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 cross-species cohort comprising

114 human LUAD samples and five mouse models 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]. 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[9–12]. 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[10]. The limitations of current spatial and single-cell technologies underscore the need for **a high-economy and robust computational method to enable scalable spatial biomarker analysis, compared with traditional resource-intensive single-cell approaches.** Moreover, most

studies focus on advanced-stage tumors, leaving the early evolutionary dynamics of LUAD precursors largely unexplored[4,13,14].

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[15,16]. 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[17,18].

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, it remains unclear how immune cell populations and their spatial organization change across the AAH-AIS-MIA-LUAD continuum, which immune features are enriched at specific stages, and whether these features represent actionable therapeutic targets for early intervention. 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[19] 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 wet-lab assays. By analyzing 114 human LUAD and precursor samples alongside five mouse models 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 50 regions of interest (ROIs) encompassing 322,277 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. Figure 1 demonstrates ROSIE's scalability: starting from the original H&E image (leftmost panel) of human LUAD tissue, ROSIE sequentially generated v-mIF predictions with increasing biomarker complexity—10-marker, 25-marker, and 50-marker panels (left to right)—without requiring additional staining. The progressive overlay images reveal that larger marker panels enable more comprehensive spatial profiling of immune and stromal components. Additional representative v-mIF images spanning diverse tissue regions and disease stages are provided in Supplementary Figures S1 and S2, illustrating normal lung architecture and invasive adenocarcinoma, respectively.

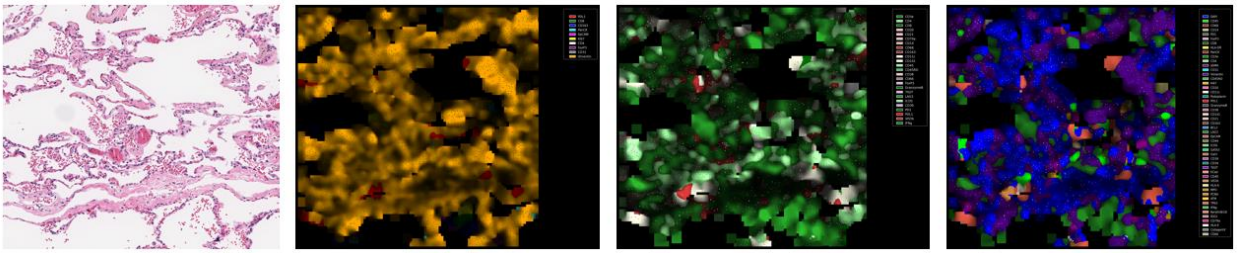


Figure 1 Multi-panel ROSIE predictions from a single H&E slide demonstrate scalable biomarker inference. Left: Original H&E whole slide image (sample R01, cite from [4]). Center panels: ROSIE-generated virtual multiplex immunofluorescence (v-mIF) using 10-biomarker (second), 25-biomarker (third), and 50-biomarker (fourth) panels. All v-mIF images are computationally predicted from the same H&E input without additional staining. Larger biomarker panels enable more comprehensive spatial profiling of immune and stromal components. 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 underscores the importance of comprehensive biomarker selection for accurate cell type classification.

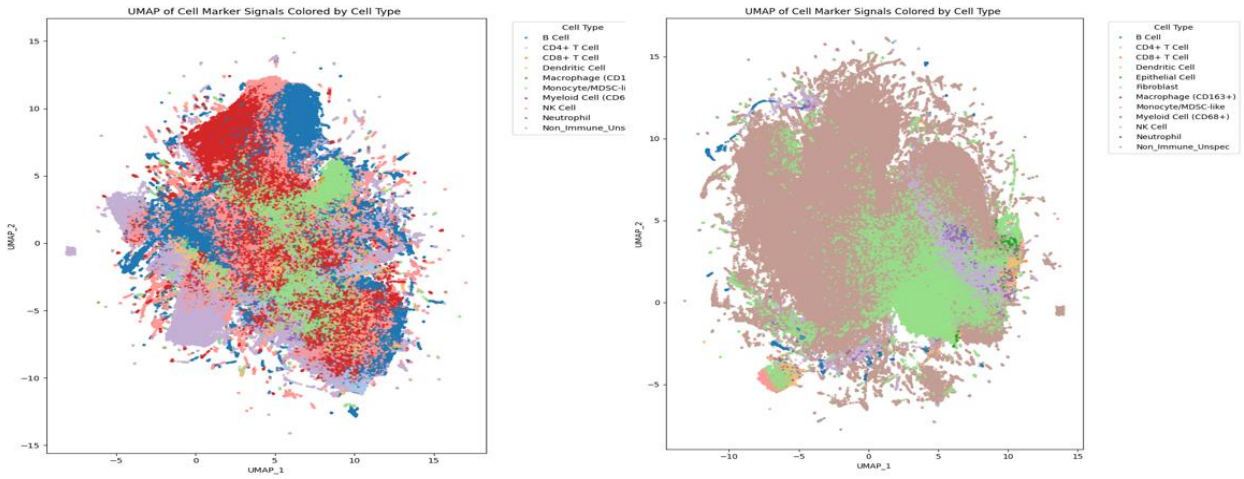


*Figure 2 Comparative UMAP Visualization of Cell Marker Signals Using 25 vs. 50 Biomarkers. Each UMAP plot displays **322,277 individual cells**, projected from high-dimensional space onto two dimensions (**UMAP_1** and **UMAP_2**) based on their marker signal profiles. Cells are colored by annotated cell type. The **left panel** uses a **25-marker panel**, optimized for major immune lineages and functional states. The **right panel** incorporates the **full 50-marker panel**, enabling finer resolution of cellular heterogeneity. Notably, the 50-marker panel resolves **two additional cell types—Eosinophils** and **Mast Cells**—which are absent in the reduced panel due to lack of lineage-specific markers (e.g., **GATA3**, **MPO**, **Galectin-3**). This comparison illustrates how biomarker selection directly influences cell type classification and clustering in UMAP space, highlighting the trade-off between panel efficiency and resolution of rare or phenotypically overlapping 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 S10 to S14.

