communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-09162-w

Single-cell and spatial transcriptomic characterization of pulmonary pleomorphic carcinoma

Atsushi Matsuoka[1], Kazuhiko Shien[1] ✉, Shuta Tomida[2], Masayoshi Ohki[1], Kazuya Hisamatsu[1], Ryota Fujiwara[1], Kosei Ishimura[1], Ryunosuke Fujii[1], Tomoaki Higashihara[1], Naohiro Hayashi[1], Kazuhiro Okada[1], Ryo Yoshichika[1], Fumiaki Mukohara[1], Mao Yoshikawa[1], Yuma Fukumoto[1], Ken Suzawa[1], Yasuaki Tomioka[1], Shin Tanaka[1], Kentaroh Miyoshi[1], Mikio Okazaki[1], Seiichiro Sugimoto[1], Yusuke Otani[3], Atsushi Tanaka[3], Hirofumi Inoue[2], Yosuke Togashi[4,5,6], Hidetaka Yamamoto[7], Daisuke Ennishi[2] & Shinichi Toyooka[1]

Pulmonary pleomorphic carcinoma (PPC) is a rare subtype of lung cancer that comprises both epithelial and sarcomatoid components. The molecular basis of PPC, including the cellular dynamics of its components, remains largely unknown. To elucidate potential therapeutic targets for PPC, we perform a multi-omics analysis incorporating digital spatial profiling and single-cell RNA sequencing (scRNA-seq). PPC exhibits diverse driver gene alterations, including MET exon 14 skipping mutation (METex14) and ALK fusion. In spatial transcriptomics, MET gene and protein are overexpressed exclusively within the epithelial component and not in the sarcomatoid component, even in patients harboring METex14. Epithelial-mesenchymal transition (EMT)-related transcriptional changes, along with extracellular matrix (ECM) remodeling between the epithelial and sarcomatoid components, are observed. scRNA-seq identifies cell populations within the epithelial component that contribute to the malignant transformation and differentiation of the sarcomatoid component. They are characterized by an intermediate EMT state with ECM remodeling signature, suggesting their potential as novel therapeutic targets for PPC.

Pulmonary pleomorphic carcinoma (PPC), a typical histologic subtype of pulmonary sarcomatoid carcinoma, is rare, accounting for approximately 0.4% of non-small cell lung cancer (NSCLC)[1]. The diagnosis of PPC can only be made using surgical specimens, as it requires histological evaluation of sarcomatoid component mixed with at least 10% of NSCLC components, such as lung adenocarcinoma (LUAD) or squamous cell carcinoma (LUSC)[2]. PPC is typically highly malignant, resistant to conventional cytotoxic chemotherapy and radiotherapy, and lacks established treatment options. Despite some advancements, the molecular basis of PPC including the cellular dynamics underlying its histologically distinct components, remains largely elusive.

In the treatment of NSCLC, molecular targeted therapeutics and immune checkpoint inhibitors have significantly improved patient prognosis. *MET* exon 14 skipping mutation (*MET*ex14), frequently found in pulmonary sarcomatoid carcinomas, presents a treatment candidate for PPC[3]. In addition, previous studies have reported high programmed death ligand 1 (PD-L1) expression in pulmonary sarcomatoid carcinomas, making immunotherapy a treatment option[4]. Nevertheless, the prognosis for PPC patients is still unsatisfactory. Conventional biomarker tests were conducted by the targeted DNA/RNA sequencing and immunohistochemistry (IHC) of the pathological specimen. In recent, several comprehensive molecular profiling of pulmonary sacromatoid carcinomas,

[1]Department of General Thoracic Surgery and Breast and Endocrinological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan. [2]Center for Comprehensive Genomic Medicine, Okayama University Hospital, Okayama, Japan. [3]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA. [4]Department of Tumor Microenvironment, Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan. [5]Department of Allergy and Respiratory Medicine, Okayama University Hospital, Okayama, Japan. [6]Faculty of Medicine, Kindai University, Osaka, Japan. [7]Department of Pathology and Oncology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan. ✉e-mail: k.shien@okayama-u.ac.jp

including PPC have been reported[5,6]. These studies revealed that epithelial-to-mesenchymal transition (EMT) plays an important role in the carcinogenesis and evolution of PPC. However, what ignites this transition and how this transition takes place remains far from being elucidated. A major limitation of these studies is that they were performed using bulk tumor specimens, and comprehensive analyses precisely separating the epithelial tumor component (Epithelial) from the sarcomatoid tumor component (Sarcomatoid) were impossible. To address these limitations and elucidate potential therapeutic targets for PPC, we performed a multi-omics analysis incorporating digital spatial profiling and single-cell RNA sequencing (scRNA-seq). Regions of interest (ROIs) were carefully defined to isolate the Epithelial, Sarcomatoid, tumor stroma, and non-tumor areas. Comprehensive gene expression data were obtained using the GeoMx digital spatial profiler (DSP). Furthermore, differentiation trajectory analysis using scRNA-seq provided insights into the cellular transitions underlying heterogeneous tumor cell populations. We revealed distinctive cell populations involved in the malignant transformations of PPC, suggesting potential therapeutic strategies targeting specific tumor components.

Results

Patient characteristics and genomic alteration profiling

As shown in Fig. 1A, we performed multi-omics analysis—including WES, bulk RNA-seq, scRNA-seq, and spatial transcriptomics—on surgically resected PPC samples. For spatial transcriptomics, formalin-fixed paraffin-embedded (FFPE) sections were profiled using the GeoMx DSP. ROIs were carefully selected based on histological features identified in HE-stained sections and morphology markers (Fig. 1B). Patient characteristics and sequenced data availability are summarized in Supplementary Table 1. Whole-exome sequencing (WES) (all nine patients) and Bulk RNA sequencing (RNA-seq) (six patients) were used to identify the somatic mutational background of this cohort. Figure 1C summarizes nonsynonymous mutations identified by WES, grouped by signaling pathways. Notably, *MET*ex14 was detected in two out of nine samples (Supplementary Fig. 1A). In contrast, no mutations were identified in *EGFR*. Regarding tumor mutational burden (TMB), there were no significant differences between PPC and the two NSCLC cohorts (LUAD and LUSC) in The Cancer Genome Atlas (TCGA) dataset (Fig. 1D). Mutations per signaling pathways revealed that RTK-RAS and Hippo pathways were mutated in all nine patients (Supplementary Fig. 1B). Single base substitution and double base substitution distributions are shown in Supplementary Fig. 1C, and mutational signature analysis is shown in Supplementary Fig. 1D. The most prominent signature observed across samples is COSMIC4, associated with tobacco smoking.

Gene fusions found in transcripts, based on RNA-seq are shown in Fig. 1E. *EML4*::*ALK* fusion, a well-known driver mutation in LUAD, was detected in one patient (OPL7) (Supplementary Fig. 1E). This patient exhibited a low TMB and no other significant driver alterations except for *SMARCA4*::*POLRMT* fusion (Fig. 1C, E, and Supplementary Fig. 1F).

Spatial transcriptomics of PPC

Overview of spatial transcriptomics and ROI characteristics. Digital spatial transcriptomics using the GeoMx Whole Transcriptome Atlas (WTA; NanoString Technologies) enables comprehensive quantification of protein-coding transcripts in spatially selected regions of tissue sections (ROIs), by combining barcoded probes with ultraviolet (UV) light-based oligo release (Fig. 2A). Profiling was performed according to the manufacturer's protocol (Supplementary Materials). After preprocessing, a total of 114 ROIs (see also Supplementary Table 2 for a summary of ROI counts) and 11327 genes were retained. ROIs ranged in area from 11115.1 to 121553.1 μm² (median: 70449.9 μm²), and were categorized into Full (without segmentation), Immune (CD45+), and Tissue (CD45−) ROIs based on segmentation strategy. Representative fluorescence images of selected ROIs are shown in Supplementary Fig. 2A, and Full ROIs were used for the statistical analyses in this section (section of spatial transcriptomics), as more Full ROIs were retained after quality control than Tissue ROIs (Supplementary Table 2).

Multivariate analysis of gene expression

Unsupervised multivariate analysis (Uniform Manifold Approximation and Projection [UMAP] and hierarchical clustering) revealed distinct clusters representing the Epithelial ($n = 14$), Sarcomatoid ($n = 18$), and normal lung ($n = 13$). The UMAP displayed distinct spatial patterns, with each component occupying separate regions, reflecting transcriptional differences (Fig. 2B and Supplementary Fig. 2B). Similarly, hierarchical clustering confirmed the presence of well-defined clusters (Fig. 2C). These findings emphasize the heterogeneity between Epithelial and Sarcomatoid. Notably, all metastatic lymph node samples ($n = 6$), pathologically adenocarcinoma, were grouped together with the epithelial adenocarcinoma component, reflecting their shared molecular characteristics. Intratumoral heterogeneity of OPL9, which had the mixed component (Mix(Sq+Sa)), containing squamous and sarcomatoid elements, is shown in Supplementary Fig. 2C.

Differentially expressed genes (DEGs) analysis between components

To further dissect the molecular differences between PPC components, we conducted differential expression analysis among normal lung, Epithelial, and Sarcomatoid Full ROIs (Fig. 2D, E, and Supplementary Fig. 2D). The Epithelial showed elevated expression of epithelial markers, including *CDH1*, *CLDN1* and *KRT8*, while the Sarcomatoid exhibited higher levels of mesenchymal-related genes such as *MMP2*, *ZEB2* and *CDH11*. These findings indicate transcriptional changes associated with EMT between Epithelial and Sarcomatoid, consistent with previous reports[5,6]. The correlation of gene expression patterns further highlights the molecular distinction between these components, reflecting PPC's diverse biology. *ERBB2* and *MET*, both receptor tyrosine kinase (RTK)s and LUAD driver genes, were significantly downregulated in the Sarcomatoid compared to the Epithelial, suggesting the loss of oncogenic signals during EMT process (Fig. 2E). Immunohistochemistry (IHC) for cMET confirmed its high expression in epithelial regions and marked reduction in sarcomatoid regions, even in cases with *MET*ex14 (Fig. 2F).

ECM components and related genes, such as integrins, exhibited subtype-specific expression across tumor components (Fig. 2D, E). *ITGA3*, highly expressed in the Epithelial, forms a heterodimer with *ITGB1* to strengthen adhesion to laminin-332, promoting tumor growth and invasion[7,8]. In contrast, *ITGA5*, highly expressed in the Sarcomatoid, binds fibronectin (*FN1*) through *ITGB1*, enhancing motility and invasion[9]. Additionally, *MMP2* and *POSTN* were co-upregulated in the Sarcomatoid (Fig. 2D, E). *MMP2*, a zinc-dependent ECM remodeler, degrades ECM and serves as a diagnostic and prognostic marker in cancer[10]. *POSTN* promotes *MMP2* expression, enhancing cancer cell migration and invasion[11]. These findings suggest that ECM remodeling through integrin-mediated interactions plays a key role in the malignant transformation of PPC.

