ROSIE-Enabled Spatial Mapping Reveals Architectural Fragmentation and Immune Reprogramming in Lung Adenocarcinoma Evolution

---

Beibei Huang[1], Bo Zhu[2]

[1] Department of Cancer Systems Imaging, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA

[2] Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Corresponding Author:

Beibei Huang

Emails: bhuang1@mdanderson.org; bbh@imdlab.org; bayjuan5@gmail.com

ORCID: 0000-0003-0555-6924

Abstract

The progression of lung adenocarcinoma (LUAD) from precancerous lesions to invasive carcinoma entails extensive remodeling of tissue architecture and immune microenvironments, driven by mechanisms that remain incompletely understood. To investigate these dynamics, we applied ROSIE (RObust in Silico Immunofluorescence), a deep learning framework that computationally infers multiplex biomarker expression directly from routine H&E-stained tissue slides. This approach enables scalable, high-resolution spatial profiling without the need for specialized staining or equipment. Using ROSIE, we analyzed a cohort comprising 114 human

LUAD samples to reconstruct the evolving immune landscape during tumorigenesis. Our analysis revealed a fundamental immunological shift: adaptive immunity progressively recedes, giving way to a dominant innate and immunosuppressive state as lesions advance. Concurrently, we observed progressive architectural fragmentation that marks the transition to malignancy. These findings highlight the power of computational pathology to map cancer evolution at scale and pinpoint critical windows for therapeutic interception, laying the groundwork for more precise and proactive strategies in oncology.

Introduction

Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer and a leading cause of cancer-related mortality worldwide[1–3]. While significant progress has been made in treating advanced-stage disease through targeted therapies and immunotherapies, the molecular and cellular events that drive progression from precancerous lesions to invasive tumors remain poorly understood. Understanding the evolutionary trajectory of LUAD precursors—including atypical adenomatous hyperplasia (AAH), Adenocarcinoma in situ (AIS), and minimally invasive adenocarcinoma (MIA)—offers a critical opportunity for early intervention and interception strategies that could prevent invasive disease altogether[4,5]. The tumor microenvironment (TME), particularly its immune composition and spatial organization, plays a pivotal role in shaping cancer progression[5–7]. Immune surveillance mechanisms can either suppress or inadvertently promote tumor evolution, and the balance between innate and adaptive immunity shifts dynamically as lesions progress. Characterizing these immunological transitions at high spatial resolution across the precancer-to-cancer continuum is essential for identifying stage-specific vulnerabilities and developing precision interception strategies.

Recent advances in multiplex immunofluorescence (mIF) technologies, such as CODEX, Imaging mass cytometry (IMC), and immunohistochemistry (IHC), have enabled spatially resolved profiling of dozens of protein biomarkers simultaneously, revealing complex immune architectures within tumors[8]. Deep learning frameworks such as PixlMap have attempted to bypass whole-cell segmentation by leveraging nuclear features for phenotyping[9], yet these approaches remain tied to resource-intensive mIF pipelines and cannot be readily scaled to archival H&E collections. Single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics like Visium, Xenium have further expanded our ability to dissect cellular heterogeneity and cell-cell interactions within the TME[10–13]. However, these approaches are costly, time-intensive, and require specialized equipment and protocols, limiting their scalability for large-cohort studies and retrospective analyses of archival tissue collections[14]. Current spatial and single-cell technologies are resource-intensive and difficult to scale. **To overcome these scalability limitations, we applied ROSIE (RObust in Silico Immunofluorescence), a deep learning framework that computationally infers multiplex biomarker expression directly from**

routine H&E-stained slides. This approach enables high-resolution spatial profiling at a fraction of the cost and time required for traditional multiplexed immunofluorescence or single-cell technologies. Moreover, most studies focus on advanced-stage tumors, leaving the early evolutionary dynamics of LUAD precursors largely unexplored[15,16].

Hematoxylin and eosin (H&E) staining remains the most ubiquitous histopathology assay in clinical practice due to its affordability and accessibility. While H&E provides rich morphological information, it cannot directly reveal molecular markers or specific immune cell subtypes—information that is critical for understanding TME dynamics and guiding therapeutic decisions[17,18]. Recent computational approaches have begun to extract biological insights from H&E images using deep learning, but their ability to predict spatially resolved, multi-marker protein expression across diverse tissue types and disease stages remains limited[19,20]. Foundation models such as TITAN have demonstrated impressive general-purpose slide-level representations across diverse cancer types[20]. However, these models are primarily optimized for diagnosis, prognosis, and report generation, rather than spatially resolved immune biomarker inference. Moreover, their reliance on massive pretraining datasets and multimodal alignment limits applicability in smaller, disease-specific cohorts.

Despite the recognized importance of the TME in cancer progression, a comprehensive, spatially resolved understanding of immune evolution during LUAD precancer development is lacking. Specifically, the spatiotemporal remodeling of immune neighborhoods across the AAH-AIS-MIA-LUAD continuum remains unmapped. It is unknown whether specific immune features represent actionable targets for early interception. Addressing this gap requires scalable technologies that can integrate spatial protein expression data with transcriptomic and functional validation across both human and mouse models—capabilities that traditional mIF approaches cannot feasibly provide at the required scale.

Research Highlights

In this study, we applied the ROSIE (RObust in Silico Immunofluorescence from H&E images) deep-learning framework[21] to systematically explore the spatial and molecular dynamics of LUAD precancer evolution. ROSIE enables protein-level inference of up to 50 biomarkers directly from routine H&E-stained slides, bypassing the need for costly mIF experiments and allowing high-throughput, spatially resolved mapping of immune and stromal components without additional experimental assays. By analyzing 114 human LUAD and precursor samples using ROSIE-predicted virtual mIF (v-mIF). Using scRNA-seq and spatial transcriptomics, we reconstructed the immune landscape across evolutionary stages and identified a transition from innate to adaptive immunity in early lesions. Critically, ROSIE-guided stratification enabled stage-specific therapeutic testing, revealing that interventions during the precancer stage—but not in advanced disease—significantly reduced tumor burden and enhanced immune activation. This cross-species, spatially resolved approach demonstrates the transformative potential of AI-powered computational pathology for biomarker discovery, therapeutic stratification, and precision oncology..

Results

ROSIE enables scalable multiplex biomarker inference from H&E slides

We applied ROSIE to generate virtual multiplex immunofluorescence (v-mIF) predictions directly from H&E-stained tissue sections. Starting with whole-slide images, ROSIE predicted expression levels of up to 50 biomarkers across 816 regions of interest (ROIs) encompassing 3,817,675 individual cells from human LUAD tissues spanning normal lung, Atypical Adenomatous Hyperplasia (AAH), Adenocarcinoma In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and invasive adenocarcinoma (IAC) stages. Larger panels enable more comprehensive spatial profiling through progressive annotation enrichment. While the initial 10-biomarker panel (PDL1, CD8, CD163, PanCK, EpCAM, Ki67, CD4, FoxP3, CD31, Vimentin) captures major immune and epithelial compartments, expansion to 50 biomarkers adds 40 additional proteins. These include immune checkpoint molecules (PD1, LAG3, TIGIT, VISTA, ICOS), myeloid markers (CD68, CD14, CD11c, HLA-DR, Gal3, MPO), B cell markers (CD20, CD21, CD38, CD79a, BCL2), and functional markers (HLA-A, HLA-E, IFNγ, IDO1, CD40). Furthermore, these newly added markers annotate previously uncharacterized stromal and immune compartments, reducing unclassified 'background' areas. As depicted in Figure 1, ROSIE's scalability is demonstrated by sequentially generating v-mIF predictions with increasing biomarker complexity—10-marker, 25-marker, and 50-marker panels (left to right)—from the original H&E image without requiring additional staining. This progressive overlay reveals that larger marker panels enable more comprehensive spatial profiling of immune and stromal components.

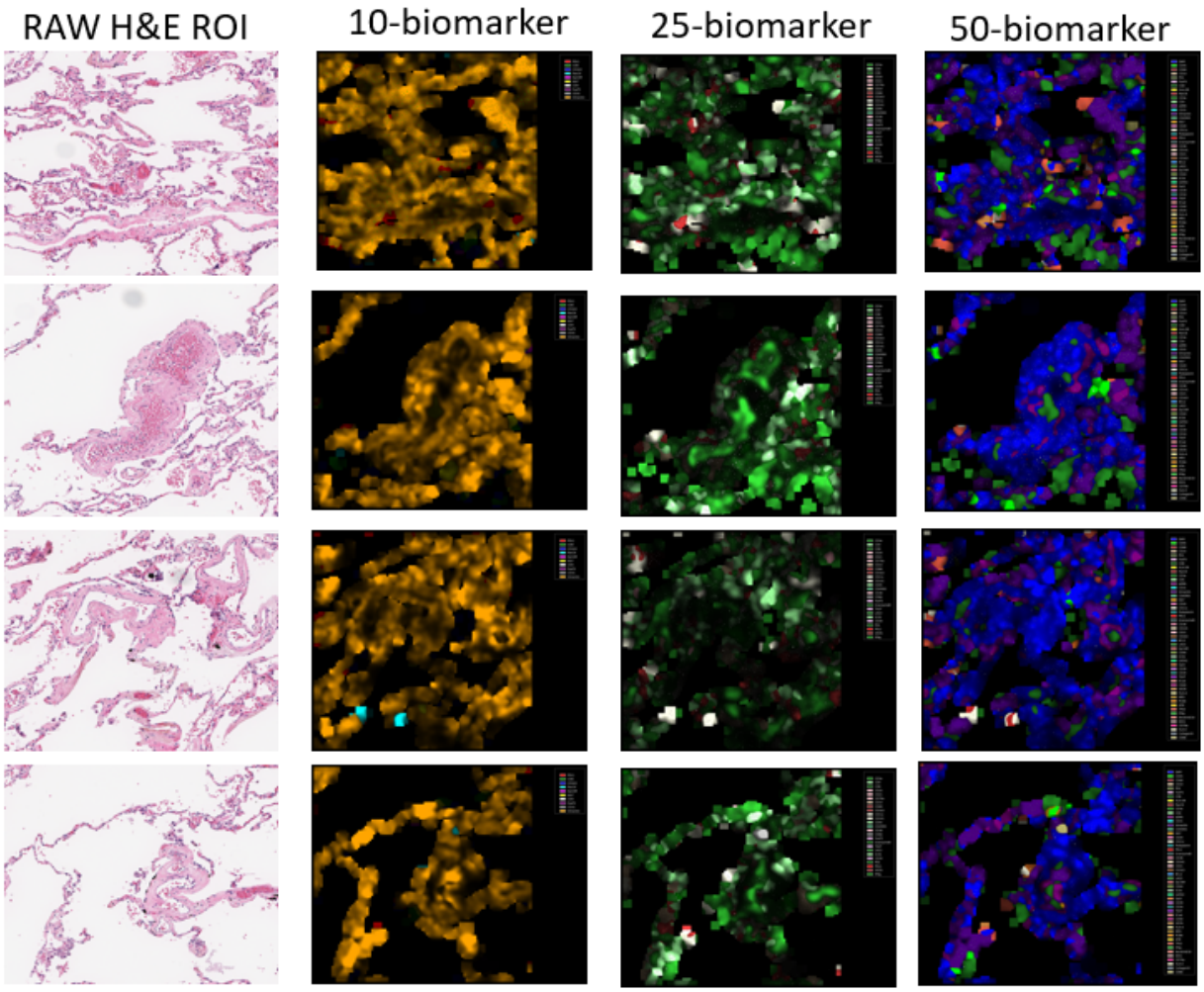

Figure 1 Multi-panel ROSIE predictions from a single H&E slide demonstrate scalable biomarker inference. Left: Original H&E whole slide image (sample R01). Center and Right: ROSIE-generated virtual multiplex immunofluorescence (v-mIF) predictions using 10-, 25-, and 50-biomarker panels derived from the same H&E input without additional staining. As panel size increases, previously unannotated tissue regions become classified (visible as increased color diversity). White arrows in the third panel highlight examples of newly annotated regions resolved by the expanded panel. Scale bar: 100 µm.