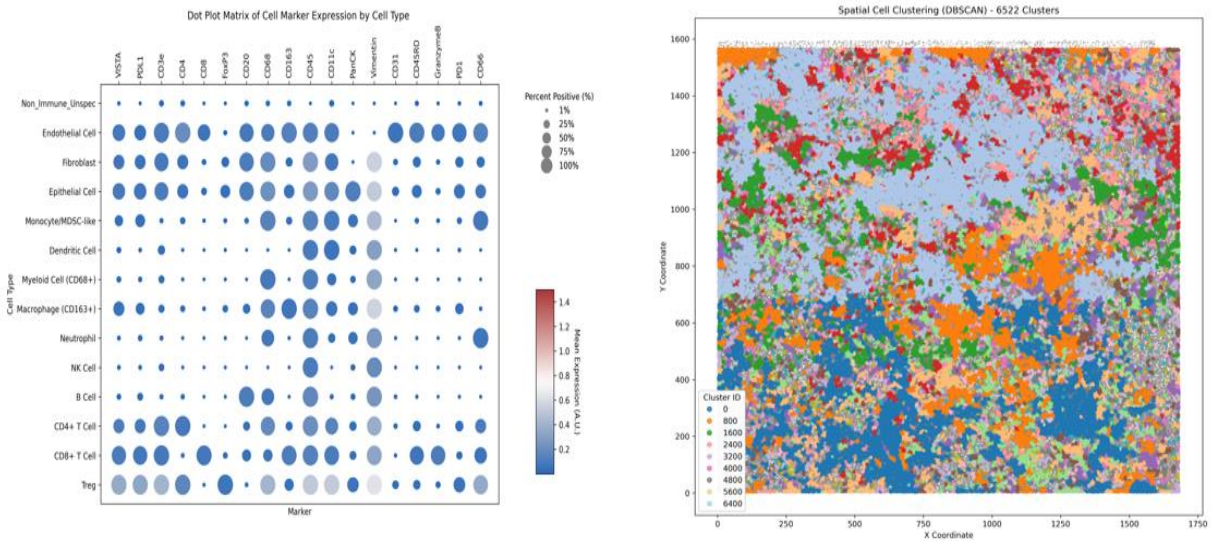


Figure 3 Cell phenotyping and spatial clustering analysis of lung adenocarcinoma tissue microenvironment. ***(Left)*** *Dot plot matrix showing marker expression profiles across 14 cell types. Dot size represents the percentage of marker-positive cells (1–100%); color intensity indicates mean expression level (0–1.4 AU). Cell types include immune populations (T cells, B cells, NK cells, myeloid cells) and structural compartments (epithelial, endothelial, fibroblasts).* ***(Right)*** *Spatial distribution of cells colored by DBSCAN-assigned clusters (6,522 clusters total). Each colored region represents a distinct spatial niche with high cell density. Gray dots indicate noise/unclustered cells in sparsely populated areas. Axes show tissue coordinates in pixels. This spatial clustering reveals the architectural organization of the tumor microenvironment and identifies discrete cellular neighborhoods*

Spatial architecture transitions from cohesive to fragmented during progression

We applied DBSCAN clustering to quantify architectural changes across disease stages (Figure 4). Normal lung tissue exhibited 18 clusters with a dominant macro-cluster containing ~87% of cells. AAH samples showed 41 clusters with increased fragmentation (~78% in the largest cluster). AIS maintained relative spatial continuity (~65% in one macro-cluster). MIA marked the onset of micro-fragmentation with 155 clusters, though one cluster still comprised ~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 (Table 1).

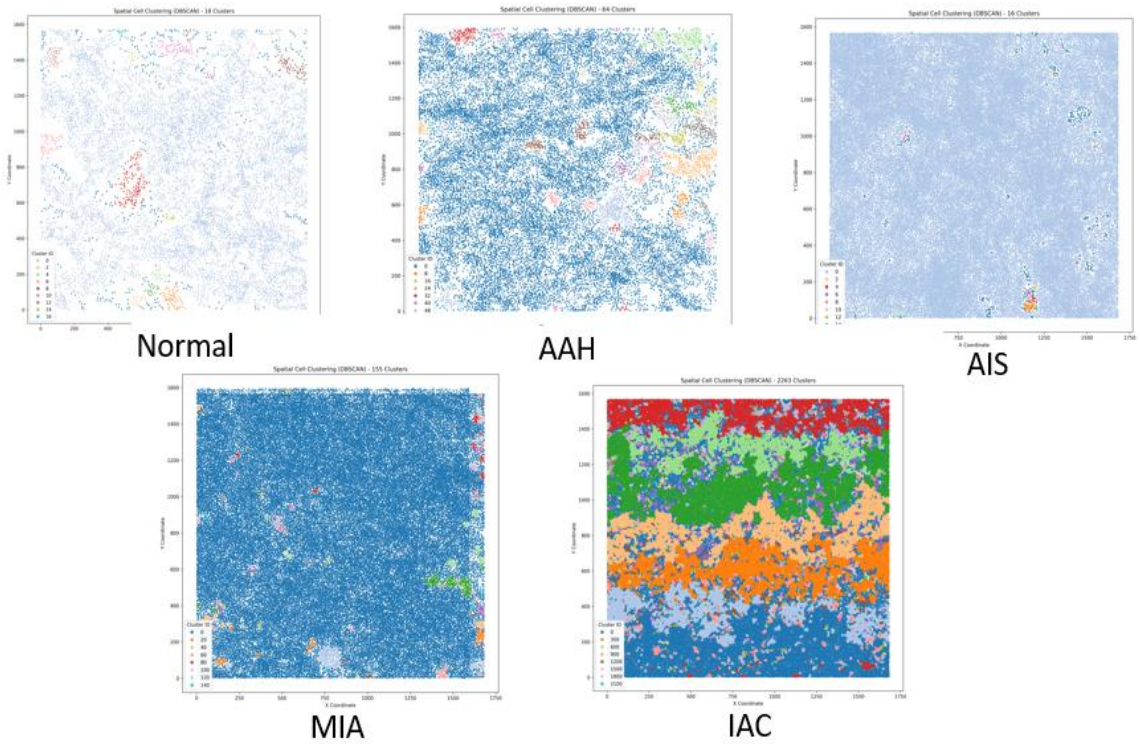


Figure 4. DBSCAN spatial clustering reveals progressive architectural disruption during LUAD development. *Spatial clustering maps across Normal lung, AAH, AIS, MIA, and IAC tissues demonstrate distinct patterns of tissue organization during malignant progression. Normal lung displays sparse, loosely distributed cells with minimal spatial clustering. AAH exhibits increased cellular density with emergence of defined spatial domains. AIS shows further architectural condensation with reduced heterogeneity. MIA maintains relatively organized spatial compartments with contiguous cellular structures. In contrast, IAC demonstrates dramatic architectural fragmentation characterized by highly heterogeneous, intermixed cellular populations and 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 clusters, minimal noise	Indicates dense and evenly distributed cells with intact tissue continuity—hallmark of healthy lung parenchyma.
Single dominant macro-structure (AAH, AIS)	One major cluster with few small clusters or noise	Reflects early-stage non-invasive lesions where cells remain spatially cohesive and localized.

Onset of micro-fragmentation (MIA)	Multiple small-to-medium clusters, macro-cluster still present	Suggests early disruption of spatial continuity, marking the transition toward micro-invasion.
Highly heterogeneous and fragmented architecture (IAC)	Numerous small clusters, high noise proportion	Captures extensive spatial heterogeneity, invasive epithelial foci, and reactive stroma—typical of malignant tumor organization.

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

Using a timed Petri net framework, we modeled biomarker-level dynamics across five histopathological stages (**Figure 5**). The analysis revealed stage-specific compositional shifts: fibroblast expansion in MIA/IAC, platelet surges in AIS, and NK-lineage enrichment in later stages (Table 2). Longitudinal trajectory analysis (**Figure 6**) demonstrated coordinated remodeling: CD4+ and CD8+ T cells exhibited progressive depletion, while NK cells and plasma cells expanded toward IAC. Myeloid compartments showed stage-specific changes, with monocytes and M2-like macrophages increasing in MIA/IAC, indicative of immunosuppressive reprogramming.

