., Fan, Y., Albertyn, Z., Huang, X., et al. (2019). Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. Nat. Commun. *10*, 1772.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal [LN2] anti-CD74	Sigma-Aldrich	RRID: AB_9514
Biological samples		
PSN tumor tissue	Guangdong Provincial People's Hospital	See Table S1 for details
Chemicals, peptides, and recombinant proteins		
MACS Tissue Storage Solution	Miltenyi	Lot#5130425019
Tumor Dissociation Kit human	Miltenyi	Cat#130095929
Phosphate Buffer Saline (PBS,1X)	Corning	Lot#21020007
Critical commercial assays		
Chromium Single Cell 3′ Library & Gel BeadKit v3.1	10x Genomics	Cat#PN-1000075
Chromium Single Cell B Chip Kit	10x Genomics	Cat#PN-1000074
Visum spatial Library construction kit	10x Genomics	Cat# PN-1000184
Deposited data		
Raw data files for scRNA-seq and ST	This study	GSA: HRA003826
Software and algorithms		
R (v4.1.0)	R Core[29]	https://www.r-project.org/
cellranger (v4.0.0)	10x Genomics[30]	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
Seurat (v4.0.2)	Hao et al.[31]	https://satijalab.org/seurat/
CopyKAT (v1.0.5)	Gao et al.[32]	https://github.com/navinlabcode/copykat/
CytoTRACE (v0.3.3)	Gulati et al.[33]	https://cytotrace.stanford.edu/
GSVA (v1.34.0)	Hänzelmann et al.[34]	https://www.bioconductor.org/packages/release/bioc/html/GSVA.html
Monocle2 (v2.12.0)	Qiu et al.[35]	http://cole-trapnell-lab.github.io/monocle-release/docs/
clusterProfiler (v4.2.2)	Wu et al.[36]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
SCENIC (v1.2.4)	Aibar et al.[37]	https://scenic.aertslab.org/
pheatmap (v1.0.12)	/	https://cran.r-project.org/web/packages/pheatmap/index.html
CellPhoneDB (v3.0.0)	Garcia-Alonso et al.[38]	https://www.cellphonedb.org/
CellChat (v1.5.0)	Jin et al.[39]	https://github.com/sqjin/CellChat
Space Ranger (v.1.0.0)	10x Genomics	https://support.10xgenomics.com/spatial-gene-expression/software/downloads/latest
stLearn (v.0.4.3)	Pham et al.[40]	https://stlearn.readthedocs.io/en/latest/index.html
STdeconvolve (v.0.99.12)	Miller et al.[41]	https://github.com/JEFworks-Lab/STdeconvolve

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author, Wen-Zhao Zhong, M.D., Ph.D. (zhongwenzhao@gdph.org.cn).

Materials availability

No new unique reagents or materials were generated in this study.

Data and code availability

- The single-cell data supporting the findings of this study including raw.fcs files have been deposited at Genome Sequence Archive (HRA003826), which is publicly available.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the corresponding author upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The clinical characteristics of these patients are listed in Table S1. This study was approved by the Institutional Review Board of Guangdong Provincial People's Hospital (approval no. 2020092H). The expedient model in this study was limited to human subjects. Written informed consent was obtained from all patients. All the patients were diagnosed with PSN lung adenocarcinoma from Guangdong Provincial People's Hospital, China, age range: 38–79 years of age, sex: n = 5 males, n = 15 females.

METHOD DETAILS

Patient enrollment and sample collection

Twenty patients with pulmonary PSNs who underwent surgery at Guangdong Provincial People's Hospital were enrolled, using the following criteria: (1) pulmonary PSNs on CT imaging; (2) tumor diameter at least 0.8 cm; (3) and pathologically diagnosed with LUAD, including minimally invasive adenocarcinoma and invasive adenocarcinoma. Four of these PSN samples (A 304-01, −02, −03, and −04) were cut into three pieces along the long axis to perform ST, scRNA-seq, and pathological annotation. The remaining 16 PSN samples underwent ST and pathological annotation (Figure S6). In addition, all samples underwent a next-generation sequencing assay of 1021 genes[42] derived from the specimens remaining after the pathological annotations.

scRNA sequencing

Using a Chromium Next GEM Single Cell 3' Reagent Kit v3.1 and Gel Bead Kit V3.1 (1000075; 10x Genomics Inc., Pleasanton, CA, USA and Chromium Single Cell B Chip Kit (1000074; 10x Genomics Inc.), the cell suspension (300–600 living cells per microliter, determined by Count Star [Alit Biotech Co., Ltd., Shanghai, China]) was loaded onto the Chromium single cell controller (10x Genomics Inc.) to generate single-cell gel beads in the emulsion according to the manufacturer's protocol. Briefly, single cells were suspended in PBS containing 0.04% bovine serum albumin. Approximately 14,000 cells were added to each channel, and the recovered target cells were estimated to amount to approximately 7,000 cells. Captured cells were lysed, and the released RNA was barcoded through reverse transcription in individual gel bead-in-emulsion (GEMs). The obtained cDNA was then amplified to generate single-cell libraries according to the manufacturer's protocol. The quality of the cDNA was assessed using an Agilent 4200 Tapestation electrophoresis system (Agilent Technologies Inc., Santa Clara, CA, USA).