To assess the impact of biomarker panel size on cellular resolution, we performed UMAP-based dimensionality reduction using predicted marker signals. The 25-marker panel captured major immune lineages but yielded compact clustering with limited resolution of rare populations (Figure 2, left). In contrast, the 50-marker panel revealed finer cellular heterogeneity, enabling identification of eosinophils and mast cells—populations absent in the reduced panel due to missing lineage-specific markers (e.g., GATA3, MPO, Galectin-3) (Figure 2, right). This comparison demonstrates that comprehensive biomarker selection is critical for accurate cell type classification.

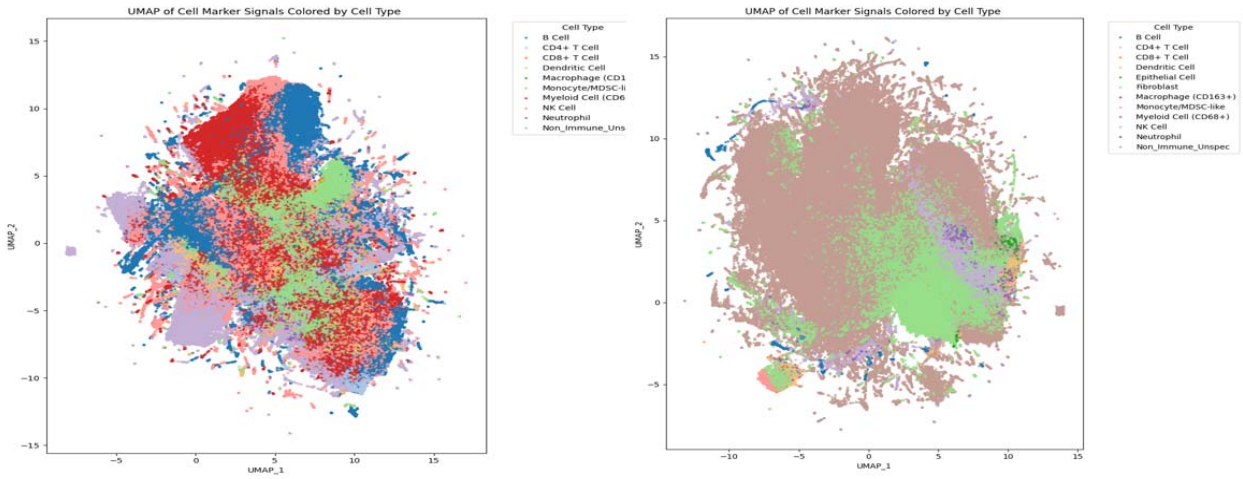

Figure 2 ***UMAP visualization comparing 25-marker versus 50-marker panels.*** *Analysis of 3,817,675 cells colored by cell type. Left: 25-marker panel captures major immune lineages with compact clustering. Right: 50-marker panel reveals finer heterogeneity, resolving eosinophils and mast cells absent in the reduced panel due to missing lineage-specific markers (GATA3, MPO, Galectin-3). This demonstrates how comprehensive biomarker selection enhances resolution of rare cell populations.*

Cell phenotyping reveals 14 distinct populations across LUAD progression

Using the 50-marker panel, we classified cells into 14 distinct types based on canonical lineage markers, including T cell subsets (CD4+, CD8+, Tregs), B cells, NK cells, myeloid populations (macrophages, dendritic cells, monocytes, MDSCs), and structural cells (epithelial, endothelial, fibroblasts) (Figure 3, left). Marker expression profiles showed expected patterns: CD3 and CD8 co-expression in cytotoxic T cells, CD163 enrichment in M2 macrophages, and pan-cytokeratin in epithelial cells. DBSCAN spatial clustering identified 6,522 microenvironmental niches, demonstrating that ROSIE preserves both molecular identity and tissue architecture (Figure 3, right). Cell–cell interaction networks derived from these niches are visualized in Supplementary Figures S1 to S5.

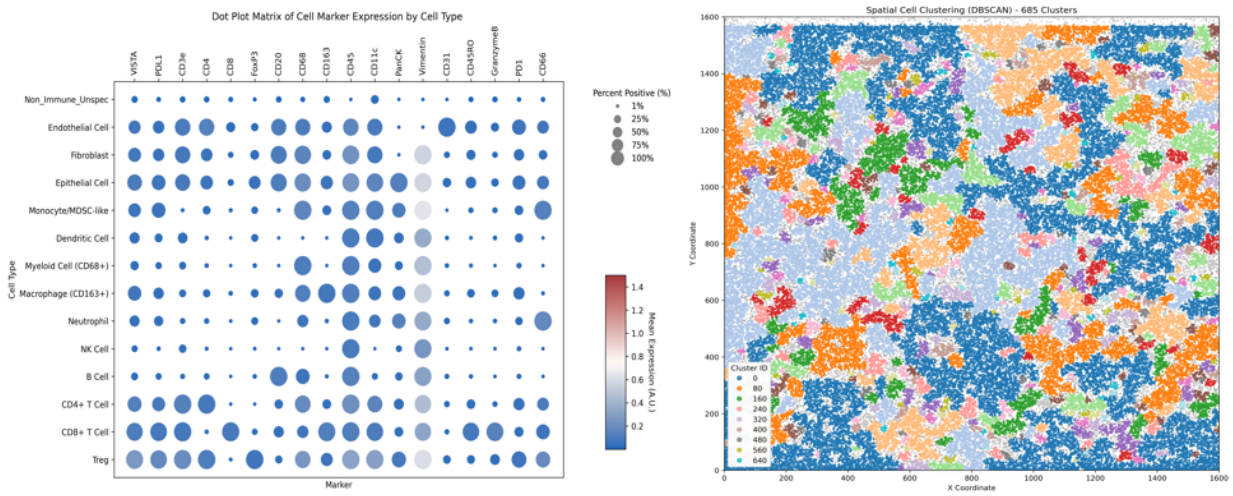

Figure 3. Cell phenotyping and spatial clustering of LUAD tissue microenvironment. *Left: Dot plot showing marker expression across 14 cell types from 3,817,675 cells. Dot size represents percentage of marker-positive cells (1-100%); color intensity*

indicates mean expression (0-1.4 AU). Cell types include immune (T cells, B cells, NK cells, myeloid) and structural populations (epithelial, fibroblast, endothelial). Right: Spatial distribution of 300,000 sampled cells colored by DBSCAN clusters (685 total). Gray dots indicate 18,893 unclustered cells. This reveals architectural organization and discrete cellular neighborhoods within the tumor microenvironment。

Spatial architecture transitions from cohesive to fragmented during progression

Having established cellular identities, we next examined how these populations organize spatially across disease progression. We applied DBSCAN clustering to quantify architectural changes across disease stages (**Figure 4**). **Normal** lung tissue (233,459 cells, ε=0.01) exhibited **991** clusters with loosely distributed spatial domains. **AAH** samples (511,157 cells, ε=0.0065) showed increased density and the emergence of defined domains, totaling **9,062** clusters. **AIS** (437,918 cells, ε=0.006) displayed architectural condensation with **12,841** identified spatial clusters. **MIA** (776,472 cells, ε=0.0055) marked the onset of structural fragmentation with **4,936** clusters. **IAC** (1,858,669 cells, ε=0.003) displayed extensive heterogeneity and loss of organized structure, characterized by **7,314** dense micro-clusters. This progressive breakdown of spatial continuity quantitatively captures the architectural collapse accompanying malignant invasion (**Table 1)**.

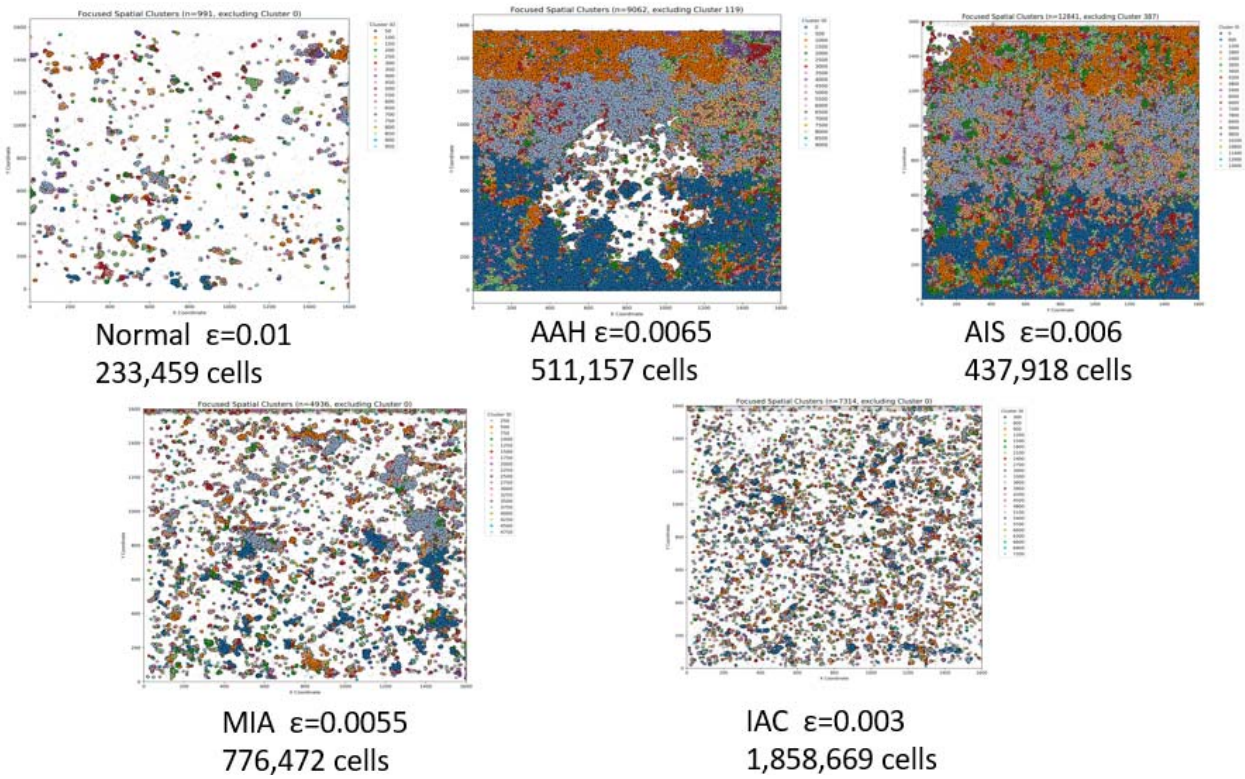

Figure 4. DBSCAN spatial clustering reveals progressive architectural disruption during LUAD development. *Spatial clustering maps across Normal lung, AAH, AIS, MIA, and IAC tissues demonstrate distinct patterns of tissue organization during malignant progression.* ***Normal*** *lung displays* ***loosely distributed spatial domains****, reflecting preserved tissue integrity.* ***AAH*** *exhibits increased cellular density with emergence of defined spatial domains.* ***AIS*** *shows further architectural condensation with a single dominant macro-structure. In contrast,* ***MIA*** *marks the* ***onset of significant structural fragmentation****, showing a loss of cohesive macro-structures and the emergence of multiple small-to-medium clusters.* ***IAC*** *demonstrates dramatic architectural fragmentation characterized by highly heterogeneous, intermixed cellular populations and total loss of organized tissue*

structure. Each point represents an individual cell colored by its DBSCAN-assigned spatial cluster, revealing how tissue architecture progressively transitions from organized macro-structures in healthy and pre-invasive states to chaotic, fragmented microenvironments in invasive adenocarcinoma.