Gene Set Enrichment Analysis (GSEA) between components

GSEA was performed to identify pathways enriched in Epithelial and Sarcomatoid components relative to normal lung (Fig. 2G). It revealed a significant normalized enrichment score (NES) difference for the EMT pathway, consistent with previous reports[5,6]. EMT-related pathways, including hypoxia signaling, were enriched in Sarcomatoid, highlighting their role as EMT-inducing factors[12]. Gene sets related to ECM, such as ECM receptor interaction and focal adhesion, were significantly enriched in Sarcomatoid, while the tight junction pathway showed a negative NES. ECM remodeling mediates adhesion and signaling, regulating processes like proliferation, differentiation, migration, and apoptosis[13]. These findings emphasize ECM remodeling's role in the aggressive nature of Sarcomatoid in PPCs. Pathways related to the cell cycle (G2M checkpoint, MYC targets V1), apoptosis, DNA repair (e.g., KEGG mismatch repair, KEGG

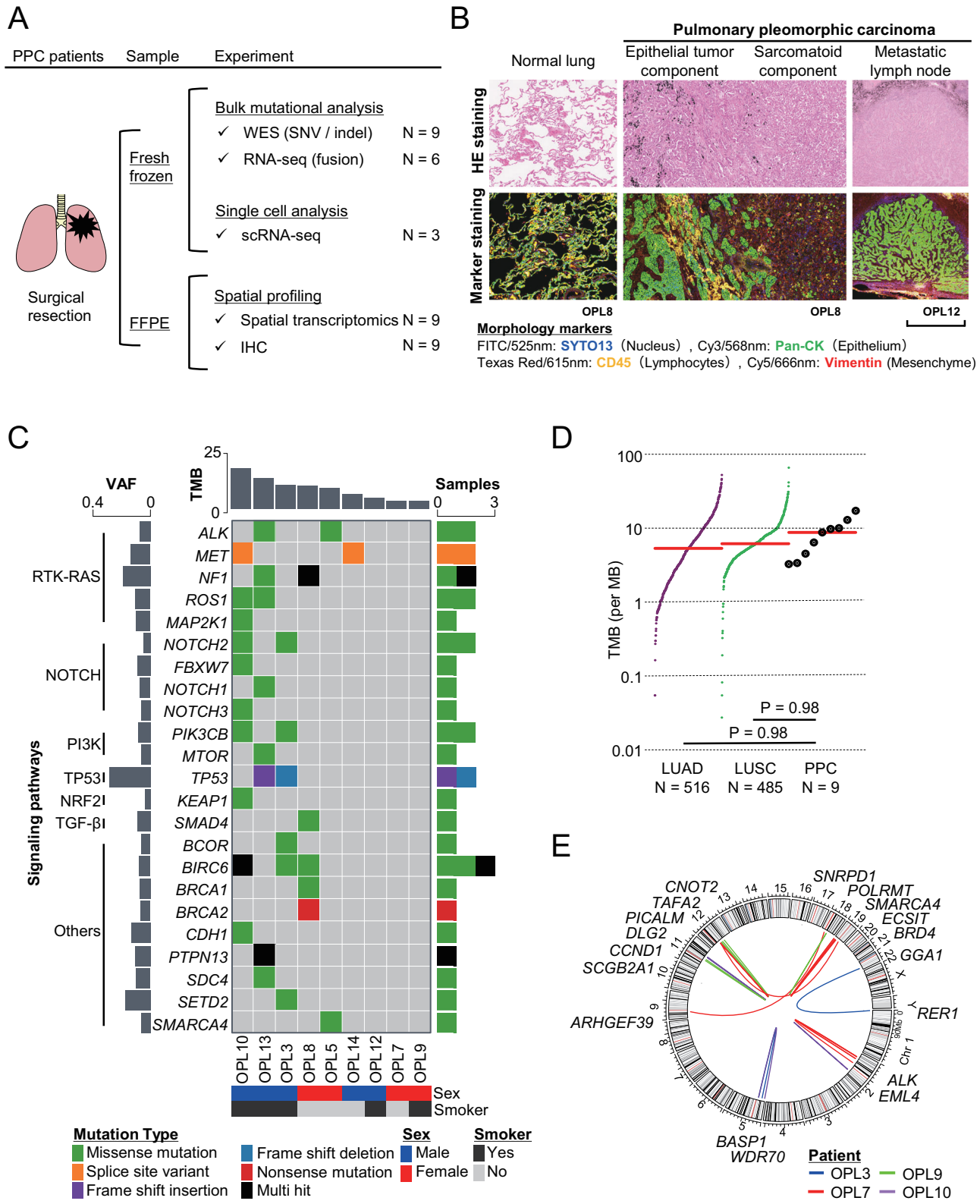
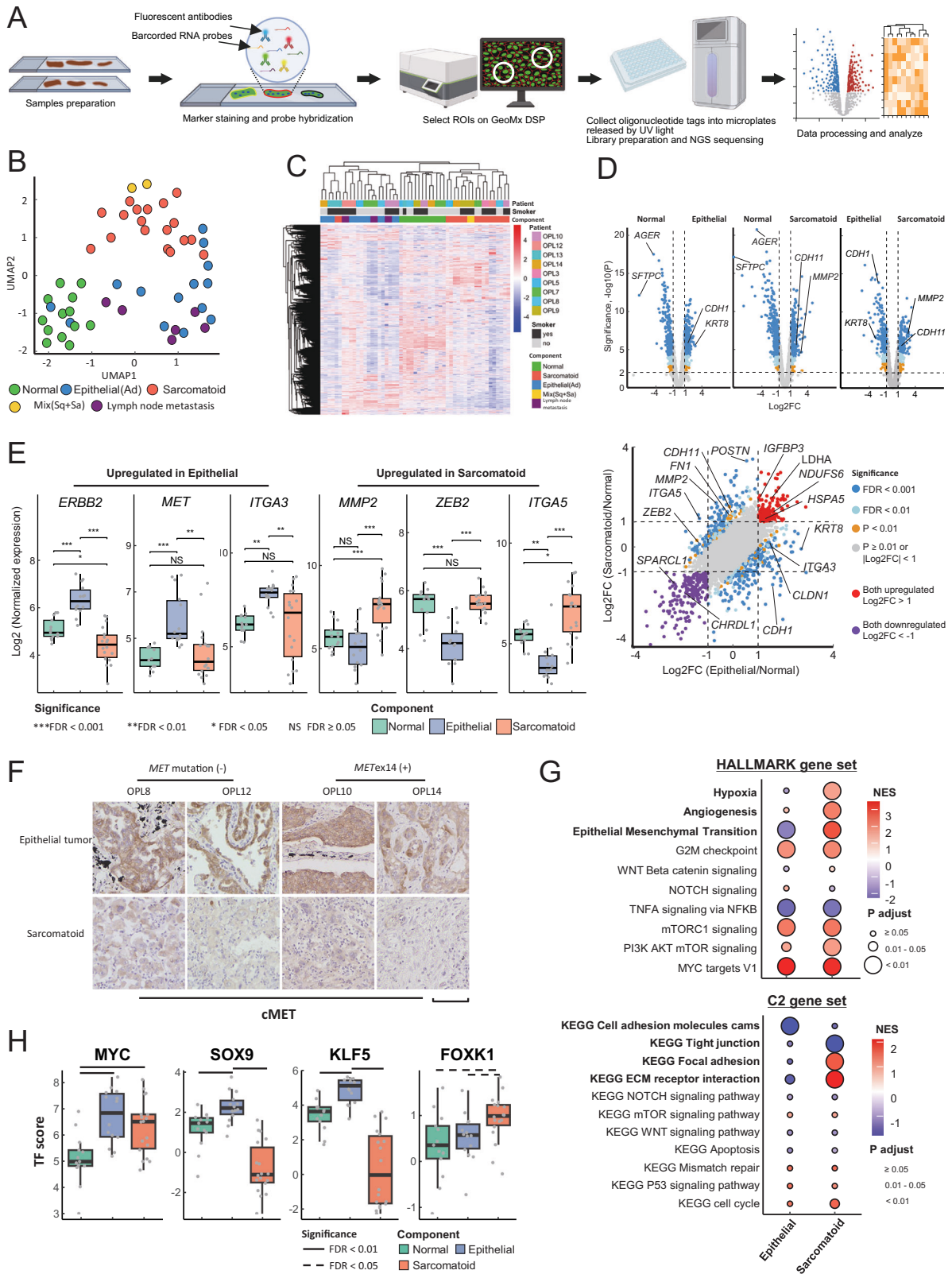


Fig. 1 | Study outline and mutational landscape of PPCs. A Study overview. Surgically resected PPC samples were analyzed by WES, RNA-seq, scRNA-seq, and spatial transcriptomics. **B** Representative images of HE-stained and morphology marker stained of normal lung, epithelial tumor, sarcomatoid tumor, and metastatic lymph node components. Scale bar, 500 μm. **C** Somatic mutational landscape in PPC. Genes are grouped by pathway; VAF is shown at left, TMB at the top, and the number of samples harboring each variant at right. **D** Comparison of TMB between PPCs (black) and TCGA NSCLC cohorts (LUAD in purple, LUSC in green). Red lines indicate median TMB. P values were calculated by t-test. **E** Circos plot of gene fusions identified by RNA-seq; colored lines represent individual patients. PP pulmonary pleomorphic carcinoma, WES whole-exome sequencing, SNV single-nucleotide variant, indel, insertion and deletion, RNA-seq bulk RNA sequencing, scRNA-seq single-cell RNA sequencing, FFPE formalin-fixed paraffin-embedded, IHC immunohistochemistry, CGC COSMIC Cancer Gene Census, VAF variant allele frequency, TMB tumor mutational burden, TCGA The Cancer Genome Atlas, LUAD lung adenocarcinoma, LUSC lung squamous cell carcinoma.



p53 signaling), and cancer cell proliferation and survival (PI3K-AKT-mTOR, WNT, NOTCH, TNFA via NFKB signaling) showed no trend differences between Epithelial and Sarcomatoid (Fig. 2G). These findings suggest shared signaling pathways supporting survival and proliferation in PPCs. Many of these pathways are linked to MYC activity[14], the observed enrichment of *MYC* target pathways reflects the strong proliferative and survival capacity in PPCs. Pathways with significant differences are shown in Supplementary Fig. 2E, F. Details of KEGG ECM receptor interaction pathways and protein-level gene expression differences are presented in Supplementary Fig. 2G. Notably, osteopontin (OPN), encoded by *SPP1*, and integrin dimer (αV-β1), upregulated in both Epithelial and Sarcomatoid, are functionally linked.

Fig. 2 | Spatial transcriptomic profiling and molecular characteristics of PPCs. **A** Schematic workflow of digital spatial transcriptomics using the GeoMx DSP. Created with BioRender.com. **B** UMAP plot of spatial transcriptomic data, showing distinct clustering of Epithelial, Sarcomatoid, and normal lung ROIs based on overall gene expression. Each dot represents an ROI, color-coded by component type. **C** Unsupervised hierarchical clustering heatmap of the top 20% most variable genes, selected based on CV. Rows represent genes, and columns represent ROIs. Clusters clearly separate the components, highlighting transcriptional heterogeneity among Epithelial, Sarcomatoid, and normal lung. Annotations at the top include patient ID and smoking status. **D** Volcano plots of DEGs between each component. Genes with statistical significance (linear model with BH correction) are color-coded: dark blue for FDR < 0.001, light blue for FDR < 0.01, and orange for $P < 0.01$ with |Log2FC | > 1. A scatter plot compares Log2FC in Epithelial vs. Normal (x-axis) and Sarcomatoid vs. Normal (y-axis), highlighting genes commonly upregulated (red, Log2FC > 1) or downregulated (purple, Log2FC < −1) in both tumor components. **E** Boxplots of normalized log2-transformed expression levels for selected DEGs. Box colors represent components; dot colors represent individual patients. Statistical significance was assessed using adjusted P values (FDR < 0.05). **F** IHC for cMET protein. Scale bar, 100 µm. **G** GSEA results comparing both tumor components to normal lung, using HALLMARK and C2 (CP) gene sets. The color scale represents NES values; P-values were adjusted using the BH method. Pathways highlighted in bold exhibit distinct trends between components. **H** TF activity inferred using the DoRothEA framework and VIPER algorithm. TF scores were calculated from normalized log2-transformed gene expression of regulons. Group-wise comparisons were performed using Wilcoxon rank-sum tests with multiple testing correction applied using the BH method. UMAP Uniform Manifold Approximation and Projection, ROI region of interest, CV coefficient of variation, DEG differentially expressed gene, Log2FC log2 fold change, FDR false discovery rate, IHC immunohistochemistry, GSEA Gene Set Enrichment Analysis, NES normalized enrichment score, BH Benjamini-Hochberg, TF transcription factor, VIPER Virtual Inference of Protein-activity by Enriched Regulon.