Single-cell RNA-seq libraries were constructed using the Single Cell 3′ Library and the Gel Bead Kit V3.1, according to the manufacturer's instructions. The libraries were sequenced using an Illumina Novaseq6000 sequencer (Illumina Inc., San Diego, CA, USA) with a sequencing depth of at least 100,000 reads per cell using a paired-end 150 bp (PE150) reading strategy. Raw BCL files were demultiplexed and mapped to the reference genome GRCh38 using Cell Ranger v.4.0.0 software (10x Genomics Inc.).

scRNA-seq data preprocessing, integration, and cluster annotation

Raw gene expression matrices were generated using CellRanger[30] and analyzed using the Seurat R package (v4.0.2)[31] in R v4.1.0 (www.r-project.org).[29] All cells expressing <300 genes were removed, as were cells with <3% ribosomal genes, >0.1% hemoglobin genes, and >20% mitochondrial counts. Genes expressed in fewer than three cells were likewise removed. The Seurat default parameters were used unless stated otherwise. For the clustering of all cell types, 2,000 variable genes were identified, and principal component analysis (PCA) was applied to the dataset to reduce dimensionality after regressing the percentage of mitochondrial genes. Data integration was performed using the Seurat functions FindIntegrationAnchors and IntegrateData. Clustering was conducted using the FindClusters function with 30 PCA components and a resolution parameter set to 1.2 and visualized using UMAP with the RunUMAP function. Cell types were annotated based on canonical cell-type markers (Table S2). Clusters that expressed two or more canonical cell type markers were classified as doublet cells and were excluded from further analyses. Differentially expressed genes (DEGs) that functionally characterized the clusters were defined using the default Wilcoxon rank-sum test implemented in the FindAllMarkers function from Seurat, with the parameters set to only.pos = TRUE, min.pct = 0.25, and logfc = 0.25. Genes were considered differentially expressed if their p values were below 0.05.

Identifying malignant cells among normal lung epithelial cells

Cells previously annotated as epithelial cells in each sample were applied to a subset and separately re-clustered, using the methods described above and the following parameters: nfeatures = 2,000, npcs = 20, and resolution = 0.5. Malignant epithelial cells were identified

based on genome-wide copy number profiles computed from the gene expression UMI matrix using the Bayesian segmentation approach, CopyKAT (v1.0.5).[32] B cells were used as references; single aneuploid cells with genome-wide copy number aberrations were predicted to be tumor cells, while diploid cells were predicted to be normal epithelial cells, and we divided the resulting cellular clusters into two molecular subtypes. CopyKat-based predictions were confirmed using CytoTRACE (v0.3.3).[33]

Gene set variation analysis

To determine which functional pathway work in heterogenic cell clusters, we used the GSVA (v1.34.0)[34] R package to perform gene set variation analyses based on four published histological subtype gene sets of LUAD taken from a previous study by Tavernari et al.[16] Pathways with p values of <0.05 were considered for further analysis. The upregulated and downregulated pathways were visualized using a bar plot or heatmap. The CancerSEA database (http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp) was used to analyze cell function in 14 different states, based on single-cell data of epithelial cells.

Pseudotime trajectory analysis and SCENIC analysis

Pseudotime analysis was performed with Monocle2 (v2.12.0)[35] to determine the key translational relationships between cell types and clusters. Further analysis using the Monocle2 plot_pseudotime_heatmap function revealed the key role of each series of genes in the differentiation process. Altered genes were identified using the differential GeneTest function in Monocle2 with a cutoff p value of <0.001. Gene ontology (GO) enrichment for genes with significantly changed expression was performed in R using clusterProfiler (v4.2.2)[36] with a p value cutoff of <0.05.

To measure the difference between cell clusters based on transcription factors or their target genes, we used the R package SCENIC (v1.2.4)[37] to run SCENIC and GENIE3 co-expression modules and analyzed them for motif enrichment with RcisTarget. The motif database for humans was downloaded from the website https://resources.aertslab.org/cistarget/databases/homo_sapiens/hg19/. The input matrix was the count expression matrix obtained from Seurat. We then visualized the cell type regulons based on their AUCell values using the R package pheatmap (v1.0.12).