Histological Feature	*DBSCAN Clustering Pattern*	*Interpretation*
Continuous and uniform architecture (Normal)	*Few large, well-defined clusters; minimal noise*	*Indicates dense and evenly distributed cells with **intact tissue continuity**—the hallmark of healthy lung parenchyma.*
Increased cellular density (AAH)	*Emergence of distinct, defined spatial domains; a few medium clusters*	*Reflects the initial increase in cell density and organization into early, but not yet dominant, lesions.*
Single dominant macro-structure (AIS)	***One major, highly cohesive cluster** with few small clusters or noise*	*Captures the **maximally cohesive** pre-invasive state where cells are spatially localized and form a single, overarching structure.*
Onset of structural fragmentation (MIA)	***Multiple small-to-medium clusters; loss of a single macro-structure***	*Marks the **transition toward micro-invasion** and suggests the early disruption of spatial continuity and the breakdown of the single large architecture.*
Highly heterogeneous and fragmented architecture (IAC)	*Numerous small, intermixed clusters; **high noise proportion***	*Captures extensive **spatial heterogeneity,** characteristic of aggressive malignant tumor organization with invasive epithelial foci and reactive stroma.*

*Table 1 **Relationship Between Histological Features and DBSCAN Clustering Patterns** This table summarizes how distinct histological architectures across lung tissue stages—ranging from normal to invasive adenocarcinoma—correlate with DBSCAN-derived spatial clustering patterns. Each row links a tissue's structural phenotype to its clustering behavior and interprets the biological significance, illustrating how spatial continuity, fragmentation, and heterogeneity evolve during disease progression.*

Immune landscape remodeling across disease progression

Beyond architectural changes, we investigated how cellular composition evolves across stages. To quantify the **rate of change and cellular flux** characterizing the tumor microenvironment (TME), we applied a timed Petri net framework to model population dynamics across the five histopathological stages of lung adenocarcinoma. This systems-level analysis unveiled coordinated, stage-specific remodeling of immune, stromal, and vascular compartments.

Figure 5 highlights broad compositional shifts: Normal tissue is dominated by epithelial cells and fibroblasts, which give way to progressive myeloid infiltration in AAH, and ultimately to diverse immune populations in advanced stages**.**

Trajectory analysis (**Figure 6**) further resolved these dynamics into distinct transition patterns:

- **Adaptive immunity:** CD8□ T cells underwent marked depletion during the AAH→AIS transition (-48.4%) before recovering in later stages. In contrast, **Regulatory T cells (Tregs)** expanded steadily through mid-stage transitions (up to +200%), establishing progressive immunosuppression. B cells infiltrated in two distinct waves (Normal→AAH and MIA→IAC), while CD4□ T cells displayed variable dynamics—initial expansion, transient depletion at AAH→AIS, and subsequent rebound.
- **Myeloid compartment:** CD163□ macrophages peaked early before contracting, whereas CD68□ myeloid cells expanded dramatically at MIA→IAC (+193.5%). Neutrophils increased initially but declined during AAH→AIS, while NK cells accumulated consistently, peaking at MIA→IAC.
- **Structural populations:** Fibroblasts enriched progressively across all transitions, reflecting sustained stromal remodeling. Epithelial cells showed biphasic expansion, consistent with early hyperplasia and late invasive growth. Endothelial cells declined sharply in early stages but underwent dramatic angiogenic activation beginning at AIS→MIA, marking the onset of vascularized disease.

Together, these transition-specific patterns reveal **critical inflection points** in TME evolution. The **immune-depleted, vascular-collapsing AAH→AIS transition** and the **angiogenic, Treg-rich AIS→MIA transition** represent synchronized microenvironmental shifts that establish a tumor-permissive state. Recognizing these stage-dependent vulnerabilities highlights potential **therapeutic windows for intervention**, particularly strategies aimed at restoring immune surveillance or disrupting stromal and vascular remodeling.

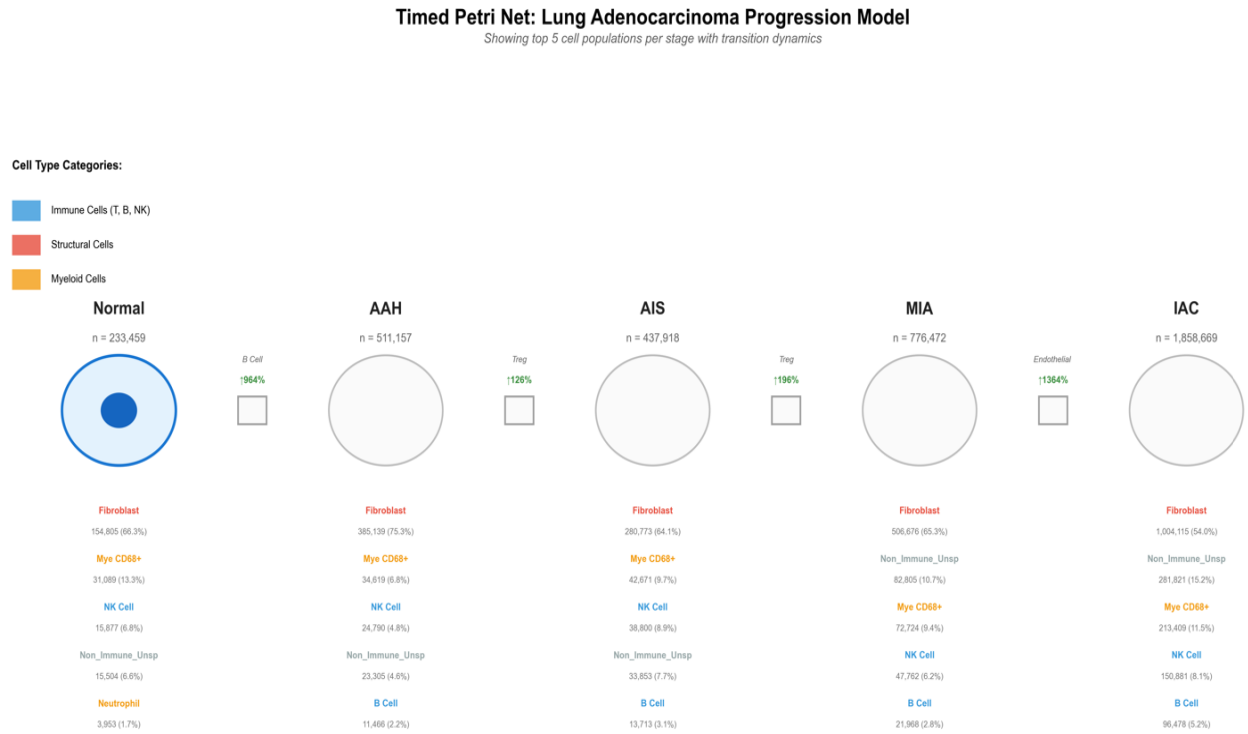

Figure 5. Timed Petri Net model of cell population dynamics across lung adenocarcinoma progression stages. *The model depicts temporal evolution of cellular compositions through five histopathological states: Normal → AAH (Atypical*

Adenomatous Hyperplasia) → AIS (Adenocarcinoma In Situ) → MIA (Minimally Invasive Adenocarcinoma) → IAC (Invasive Adenocarcinoma). Each circle represents a disease state with associated cell counts (n) and the top five cell populations displayed below. Cell types are color-coded by functional category: blue (Immune Cells: T, B, NK), red (Structural Cells: epithelial, fibroblasts, endothelial), and orange (Myeloid Cells: macrophages, neutrophils, dendritic cells). Transition boxes (t1-t4) between states indicate progression events, with annotations showing the most significantly changing cell type and its percentage change (↑ increase, ↓ decrease). Notable trends include progressive fibroblast enrichment, declining epithelial cell proportions, dynamic myeloid cell fluctuations, and emerging immune cell populations. The model integrates experimental single-cell data to quantify stage-specific microenvironment remodeling during adenocarcinoma development.