Inferring transcription factor (TF) activity scores

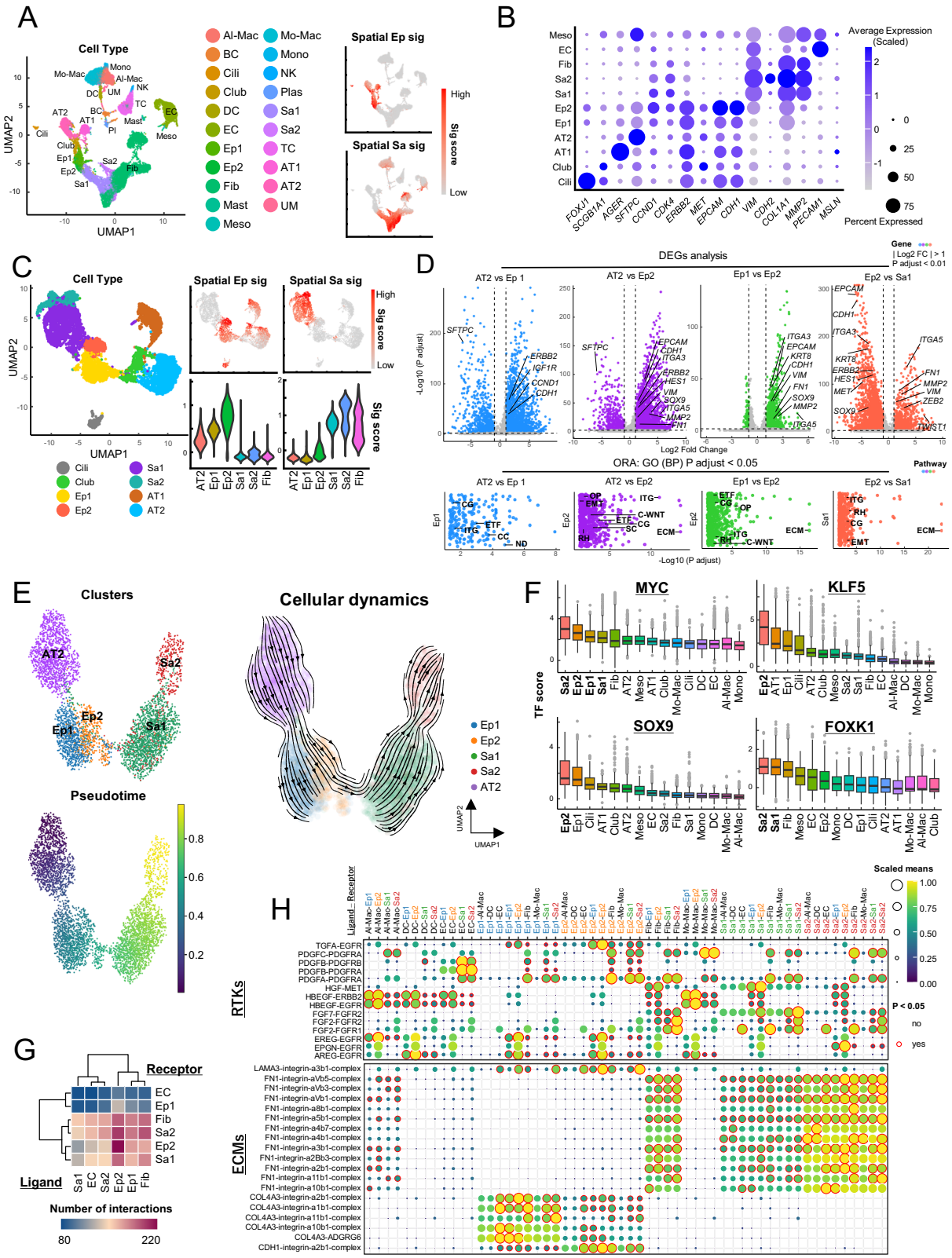
To understand transcriptional regulators underlying the observed gene expression patterns, we inferred TF activity from spatial gene expression (Fig. 2H, Supplementary Fig. 2H, I). Similar to GSEA, TF activity scores (TF scores) showed a significant increase in MYC activity in both tumor components, indicating a strong influence of MYC on oncogenic processes in PPCs (Fig. 2H). In the Epithelial, KLF5 and SOX9 showed significantly higher activity. KLFs are important for the establishment and/or maintenance of pluripotency[15]. Similarly, SOX9 is known to play a role in the formation and maintenance of cancer stem cells in various solid tumors[16,17]. In contrast, FOXK1 activity was significantly higher in Sarcomatoid. FOXK1 is known to induce EMT and promote invasion and metastasis[18,19]. These results indicate that while both tumor components share MYC-related pathways as a central oncogenic driver, the Epithelial acquires stem cell-like characteristics. In contrast, Sarcomatoid loses stem cell traits and exhibits EMT induction, consistent with its mesenchymal phenotype and histological appearance.

Single-cell analysis of PPC tumor cell heterogeneity

Classification of cell types. To dissect the cellular heterogeneity of PPC tumors, we performed single-cell transcriptomic profiling. To improve annotation accuracy, our scRNA-seq data were initially integrated with a publicly available lung adenocarcinoma dataset (GSE131907)[20]. After normalization and data integration, cells were clustered and annotated based on reference cell identities (Supplementary Fig. 3A–C). Tumor cells were further validated by inferCNV (Broad Institute) analysis (Supplementary Fig. 3D), and PPC-derived cells were extracted for downstream analysis. UMAP dimensionality reduction and cell type annotation for all PPC samples are shown in Fig. 3A. The population of cells bridging epithelial clusters and fibroblast clusters was identified and classified as sarcomatoid cells. In the UMAP space, these sarcomatoid clusters appeared positioned closer to fibroblasts than to other mesenchymal-derived cell types, such as endothelial cells or mesothelial cells, which may suggest a degree of mesenchymal similarity to fibroblasts. Additionally, genes specifically upregulated in spatial transcriptomic analysis in the Epithelial and Sarcomatoid were consolidated into spatial signatures (Spatial Epithelial [Ep] sig and Spatial Sarcomatoid [Sa] sig) (Supplementary Table 3). Average expression levels of these signatures at the single-cell level matched their respective tumor cells (Fig. 3A, right), confirming that spatial transcriptomic profiles accurately reflect tumor cell gene expression. Known marker genes supported annotation accuracy (Fig. 3B and Supplementary Fig. 3E). Subsetting epithelial and tumor cells for UMAP revealed tumor cells form a continuum distinct from differentiated epithelial cells (Fig. 3C). Spatial signature score distributions aligned closely, further supporting consistency between spatial and single-cell data. Although UMAPs color-coded by patient ID (Supplementary Fig. 3F, G) showed that sarcomatoid and epithelial tumor cells predominantly originated from distinct individual patients, Spatial Ep and Sa sig scores—derived from spatial transcriptomic DEG analyses across multiple patients—were well-aligned with single-cell annotations and smoothly distributed across clusters, supporting representative characterization of PPC cell states across patients. Violin plots demonstrated that Sa1 and Sa2 clusters had higher distributions of Spatial Sa gene signature scores (sig scores) compared to fibroblasts, suggesting that Sarcomatoid ROIs better represent sarcomatoid cell properties than fibroblasts.

DEGs analysis and Over-Representation Analysis (ORA) between tumor cell subtypes

To identify molecular programs underlying phenotypic transitions in PPC, we performed differential expression and enrichment analyses comparing Type II alveolar cell (AT2) cells with tumor subtypes (Fig. 3D). In comparisons between AT2 and Ep1 clusters (blue), upregulated genes in Ep1 included *IGF1R* and *CCND1*, involved in tumor proliferation, and epithelial tumor markers such as *ERBB2* and *CDH1*. ORA revealed enrichment of tumorigenesis-related pathways. Most importantly, in comparison with AT2 (purple) and Ep1 (green), Ep2 showed upregulation of these genes as well as *VIM*, indicating co-expression of epithelial and mesenchymal markers, consistent with a hybrid EMT phenotype. Such "intermediate EMT state" has been reported to enhance metastatic and invasive potential prior to complete EMT[21]. In addition, ECM remodeling genes, such as *MMP2* and *FN1*, were significantly upregulated in Ep2, along with both *ITGA3* and *ITGA5*, which displayed a trade-off relationship in spatial transcriptomics. Furthermore, stemness-related genes, such as *SOX9*[16,17] and *HES1*[22,23], were highly expressed in Ep2. ORA also highlighted enrichment of hypoxia, EMT, and ECM interaction pathways in Ep2, characteristic of Sarcomatoid in spatial transcriptomics. Strong ECM and integrin-related pathway enrichment supported sarcomatoid-like changes, including ECM remodeling, within this epithelial tumor subset. Enrichment of stem cell development and canonical WNT signaling pathways underscores the stem cell-like and multipotent nature of this subset[24]. The intermediate EMT state is also reported to exhibit higher plasticity and stemness[21]. Comparisons between Ep2 and Sa1 (red) revealed loss of stemness and epithelial characteristics in Sa1, adopting a fully mesenchymal phenotype. Classical EMT markers, such as *ZEB2* and *TWIST1*, were significantly differentially expressed only between Ep2 and Sa1, indicating progression to an "extreme EMT state"[21]. Thus, ECM remodeling in Ep2 appears to initiate earlier in the intermediate EMT state, highlighting a detectable process of sarcomatoid differentiation. Major ECM-related gene expressions are shown in Supplementary Fig. 3H. Overall, Ep2 occupies the intermediate EMT state, retaining both epithelial and mesenchymal traits. Active ECM remodeling in this state likely drives tumor cell differentiation into a mesenchymal phenotype while maintaining stemness.



Differentiation trajectories of PPC tumor cells

To investigate tumor cell dynamics and differentiation trajectories, scTour[25] was used for high-dimensional visualization and pseudotime inference (Fig. 3E and Supplementary Fig. 3I). This analysis revealed a directional trajectory originating from Ep2 and progressing toward Sa1, with no apparent transitions observed from Ep1. This suggests that differentiation toward the sarcomatoid phenotype originates primarily from Ep2. These findings align with DEGs and ORA results, highlighting the intermediate EMT characteristics of Ep2 and the heterogeneity within epithelial tumor cells. Ep2 plays a distinct role in driving the transition to sarcomatoid cells, likely through ECM remodeling and EMT-related pathways.

Fig. 3 | scRNA-seq reveals cellular trajectories and interactions in PPCs. A UMAP of scRNA-seq data from PPC samples. Sarcomatoid cells bridge epithelial and fibroblast clusters. Spatial Ep and Sa signature scores (sig scores) are overlaid on tumor cell clusters. **B** Dot plot of marker gene expression for epithelial and mesenchymal clusters. **C** UMAP of epithelial and tumor cells highlighting a continuum from epithelial to sarcomatoid states. Spatial Ep and Spatial Sa sig scores are shown on UMAPs and violin plots. **D** Volcano plots of DEGs among tumor cell subtypes. Genes meeting |Log2FC | > 1 and Bonferroni-adjusted $P < 0.01$ are considered significant. ORA results show enriched GO biological process terms (BH-adjusted $P < 0.05$). Abbreviations for GO terms are listed below. CC Cell cycle checkpoint signaling, CG Cell growth, C-WNT Canonical WNT signaling pathway, ECM Extracellular matrix organization, EMT Epithelial to mesenchymal transition, ETF Epithelial tube formation, ITG Integrin-mediated signaling pathway, ND Nuclear division, OP Oxidative phosphorylation, RH Response to hypoxia, SC Stem cell development. **E** UMAP with pseudotime inference illustrating differentiation trajectory from Ep2 to Sa1. Vector fields (arrows) and streamlines (curves) indicate a directional transition toward a fully mesenchymal phenotype, highlighting progression from intermediate to extreme EMT states. **F** Boxplots show TF scores in non-lymphocyte cell types. **G** Heatmap of ligand-receptor interactions across cell types. Columns are ligand-releasing cell types, rows are receptor-expressing cell types; the color scale represents the number of observed interactions. **H** Dot plot of ligand-receptor interactions focusing on major RTKs and ECM-related pathways. Dot size and color represent scaled mean expression, with significant interactions ($P < 0.05$) outlined in red. sig score signature score< Al-Mac Alveolar Macrophage, AT1 Type I alveolar cell, AT2 Type II alveolar cell, BC B cell, Cili Ciliated cell, Club Club cell, DC Dendritic cell, EC Endothelial cell, Ep1 Epithelial tumor cell 1, Ep2 Epithelial tumor cell 2, Fib Fibroblast, Mast Mast cell, Meso Mesothelial cell, Mo-Mac Monocyte-Derived Macrophage, Mono Monocyte, NK Natural-Killer cell, Plas Plasma cell, Sa1 Sarcomatoid cell 1, Sa2 Sarcomatoid cell 2, TC T cell, UM Undetermined Myeloid, ORA Over-Representation Analysis, GO Gene Ontology, BP Biological Process, RTK receptor tyrosine kinase, ECM extracellular matrix.

Inferring TF scores

TF score distributions of scRNA-seq were largely consistent with spatial transcriptomics (Fig. 3F and Supplementary Fig. 3J). Among all TFs, MYC showed the highest activity across tumor clusters, highlighting its central role in promoting PPC cell proliferation and survival. KLF5 and SOX9 activity peaked in epithelial tumor cells, particularly in Ep2, consistent with the stem-like transcriptional profile of this subset. In contrast, FOXK1 was most active in sarcomatoid cells, with intermediate activity in Ep2, suggesting Ep2 as a transitional state between epithelial and sarcomatoid phenotypes.