Cell-cell interaction network analysis

The CellPhoneDB v3.0.0 database (https://www.cellphonedb.org)[38] was used to investigate cell-cell interactions between different cell types, especially between tumor cells and other cell types based on single-cell data. The gene-cell normalized count matrix data, cell type annotation information, and ligand-receptor pairs were input to CellPhoneDB with a threshold of 0.1. Significant ligand-receptor pairs with p values of <0.05 and mean values of ≥ 1 were selected for subsequent analyses. To capture more significant interactions, we used the CellPhoneDB method in degs_analysis mode to filter the ligand-receptor pairs. We then visualized ligand-receptor pairs based on the CellPhoneDB results using the netVisual_circle function included in the R package CellChat (v1.5.0).[39]

Spatial transcriptomics

Tissue samples were embedded in an optimal cutting temperature compound and stored at −80°C. Tissue blocks were cut into 10-μm sections and processed using the Visium Spatial Gene Expression Kit (10x Genomics) according to the manufacturer's instructions. First, the tissue permeabilization conditions were optimized using the Visium Spatial Tissue Optimization Kit, which yielded ideal results at 12 min. The sections were stained with hematoxylin and eosin (H&E) and imaged using a Leica DM6000 microscope (Leica Camera AG, Wetzlar, Germany) at a 20x lens magnification. They were then annotated by pathologists based on the H&E staining, processed for spatial transcriptomics, and the constructed complementary DNA library was accessed for quality control and sequenced using an Illumina NovaSeq 6000 system (Illumina Inc.).

ST data processing

Reads were demultiplexed and mapped to the reference genome GRCh38 using Space Ranger software (v.1.0.0) (10x Genomics). Count matrices were loaded into Seurat for subsequent data filtering, normalization, dimensional reduction, and visualization. Data from independent tissue sections were normalized using the variance-stabilizing transformation method implemented in the SCTransform function of Seurat. Dimensionality reduction and clustering were performed using PCA at a resolution of 0.8 for the first 30 PCAs.

ST spot scoring and spatial distribution analysis of malignant epithelial cells

The top 30 specifically highly-expressed genes of MC1 and MC2, according to the scRNA-seq results, were selected according to avg_log2FC. Four cases of spatial transcriptome data with paired single-cell data were used to score each spot in the space for the top 30 genes using the AddModuleScore function of the Seurat package. The top 25% of spots with higher scores were defined as MC1 or MC2 areas.

Gene signature score evaluation

For 20 cases of spatial transcriptome data, three gene sets were scored using the GSVA function, which included gene signature sets of 28 immune cell subpopulations from Charoentong et al.,[43] the CancerSEA database pathway gene set, and four published histological subtype gene sets of LUAD taken from a previous report by Tavernari et al.[16]

Spatial cell-cell interactions

To study cell-cell interactions, we used stLearn (v.0.4.3)[40] to combine cell type diversity and L-R co-expression into an interaction measure that can be used to automatically scan a whole tissue section. We identified regions with high L-R interactions and diverse cell types. These regions with high cell diversity and L-R activities are considered hotspots in the tissue, most likely exhibiting cell-cell interaction activity. To investigate biological inferences drawn from the cell-cell interaction analysis, we checked the co-expression patterns of a series of ligand-receptor interaction pairs from the CellPhoneDB results in the scRNA-seq data, the stLearn co-localization results were used for spatial visualization.

STdeconvolve

We used the STdeconvolve (v.0.99.12)[41] method to deconvolute each spatial transcriptome sample and obtain the spatially deconvolved cell types. To annotate the deconvolved cell types, we calculated Pearson correlations between the transcriptional profiles of the deconvolved cell types and those of the annotated scRNA-seq data. Each deconvolved cell type was paired with the reference cell type with the highest correlation. For four samples with scRNA-seq data, we used the data from the samples themselves for this analysis. For 16 samples without scRNA-seq data, we combined the data of four single-cell samples as the cell type reference.

Immunohistochemistry and hematoxylin and eosin staining

The tumor tissues were fixed in 10% formalin and embedded in paraffin. Paraffin-embedded specimens were sliced, deparaffinized, rehydrated, and stained with H&E. For immunohistochemical staining, deparaffinized slides were subjected to antigen retrieval and probed with mouse anti-human CD74 antibody (1:300, ab9514; Abcam, Cambridge, UK) overnight. On the second day, the sections were incubated with secondary antibodies for 1 h, washed, and incubated with horseradish peroxidase streptavidin.

QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical method was used to pre-determine the sample size. Statistical significance for DEGs was determined using the Wilcoxon rank-sum test, with all p values adjusted using the Bonferroni correction. All boxplots depict the first and third quartiles as lower and upper bounds, respectively. The whiskers represent 1.5x the interquartile range (IQR), and the center depicts the median. All statistical tests are defined in the figure legends. Survival curves were drawn using the Kaplan–Meier method from the TCGA LUAD and the CTONG 1308/CHOICE datasets.[44] The log rank test was used to evaluate the differences between the survival curves.