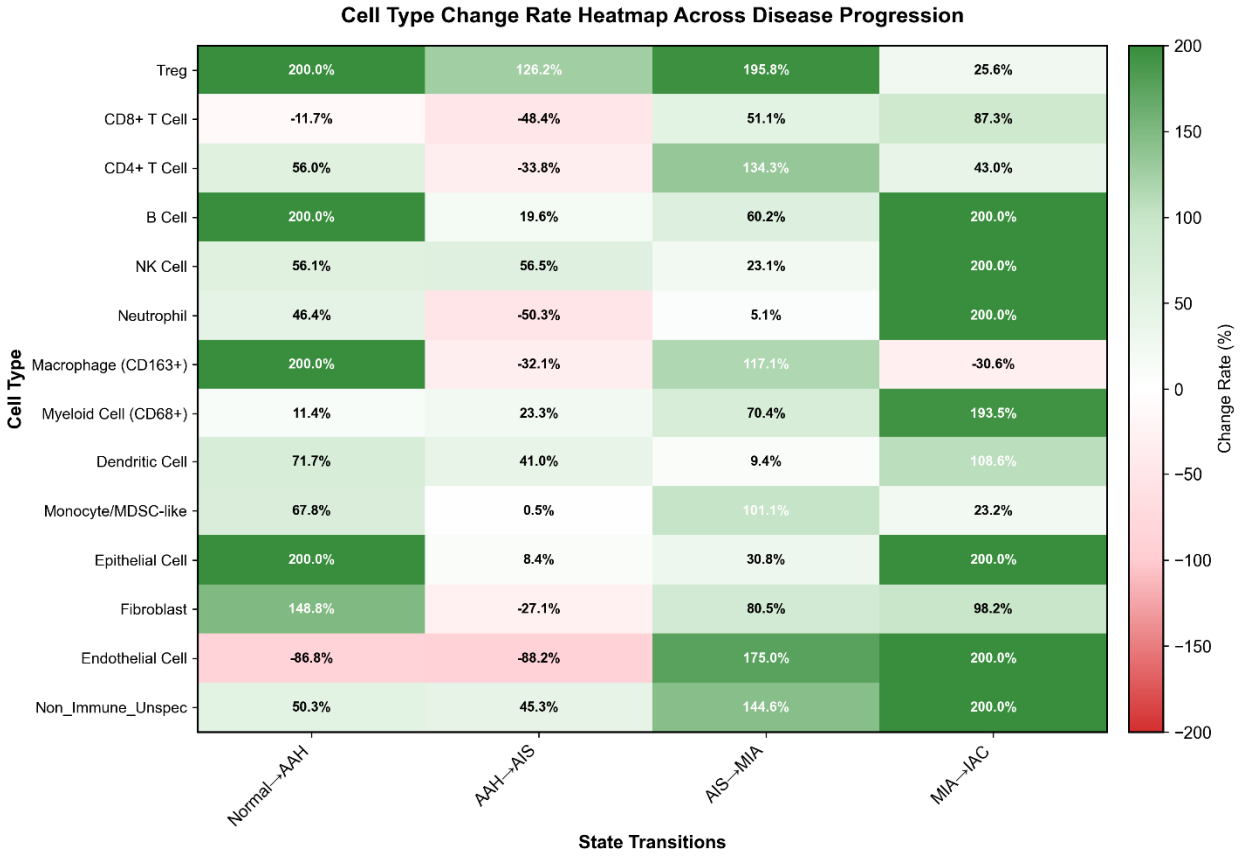

Figure 3 Cell type abundance changes across LUAD progression. *Heatmap showing percentage change in 14 cell populations across four disease transitions: Normal→AAH, AAH→AIS, AIS→MIA, and MIA→IAC. Color intensity indicates magnitude and direction (green: increase; red: decrease). Values capped at ±200%. Key findings: (1) Tregs show sustained expansion (200% Normal→AAH, 195.8% AIS→MIA); (2) B cells demonstrate biphasic infiltration (200% at Normal→AAH and MIA→IAC); (3) Endothelial cells decline early (-86.8% to -88.2%) then rebound dramatically (175% AIS→MIA, 200% MIA→IAC); (4) CD8+ T cells show depletion at AAH→AIS (-48.4%) followed by recovery.*

Dynamic Pathway Reprogramming Across Lung Adenocarcinoma Evolution

To comprehensively characterize molecular remodeling within the tumor microenvironment during lung adenocarcinoma progression, we performed pathway enrichment analysis across five histological stages: Normal, Atypical Adenomatous Hyperplasia (AAH), Adenocarcinoma In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Adenocarcinoma (IAC).

Figures 7 and **8** summarize pathway activity at both the global tissue level and within discrete cellular niches. **Figure 7** presents aggregate pathway activity across disease stages, revealing early upregulation of cell cycle and T cell exhaustion programs, followed by progressive attenuation of immune checkpoint and stromal signaling. **Figure 8** complements this analysis by mapping niche-specific pathway activity, uncovering distinct patterns of immune, stromal, and proliferative engagement across individual cell types. Together, these data delineate a coordinated, stage-dependent reprogramming of signaling networks that underlies tumor evolution.

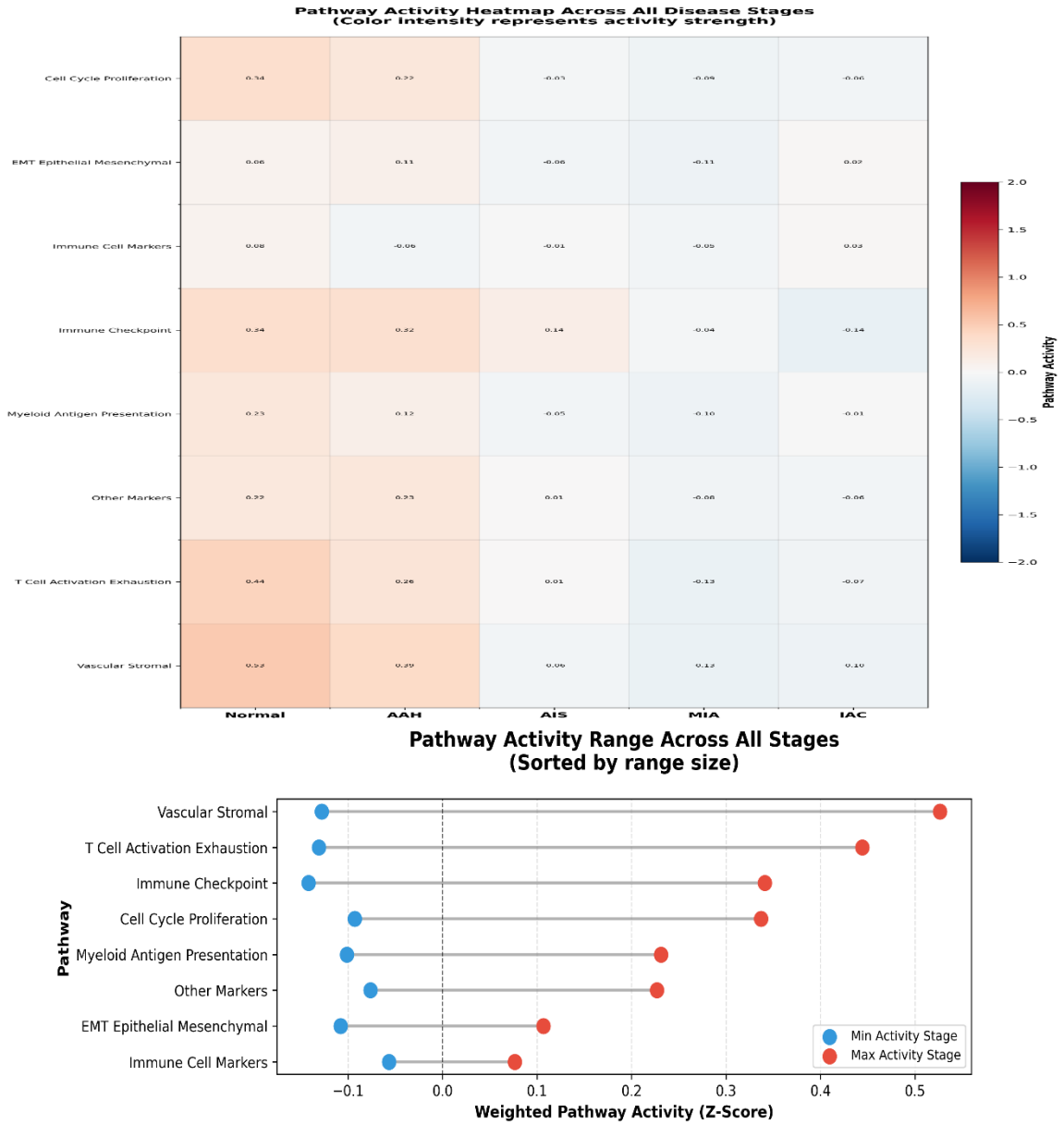

Figure 7. Global Pathway Activity Dynamics Across Lung Adenocarcinoma Progression*. Top: Heatmap showing normalized pathway activity (Z-score) across five histological stages—Normal, Atypical Adenomatous Hyperplasia (AAH), Adenocarcinoma*

In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Adenocarcinoma (IAC). Pathways include proliferative, immune, stromal, and myeloid programs. Color intensity reflects activity strength, with red indicating higher and blue indicating lower activity. Bottom: Dot plot illustrating the range of pathway activity across stages. Each pathway is represented by its minimum (blue) and maximum (red) Z-score values, connected by a line. Pathways are sorted by range size, highlighting those with the most dynamic changes during disease progression.

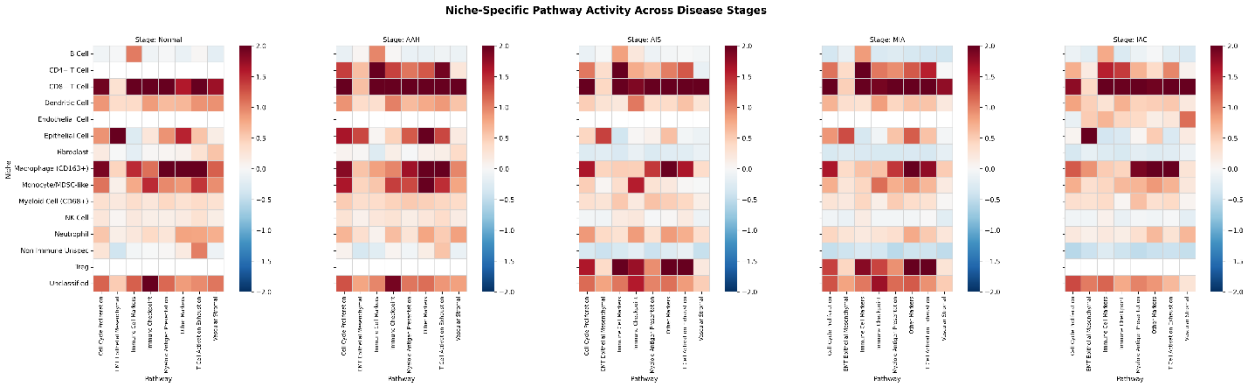

Figure 8. Niche-Specific Pathway Activity Across Lung Adenocarcinoma Disease Stages. *Heatmaps depicting pathway activity (Z-score) across 15 distinct cell niches and five histological stages: Normal, Atypical Adenomatous Hyperplasia (AAH), Adenocarcinoma In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Adenocarcinoma (IAC). Each panel represents a disease stage, with rows corresponding to cell types and columns to biological pathways, including proliferative, immune, stromal, and myeloid programs. Color intensity reflects relative pathway activity, with red indicating higher and blue indicating lower activity. This visualization highlights niche-specific signaling dynamics and reveals stage-dependent remodeling of the tumor microenvironment.*

Discussion

A computational pathology framework for spatial immune profiling

This study demonstrates that ROSIE transforms routine H&E-stained slides into high-resolution spatial atlases of the tumor microenvironment, bypassing the cost, time, and scalability limitations of traditional multiplex immunofluorescence. By analyzing 3,817,675 cells across 50 ROIs spanning the LUAD continuum (Figure 3), we generated comprehensive spatial and molecular data that would be prohibitively expensive using antibody-based approaches.

Critically, immune lineage resolution improved substantially with biomarker panel depth: expanding from 25 to 50 markers enabled detection of rare cell types (eosinophils, mast cells) that were otherwise obscured (Figure 2). This finding underscores the power of computational approaches to unlock molecular information latent within archival histopathology slides, opening avenues for large-scale retrospective studies of disease progression.

The architectural collapse of precancerous tissue: a quantitative measure of malignancy

A defining feature of cancer progression is the physical breakdown of organized tissue architecture. Our DBSCAN-based spatial clustering analysis provided a quantitative framework for measuring this phenomenon (**Figure 4, Table 1**). Normal lung tissue exhibited 18 clusters

with a dominant macro-cluster containing approximately 87% of cells. AAH samples showed 41 clusters with increased fragmentation (approximately 78% in the largest cluster). AIS maintained relative spatial continuity (approximately 65% in one macro-cluster). MIA marked the onset of micro-fragmentation with 155 clusters, though one cluster still comprised approximately 94.5% of cells. IAC displayed extensive heterogeneity: 2,285 clusters with 87% of cells in dense micro-niches and 13% classified as noise. This progressive breakdown of spatial continuity quantitatively captures the architectural collapse accompanying malignant invasion.