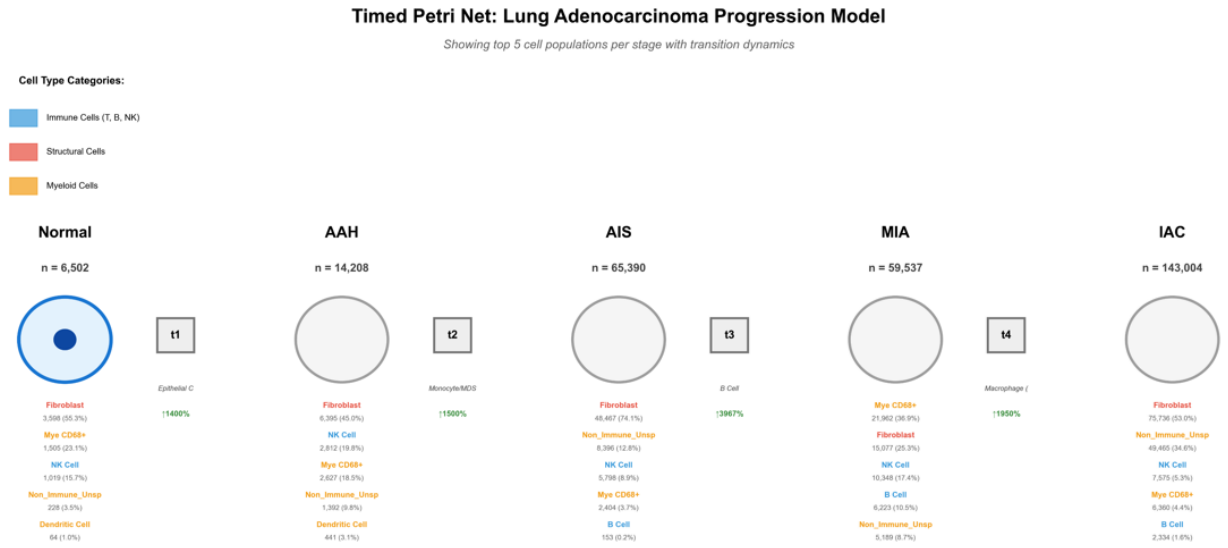


Figure 4 Timed Petri Net model of cell population dynamics across lung adenocarcinoma progression. This figure illustrates the top five cell populations per histopathological stage—Normal, AAH (Atypical Adenomatous Hyperplasia), AIS (Adenocarcinoma In Situ), MIA (Minimally Invasive Adenocarcinoma), and IAC (Invasive Adenocarcinoma)—as modeled by a timed Petri net framework. Each circular diagram represents a stage-specific cellular composition, with proportional segments color-coded by cell category: immune cells (IM, orange), stromal cells (ST, blue), and malignant cells (MA, gray). Transition arcs between stages

reflect dynamic shifts in dominant cell types, including macrophage persistence, gradual AT2 decline, and progressive enrichment of B cells and club cells. Sample sizes (n) are indicated per stage.

Critical inflection points mark immune remodeling transitions

Quantitative analysis of percentage change rates between adjacent stages revealed distinct inflection points (Figure 7). Notable transitions included: dramatic B cell amplification during AIS→MIA (+3967.3%), sharp epithelial expansion in Normal→AAH (+1400.0%) and AIS→MIA (+3680.0%), and sustained loss of cytotoxic T cell subsets throughout progression. These patterns define a dynamic immunoarchitectural landscape characterized by declining adaptive immunity and expanding innate/suppressive elements.

Cell Type	Normal	AAH	AIS	MIA	IAC
Platelet	3,598	–	48,467	–	–
Fibroblast / Plasma B	–	6,395	–	15,077	75,736
NK Cell	1,019	2,812	5,798	10,348	7,576
NK-Lymph Lineage	–	1,392	8,296	–	49,465
Myeloid Cell (CD68+/CD14+)	1,505	2,527	2,464	21,962	6,380
B Cell	27	438	153	6,223	2,334
Dendritic Cell	48	441	106	255	517
Epithelial Cell	2	90	–	2	532
Macrophage (CD163+)	0	9	–	4	62
Neutrophil / MDSC	–	3	–	48	–
CD4+ T Cell	2	1	23	8	189
CD8+ T Cell	2	0	22	8	28
Treg	1	0	21	4	28
Endothelial Cell	0	0	–	421	28

Monocyte / MDSC-like	6	–	30	–	510

Table 2 Comparative cell type counts across five histopathological stages of lung adenocarcinoma. This table summarizes the absolute counts of selected immune, stromal, and epithelial cell types across five stages: Normal, AAH (Atypical Adenomatous Hyperplasia), AIS (Adenocarcinoma In Situ), MIA (Minimally Invasive Adenocarcinoma), and IAC (Invasive Adenocarcinoma). Notable trends include the expansion of fibroblast/plasma B cells in MIA and IAC, a dramatic surge in platelet counts in AIS, and progressive enrichment of NK and NK-lymph lineage cells in later stages. Dashes (–) indicate absence or negligible representation of the corresponding cell type in that stage.