Cellular interactions between different cell types

To investigate cell–cell communication within the PPC tumor microenvironment, we applied CellPhoneDB[26] to infer ligand–receptor interactions across cell types (Fig. 3G). Ep2 emerged as the most prominent ligand-releasing cell type, extensively interacting with fibroblasts and sarcomatoid cells. These findings align with prior analyses, indicating that pluripotent and stem-like traits of Ep2 drive differentiation and tumor heterogeneity[27]. Fibroblasts also played a central role, strongly interacting with Ep2, Sa1, and Sa2, while Ep1 showed fewer interactions. All cell interactions are detailed in Supplementary Fig. 3K. To further dissect signaling dynamics, we examined RTK and ECM-related interactions, which revealed distinct patterns across tumor subtypes (Fig. 3H). Ep1 and Ep2 exhibited strong RTK signaling through *EGFR*, *ERBB2*, and *MET*, supporting proliferation and survival. In contrast, Sa1 and Sa2 shifted towards *FGFR* and *PDGFR* signaling, consistent with mesenchymal phenotypes. ECM remodeling was evident, with interactions transitioning from epithelial-associated ECM components like *LAMA3* (a subunit of laminin-332), *COL4A3*, and *CDH1* to fibronectin (*FN1*)-integrin complexes. Integrins, particularly β1-containing heterodimers, facilitate RTK signaling and support tumor survival and therapy resistance[28]. The shift to ECM-integrin interactions in Sa1 and Sa2 indicates enhanced mechanotransduction and RTK crosstalk, driving sarcomatoid progression. Together, these results underscore the importance of integrin–ECM interactions in promoting phenotypic transitions and sustaining RTK signaling in PPC. Integrin-RTK crosstalk may underline the functional diversity and treatment resistance in PPC. Significant integrin-related interactions are detailed in Supplementary Fig. 3L.

External validation using an independent dataset

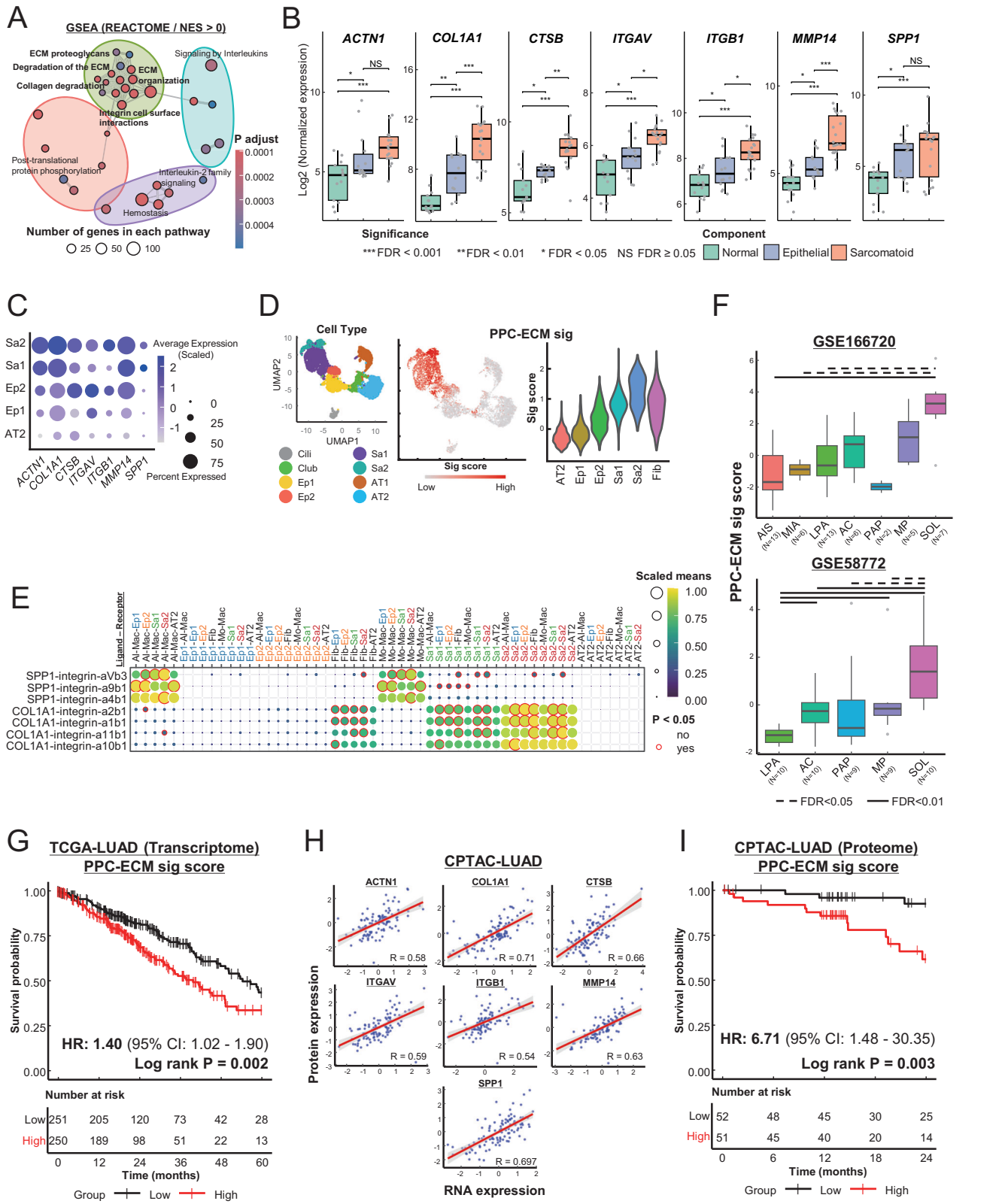
To evaluate the robustness of our cell-type annotation, we applied automated annotation to the scRNA-seq dataset of Bischoff et al.[29], from which we analyzed the 10 NSCLC cases, including one diagnosed as sarcomatoid carcinoma (histological subtypes shown in Supplementary Fig. 3M). Using our PPC data as the SingleR[30] reference, we annotated all QC-passed cells in each patient (Supplementary Fig. 3N). We then examined the relative composition of the four tumor-cell categories after normalizing by the total number of tumor cells per sample and found that Sa1 was most enriched in the patient with sarcomatoid histology (Supplementary Fig. 3O). These results indicate that the annotation approach using our PPC dataset was able to assign cell identities to other sarcomatoid carcinoma cases with reasonable accuracy. However, since Bischoff's cohort is not composed exclusively of PPC cases, our PPC-derived reference may not perform equally well across all samples.

PPC-ECM signature related to poor prognosis of LUAD

Definition of the PPC-ECM signature. Building on the observation that ECM remodeling accompanies the phenotypic transition of PPC, we defined a PPC-ECM signature that may hold diagnostic and therapeutic value. Using Reactome Pathways Gene Set, GSEA was performed on spatial transcriptomics, comparing Epithelial and Sarcomatoid. An enrichment map with clustering identified a densely interconnected ECM and integrin-related cluster (highlighted in yellow-green in Fig. 4A). From the core enrichment genes in this cluster, we extracted genes consistently upregulated in both tumor components compared to the normal region (Log2FC > 0.5, false discovery rate [FDR] < 0.05). Seven genes were identified, collectively forming the PPC-ECM signature (Supplementary Table 3). Spatial transcriptomics demonstrated elevated expression of these genes in tumor regions (Fig. 4B), corroborated by scRNA-seq, which revealed higher expression in tumor clusters than AT2 cells (Fig. 4C). The PPC-ECM sig score increased progressively along the cellular trajectory (Ep2-Sa1-Sa2) (Fig. 4D). The analysis of cell-cell interactions associated with the PPC-ECM signature revealed significantly enhanced ligand-receptor interactions in tumor cells compared to AT2 cells (Fig. 4E). Additionally, *SPP1*, known to be secreted by tumor-associated macrophages[31], and *COL1A1*, expressed by fibroblasts[32], showed stronger ligand activity in sarcomatoid cells. These observations suggest that, during phenotypic transition, sarcomatoid cells may acquire autocrine signaling properties associated with the PPC-ECM signature, potentially contributing to their aggressive behavior.

Relation between PPC-ECM signatures and prognosis of LUAD

To assess the prognostic value of PPC-ECM signatures in LUAD, we analyzed gene expression datasets (GSE166720[33], GSE58772[34]) annotated with histological subtypes (Fig. 4F). Higher PPC-ECM sig scores were observed in histologically aggressive subtypes, such as solid adenocarcinoma, indicating that ECM remodeling may also characterize poorly differentiated LUAD subtypes without overt sarcomatoid features. Next, PPC-ECM sig scores for TCGA-LUAD samples were calculated from RNA expression data. Patients were stratified into High and Low groups by the median score. The higher PPC-ECM sig scores were significantly associated with poorer overall survival (OS) (Fig. 4G). Using Clinical Proteomic Tumor Analysis Consortium (CPTAC) proteomics data, RNA, and protein expression of PPC-ECM signature genes showed a positive correlation (all R > 0.5, Fig. 4H and Supplementary Fig. 4A). Protein-based signature scores were significantly associated with worse short-term OS (2 years) (Fig. 4I), and a similar trend was observed for 5-year OS, although statistical significance was marginal (Supplementary Fig. 4B). These results identify the PPC-ECM signature as a prognostic marker for LUAD.



Tumor immune microenvironment (TiME) and mediastinal lymph node metastasis of PPCs

Digital cytometry of TiME. Immune ROIs (Supplementary Fig. 2A) were deconvoluted using CIBERSORTx[35] to estimate immune cell proportions (Fig. 5A). Plasma cells and total B cells (sum of naive, memory, and plasma cells) were more abundant in tumor components than in normal lung. Detailed proportions per ROI and component are shown in Supplementary Fig. 5A.

DEGs analysis of TiME

To reduce background noise, DEGs analysis compared immune and tissue ROIs within the same components, identifying 1718 upregulated immune-

Fig. 4 | The PPC-ECM signature is associated with poor prognosis in LUAD. **A** Emapplot of GSEA (Reactome pathways) comparing Epithelial vs Sarcomatoid in spatial transcriptomics. Node size indicates the number of genes, and color intensity reflects adjusted P-values. Yellow-green cluster includes ECM and integrin-related pathways with dense edges. **B** Boxplots of spatial transcriptomics expression for seven PPC-ECM genes in normal lung, Epithelial, and Sarcomatoid. **C** Dot plot showing scRNA-seq average scaled expression of PPC-ECM genes in tumor clusters (AT2 as control). **D** UMAP and violin plots of PPC-ECM sig scores in scRNA-seq data. **E** Dot plot of ligand-receptor interactions for PPC-ECM genes across cell types. Dot size and color represent scaled mean expression, with significant interactions ($P < 0.05$) are outlined in red. **F** Boxplots of PPC-ECM sig scores in different LUAD subtypes from two public datasets. Pairwise comparisons between subtypes were performed using the Wilcoxon rank-sum test with BH correction. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LPA lepidic predominant adenocarcinoma, AC acinar predominant adenocarcinoma, PAP papillary predominant adenocarcinoma, MP micropapillary predominant adenocarcinoma, SOL solid predominant adenocarcinoma. **G** Kaplan-Meier curves of OS in TCGA-LUAD, stratified by median PPC-ECM sig scores (High vs Low). This analysis was restricted to a follow-up period of 60 months, with all events beyond this point censored. Multivariable Cox proportional hazard model was adjusted for pathological stage and age at diagnosis. **H** Scatter plots correlating RNA (x-axis) and protein (y-axis) expression of PPC-ECM genes in CPTAC-LUAD. RNA and protein expression values were normalized to Z-scores across samples. Linear regression lines (red) with 95% CI (gray shading) are shown. Pearson's correlation coefficients are indicated for each gene is shown as "R", all exceeding 0.5, demonstrating a positive correlation. **I** Kaplan-Meier curves of short-term OS in CPTAC-LUAD, stratified by protein-based PPC-ECM sig scores (High vs Low). This analysis was restricted to a follow-up period of 24 months. Multivariable Cox proportional hazards model was adjusted for pathological stage and age at diagnosis. emapplot enrichment map plot, OS overall survival, HR Hazard Ratio, CI Confidence Interval, CPTAC Clinical Proteomic Tumor Analysis Consortium.