The architectural collapse parallels findings in other solid tumors where spatial disorganization correlates with poor prognosis and treatment resistance. DBSCAN clustering parameters proved critical for capturing stage-specific patterns: whereas large epsilon values (ε=0.0055) in MIA detected early fragmentation, small epsilon values (ε=0.003) in IAC resolved fine-scale heterogeneity. This methodological flexibility enables quantitative assessment of tumor architecture across disease stages, potentially serving as a computational biomarker for invasion risk stratification.

Immune evolution: from surveillance to suppression

Our spatially resolved analysis revealed clear, stage-dependent remodeling of the immune microenvironment, characterized by a fundamental shift from adaptive immunity toward innate and ultimately immunosuppressive states (**Figures 5 and 6**). This transition represents a central narrative in LUAD precancer evolution with profound implications for interception strategies.

Adaptive immune exhaustion. CD8+ T cells underwent marked depletion during the AAH→AIS transition (-48.4%, Figure 6) before recovering in later stages. This sharp decline suggests active immune evasion or exhaustion mechanisms emerging early in premalignancy. CD4+ T cells displayed variable dynamics—initial expansion, transient depletion at AAH→AIS, and subsequent rebound (Figure 6). Regulatory T cells (Tregs) expanded steadily through mid-stage transitions, with increases of 200% at Normal→AAH and 195.8% at AIS→MIA (Figure 6), establishing progressive immunosuppression.

Innate immune expansion and plasticity. Concurrent with adaptive immune decline, NK cells accumulated consistently, peaking at MIA→IAC (**Figure 6**). NK cells provide rapid, MHC-independent tumor surveillance and may initially compensate for T cell dysfunction. However, their sustained elevation into invasive stages suggests incomplete tumor control, possibly reflecting NK cells exhaustion or tumor immune editing. B cells infiltrated in two distinct waves, with dramatic expansion at Normal→AAH and MIA→IAC transitions (200% increases at both transitions, **Figure 6**). This biphasic pattern may indicate formation of tertiary lymphoid structures (TLS), which can paradoxically promote either anti-tumor immunity or tumor-supporting inflammation depending on context.

Myeloid reprogramming toward immunosuppression. Myeloid compartments underwent significant stage-specific remodeling (**Figure 6**). CD163+ macrophages peaked early before contracting, whereas CD68+ myeloid cells expanded dramatically at MIA→IAC (+193.5%). M2-polarized macrophages produce immunosuppressive cytokines (IL-10, TGF-β) and promote angiogenesis, tissue remodeling, and tumor progression. Neutrophils increased initially but

declined during AAH→AIS. This myeloid reprogramming likely creates a permissive niche for malignant transformation, highlighting potential therapeutic targets for myeloid checkpoint blockade or CSF-1R inhibition.

Stromal activation and fibroblast expansion. Fibroblasts enriched progressively across all transitions, reflecting sustained stromal remodeling (**Figure 6**). Cancer-associated fibroblasts (CAFs) secrete extracellular matrix proteins, growth factors, and chemokines that remodel tissue architecture and recruit immunosuppressive cells. The coordinated rise in fibroblasts and immunosuppressive myeloid cells suggests reciprocal crosstalk driving TME evolution.

Angiogenic activation. Endothelial cells showed consistent decline in early stages (-86.8% to -88.2%) followed by dramatic rebound (175.0% at AIS→MIA, 200% at MIA→IAC, **Figure 6**), indicating angiogenic activation in invasive phases and marking the onset of vascularized disease.

Biological insights into immune-epithelial dynamics

Epithelial cells exhibited biphasic expansion with 200% surge at Normal→AAH, followed by stability, then another 200% increase at MIA→IAC (Figure 6), reflecting epithelial compartment expansion during malignant transformation. This dynamic pattern suggests coordinated immune-epithelial interactions at critical transition points. Several biological mechanisms may underlie this phenomenon:

Epithelial proliferation and immune recognition. Accelerated epithelial proliferation during malignant transformation likely generates replication stress, DNA damage, and neoantigen expression. This antigenic burden may trigger B cells recruitment and antibody responses. However, if these responses are ineffective or immunosuppressive (e.g., regulatory B cells), they may inadvertently promote tumor progression.

Inflammation-driven selection. Chronic B cells infiltration and antibody production can create selective pressure favoring immune-evasive clones. Tumors may exploit B cell-derived cytokines (IL-10, TGF-β) to suppress T cell responses while evading antibody-mediated killing through antigen loss or Fc receptor modulation.

Future studies should characterize B cells clonality, antibody specificities, and functional states (pro- vs. anti-tumor) to determine whether B cells infiltration represents a therapeutic opportunity or obstacle.

Limitations and future directions

While our study provides unprecedented spatial and molecular resolution of LUAD immune evolution, several limitations warrant consideration:

Computational prediction validation. ROSIE predictions are derived from deep learning models trained on CODEX multiplex immunofluorescence data. Although extensive validation demonstrated high accuracy for major cell types, rare populations and subtle phenotypic states

may be less reliably predicted. Orthogonal validation using targeted antibody staining or spatial transcriptomics in representative samples would strengthen confidence in cell type assignments.

Static snapshots vs. dynamic processes. Our analysis relies on cross-sectional sampling of different disease stages rather than longitudinal tracking of individual lesions. While this approach reveals stage-specific patterns, it cannot capture the temporal dynamics, clonal selection events, or stochastic transitions that occur within individual patients. Longitudinal studies following patients with screen-detected precursor lesions would provide complementary insights.

Clinical translation pathways. Implementing ROSIE-guided interception strategies in clinical practice requires prospective validation. Key questions include: Which imaging or biomarker criteria should trigger ROSIE profiling? What spatial features or cell type ratios optimally predict interception benefit? How should patients be monitored following intervention? Designing clinical trials addressing these questions is a priority.

Broader implications for computational pathology and precision oncology

This work exemplifies a paradigm shift in how we extract biological insights from routine clinical specimens. By applying AI-powered computational pathology at scale, we transformed ubiquitous H&E slides—often relegated to diagnostic archives—into quantitative, spatially resolved datasets rivaling expensive molecular assays. This democratizes access to high-dimensional tissue profiling, enabling retrospective studies of archival cohorts, real-time clinical decision support, and large-scale biomarker discovery.

Beyond LUAD, ROSIE and similar approaches hold promise for other malignancies where immune evolution drives progression: colorectal adenomas, Barrett's esophagus, cervical intraepithelial neoplasia, and pancreatic precursors. Each of these represents an interception opportunity where spatially informed immune profiling could guide preventive strategies.

Moreover, the stage-specific vulnerabilities we identified—particularly the immune-depleted, vascular-collapsing AAH→AIS transition and the angiogenic, Treg-rich AIS→MIA transition—challenge the "one-size-fits-all" checkpoint inhibitor paradigm. Our findings suggest that therapeutic efficacy windows may be narrow and context-dependent, necessitating precision matching of interventions to disease stage, immune state, and spatial architecture. ROSIE-enabled profiling could facilitate this precision approach, enabling clinicians to stratify patients and select optimal intervention timing.

Toward a new era of cancer interception

The ultimate goal of oncology is not merely to treat cancer, but to prevent it. Our study demonstrates that this vision is becoming achievable through convergence of AI-powered pathology, spatial biology, and cross-species validation. By defining the immune evolutionary trajectory from normal tissue to invasive carcinoma and identifying critical inflection points in TME evolution, this work establishes a generalizable framework: (1) use computational pathology to map spatial immune dynamics across disease stages; (2) identify stage-specific

vulnerabilities through integrative analysis; (3) validate therapeutic efficacy in preclinical models; and (4) translate findings into clinical interception trials. Applying this framework systematically across cancer types could transform early detection programs from surveillance-only strategies into active intervention platforms, ultimately shifting the balance from cancer treatment to cancer prevention.

Method

Human Tissue Samples and Cohort Design

We analyzed H&E-stained formalin-fixed paraffin-embedded (FFPE) tissue sections from 114 human lung adenocarcinoma specimens, spanning five histopathological stages: normal lung tissue, atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC). All samples were obtained with institutional review board approval and informed consent. Pathological staging was performed by board-certified thoracic pathologists according to WHO 2015 criteria. Whole-slide images (WSI) were acquired at 20× magnification using a digital slide scanner.

ROSIE Framework and Virtual Multiplex Immunofluorescence Prediction

We applied the ROSIE (Robust in Silico Immunofluorescence from H&E images) model[1], a deep convolutional neural network trained on paired H&E and CODEX multiplex immunofluorescence data, to predict spatially resolved expression of 50 protein biomarkers directly from H&E images. For each WSI, 50 regions of interest (ROIs) were selected to capture diverse architectural features across the tissue, yielding a total of 322,277 single-cell profiles for human LUAD analysis (**Figure 1**). The analysis workflow proceeded as follows: QuPath was first used to perform precise annotation and segmentation of individual cells within the H&E ROIs based on nuclear morphology. ROSIE inference was then applied to generate pixel-level predictions for each biomarker in these segmented cells. These predictions were further refined using a watershed-based algorithm informed by nuclear and cytoplasmic channels to delineate single-cell compartments and assign virtual multiplex immunofluorescence (v-mIF) classifications.

Biomarker Panel Design

Our 50-marker panel was designed to comprehensively profile immune, stromal, and epithelial compartments based on established lineage specificity and functional relevance (**Figure 1**). The initial 10-biomarker panel included PDL1, CD8, CD163, PanCK, EpCAM, Ki67, CD4, FoxP3, CD31, and Vimentin to capture major immune and epithelial compartments. The full 50-marker panel added 40 additional markers including: **immune checkpoint molecules** (PD1, LAG3, TIGIT, VISTA, ICOS), **myeloid markers** (CD68, CD14, CD11c, HLA-DR, Gal3, MPO), **B cells markers** (CD20, CD21, CD38, CD79a, BCL2), **T cell subsets** (CD3e, CD45RO, GranzymeB), **stromal markers** (αSMA, Podoplanin, CD34, CollagenIV), **functional markers** (HLA-A, HLA-E, IFNγ, IDO1, CD40), **proliferation markers** (PCNA, ATM), **epithelial**

markers (ECad, TP63, Keratin8/18, CD66, CD44), and **nuclear stain** (DAPI). To assess the impact of panel size on resolution, we analyzed marker expression across the dataset using both 25-marker and 50-marker panels (**Figure 2**).

Cell Phenotyping and Classification

Using the 50-marker panel, we classified 3,817,675 total cells into 14 distinct cell types based on canonical lineage markers (**Figure 3, left; Table 2**). We assigned cell type identities using a custom fuzzy scoring algorithm guided by a manually curated rule set (**Figure 9**). This rule set was developed by synthesizing canonical marker expression patterns from established immunophenotyping literature[22,23]. For each cell type—including T cell subsets (CD4□, CD8□, Tregs), B cells, NK cells, myeloid populations (macrophages, dendritic cells, monocytes, MDSCs), and structural cells (epithelial, endothelial, fibroblasts) — we defined required positive markers and exclusionary negative markers based on authoritative lineage definitions[24–26]. The algorithm applies these rules through a flexible scoring system that accommodates biological and technical variability in multiplexed imaging signals, ensuring robust and interpretable classifications. Cell type assignment proceeds as follows:

1. **Candidate Screening**: For each cell type, we calculate the proportion of positive markers expressed (posRatio). Cell types with posRatio below a minimum threshold are excluded.
2. **Fuzzy Scoring**: For qualifying candidates, we compute a score: Score = posRatio − (0.2 × number of negative markers) This balances specificity with tolerance for incomplete marker expression.
3. **Edge Case Handling**: If no positive markers are detected, the score is set to 1.0 if no negative markers are present, or −1.0 if any negative marker is detected.
4. **Label Assignment**: Each cell is assigned the identity with the highest score exceeding a final acceptance threshold. Cells below this threshold are labeled as "Unclassified_By_New_Rules."