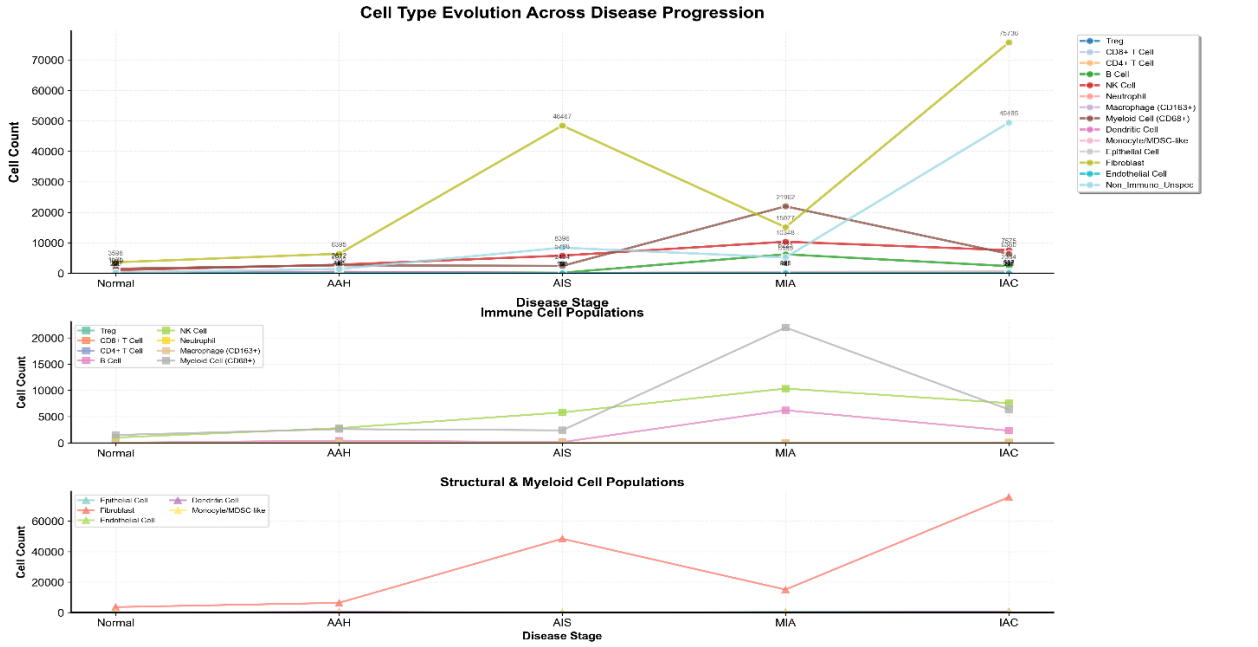


Figure 5 ***Cell type evolution across lung adenocarcinoma progression stages****. Following the Petri net–based modeling of LUAD dynamics, this figure presents a quantitative view of cell population trajectories across five histopathological stages: Normal, AAH, AIS, MIA, and IAC. The composite panel includes three line plots: (top) global trends of immune, stromal, and epithelial cell types; (middle) focused dynamics of key immune subsets including CD4⁺ T cells, CD8⁺ T cells, NK cells, B cells, plasma cells, and Tregs; and (bottom) structural and myeloid populations such as monocytes, macrophages (M1/M2), dendritic cells, mast cells, and epithelial cells. Notable transitions include a progressive decline in CD8⁺ T cells, expansion of plasma cells and NK subsets, and a sharp rise in epithelial cells at the IAC stage, reflecting immune remodeling and malignant transformation.*

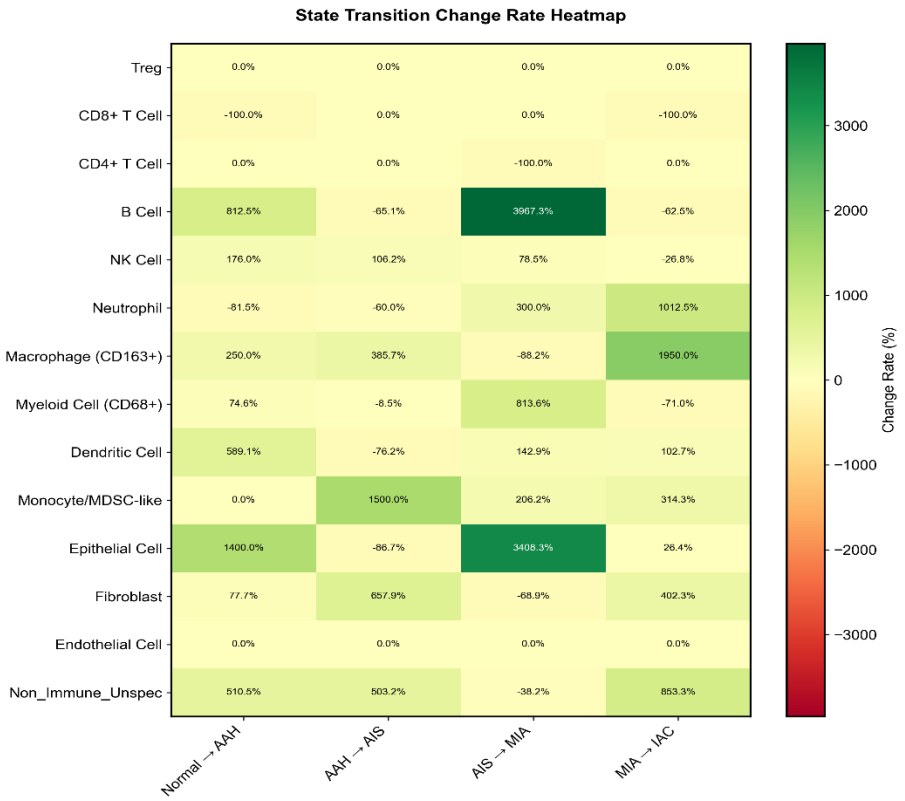


Figure 6 Heatmap of cell type transition rate across LUAD progression stages*. This figure visualizes the percentage change rates of selected immune, stromal, and epithelial cell populations across sequential histopathological transitions: Normal → AAH, AAH → AIS, AIS → MIA, and MIA → IAC. Each cell in the heatmap represents the relative change in abundance for a given cell type between adjacent stages, with color gradients indicating direction and magnitude (green: positive; red: negative). Notable transitions include a dramatic surge in B cells during AIS → MIA (+3967.3%), a sharp epithelial expansion in Normal → AAH (+1400.0%) and AIS → MIA (+3680.0%), and a marked decline in several immune subsets during MIA → IAC. These patterns highlight stage-specific remodeling of the tumor microenvironment and suggest potential inflection points for therapeutic intervention.*

Discussion

A computational pathology framework for spatial immune profiling

This study demonstrates that ROSIE can transform 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 322,277 cells across 50 ROIs spanning the LUAD continuum, 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. 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. In normal and early precancerous lesions (AAH, AIS), cells organized into large, contiguous macro-clusters reflecting intact tissue structure. However, at the onset of micro-invasion (MIA), this cohesion began fragmenting, and by the invasive stage (IAC), the landscape was dominated by thousands of discrete micro-niches. This transition from uniformity to heterogeneous fragmentation captures the essence of malignant invasion, where tumor cells break from their original focus and create spatially distinct microenvironments that interact with stromal and immune elements.

The architectural collapse we observed 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=0.215) in MIA detected early fragmentation, small epsilon values (ε=0.015) 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 a clear, stage-dependent remodeling of the immune microenvironment, characterized by a fundamental shift from adaptive immunity toward innate and ultimately immunosuppressive states. This transition represents a central narrative in LUAD precancer evolution and has profound implications for interception strategies.

Adaptive immune exhaustion.The most striking observation was the progressive and consistent depletion of CD8+ T cells throughout disease progression. CD8+ cytotoxic T lymphocytes are primary effectors of anti-tumor immunity, and their gradual disappearance suggests active immune evasion or exhaustion mechanisms emerging early in premalignancy. This depletion occurred in parallel with CD4+ T cell loss, indicating broad adaptive immune dysfunction rather than subset-specific suppression. Notably, regulatory T cells (Tregs) showed variable dynamics, suggesting that immunosuppression involves mechanisms beyond Treg-mediated tolerance.

Innate immune expansion and plasticity. Concurrent with adaptive immune decline, we observed steady expansion of innate populations, particularly NK cells and their lymphoid lineage derivatives. NK cells provide rapid, MHC-independent tumor surveillance and may compensate initially for T cell dysfunction. However, their sustained elevation into invasive stages suggests incomplete tumor control, possibly reflecting NK cell exhaustion or tumor immune editing. The dramatic increase in plasma cells and B cells—particularly the +3967% surge during AIS→MIA—was unexpected and warrants further investigation. This 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. We observed increased M2-polarized macrophages (CD163+) and monocyte/MDSC-like populations in MIA and IAC stages. M2 macrophages produce immunosuppressive cytokines (IL-10, TGF-β) and promote angiogenesis, tissue remodeling, and tumor progression. The rise in MDSC-like cells further suggests active recruitment and education of myeloid suppressors within evolving lesions. 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 showed marked expansion in MIA and IAC, reaching 75,736 cells in invasive tumors. 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 a reciprocal crosstalk driving TME evolution. Disrupting CAF-immune cell interactions may represent an orthogonal strategy for precancer interception.