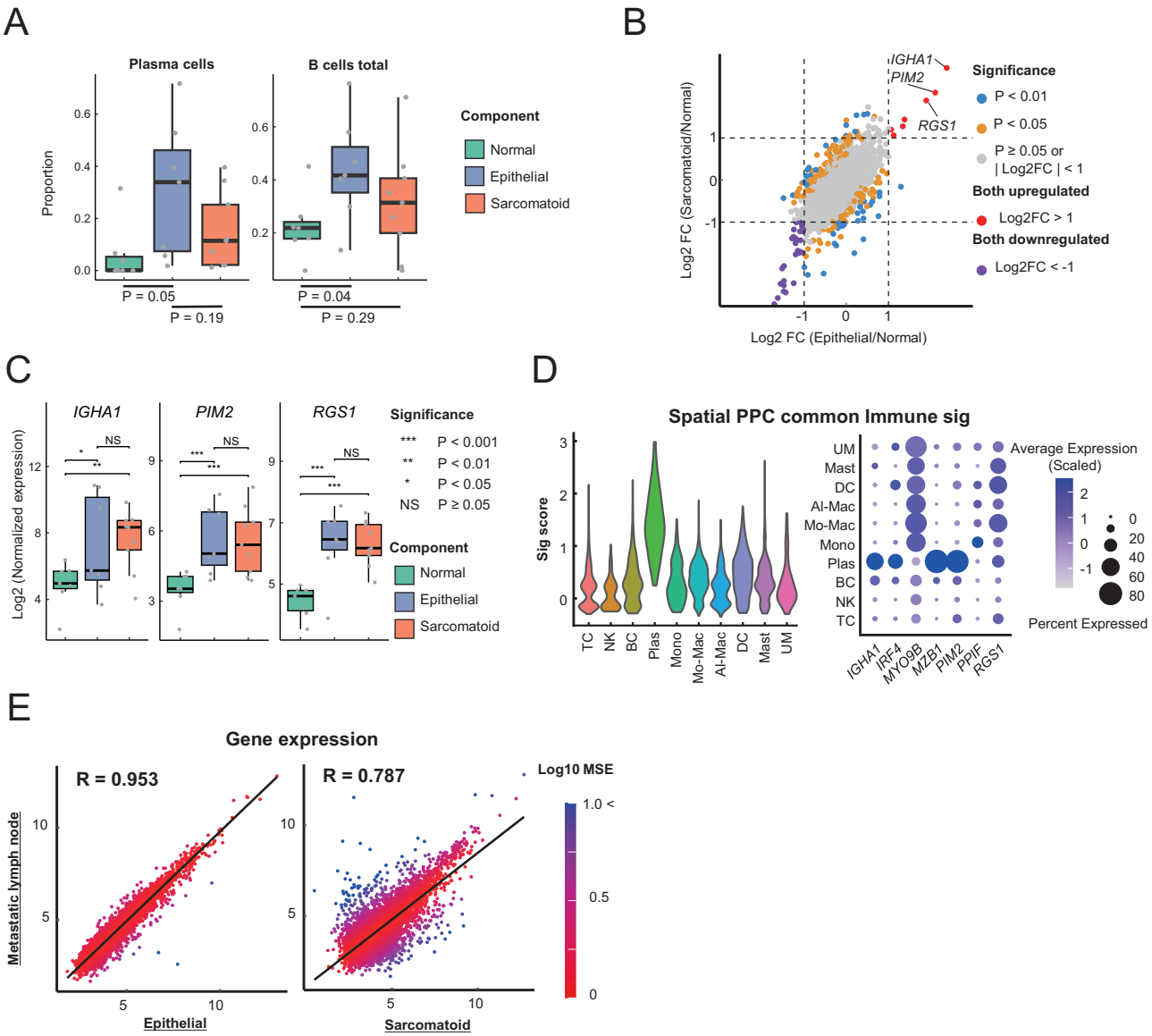


Fig. 5 | Tumor immune microenvironment and mediastinal lymph node metastasis of PPCs. **A** Boxplots showing proportions of plasma cells and total B cells (naive, memory, plasma) in immune ROIs (CD45+ regions) from normal lung ($n = 7$), Epithelial ($n = 7$), and Sarcomatoid ($n = 9$). Immune cell fractions were estimated by CIBERSORTx using normalized log2-transformed data and the LM22 signature matrix (officially distributed). Pairwise t-tests were used for statistical comparisons. **B** Scatter plot comparing Log2FC of immune-related genes in Epithelial vs Normal (x-axis) and Sarcomatoid vs Normal (y-axis). Genes upregulated in both tumor components (Log2FC > 1, $P < 0.05$) are highlighted. **C** Boxplots of normalized log2-transformed expression levels for selected immune-related genes. **D** Violin plots of the Spatial PPC common immune signature across single-cell immune populations, and a dot plot showing expression of each gene (signature genes highlighted in red in **B**). **E** Scatter plots comparing average gene expression in metastatic lymph nodes (y-axis) with Epithelial or Sarcomatoid components (x-axis). Pearson's correlation coefficients (R) indicate a stronger association with Epithelial. Color intensity shows Log10 MSE. MSE mean squared error.

related genes (Supplementary Fig. 5B). Subsequent DEGs analysis between immune ROIs of different components identified significant upregulation of *IGHA1*, *PIM2*, and *RGS1* in Epithelial and Sarcomatoid components compared to the normal region (Fig. 5B, C, and Supplementary Fig. 5C). *IGHA1*, encoding IgA, and *PIM2*, which regulates plasma cell differentiation and prevents apoptosis[36], suggest plasma cell enrichment. *RGS1* modulates immune cell migration and is implicated in inhibiting effector T cell infiltration[37] and contributing to T cell exhaustion[38], suggesting immune evasion mechanisms in PPCs. We defined a common immune signature by identifying genes that were consistently upregulated in both tumor components relative to normal lung (Supplementary Table 3). Single-cell analysis revealed high signature scores in plasma cells, with *IGHA1* and *PIM2* specifically expressed in plasma cells, while *RGS1* was broadly expressed across immune cells (Fig. 5D). Genes uniquely enriched in the Epithelial and Sarcomatoid are detailed in Supplementary Table 3 and Supplementary Fig. 5D.

Mediastinal lymph node metastasis of PPCs

Mediastinal lymph node metastasis (N2) was observed in 3 out of 9 cases, all histologically classified as adenocarcinoma (Fig. 1B and Supplementary Fig. 2A). Pearson's correlation coefficients using overall gene expression data confirmed a significantly higher correlation between metastatic lesions and the Epithelial compared to the Sarcomatoid (Fig. 5E), suggesting that the Epithelial may contribute to lymph node metastasis at the molecular level.

Discussion

*MET*ex14 occurs in 1–3% of NSCLC but is more prevalent in PPC (13–32%)[39,40]. MET inhibitors have shown clinical benefits in NSCLC with *MET*ex14[41,42]. In our cohort, *MET*ex14 was found in 22% of cases. However, MET gene and protein expression were restricted to the epithelial region, with no expression in the sarcomatoid. Regarding TMB, PPC was comparable to other NSCLC subtypes, and did not exhibit high PD-L1 by IHC. These findings differ from earlier studies[4,43]. Furthermore, spatial immune analysis revealed molecules like *RGS1*, associated with immune evasion and exhaustion[37,38], in both epithelial and sarcomatoid tumor components. Thus, it is suggested that conventional molecular targeted therapy and immune checkpoint inhibition monotherapy may be inadequate for PPC, especially for the sarcoma component, that exhibits heterogenous intratumor expression profiles. As a shared signaling pathway, MYC-related pathways and the TF activity of MYC were emphasized. Notably, the PI3K-AKT-mTOR signaling pathway was commonly enriched in this cohort. Previous studies have linked *PIK3CA* mutations to poor postoperative outcomes in PPC[44], although no patients in this cohort had its mutations.

Spatial transcriptomics revealed EMT-related transcriptional changes and ECM remodeling between epithelial and sarcomatoid components. The PPC-ECM signature genes were consistently upregulated across tumor components, highlighting their role in tumor progression. Previous studies have reported that ECM remodeling was a key pathway in lung cancer progression and prognosis[45]. In our findings, high expression of *ITGA3* was observed in epithelial component, while *ITGA5* was highly expressed in sarcomatoid component. 'Integrin switching', as the tumor matrix biochemistry evolves, has been reported to contribute to progression and chemoresistance in solid tumors[46,47]. Integrin β1-containing heterodimers play pivotal roles in EMT, cancer metastasis, and drug resistance, particularly to EGFR TKIs[48,49]. Overexpression of *ITGB1* has been linked to poor prognosis in various tumors[50–52]. scRNA-seq provided higher-resolution insights into PPC's cellular heterogeneity. Among epithelial tumor cells, the Ep2 subclone displayed an "intermediate EMT state", characterized by ECM remodeling, stemness, and plasticity. This subclone also functioned as a signaling hub, interacting extensively with fibroblasts and sarcomatoid cells. The limited interactivity of Ep1 suggests a more static role in the TME compared to Ep2. Collectively, these findings reveal the pivotal roles of Ep2 and fibroblasts in shaping the TME and identify potential therapeutic targets aimed at disrupting tumor-stroma interactions mainly intermediary by ECM and integrins. In addition, we revealed that PPC-ECM signature genes present promising targets for PPC. ITGAV inhibitors cilengitide have shown efficacy in reducing lung metastasis in breast cancer animal model[53]. CTSB, which degrades ECM proteins, reduced tumorigenesis by knockdown in renal carcinoma models[54]. ACTN1, a cytoskeletal protein, regulates cell-matrix adhesion via PI3K signaling, and its knockdown reduced tumor growth in vivo[55]. These reports highlight the potential of targeting ECM and integrin-mediated pathways for PPC and highly malignant LUAD. The observed transcriptional and histological similarity between metastatic lymph nodes and Epithelial may suggest that Epithelial, rather than Sarcomatoid, contributes to lymph node metastasis. Further investigation will be required to clarify this relationship, including the underlying molecular mechanisms.

This study has limitations, including a small sample size due to PPC's rarity and its descriptive focus on transcriptomic data. Although the scRNA-seq analysis included only three patients and certain tumor cell types primarily derived from individual samples, spatial gene signature scores were smoothly distributed across single-cell clusters, supporting representative characterization of PPC cell states. Nonetheless, to the best of our knowledge, this is the first study to integrate spatial transcriptomics and scRNA-seq for PPC, providing novel insights into its molecular pathology. Most epithelial tumor regions in this cohort were LUAD, whereas prior studies included squamous cell carcinoma components, suggesting potential racial or regional differences[6].

In this study, we integrated spatial transcriptomic analysis and scRNA-seq to elucidate the pathogenesis of PPC with heterogeneous tissues and cell populations, which could inform future efforts toward personalized therapeutic approaches. PPC exhibits highly heterogeneous gene expression profiles within tumor components. The intermediate EMT state, characterized by ECM remodeling, plays a significant role in PPC malignancy. Several treatment options for PPC targeting specific tumor components were suggested. Further studies are warranted to translate these findings into clinical applications and evaluate their impact on patient outcomes.

Methods

Study design and sample collection

This retrospective study includes nine patients who underwent surgical resection for primary lung cancer and pathologically diagnosed as PPC at Okayama University Hospital (Okayama, Japan) from 2006 to 2023. All patients included in this study satisfied the conditions for observational study inclusion and had appropriately stored surgical specimens. The study received approval from the Okayama University Clinical Research Ethics Committee (approval number: 2411-001). Participant consent was obtained using an opt-out method. All ethical regulations relevant to human research participants were followed. Tumor and adjacent normal lung tissues were collected at the time of surgery and stored accordingly. Fresh-frozen samples were processed for WES, RNA-seq, and scRNA-seq, while FFPE blocks were used for spatial transcriptomics and IHC (Fig. 1A). All key analytical procedures required to reproduce the results are described below, with additional supporting details provided in the Supplementary Information.

WES and variant analysis

Fresh-frozen tumor and matched normal lung tissues (20–30 mg each) were homogenized using BioMasher II (Nippi Inc., Tokyo, Japan), and genomic DNA was extracted with the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. DNA quantity and integrity were assessed with Qubit (Thermo Fisher Scientific) and TapeStation (Agilent Technologies). All samples passed quality control prior to library preparation.

Whole-exome libraries were prepared using the SureSelect Human All Exon V6 kit (Agilent Technologies). After PCR amplification, 250 ng of each library was hybridized with capture baits at 65 °C for 24 h, washed, and re-amplified. Sequencing was performed on a NovaSeq X platform (Illumina) with 150 bp paired-end reads, targeting approximately 40 million reads per sample.