Spatial Clustering and Microenvironment Analysis

We applied DBSCAN clustering to cell centroid coordinates to identify spatially organized neighborhoods. Parameters were optimized per tissue stage:

- Normal lung: epsilon = 50 pixels, min_samples = 50 (to capture loosely distributed spatial domains)
- AAH: epsilon = 50 pixels, min_samples = 50 (to detect emerging defined spatial domains)
- AIS: SmartDensity mode with density threshold = 0.005, min_samples = 100 (to identify the single dominant macro-structure)
- MIA: epsilon = 215 pixels, min_samples = 105 (to capture early structural fragmentation)
- IAC: epsilon = 15 pixels, min_samples = 105 (to resolve fine-scale heterogeneity and extensive fragmentation)

SmartDensity mode adaptively selects epsilon based on local cell density, enabling detection of both compact and dispersed cellular organizations. Cells not assigned to clusters were classified as spatial "noise" and quantified as a percentage of total cells. For each identified cluster (6,522

total microenvironmental niches across all samples), we calculated cell type proportions, spatial metrics (area, perimeter, convexity), and neighborhood enrichment scores. Stage-specific clustering results are visualized in Supplementary **Figures S1-S5**, demonstrating cell-cell interactions within identified spatial niches.

Dimensionality Reduction and Visualization

To visualize cellular heterogeneity and assess the impact of biomarker panel size on cell type resolution, we applied UMAP to the 3,817,675-cell dataset (**Figure 2**). Analysis was performed using the following parameters: n_neighbors = 30, min_dist = 0.1, n_components = 2, metric = 'euclidean'. Marker expression values were log-transformed and standardized (Z-score normalization) prior to UMAP. We generated separate UMAP projections using 25-marker and 50-marker panels to demonstrate how comprehensive biomarker selection enables identification of rare populations (eosinophils, mast cells) that were absent in the reduced panel due to missing lineage-specific markers (e.g., GATA3, MPO, Galectin-3).

Temporal Modeling with Timed Petri Nets

To capture the temporal evolution of cell population dynamics across disease progression, we constructed a timed Petri net model (**Figure 5**). Each histopathological stage (Normal, AAH, AIS, MIA, IAC) was modeled as a discrete state, with cell type abundances represented as tokens. Transitions between stages (t1-t4) were governed by empirically derived rate constants proportional to observed fold-changes in cell populations. For each transition, we identified the most significantly changing cell type and calculated percentage changes to reveal stage-specific remodeling patterns (**Figure 6**). This systems-level framework enabled quantitative analysis of coordinated immune, stromal, and vascular compartment changes throughout LUAD evolution.

Pathway Enrichment Analysis

To characterize molecular remodeling within the tumor microenvironment during progression, we performed pathway enrichment analysis across all five histological stages. We analyzed both global tissue-level pathway activity (**Figure 7**) and niche-specific pathway activity within discrete cellular neighborhoods (**Figure 8**). Pathway scores were calculated using gene set enrichment methods and normalized (Z-score) to enable cross-stage comparisons. Analyzed pathways included proliferative, immune checkpoint, stromal remodeling, myeloid, and vascular programs. For global pathway analysis, we computed minimum and maximum Z-score values across stages to identify pathways with the most dynamic changes during disease progression. For niche-specific analysis, pathway activity was calculated separately for each of the 15 cell type categories to reveal cell type-specific signaling dynamics.

Statistical Analysis

Cell type abundance comparisons between stages were performed using Kruskal-Wallis tests followed by Dunn's post-hoc tests with Benjamini-Hochberg correction. Spatial enrichment analyses used permutation tests (10,000 iterations). Percentage change rate for cell populations across transitions (Figure 6) was calculated as:

$$\text{Change} = [(\text{Count_stage2} - \text{Count_stage1}) / \text{Count_stage1}] \times 100\%$$

Values were capped at ±200% for visualization clarity in heatmap representations. All analyses were performed in Python 3.9 using scipy, statsmodels, and scikit-learn. Visualizations were generated with matplotlib, seaborn, and plotly. P-values < 0.05 were considered statistically significant unless otherwise noted.

Data and Code Availability

The lung adenocarcinoma H&E whole slide images and ROSIE-predicted virtual multiplex immunofluorescence data for 816 ROIs (3,817,675 cells) analyzed in this study are available from the corresponding author upon reasonable request. The ROSIE deep-learning framework, including model architecture, training methodology, and inference code, is publicly available at https://gitlab.com/enable-medicine-public/rosie. Original CODEX training data used to develop ROSIE are subject to licensing restrictions and patient privacy protections but may be available through Enable Medicine with appropriate data use agreements. Processed datasets including cell phenotype annotations (14 cell types), marker expression profiles (50-marker panel), DBSCAN spatial clustering assignments (6,522 clusters), pathway enrichment scores, and all quantitative data underlying **Figures 1-9** and Supplementary **Figures S1-S5** are provided as Supplementary Data files. Custom analysis scripts for spatial clustering, temporal modeling, and pathway analysis are available from the corresponding author upon reasonable request.

Conclusion

In this study, we establish that ROSIE—a **high-economy, robust, and scalable deep-learning framework**—can transform routine H&E-stained slides into high-resolution spatial maps of the tumor microenvironment. By applying this approach across the LUAD precancer-to-cancer continuum, we systematically charted the progressive breakdown of tissue architecture and a fundamental immune shift from adaptive surveillance to an innate-dominant, immunosuppressive state. Our findings validate a new computational pathology paradigm that leverages ubiquitous H&E images to enable spatially resolved biomarker discovery, overcoming the cost and scalability limitations of traditional single-cell and multiplexed imaging methods. This approach opens avenues for large-cohort retrospective studies and provides a scalable strategy for uncovering biological mechanisms and interception targets, moving the field closer to cancer prevention, not just treatment.

Cell Classification Rules (CLASSIFICATION RULES)

Cell Type	Positive Markers	Negative Markers
Treg	CD45, CD3e, CD4, FoxP3	CD8, CD20, CD68, PanCK
CD8+ T Cell	CD45, CD3e, CD8	CD4, FoxP3, CD20, CD68, PanCK
CD4+ T Cell	CD45, CD3e, CD4	CD8, FoxP3, CD20, CD68, PanCK
B Cell	CD45, CD20	CD3e, CD68, CD163, CD66, CD11c, PanCK
NK Cell	CD45	CD3e, CD20, CD68, CD163, CD66, CD11c, PanCK, CD31
Neutrophil	CD45, CD66	CD3e, CD20, CD68, CD163, CD11c, PanCK
Macrophage (CD163+)	CD45, CD163	CD3e, CD20, CD66, CD11c, PanCK
Myeloid Cell (CD68+)	CD45, CD68	CD3e, CD20, CD66, CD163, CD11c, PanCK
Dendritic Cell	CD45, CD11c	CD3e, CD20, CD66, CD68, CD163, PanCK
Monocyte/MDSC-like	CD45, CD68	CD3e, CD20, CD31, PanCK, Vimentin
Epithelial Cell	PanCK	CD45, Vimentin
Fibroblast	Vimentin	CD45, PanCK
Endothelial Cell	CD31	CD45, PanCK, Vimentin
Non-Immune Unspec	—	CD45, PanCK, Vimentin, CD31
Background/Junk	—	CD45, PanCK, Vimentin, CD31

Table 2. Marker-based cell classification rules used for fuzzy labeling. *This table defines the positive and negative marker sets used to assign primary cell type labels via the fuzzy classification algorithm. Each cell type is characterized by canonical marker combinations, with positive markers required for identity confirmation and negative markers used to exclude conflicting phenotypes. The rule set spans immune subsets (e.g., Tregs, $CD4^+/CD8^+$ T cells, B cells, NK cells), myeloid populations (e.g., macrophages, dendritic cells, monocytes), and structural cells (e.g., epithelial, fibroblast, endothelial), ensuring biologically interpretable and reproducible annotation across LUAD tissue samples.*

Pseudocode: Fuzzy Cell Type Classification

Algorithm 1 FuzzyClassifyPrimaryLabel(cell)

```
Input: cell with marker signals
Output: primary cell type label
signals ← cell["Signals"]
bestScore ← −∞
bestType ← "Unclassified_By_New_Rules"
for each cellType, rule in CLASSIFICATION_RULES do
    posMarkers ← rule["pos"]
    negMarkers ← rule["neg"]
    posCount ← number of posMarkers with signal > threshold
    negCount ← number of negMarkers with signal > threshold
    if posMarkers is not empty then
        posRatio ← posCount / length(posMarkers)
        if posRatio ≥ MIN_POS_PROPORTION_V9 then
            score ← posRatio − (negCount × 0.2)
        else
            score ← −∞
        end if
    else
        if all negMarkers have signal ≤ threshold then
            score ← 1.0
        else
            score ← −1.0
        end if
    end if
    if score > bestScore then
        bestScore ← score
        bestType ← cellType
    end if
end for
if bestScore ≥ MIN_CLASSIFICATION_SCORE_V9 then
    return bestType
else
    return "Unclassified_By_New_Rules"
end if
```

Figure 9. Pseudocode for fuzzy cell type classification algorithm. *This figure outlines the logic of the FuzzyClassifyPrimaryLabel algorithm used to assign primary cell type labels based on multiplexed marker signals. The algorithm iteratively evaluates each candidate cell type defined in the classification rule set, computing a fuzzy score based on the proportion of positive markers detected and penalizing the presence of negative markers. Cells are assigned the label with the highest score exceeding a minimum classification threshold; otherwise, they are marked as Unclassified_By_New_Rules. This approach enables robust and interpretable annotation of diverse immune, stromal, and epithelial populations in LUAD tissue samples.*

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).

2. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors. *J. Thorac. Oncol.* **10**, 1243–1260 (2015).

3. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).

4. Zhu, B. *et al.* Spatial and multiomics analysis of human and mouse lung adenocarcinoma precursors reveals TIM-3 as a putative target for precancer interception. *Cancer Cell* **43**, 1125-1140.e10 (2025).

5. Balasubramanian, A. & Asselin-Labat, M.-L. Immune evolution in pre-invasive lung adenocarcinoma. *Cancer Cell* **43**, 1004–1006 (2025).

6. Fu, T. *et al.* Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. *J. Hematol. Oncol.J Hematol Oncol* **14**, 98 (2021).

7. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

8. Kuswanto, W., Nolan, G. & Lu, G. Highly multiplexed spatial profiling with CODEX: bioinformatic analysis and application in human disease. *Semin. Immunopathol.* **45**, 145–157 (2023).

9. Baird, R. L. *et al.* PixlMap: A generalisable pixel classifier for cellular phenotyping in multiplex immunofluorescence images. *PLOS One* **20**, e0317865 (2025).