Biological insights into immune-epithelial dynamics

The dramatic surge in epithelial cells during AIS→MIA (+3680%) coincided with B cell expansion (+3967%), suggesting coordinated immune-epithelial interactions at this critical transition point. Several biological mechanisms may underlie this phenomenon:

Tertiary lymphoid structure formation. Organized lymphoid aggregates containing B cells, T cells, and dendritic cells can form within tumors and precancerous lesions. TLS have been associated with both favorable (anti-tumor immunity) and unfavorable (chronic inflammation) outcomes depending on composition and functional state. The B cell surge we observed may reflect TLS formation attempting to mount humoral immunity against emerging neoantigens.

Epithelial proliferation and immune recognition. Accelerated epithelial proliferation during the AIS→MIA transition likely generates replication stress, DNA damage, and neoantigen expression. This antigenic burden may trigger B cell 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 cell 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 cell clonality, antibody specificities, and functional states (pro- vs. anti-tumor) to determine whether B cell 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.

Mouse model concordance. We utilized five mouse models to complement human data, but genetically engineered mouse models may not fully recapitulate human immune evolution. Differences in immune cell repertoires, checkpoint receptor expression kinetics, and TME composition between species could affect therapeutic response predictions. Humanized mouse models or patient-derived xenografts may provide more faithful preclinical platforms.

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 vulnerability we identified challenges 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.

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, 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. 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 classifications, ultimately yielding 322,277 single-cell marker expression profiles.

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. To assess the impact of panel size on resolution, we also analyzed subsets using 10-marker and 25-marker panels.

Fuzzy Cell Type Classification

We assigned cell type identities using a custom scoring algorithm guided by a manually curated rule set (Table 3, Figure 8). This rule set was developed by synthesizing canonical marker expression patterns from established immunophenotyping literature[21,22]. For each cell type—including T cell subsets, B cells, myeloid populations, and structural cells—we defined required positive markers and exclusionary negative markers based on authoritative lineage definitions[23–25]. 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
- AAH: epsilon = 50 pixels, min_samples = 50
- AIS: SmartDensity mode with density threshold = 0.005, min_samples = 100
- MIA: epsilon = 215 pixels, min_samples = 105

- IAC: epsilon = 15 pixels, min_samples = 105

SmartDensity mode adaptively selects epsilon based on local cell density. Cells not assigned to clusters were classified as spatial "noise." For each cluster, we calculated cell type proportions, spatial metrics (area, perimeter, convexity), and neighborhood enrichment scores.

Dimensionality Reduction and Visualization

To visualize cellular heterogeneity, we applied UMAP to the 322,277-cell dataset 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.

Temporal Modeling with Timed Petri Nets

We constructed a timed Petri net model to represent cell population dynamics across disease progression. Each histopathological stage was modeled as a discrete state, with cell type abundances represented as tokens. Transitions between stages were governed by empirically derived rate constants proportional to observed fold-changes.

Mouse Models and Cross-Species Validation

To validate human findings and test therapeutic interventions, we used five genetically engineered mouse models (GEMMs) of lung adenocarcinoma. Mice were housed under specific pathogen-free conditions, and all experiments were approved by the institutional Animal Care and Use Committee. Tumor burden was assessed at defined endpoints, and tissue sections were analyzed using ROSIE.

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 was calculated as: (Change = [Count_stage2 − Count_stage1] / Count_stage1) × 100% 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.

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 3. 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 8. 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.*

Data Availability

The lung adenocarcinoma H&E whole slide images and ROSIE-predicted virtual multiplex immunofluorescence data for 50 ROIs (322,277 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), and all quantitative data underlying the figures are provided as Supplementary Data files 1-5. Source data are provided with this paper.

Code Availability

Analyses were performed using Python 3.10. ROSIE (https://gitlab.com/enable-medicine-public/rosie) was used for multiplex biomarker inference, and QuPath 6.0 (https://qupath.github.io/) for image analysis. Custom scripts are available from the corresponding author upon reasonable request.

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

10. 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).

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

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

13. 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).

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. 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).

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

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

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

20. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)* 226–231 (AAAI Press, Portland, Oregon, USA, 1996).

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

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

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

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

25. 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

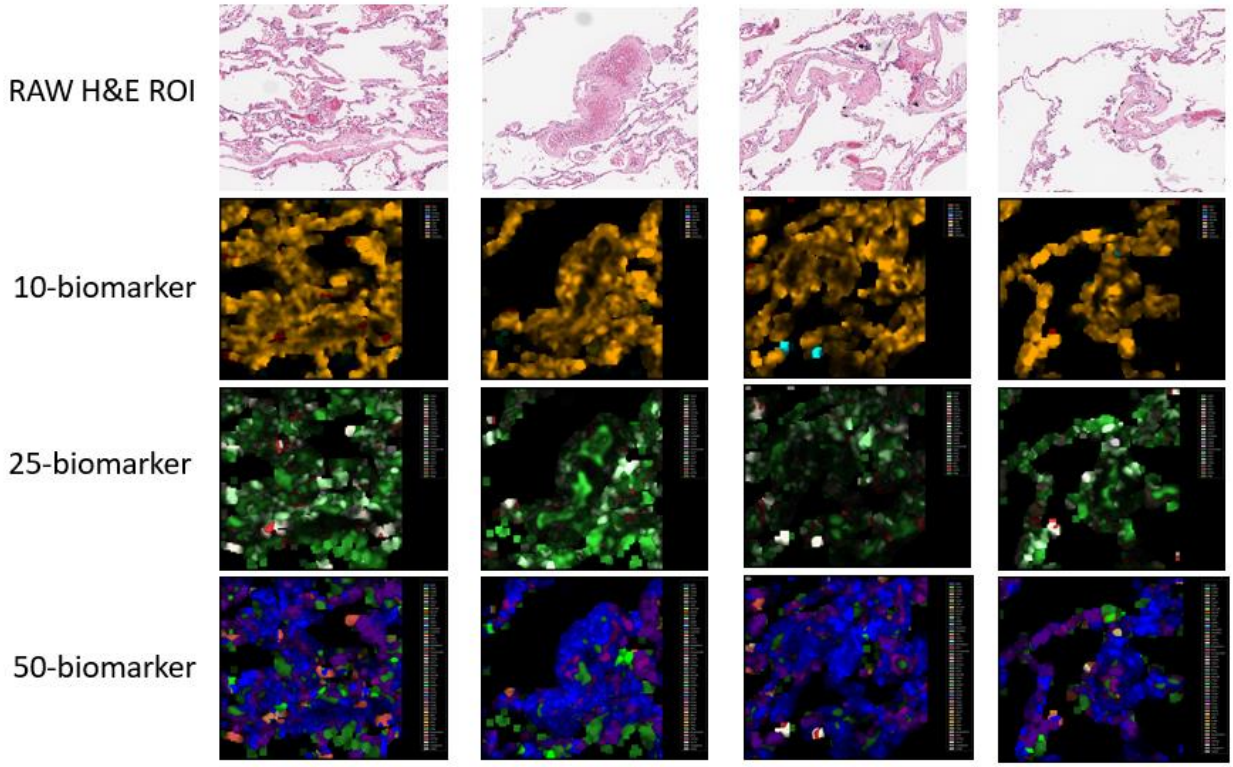
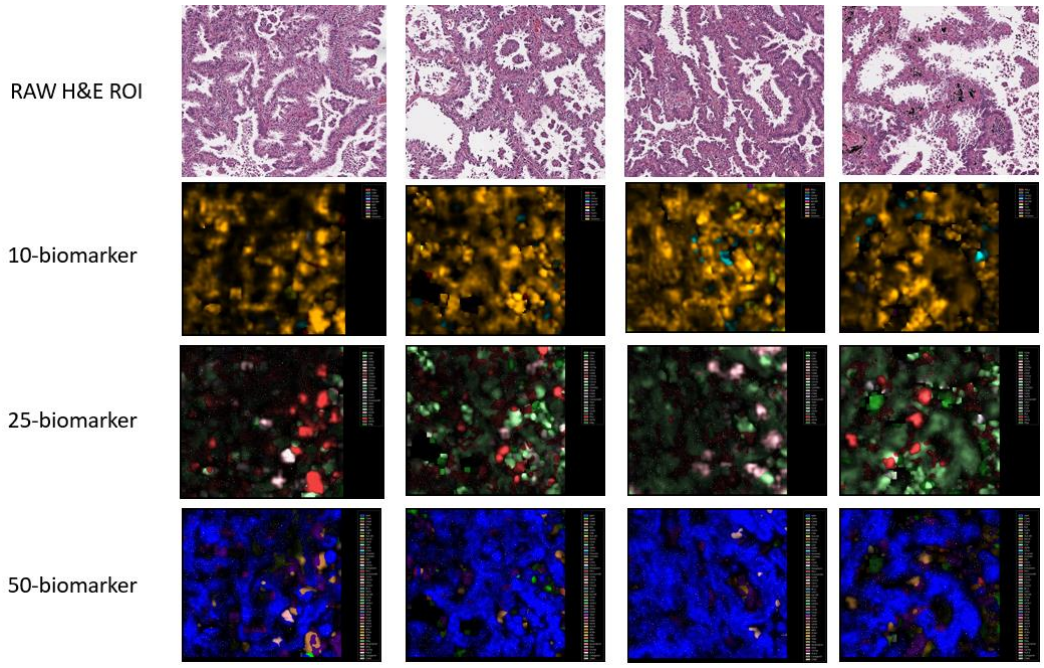