FASTQ files were trimmed using Trimmomatic (v0.39)[56] to remove low-quality bases and reads. Specifically, paired-end reads were trimmed with a leading and trailing base quality cutoff of Phred quality score 30, a sliding window quality cutoff of Phred quality score 15 across 4 bases, and a minimum read length of 20 bases. Cleaned reads were aligned to the GRCh38 reference genome using BWA-MEM alignment algorithm in BWA (v0.7.17-r1188)[57] with default parameters, and BAMs were sorted with Samtools (v1.13)[58]. PCR duplicates were marked, and base-quality score recalibration (BQSR) was performed with GATK (v4.5.0.0)[59]. Duplicate reads were marked using MarkDuplicates. BQSR was performed using BaseRecalibrator, leveraging known sites of variation from the file (Homo_sapiens_assembly38.known_indels.vcf.gz) (https://console.cloud.google.com/storage/browser/gcp-public-data--broad-references/hg38/v0). The recalibrated BAM files were then processed using ApplyBQSR. Somatic variants were called with GATK Mutect2[60] on paired tumor–normal data, using a panel of normals (1000g_pon.hg38.vcf.gz) (https://console.cloud.google.com/storage/browser/gcp-public-data--broad-references/hg38/v0) and a germline resource (af-only-gnomad.hg38.vcf.gz) (https://console.cloud.google.com/storage/browser/gatk-best-practices/somatic-hg38) files; calling was restricted to the SureSelect V6 capture BED to reduce off-target noise.

Variants were annotated with SnpEff (v5.2a)[61] and SnpSift (v5.2)[62], referencing allele frequencies from the Japanese Multi-Omics Reference Panel (ToMMo 38KJPN-SNV/INDEL; v20220929) (https://jmorp.megabank.tohoku.ac.jp/docs/guide-ja/dataset_detail/tommo-38kjpn_snvindel). Variants with read depth <25, minimum base quality <20, < 3 tumor-supporting reads or >3 reads in normal, or allele frequency ≥0.01 in ToMMo were excluded to minimize false positives. Filtered VCFs were converted to MAF format using vcf2maf (v1.6.22)[63]. TMB was computed and compared with TCGA LUAD and LUSC cohorts using maftools (v2.2.0)[64] in R (v4.4.0), assuming a 60 Mb capture size for SureSelect V6 (maftools default for TCGA cohorts: 35.8 Mb). Mutational signature analysis of PPC samples was calculated using the sig_fit function from the sigminer (v2.3.1)[65] in R. Any signature with a contribution below 5% in each sample was excluded from the result by setting relative threshold (0.05). The default 'legacy' database, which includes COSMIC v2 SBS mutational signatures[66], was used to fit the mutational profiles of each sample.

RNA-seq and gene fusion detection

Total RNA was extracted from frozen tumor tissues using either the AllPrep DNA/RNA Mini Kit (Qiagen) or Direct-zol DNA/RNA Miniprep Kit (Zymo Research). RNA quantity and integrity were evaluated; samples with RIN <3 or DV200 below institutional thresholds were considered failures. rRNA depletion used the NEBNext rRNA Depletion Kit (Human/Mouse/Rat), and libraries were prepared with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs); sequencing was on a NovaSeq X (Illumina) with 150 bp paired-end reads targeting approximately 60 million reads per sample.

Gene fusion events were identified using the nf-core/rnafusion pipeline (v3.0.2)[67]. Only fusions detected by at least two tools applied in the pipeline (Arriba, STAR-Fusion, or FusionCatcher) and supported by at least three junctional reads were retained as high-confidence calls.

Spatial transcriptomics

Spatial transcriptomic profiling was performed using the GeoMx DSP (NanoString Technologies) according to the manufacturer's instructions. Nine FFPE surgical resection specimens from patients with pulmonary pleomorphic carcinoma were analyzed. A pathologist identified regions corresponding to epithelial tumor, sarcomatoid, and adjacent normal lung areas on HE-stained slides. Based on these annotations, serial 5 µm FFPE sections were macrodissected and mounted on two experimental slides for GeoMx processing.

Slides were deparaffinized, subjected to antigen retrieval, and digested enzymatically using the BOND RX system (Leica Biosystems) with reagents including BOND Wash Solution, 10% neutral-buffered formalin (NBF), NBF stop buffer (6.06 g Tris base and 3.75 g glycine in 500 mL DEPC-treated water), and Enzyme 1 (Proteinase K). After washing in PBS, slides were hybridized overnight at 37 °C with the Human Whole Transcriptome Atlas (WTA) (NanoString Technologies), which targets human protein-coding transcripts. Stringent SSC washes were performed to remove unbound probes. Slides were stained with morphology markers, including GeoMx Solid Tumor TME Morphology Kit Human RNA Compatible (Immunofluorescence assays for Pan-CK, CD45, and SYTO13) (NanoString Technologies) and Vimentin Antibody (RV203) [Alexa FluorR 647] (Novus Biologicals) for sarcomatoid component, to visualize and distinguish different tissue components under the GeoMx DSP. After staining, the slides were washed and prepared for imaging. The prepared slides were loaded onto the GeoMx DSP, designed to visualize and analyze specific region of interest (ROI) in tissue sections. ROI within the tissue were selected based on the experimental design, focusing on areas such as Epithelial, Sarcomatoid, and normal lung regions in the PPC samples identified by pathologist. In cases where some components were not present, only ROIs of the present components were selected. For each region, three types of ROIs were created: Full ROI (without segmentation), immune ROI (segmented by CD45+), and tissue ROI (segmented by CD45-). If the number of nuclei is insufficient by segmentation (nuclei < 100), the selection of ROI was abandoned. Overall, 125 ROIs were selected (Full ROI: 59, immune ROI: 26, tissue ROI: 40). After ROI selection, UV light was used to cleave oligonucleotide tags corresponding to target transcripts. Released tags were collected into microplates, dried, and rehydrated with DEPC water. Library amplification was performed with SeqCode primers and master mix (NanoString Technologies) using a Bio-Rad thermal cycler under manufacturer-specified cycling conditions, including positive and negative PCR controls. Amplified libraries were purified with AMPure XP beads (Beckman Coulter) and eluted in 10 mM Tris-HCl, pH 8.0, containing 0.05% Tween-20. Library quality was verified by TapeStation, and sequencing was conducted on NovaSeq X Plus platform (Illumina) at Novogene (Beijing, China) Japan laboratory, with a read length of 150 bp paired-end, targeting approximately 2.5 billion reads (1.25 billion read pairs). Data trimming was performed to retain the first 27 bp of both R1 and R2 reads and generated raw FASTQ files.

Following sequencing, FASTQ files were processed into Digital Count Conversion (DCC) files using GeoMx NGS Pipeline (v3.1.1.6; NanoString Technologies), with the generated files then analyzed in R using GeomxTools (v3.8.0)[68] as per NanoString's official workflow. Segment-level QC was first applied to ensure the reliability of each ROI. Segments with fewer than 1000 total reads, alignment rate <80%, trimming/stitching rate < 80%, sequencing saturation <50%, <100 nuclei, or <5000 µm² area were removed. Probe-level QC was then performed to exclude low-quality probes; probes were globally removed if their geometric mean count across segments was <10% of the target's mean or flagged as outliers by Grubbs' test in >20% of segments. After probe QC, data were summarized into a gene-level matrix by geometric averaging of multiple probes per target. A limit of quantification (LOQ) was computed per segment as the higher of twice the SD of negative-control probes or a minimum value of 2. Segments with gene detection rates <5% and genes detected in <10% of all segments were excluded. After QC, 11,327 genes and 114 ROIs (Full ROI: 53, immune ROI: 24, tissue ROI: 37) were retained. Normalization was performed using the Q3 (75th percentile) method.

scRNA-seq

Fresh-frozen PPC tissue samples (50 mg) from four patients (OPL7, OPL8, OPL10, and OPL14) were processed using the Chromium Next GEM Single Cell Fixed RNA Sample Preparation Kit (10x Genomics) according to the manufacturer's instructions. Briefly, tissues were minced and incubated in fixation buffer for 24 h at 4 °C. Fixed tissues were then dissociated using the gentleMACS Octo Dissociator (Miltenyi Biotec). Library preparation was performed using the Chromium Fixed RNA Profiling Reagent Kit for Multiplexed Samples (10x Genomics). Fixed single-cell suspensions were hybridized with distinct probe barcodes from the Chromium Fixed RNA

Kit, Human Transcriptome (10x Genomics) at 42 °C for 19 h. After washing to remove unbound probes, equal numbers of cells from each sample were pooled. Gel Beads-in-Emulsion (GEMs) were generated using the Chromium iX (10x Genomics) with the Chromium Next GEM Chip Q Single Cell Kit and 10x GEM Barcodes. Each GEM contained a barcoded Gel Bead with a unique molecular identifier (UMI) and a sample-specific barcode. Following GEM generation, probes were ligated and pre-amplified by PCR. The amplified products were purified using SPRIselect reagent (Beckman Coulter) and used to construct gene expression libraries. Index PCR was performed using the Dual Index Kit TS Set A (10x Genomics) to add Illumina adapters and sample indices. The prepared libraries were sequenced on a NovaSeq 6000 platform (Illumina) at Rhelixa (Tokyo, Japan) with 150 bp paired-end reads, targeting approximately 800 million reads (400 million read pairs), and raw FASTQ files were generated.

The resulting FASTQ files were processed using cellranger-multi in Cell Ranger (v8.0; 10x Genomics) with the human transcriptome reference (GRCh38, 2024-A) and the Human Transcriptome Probe Set v1.0, following the official 10x Genomics documentation. Quality metrics were evaluated for all samples. Samples OPL7, OPL8, and OPL14 showed sufficient cell counts (2306–11,259 cells) and high confidently mapped read percentages (92.9%, 93.9%, and 52.8%, respectively). Sample OPL10 exhibited poor quality (527 cells, 48.8% mapped reads) and was excluded from downstream analyses. QC-passed data were analyzed using Seurat (v5.1.0)[69]. For data integration, PPC samples were combined with publicly available lung adenocarcinoma scRNA-seq data from GSE131907[20]. Data from primary tumors (tLung, $n = 11$) and normal lungs (nLung, $n = 11$) were extracted for integration. Cells with fewer than 200 or more than 5000 detected genes, or with >10% mitochondrial gene content, were filtered out. After QC, normalization was performed using the LogNormalize function, and scaling was applied with ScaleData function using default parameters. Integration was performed using the FindIntegrationAnchors and IntegrateData functions in Seurat. The top 2000 variable genes were selected using the vst method in FindVariableFeatures function. Dimensionality reduction was performed by principal component analysis (PCA) using the first 25 principal components. A K-nearest neighbor (KNN) graph was constructed with FindNeighbors function, and clustering was performed using FindClusters function. Cell clusters were visualized using UMAP. Cell-type annotation was based on the GSE131907 reference dataset, and cluster validation was performed using FindAllMarkers function. Among epithelial cells, clusters lacking nLung-derived cells were defined as epithelial tumor cells. A cluster connecting epithelial and fibroblast populations, consisting exclusively of PPC-derived cells, was defined as sarcomatoid cells. To validate tumor identity, copy-number variation (CNV) inference was performed using inferCNV (v1.3.3) (https://github.com/broadinstitute/InferCNV). Epithelial cells from nLung served as reference. CNV states were inferred using a Hidden Markov Model (HMM) with denoising, and subclustering was conducted with the Leiden algorithm to identify distinct CNV patterns. The standard deviation (SD) of inferred CNV values was used to confirm tumor-associated genomic alterations.

Cells derived from PPC samples were extracted for further analysis. Variable gene selection, dimensionality reduction, and clustering were repeated using the first 20 principal components. To focus on tumor cell trajectories, epithelial and tumor cells (Ciliated, Club, AT1, AT2, Ep, and Sa) were isolated. The number of variable features was set to 500, and 15 principal components were used for dimensionality reduction. Re-clustering was performed at a resolution of 0.2 using FindClusters function, resulting in four major subclusters: epithelial tumor cells (Ep1, Ep2) and sarcomatoid cells (Sa1, Sa2). Cluster annotations were refined accordingly.

Unsupervised analysis for spatial transcriptomics

Unsupervised analyses were performed using GeomxTools. Normalized gene expression values were log2-transformed before analysis. For dimensionality reduction based on overall gene expression across ROIs, UMAP was applied. Unsupervised hierarchical clustering was conducted by calculating the coefficient of variation (CV) for each gene (g) using the formula CV(g) = SD(g) / mean(g). The top 20% of genes with the highest variability were selected for clustering. Hierarchical clustering was performed using the average linkage method and correlation distance for both rows (genes) and columns (ROIs).