10. Wang, Z. *et al.* Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nat. Commun.* **12**, 6500 (2021).

11. Zhu, J. *et al.* Progress of single-cell RNA sequencing combined with spatial transcriptomics in tumour microenvironment and treatment of pancreatic cancer. *J. Transl. Med.* **22**, 563 (2024).

12. Ahmed, R. *et al.* Single-Cell RNA Sequencing with Spatial Transcriptomics of Cancer Tissues. *Int. J. Mol. Sci.* **23**, 3042 (2022).

13. Molla Desta, G. & Birhanu, A. G. Advancements in single-cell RNA sequencing and spatial transcriptomics: transforming biomedical research. *Acta Biochim. Pol.* **72**, 13922 (2025).

14. Kim, N. *et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).

15. Odate, T. *et al.* Diagnostic challenges of faded hematoxylin and eosin slides: limitations of re-staining and re-sectioning and possible reason to go digital. *Virchows Arch.* https://doi.org/10.1007/s00428-025-04209-z (2025) doi:10.1007/s00428-025-04209-z.

16. Dai, Y. *et al.* Abstract 759: Spatial transcriptomics identifies unique tumor and microenvironment pathomic programs that are associated with the lung premalignancy and adenocarcinoma continuum. *Cancer Res.* **85**, 759–759 (2025).

17. Dunn, C. *et al.* Quantitative assessment of H&E staining for pathology: development and clinical evaluation of a novel system. *Diagn. Pathol.* **19**, 42 (2024).

18. Ruiz, D. *et al.* Completing Spatial Transcriptomics Data for Gene Expression Prediction Benchmarking. Preprint at https://doi.org/10.48550/ARXIV.2505.02980 (2025).

19. Wu, Y. *et al.* DANet: spatial gene expression prediction from H&E histology images through dynamic alignment. *Brief. Bioinform.* **26**, bbaf422 (2025).

20. Ding, T. *et al.* A multimodal whole-slide foundation model for pathology. *Nat. Med.* **31**, 3749–3761 (2025).

21. Wu, E. *et al.* ROSIE: AI generation of multiplex immunofluorescence staining from histopathology images. *Nat. Commun.* **16**, 7633 (2025).

22. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

23. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

24. McKee, A. S., Munks, M. W. & Marrack, P. How Do Adjuvants Work? Important Considerations for New Generation Adjuvants. *Immunity* **27**, 687–690 (2007).

25. Müschen, M. Origins of the human B-cell lineage. *Blood* **134**, 998–999 (2019).

26. Abbas, A. K., Lichtman, A. H., Pillai, S. & Baker, D. L. *Cellular and Molecular Immunology*. (Elsevier, Philadelphia, Pennsylvania, 2022).

Acknowledgements (optional)

Keep acknowledgements brief and do not include thanks to anonymous referees or editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Ethics declarations

Competing interests

Submission of a competing interests statement is required for all content of the journal.

Supplementary Information

S1.

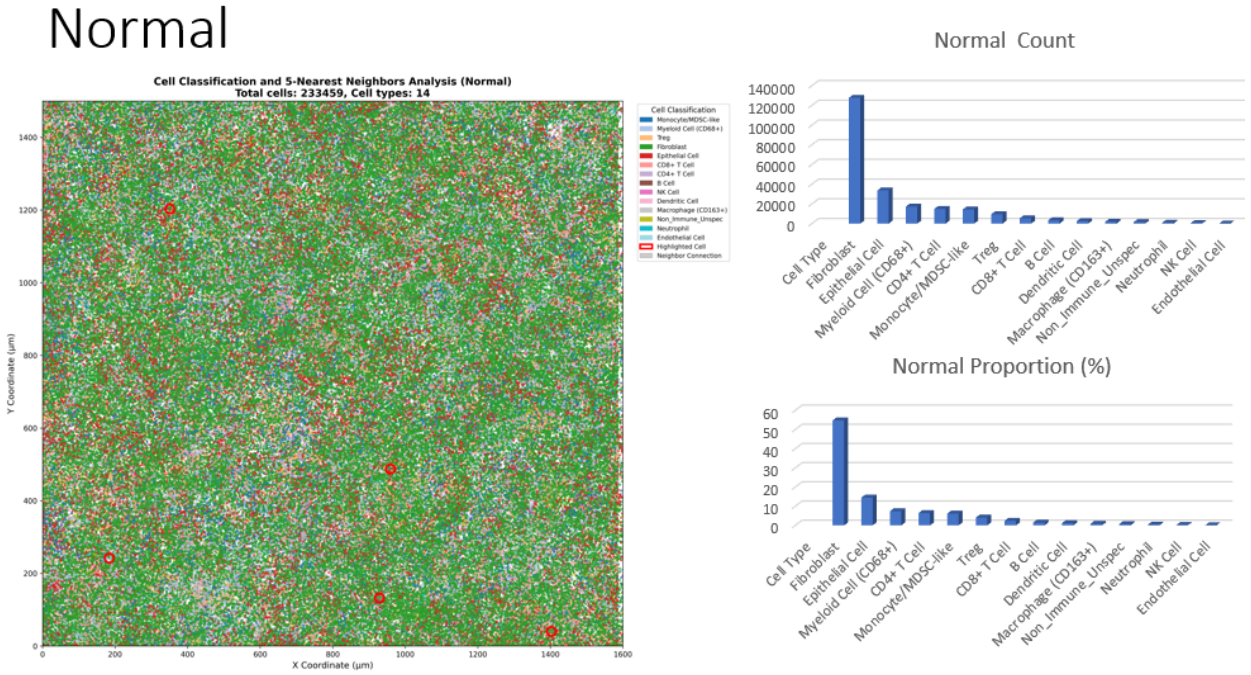

Figure 4 ***Spatial distribution and quantification of cell types in normal lung tissue.*** *Left: Spatial cell classification map displaying 233,459 individual cells across 14 cell types within normal lung tissue. Each point represents a single cell positioned according to its X-Y coordinates and colored by predicted cell type. The tissue shows relatively uniform spatial distribution of multiple cell populations throughout the analyzed region. Right panels: Quantitative analysis showing absolute cell counts (top) and relative proportions (bottom) for each cell type. Cell 1 type represents the predominant population with approximately 140,000 cells (~60% of total), followed by Epithelial Cell and Endothelial Cell populations. Additional cell types including Fibroblast, Myeloid Cell (CD68+), CD8+ T Cell, CD4+ T Cell, Treg, B Cells, Dendritic Cells, Monocyte/MDSC Cells, Macrophage (CD163+), Neutrophil, and NK Cell are present at progressively lower frequencies. The 5-Nearest Neighbors analysis reveals complex cellular interactions with neighbor compositions indicated in the spatial plot. This cellular architecture establishes the baseline tissue organization of normal lung parenchyma for comparative analysis with pre-malignant and malignant stages of lung adenocarcinoma (LUAD) progression.*

S2.

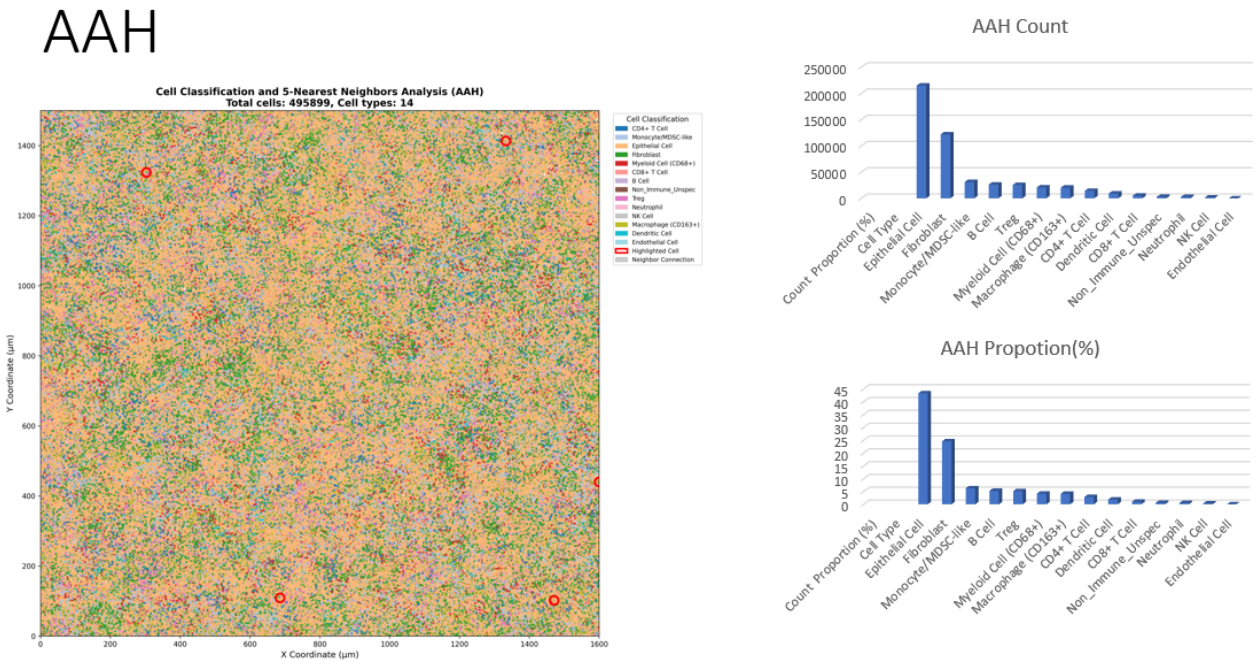

Figure 5 ***Spatial distribution and quantification of cell types in Atypical Adenomatous Hyperplasia (AAH) tissue.*** *Left: Spatial cell classification map displaying 495,899 individual cells across 14 cell types within AAH tissue, representing early pre-malignant lung lesions. Each point represents a single cell positioned according to its X-Y coordinates and colored by predicted cell type. Red circles indicate regions of neighbor composition analysis using the 5-Nearest Neighbors method. The spatial distribution reveals a more heterogeneous cellular architecture compared to normal tissue, with distinct clustering patterns emerging across different regions. Right panels: Quantitative analysis showing absolute cell counts (top) and relative proportions (bottom) for each cell type. Count Proportion (%) dominates with approximately 220,000 cells (~45% of total), followed by Cell 1 type (~25%) and Epithelial Cell populations. The cellular composition shows notable shifts from normal tissue, with increased representation of Fibroblast, Monocyte/MDSC-like cells, B Cells, and T Cell populations (CD4+ T Cells, CD8+ T Cells). Additional immune cell types including Myeloid Cell (CD68+), Macrophage (CD163+), Treg, Neutrophil, Dendritic Cells, Non-Immune/Tumor Cells, NK Cells, and Endothelial Cells are present at varying frequencies. This altered cellular landscape reflects the transitional nature of AAH as an early pre-malignant lesion in lung adenocarcinoma (LUAD) progression, characterized by increased cellular heterogeneity and emerging immune infiltration.*

S3.