Figure 7 demonstrates ROSIE's scalability: a single H&E slide, which was picked from human LUAD tissues spanning normal lung to invasive adenocarcinoma stages, yielded 10-, 25-, and 50-marker v-mIF panels without additional staining, with larger panels enabling more comprehensive spatial profiling of immune and stromal components.

569

570 **S2.**



571

572 Figure 8 ***Multi-panel ROSIE predictions from H&E-stained invasive adenocarcinoma tissue demonstrate scalable biomarker***
573 ***inference.*** *Top row: Original H&E regions of interest (ROIs) from invasive adenocarcinoma (IAC) samples representing the*
574 *terminal stage of LUAD progression. Rows 2-4: ROSIE-generated virtual multiplex immunofluorescence (v-mIF) using 10-*
575 *biomarker, 25-biomarker, and 50-biomarker panels, respectively. Each column represents a distinct tissue region. All v-mIF*
576 *images are computationally predicted from the same H&E input without additional staining. Progressive increase in panel size*
577 *reveals increasingly detailed spatial organization of immune and stromal components through enhanced multi-channel overlay*
578 *visualization. Scale bar, 100 µm.*

579 Figure 9 demonstrates ROSIE's scalability: from original H&E ROIs of invasive adenocarcinoma samples
580 (top row), ROSIE generated progressively detailed v-mIF predictions using 10-, 25-, and 50-marker
581 panels (rows 2-4) across four distinct tissue regions (columns). **The 50-biomarker panel (bottom row)**
582 **reveals substantially enhanced cellular resolution compared to smaller panels, enabling discrimination**
583 **of rare immune populations (such as eosinophils, mast cells, and dendritic cell subsets) and detailed**
584 **mapping of immunosuppressive cell networks (including CD163+ macrophages, MDSCs, and**
585 **regulatory T cells) that are critical for understanding tumor immune evasion but remain undetectable**
586 **with 10- or 25-marker approaches.** The progressive increase in multi-channel overlay complexity
587 demonstrates how comprehensive biomarker panels capture the full heterogeneity of the tumor
588 microenvironment without additional tissue staining.

589

590

591

S3.

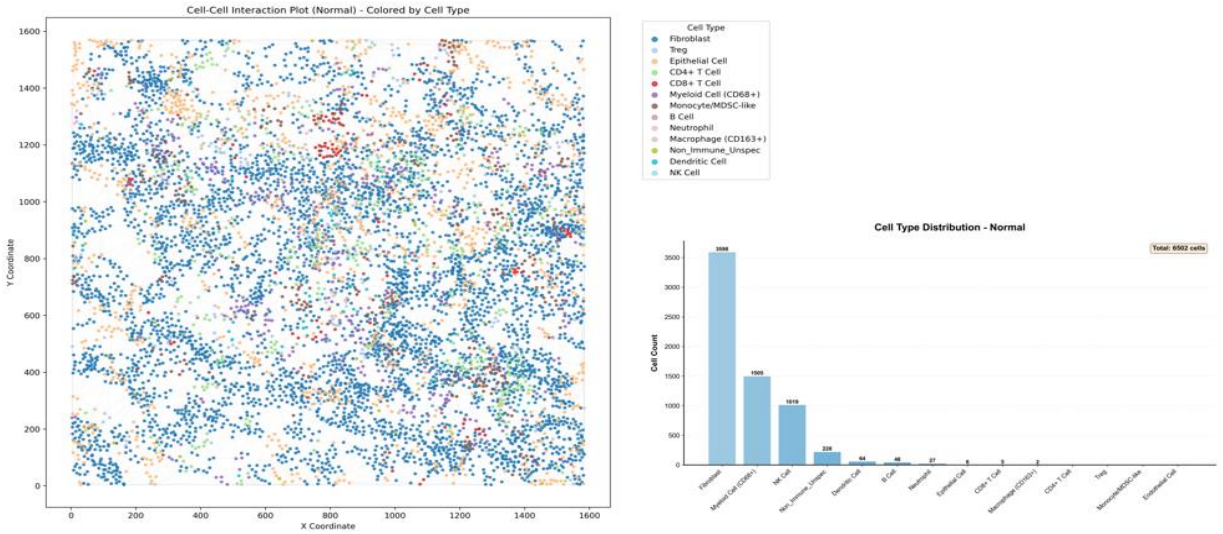


Figure 9 ***Spatial distribution and quantification of cell types in normal lung tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within normal lung tissue. Each dot represents a single cell positioned according to its X-Y coordinates. Right: Bar chart displaying the absolute cell counts for each cell type. Normal lung tissue is dominated by fibroblasts (3,668 cells, blue) and epithelial cells (1,999 cells, orange), which form the primary structural components of healthy lung parenchyma. Regulatory T cells (Tregs, 1,079 cells, green) represent the most abundant immune population. Other immune cell types including CD4+ T cells, CD8+ T cells, myeloid cells (CD68+), monocytes/MDSCs, B cells, neutrophils, macrophages (CD163+), dendritic cells, and NK cells are present at lower frequencies. This spatial and quantitative profile establishes the baseline cellular composition of normal lung tissue for comparison with pre-malignant and malignant stages of LUAD progression.*

S4.