DEGs analysis

For spatial transcriptomics, differential expression analysis was performed using the limma package (v3.60.4)[70] in R. Normalized gene expression values were log2-transformed prior to analysis. A linear model was fitted to each gene to compare expression between defined regions. The empirical Bayes method was applied to moderate the standard errors of the estimated Log2FC. Genes were identified as differentially expressed based on Log2FC values and statistical significance. P-values were adjusted using the Benjamini–Hochberg (BH) method to control FDR. The epithelial component of sample OPL9 was excluded from DEGs analysis because it contained a mixed population of squamous and sarcomatoid elements that could not be separated, potentially confounding comparisons with other epithelial regions.

For scRNA-seq, DEGs analysis between tumor subclusters was performed using the FindMarkers function in Seurat. The test used the limma implementation of the Wilcoxon rank-sum test (argument: test.use = "wilcox_limma"). *P*-values were adjusted using the Bonferroni correction method.

GSEA and ORA

GSEA was performed using the clusterProfiler (v4.12.2)[71] and enrichplot (v1.24.2)[71] packages in R. Genes were ranked in descending order of Log2FC and converted to Entrez Gene IDs prior to analysis. Gene sets were obtained from the Molecular Signatures Database (MSigDB) (v2024.1.Hs)[72], including HALLMARK[73], C2[72], and the KEGG_LEGACY subset of canonical pathways. Additionally, the Reactome Pathway Gene Set was retrieved from the Reactome database[74]. All gene sets were downloaded as GMT files. GSEA was performed using the GSEA function in clusterProfiler with default parameters, and *P*-values were adjusted using the BH method.

ORA was conducted using the enrichGO function in clusterProfiler, focusing on the Gene Ontology (GO)[75] Biological Process (BP) category with default parameters. The background gene set consisted of all genes included in the DEGs analysis. P-values were adjusted using the BH method.

TF activity analysis

TF activities were inferred using the DoRothEA (v1.0.0) resource[76] in combination with the Virtual Inference of Protein-activity by Enriched Regulon analysis (VIPER) algorithm, as implemented in the decoupleR package (v2.9.7)[77] in R. Normalized log2-transformed gene expression values were used as input. The DoRothEA gene regulatory network was filtered with confidence levels A, B, and C, as recommended. TF activity scores were computed using the run_viper function in decoupleR with default parameters. Group-wise comparisons of TF activity scores were performed using the Wilcoxon rank-sum test, and *P*-values were adjusted for multiple testing using the BH method.

Trajectory analysis

Single-cell trajectory inference and pseudotime analyses were performed using the scTour (v1.0.0)[25], in combination with Scanpy (v1.10.4)[78] in Python (v3.10.16). Gene expression and cell annotation data were extracted from the Seurat object after normalization. The top 1000 highly variable genes were selected using the Seurat_v3 method. scTour transformed the dataset into a latent space representation using mean squared error (MSE) loss and the implicit Adams method for solving ordinary differential equations. The model was trained for 400 epochs. The latent representation was optimized by combining inferred latent variables (argument: alpha_z = 0.5) and predicted latent variables (argument: alpha_predz = 0.5), achieving a balanced integration of cellular states and trajectory predictions. UMAP was applied to the latent space to display distinct cellular clusters and

pseudotime trajectories. Vector fields were overlaid to illustrate the directionality of cellular transitions. The UMAP neighborhood graph was computed using 15 neighbors, minimum distance = 0.5, and spread = 0.2. Vector orientations were reversed following the official scTour documentation (https://sctour.readthedocs.io/en/latest/notebook/scTour_inference_PostInference_adjustment.html).

Inferring ligand–receptor interactions between single cells

Cell–cell communication analysis was performed using CellPhoneDB (v5.0.1)[26]. The pipeline was implemented in Python (v3.8.20), and the statistical inference method integrated in CellPhoneDB was applied. Normalized gene expression and cell annotation data were used as input. Ligand–receptor interaction predictions were performed using the CellPhoneDB v5 database. Statistical analysis was executed with the cpdb_statistical_analysis_method function using default parameters. Visualization of the inferred interaction results (heatmap and dot plot) was performed using the ktplots (v2.4.0) package[26] in R.

Automated cell annotation of an independent dataset

Processed count matrices from 10 tumor-derived samples reported by Bischoff et al.[29] were obtained. For each sample, Seurat objects were constructed and subjected to standard quality control, including the number of detected genes per cell, total UMI counts per cell, and the fractions of mitochondrial and hemoglobin transcripts, following the authors' published scripts (https://doi.org/10.24433/CO.0121060.v1). Expression values were log-normalized using NormalizeData function to enable identification of highly variable genes (HVGs). The two datasets shared 19436 genes. HVGs were independently identified using the FindVariableFeatures function ($n = 3000$ per dataset), and the intersection of 1592 HVGs was retained for downstream analysis. Both datasets were converted into SingleCellExperiment objects. The PPC dataset served as the reference, with prior cluster-based cell-type annotations. SingleR[30] was applied at the single-cell level using the 1592 intersected HVGs, employing the Wilcoxon rank-sum test as the differential expression method. Four tumor-cell subtypes (Ep1, Ep2, Sa1, and Sa2) defined in the PPC reference were then mapped onto Bischoff's cohort. SingleR-assigned labels corresponding to these categories were extracted for each sample. The number of cells per subtype was divided by the total number of tumor cells in each sample to normalize for differences in tumor-cell yield. The resulting proportions were rescaled within each subtype so that the total contribution across all samples summed to 1, ensuring equal sample weighting regardless of cell number. Relative subtype proportions were visualized as stacked bar plots annotated with total cell counts per subtype

Gene signature score calculation

For spatial transcriptomics and public datasets, gene signature scores were calculated using the GSVA package (v1.52.3)[79] in R. Normalized log2-transformed gene expression data were used as input. Enrichment scores were computed using the gsva function with parameters defined by the zscoreParam function[80] with default parameters. The resulting gene set–by–sample matrix contained NES, which were used as signature scores for downstream analyses.

For single-cell datasets, signature scores were computed using the AddModuleScore function in Seurat with default parameters. Log-normalized expression data were used as input, and the resulting scores were used as signature scores in subsequent analyses.

Analysis of PPC–ECM signature scores across LUAD histological subtypes

To examine whether PPC–ECM sig scores varied among histological subtypes of LUAD, we analyzed two publicly available datasets from the GEO repository. GSE166720[33] includes RNA sequencing data from early-stage LUAD samples, provided as FPKM-normalized and log2-transformed expression matrices with corresponding clinical metadata. GSE58772[34] comprises microarray data from histologically microdissected LUAD samples, quantile-normalized and variance-stabilized, accompanied by detailed metadata. Data were retrieved using the GEOquery package (v2.72.0)[81] in R. Expression matrices and clinical metadata were merged using common sample identifiers. Samples lacking histological subtype information ($n = 1$ in GSE166720) were excluded. Histological subtypes were categorized according to metadata. Sig scores were calculated as described in *Gene signature score calculation*. Pairwise comparisons of PPC–ECM sig scores between histological subtypes were performed using Wilcoxon rank-sum tests with BH correction.

Survival analyses using TCGA–LUAD and CPTAC–LUAD datasets

For TCGA–LUAD cohort, RNA sequencing and clinical data were obtained from the Genomic Data Commons (GDC) using the GDCquery, GDCdownload, and GDCprepare functions in the TCGAbiolinks package (v2.32.0)[82]. RNA-seq data (STAR–Counts) were transformed to TPM and log2-normalized. Survival time was defined as the number of days to death or last follow-up, whichever occurred first, with a 60-month cutoff. Pathological stages were categorized into Stage I–IV. Sig scores were calculated as described in *Gene signature score calculation*. Samples were stratified into "High" and "Low" groups based on the median sig score. Kaplan–Meier survival curves were generated, and log-rank tests were performed to compare OS between groups. Multivariable Cox proportional hazards models were applied to estimate HR and 95% confidence intervals (CI), adjusting for age and pathological stage. Survival analysis was performed using the survival (v3.7-0)[83] and survminer (v0.5.0)[84] packages in R.

For CPTAC–LUAD cohort, transcriptomic, proteomic, and clinical data were downloaded using the cptac package (v1.5.14)[85] in Python (v3.10.16). Data preprocessing followed the Baylor College of Medicine (BCM) standard pipelines. Samples annotated as normal (suffix ".N") or lacking key clinical variables (age, pathological stage, survival information) were excluded. Only genes shared between RNA and proteome datasets were retained, yielding 103 tumor samples. First, median RNA and protein expression values across all genes were compared using Pearson's correlation to assess overall concordance. Second, RNA and protein expression values were standardized to Z-scores across samples, and Pearson's correlation coefficients were calculated individually for each gene. For survival analysis of protein expression, normalized log2-transformed protein expression data were used to compute PPC–ECM sig scores, which were then dichotomized by the median. OS was evaluated using Kaplan–Meier curves and log-rank tests at 24- and 60-month cutoffs. Cox proportional hazards models adjusted for pathological stage and age were used to estimate HR and 95% CI. Survival analyses were conducted using the survival and survminer packages. The results are based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga) and CPTAC (https://pdc.cancer.gov).

Deconvolution of immune cell subtypes

The immune cell composition within each immune-region ROI was estimated using CIBERSORTx[35]. Normalized $\log_2$-transformed gene expression data were used as input to stabilize variance. Deconvolution was performed with the LM22 signature matrix provided by the official CIBERSORTx platform. The analysis output included the relative proportions of each immune cell type within individual ROIs and used for statistical tests (Welch's t-tests) across components.

Correlation analysis between metastatic lymph nodes and primary tumor components

For each sample, average expression values were calculated per gene within each component. Pearson's correlation coefficients were computed between metastatic and Epithelial or Sarcomatoid components using the cor() function in R.

IHC

FFPE blocks were sectioned at 4 µm thickness, deparaffinized, and rehydrated through graded ethanol. Endogenous peroxidase activity was quenched by incubation in 3.0% H_2O_2 for 5 min. Sections were then blocked with 2.5% Normal Horse Serum (Vector Laboratories) for 20 min and incubated overnight at 4 °C with a c-MET primary antibody (Cell Signaling Technology; Cat. #8198; dilution 1:150). Sections were incubated for 30 min with the ImmPRESS VR Polymer HRP Anti-Rabbit IgG Reagent (Vector Laboratories). Immunoreactivity was detected using the ImmPACT DAB Peroxidase Substrate Kit (Vector Laboratories) and counterstained with Mayer's hematoxylin.

Statistics and reproducibility

For differential expression analyses, genes were considered significant when FDR or adjusted $P < 0.01$ and $|\log2FC| > 1$. For GSEA and ORA, adjusted $P < 0.05$ was regarded as significant. For TF activity inference, regulons were compared between groups, with FDR < 0.05 considered significant. For ligand–receptor interaction analyses, a permutation $P < 0.05$ was regarded as significant. For comparisons of PPC–ECM sig scores across LUAD histological subtypes, FDR < 0.05 was used as the significance threshold. For survival analyses, log-rank $P < 0.05$ was considered statistically significant, and statistical significance in Cox proportional hazards models was defined as 95% CI that did not cross 1.0. For immune-cell deconvolution analyses, $P < 0.05$ considered significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The spatial transcriptomics data have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE289483. Whole slide images used for spatial transcriptomics (H&E-stained) are available in Zenodo (https://doi.org/10.5281/zenodo.16736588). The scRNA-seq data have been deposited in GEO under accession code GSE289611. The source data underlying the figures of this study are provided in the Supplementary Data files. All other datasets are available from the corresponding author upon reasonable request.