AIS

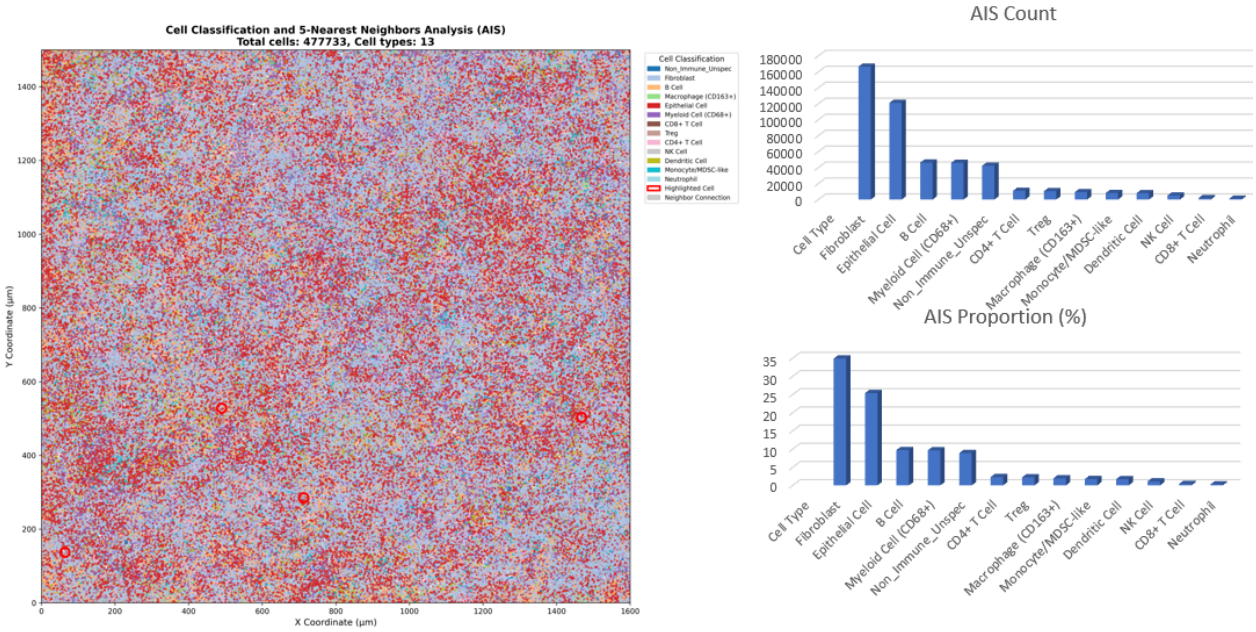

Figure 6 ***Spatial distribution and quantification of cell types in Adenocarcinoma In Situ (AIS) tissue.*** *Left: Spatial cell classification map displaying 477,733 individual cells across 13 cell types within AIS tissue, representing a more advanced pre-malignant stage in lung adenocarcinoma progression. Each point represents a single cell positioned according to its X-Y coordinates and colored by predicted cell type. Red circles indicate regions of 5-Nearest Neighbors analysis for cellular interaction assessment. The spatial architecture shows marked clustering with distinct red-dominant regions interspersed throughout a blue-purple background, indicating increased segregation of cell populations compared to earlier stages. Right panels: Quantitative analysis showing absolute cell counts (top) and relative proportions (bottom) for each cell type. Cell 1 type remains the predominant population with approximately 175,000 cells (~35% of total), followed by Fibroblast (~27%) and Epithelial Cell populations. Notably, the cellular composition reveals further shifts characteristic of pre-malignant progression, with substantial representation of B Cells, Myeloid Cells (CD68+), and Non-Immune/Tumor Cell populations. Additional cell types including CD4+ T Cell, Treg, Macrophage (CD163+), Monocyte/MDSC-like, Dendritic Cell, NK Cell, CD8+ T Cell, and Neutrophil are present at varying frequencies. The pronounced spatial segregation and altered proportions of stromal and immune cell populations reflect the evolving tumor microenvironment as lesions progress from AAH through AIS toward invasive adenocarcinoma.*

S4.

MIA

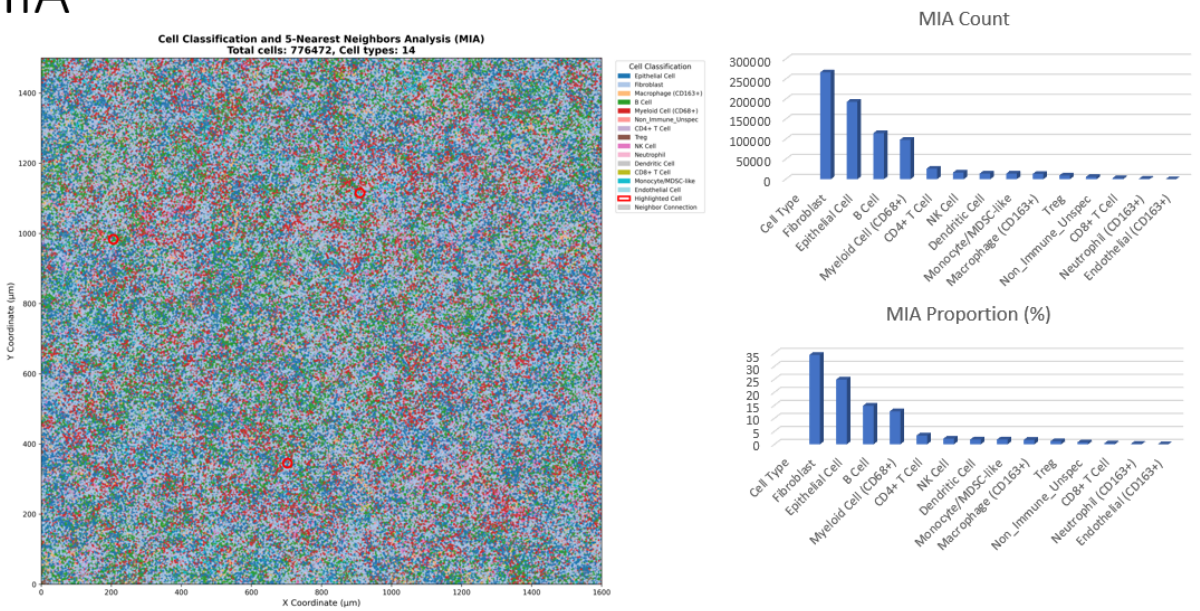

Figure 7 ***Spatial distribution and quantification of cell types in Minimally Invasive Adenocarcinoma (MIA) tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within MIA tissue. Each dot represents a single cell positioned according to its X-Y coordinates. Right: Bar chart displaying the absolute cell counts for each cell type (total: 59,027 cells). MIA tissue exhibits a diversified cellular landscape with B cells (21,982 cells, tan) emerging as the most abundant population, followed by fibroblasts (16,077 cells, blue) and myeloid cells (CD68+, 10,334 cells, pink). Regulatory T cells (Tregs, 6,215 cells, green) and monocytes/MDSCs (5,766 cells, brown) are notably increased compared to AIS. CD4+ T cells (red), epithelial cells (orange), neutrophils, macrophages (CD163+), dendritic cells, NK cells, CD8+ T cells, and non-immune unspecified cells are present at lower frequencies. This cellular composition reflects the MIA stage where early stromal invasion begins, characterized by substantial B cells infiltration, restoration of stromal components, and a more complex immune microenvironment compared to the CD4+ T cell-dominated AIS stage.*

S5.

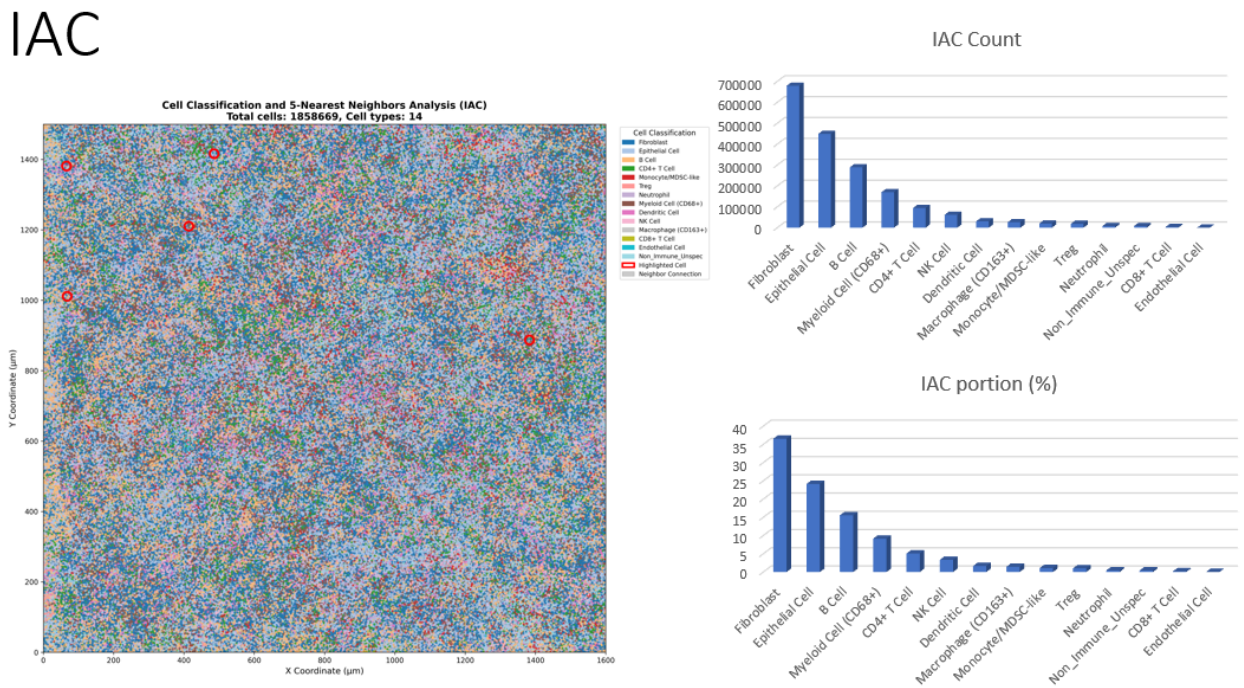

Figure 8 ***Spatial distribution and quantification of cell types in Invasive Adenocarcinoma (IAC) tissue.*** *Left: Spatial cell classification map displaying 1,858,669 individual cells across 14 cell types within IAC tissue, representing fully invasive lung adenocarcinoma. Each point represents a single cell positioned according to its X-Y coordinates and colored by predicted cell type. Red circles indicate regions of 5-Nearest Neighbors analysis for cellular interaction assessment. The spatial architecture demonstrates extensive cellular heterogeneity with complex mixing of multiple cell populations throughout the tissue, displaying a highly diverse mosaic of colors that reflects the chaotic cellular organization characteristic of invasive cancer. Right panels: Quantitative analysis showing absolute cell counts (top) and relative proportions (bottom) for each cell type. Fibroblast emerges as the dominant population with approximately 700,000 cells (~38% of total), representing a dramatic expansion of the stromal compartment compared to earlier stages. Epithelial Cell (~26%) and B Cells (~17%) constitute the next most abundant populations. The invasive stage shows robust immune cell infiltration with substantial representation of Myeloid Cell (CD68+), CD4+ T Cell, NK Cells, Dendritic Cells, Macrophage (CD163+), Monocyte/MDSC-like, Neutrophil, Treg, CD8+ T Cells, Non-Immune/Unique, and Endothelial Cell populations. This marked stromal expansion, combined with diverse immune infiltration and loss of organized tissue architecture, exemplifies the complex tumor microenvironment of fully invasive adenocarcinoma, representing the culmination of progressive cellular and architectural changes throughout lung adenocarcinoma development.*