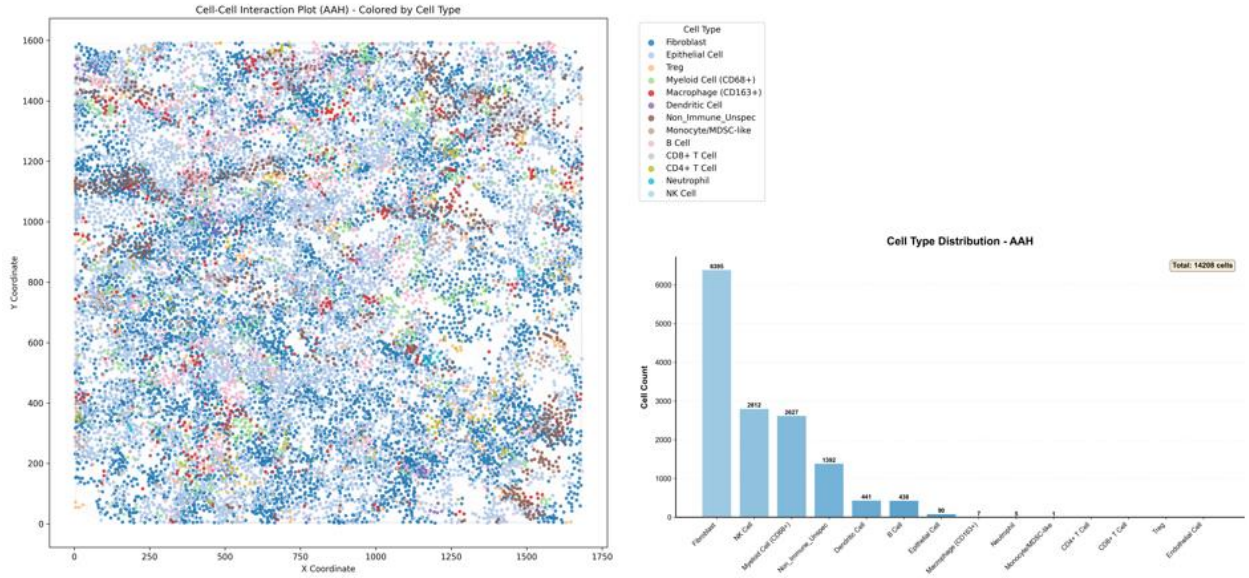


Figure 10 ***Spatial distribution and quantification of cell types in Atypical Adenomatous Hyperplasia (AAH) tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within AAH tissue. Each dot represents a single cell positioned according to its X-Y coordinates. Right: Bar chart displaying the absolute cell counts for*

each cell type (total: 14,069 cells). AAH tissue exhibits a marked increase in fibroblasts (6,461 cells, blue) compared to normal lung, alongside substantial populations of epithelial cells (2,914 cells, orange) and myeloid cells (CD68+, 2,607 cells, pink). CD4+ T cells (1,242 cells, red) are more prominent than in normal tissue, while regulatory T cells (Tregs, 631 cells, green) remain present. Other immune populations including CD8+ T cells, macrophages (CD163+), monocytes/MDSCs, B cells, neutrophils, dendritic cells, and NK cells are detected at lower frequencies. This cellular composition reflects the early pre-malignant stage of LUAD progression, characterized by stromal expansion and increased immune infiltration compared to normal lung tissue.

S5.

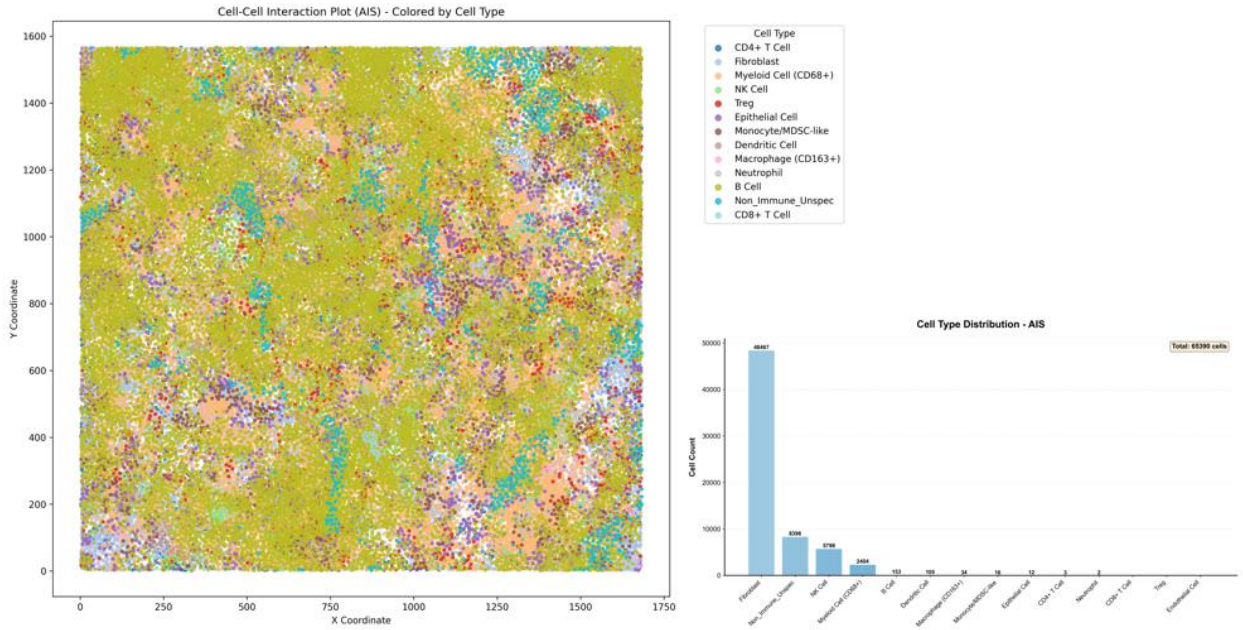


Figure 11 ***Spatial distribution and quantification of cell types in Adenocarcinoma In Situ (AIS) tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within AIS tissue. Each dot represents a single cell positioned according to its X-Y coordinates. Right: Bar chart displaying the absolute cell counts for each cell type (total: 63,080 cells). AIS tissue shows a dramatic shift in cellular composition, with CD4+ T cells (48,977 cells, red) becoming overwhelmingly dominant, representing the most abundant population. Myeloid cells (CD68+, 9,883 cells, pink) constitute the second largest population, followed by NK cells (1,917 cells, cyan) and fibroblasts (878 cells, blue). Epithelial cells, monocytes/MDSCs, B cells, macrophages (CD163+), neutrophils, regulatory T cells (Tregs), dendritic cells, non-immune unspecified cells, and CD8+ T cells are present at markedly lower frequencies. This pronounced T cell infiltration and altered stromal-immune balance characterizes the AIS stage, representing a critical transition point in LUAD progression where the tumor remains non-invasive but exhibits significant immunological remodeling.*

S6.