Received: 25 March 2025; Accepted: 30 October 2025;
Published online: 16 December 2025

References

1. Chang, Y. L., Lee, Y. C., Shih, J. Y. & Wu, C. T. Pulmonary pleomorphic (spindle) cell carcinoma: peculiar clinicopathologic manifestations different from ordinary non-small cell carcinoma. *Lung Cancer* **34**, 91–97 (2001).
2. Travis, W. D. et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J. Thorac. Oncol.* **10**, 1243–1260 (2015).
3. Liu, X. et al. Next-generation sequencing of pulmonary sarcomatoid carcinoma reveals high frequency of actionable MET gene mutations. *J. Clin. Oncol.* **34**, 794–802 (2016).
4. Domblides, C. et al. Efficacy of immune checkpoint inhibitors in lung sarcomatoid carcinoma. *J. Thorac. Oncol.* **15**, 860–866 (2020).
5. Yang, Z. et al. Integrated molecular characterization reveals potential therapeutic strategies for pulmonary sarcomatoid carcinoma. *Nat. Commun.* **11**, 4878 (2020).
6. Roma, L. et al. Tracing tumor heterogeneity of pleomorphic carcinoma of the lung. *J. Thorac. Oncol.* **19**, 1284–1296 (2024).
7. Ndoye, A., Miskin, R. P. & DiPersio, C. M. Integrin α3β1 represses reelin expression in breast cancer cells to promote invasion. *Cancers* **13**, 344 (2021).
8. Longmate, W. M. et al. Integrin α3β1 on tumor keratinocytes is essential to maintain tumor growth and promotes a tumor-supportive keratinocyte secretome. *J. Invest. Dermatol.* **141**, 142–151.e6 (2021).
9. Hamidi, H. & Ivaska, J. Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer* **18**, 533–548 (2018).
10. Cabral-Pacheco, G. A. et al. The roles of matrix metalloproteinases and their inhibitors in human diseases. *Int. J. Mol. Sci.* **21**, 9739 (2020).
11. Ratajczak-Wielgomas, K., Kmiecik, A. & Dziegiel, P. Role of periostin expression in non-small cell lung cancer: periostin silencing inhibits the migration and invasion of lung cancer cells via regulation of MMP-2 expression. *Int. J. Mol. Sci.* **23**, 1240 (2022).
12. Hapke, R. Y. & Haake, S. M. Hypoxia-induced epithelial to mesenchymal transition in cancer. *Cancer Lett.* **487**, 10–20 (2020).
13. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J. & Werb, Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* **11**, 5120 (2020).
14. Weber, L. I. & Hartl, M. Strategies to target the cancer driver MYC in tumor cells. *Front. Oncol.* **13**, 1142111 (2023).
15. Tetreault, M. P., Yang, Y. & Katz, J. P. Krüppel-like factors in cancer. *Nat. Rev. Cancer* **13**, 701–713 (2013).
16. Domenici, G. et al. A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. *Oncogene* **38**, 3151–3169 (2019).
17. Chen, Q. et al. SOX9 modulates the transformation of gastric stem cells through biased symmetric cell division. *Gastroenterology* **164**, 1119–1136.e12 (2023).
18. Zhang, H. et al. Coexpression of FOXK1 and vimentin promotes EMT, migration, and invasion in gastric cancer cells. *J. Mol. Med.* **97**, 163–176 (2019).
19. Wu, M. et al. FOXK1 interaction with FHL2 promotes proliferation, invasion and metastasis in colorectal cancer. *Oncogenesis* **5**, e271 (2016).
20. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).
21. Celià-Terrassa, T. & Kang, Y. How important is EMT for cancer metastasis?. *PLoS Biol.* **22**, e3002487 (2024).
22. Gao, F. et al. Hes1 is involved in the self-renewal and tumourigenicity of stem-like cancer cells in colon cancer. *Sci. Rep.* **4**, 3963 (2014).
23. Goto, N. et al. Distinct roles of HES1 in normal stem cells and tumor stem-like cells of the intestine. *Cancer Res* **77**, 3442–3454 (2017).
24. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434**, 843–850 (2005).
25. Li, Q. ScTour: a deep learning architecture for robust inference and accurate prediction of cellular dynamics. *Genome Biol.* **24**, 149 (2023).
26. Troulé, K. et al. CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. *Nat. Protoc*. https://doi.org/10.1038/s41596-024-01137-1 (2025).
27. Malta, T. M. et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354.e15 (2018).
28. Cruz da Silva, E., Dontenwill, M., Choulier, L. & Lehmann, M. Role of integrins in resistance to therapies targeting growth factor receptors in cancer. *Cancers* **11**, 692 (2019).
29. Bischoff, P. et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* **40**, 6748–6758 (2021).
30. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
31. Bill, R. et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515–524 (2023).

32. Thorlacius-Ussing, J. et al. The collagen landscape in cancer: profiling collagens in tumors and in circulation reveals novel markers of cancer-associated fibroblast subtypes. *J. Pathol.* **262**, 22–36 (2024).
33. Yoo, S. et al. Integrative network analysis of early-stage lung adenocarcinoma identifies aurora kinase inhibition as interceptor of invasion and progression. *Nat. Commun.* **13**, 1592 (2022).
34. Zabeck, H. et al. Molecular signatures in IASLC/ATS/ERS classified growth patterns of lung adenocarcinoma. *PLoS One* **13**, e0206132 (2018).
35. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
36. Haas, M. et al. PIM2 kinase has a pivotal role in plasmablast generation and plasma cell survival, opening up novel treatment options in myeloma. *Blood* **139**, 2316–2337 (2022).
37. Huang, D. et al. Targeting regulator of G protein signaling 1 in tumor-specific T cells enhances their trafficking to breast cancer. *Nat. Immunol.* **22**, 865–879 (2021).
38. Bai, Y., Hu, M., Chen, Z., Wei, J. & Du, H. Single-cell transcriptome analysis reveals RGS1 as a new marker and promoting factor for T-cell exhaustion in multiple cancers. *Front. Immunol.* **12**, 767070 (2021).
39. Karim, N. A. et al. Pulmonary sarcomatoid carcinoma: University of Cincinnati experience. *Oncotarget* **9**, 4102–4108 (2017).
40. Baldovini, C., Rossi, G. & Ciarrocchi, A. Approaches to tumor classification in pulmonary sarcomatoid carcinoma. *Lung Cancer* **10**, 131–149 (2019).
41. Pruis, M. A. et al. Highly accurate DNA-based detection and treatment results of MET exon 14 skipping mutations in lung cancer. *Lung Cancer* **140**, 46–54 (2020).
42. Organ, S. L. & Tsao, M. S. An overview of the c-MET signaling pathway. *Ther. Adv. Med. Oncol.* **3**, S7–S19 (2011).
43. Schrock, A. B. et al. Pulmonary sarcomatoid carcinomas commonly harbor either potentially targetable genomic alterations or high tumor mutational burden as observed by comprehensive genomic profiling. *J. Thorac. Oncol.* **12**, 932–942 (2017).
44. Kojima, K. et al. PIK3CA mutations associated with a poor postoperative prognosis in patients with pulmonary pleomorphic carcinoma: a retrospective cohort study. *BMC Cancer* **22**, 1066 (2022).
45. Cox, T. R. The matrix in cancer. *Nat. Rev. Cancer* **21**, 217–238 (2021).
46. Samaržija, I. et al. Integrin crosstalk contributes to the complexity of signalling and unpredictable cancer cell fates. *Cancers* **12**, 1910 (2020).
47. Madamanchi, A., Zijlstra, A. & Zutter, M. M. Flipping the switch: integrin switching provides metastatic competence. *Sci. Signal.* **7**, pe9 (2014).
48. Rabie, E. M. et al. Substratum stiffness signals through integrin-linked kinase and β1-integrin to regulate midbody proteins and abscission during EMT. *Mol. Biol. Cell* **32**, 1664–1676 (2021).
49. Iwatate, Y. et al. Transcriptomic analysis reveals high ITGB1 expression as a predictor for poor prognosis of pancreatic cancer. *PLoS One* **17**, e0268630 (2022).
50. Zhang, J. et al. Upregulation of nectin-4 is associated with ITGB1 and vasculogenic mimicry and may serve as a predictor of poor prognosis in colorectal cancer. *Oncol. Lett.* **18**, 1163–1170 (2019).
51. Chang, R. et al. ITGB1-DT facilitates lung adenocarcinoma progression via forming a positive feedback loop with ITGB1/Wnt/β-catenin/MYC. *Front. Cell Dev. Biol.* **9**, 631259 (2021).
52. Rana, P. S. et al. The WAVE2/miR-29/Integrin-β1 oncogenic signaling axis promotes tumor growth and metastasis in triple-negative breast cancer. *Cancer Res. Commun.* **3**, 160–174 (2023).
53. Cheuk, I. W. et al. ITGAV targeting as a therapeutic approach for treatment of metastatic breast cancer. *Am. J. Cancer Res.* **10**, 211–223 (2020).
54. Chen, C. H. et al. Study of cathepsin B inhibition in VEGFR TKI treated human renal cell carcinoma xenografts. *Oncogenesis* **8**, 15 (2019).
55. Chen, Q., Zhou, X. W., Zhang, A. J. & He, K. ACTN1 supports tumor growth by inhibiting Hippo signaling in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **40**, 23 (2021).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
57. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arXiv.org/1303.3997 (2013).
58. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
60. Benjamin, D. et al. Calling somatic SNVs and indels with Mutect2. Preprint at https://doi.org/10.1101/861054 (2019).
61. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
62. Cingolani, P. et al. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. *SnpSift. Front. Genet.* **3**, 35 (2012).
63. Kandoth, C. et al. mskcc/vcf2maf: vcf2maf v1.6.14 (v1.6.14). Zenodo. https://doi.org/10.5281/zenodo.840297 (2017).
64. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747–1756 (2018).
65. Wang, S., Tao, Z., Wu, T. & Liu, X. S. Sigflow: an automated and comprehensive pipeline for cancer genome mutational signature analysis. *Bioinformatics* **37**, 1590–1592 (2021).
66. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
67. Renevey, A. et al. nf-core/rnafusion: 3.0.2. Zenodo. https://doi.org/10.5281/zenodo.10952554 (2024).
68. Griswold, M., Ortogero, N., Yang, Z., Vitancol, R. & Henderson, D. GeomxTools: NanoString GeoMx Tools. R package version 3.8.0, https://doi.org/10.18129/B9.bioc.GeomxTools (2024).
69. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
70. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
71. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
72. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
73. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
74. Milacic, M. et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* **52**, D672–D678 (2024).
75. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
76. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* **29**, 1363–1375 (2019).
77. Badia-I-Mompel, P. et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).
78. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

79. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
80. Lee, E., Chuang, H. Y., Kim, J. W., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.* **4**, e1000217 (2008).
81. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
82. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71 (2016).
83. Therneau, T. A Package for Survival Analysis in R. R package version 3.7-0, https://CRAN.R-project.org/package=survival (2024).
84. Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.5.0, https://CRAN.R-project.org/package=survminer (2024).
85. Lindgren, C. M. et al. Simplified and unified access to cancer proteogenomic data. *J. Proteome Res.* **20**, 1902–1910 (2021).

Acknowledgements

We thank Yumiko Kawase, Yayoi Kubota (Center for Comprehensive Genomic Medicine, Okayama University Hospital, Okayama, Japan) and Fumiko Isobe (Department of Thoracic, Breast and Endocrinological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan) for their technical support. This work was supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS KAKENHI grant numbers: JP23K27687 and JP23K18317 to K.Shi., JP23K27688 to S.Tom., JP23K24421 to S.Toy). Figure 2A was created in BioRender. A.M. (2025) https://BioRender.com/z3y8bwc.

Author contributions

A.M. conceived the study, managed project administration, performed data curation, formal analysis, investigation, developed the methodology, implemented software, validated and visualized the results, and contributed to writing the original draft and subsequent revisions. K.Shi. contributed to conceptualization, funding acquisition, methodology, project administration, resources, supervision, and participated in writing the original draft and subsequent revisions. S.Tom. contributed to conceptualization, data curation, funding acquisition, methodology, resources, software implementation, validation, and assisted in manuscript revision. M.Ohk., K.H., R.Fujiw., K.I., R.Fujii., T.H., N.H., K.O., R.Y., F.M., M.Y., Y.F., Y.Tom., S.Tan., K.Suz., K.M., M.Oka., and S.S. provided essential resources. Y.O., A.T., D.E., and Y.Tog. contributed to manuscript review and editing. H.Y. evaluated the FFPE sections used for spatial transcriptomics as a pathologist. H.I. performed part of the experimental work. S.Toy. contributed to funding acquisition, provided resources, supervised the project, and participated in manuscript revision. All authors approved the final 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/s42003-025-09162-w.

Correspondence and requests for materials should be addressed to Kazuhiko Shien.

Peer review information *Communications Biology* thanks Daniele Tavernari, Robert Homer, and Leonhard Müllauer, who co-reviewed the paper with Alissa Blessing for their contribution to the peer review of this work. Primary Handling Editor: Mengtan Xing. A peer review file is available.