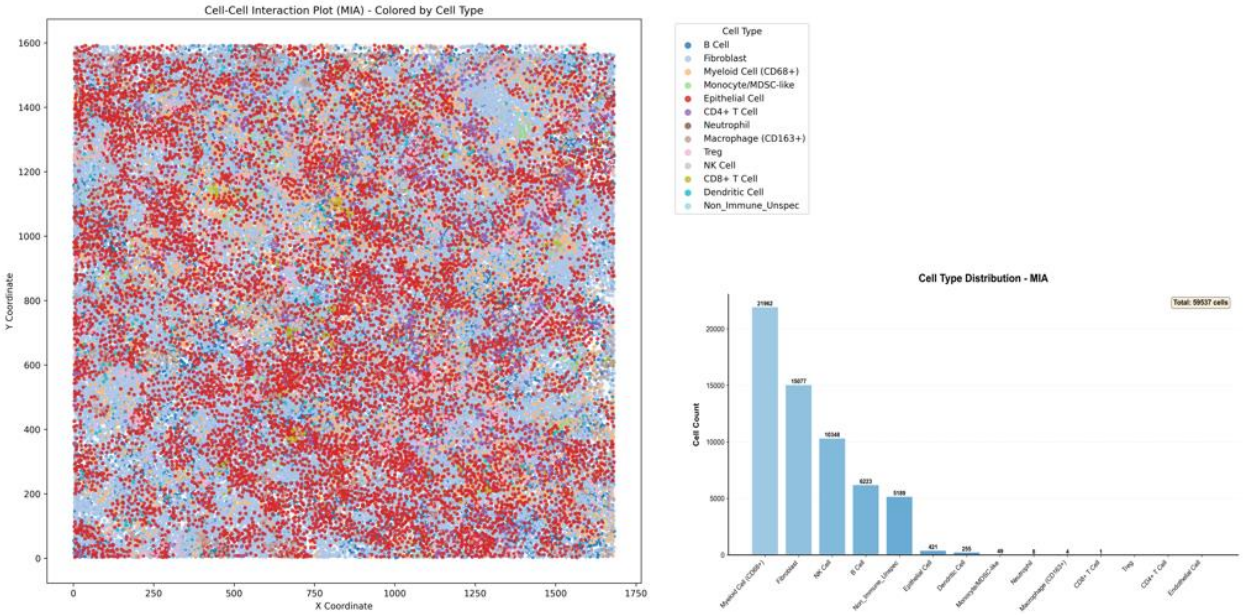


Figure 12 ***Spatial distribution and quantification of cell types in Minimally Invasive Adenocarcinoma (MIA) tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within MIA tissue. Each dot represents a single cell positioned 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 cell infiltration, restoration of stromal components, and a more complex immune microenvironment compared to the CD4+ T cell-dominated AIS stage.*

S7.

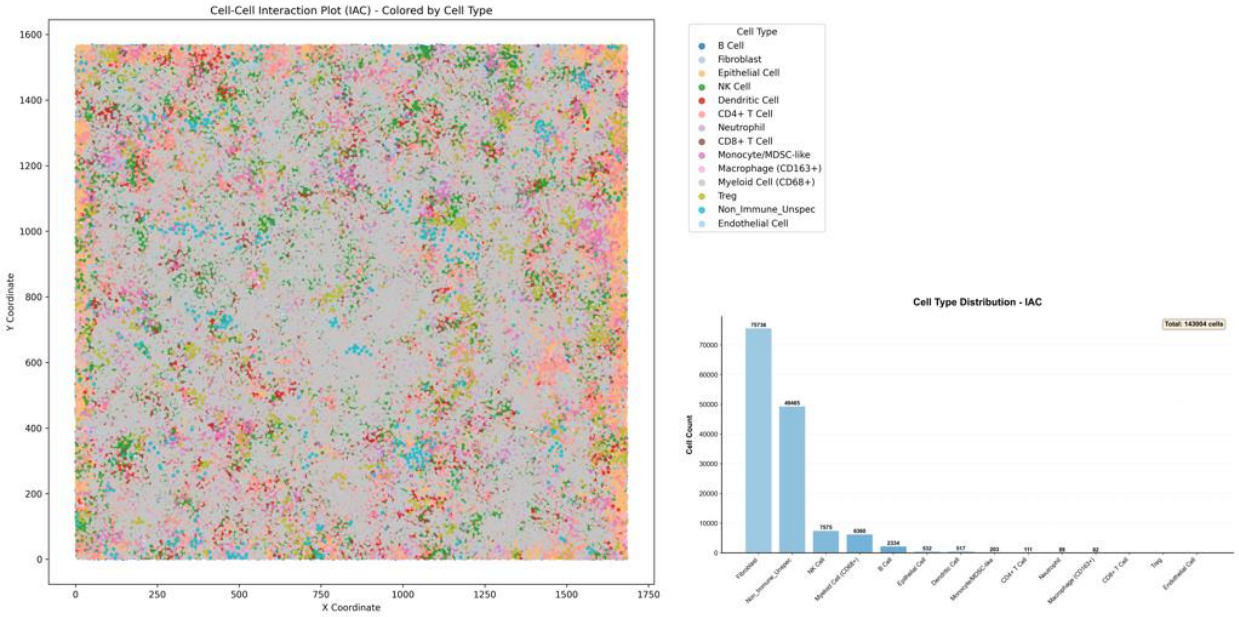


Figure 13 ***Spatial distribution and quantification of cell types in Invasive Adenocarcinoma (IAC) tissue.*** *Left: Cell-cell interaction plot showing the spatial coordinates of individual cells colored by their predicted cell types within IAC tissue. Each dot represents a single cell positioned according to its X-Y coordinates. Right: Bar chart displaying the absolute cell counts for each cell type (total: 153,084 cells). IAC tissue, representing the terminal stage of LUAD progression, is dominated by B cells (73,746 cells, tan) and fibroblasts (49,691 cells, blue), which together constitute over 80% of the total cell population. Epithelial cells (7,976 cells, orange) and myeloid cells (CD68+, 5,866 cells, pink) are present at moderate frequencies. Regulatory T cells (Tregs, 3,468 cells, green), NK cells (cyan), dendritic cells (yellow), CD8+ T cells (pink), neutrophils (light pink), monocytes/MDSCs (brown), macrophages (CD163+), CD4+ T cells (red), and non-immune unspecified cells are detected at relatively low frequencies. This cellular composition reflects extensive stromal remodeling and sustained B cell infiltration characteristic of fully invasive adenocarcinoma, with a notable shift from the T cell-dominated profiles observed in earlier pre-malignant stages.*