Multi-omics Analyses Identify ANLN as a Prognostic Biomarker for Recurrence and Metastasis in Non–Small Cell Lung Cancer

Haiwei Quan[1#], Zhiguang Xu[2], Zhibin Wang[2,3*]

[1]*Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen, Guangdong, PR China;*

[2]*Department of Biopharmaceutical Sciences, Faculty of Pharmaceutical Sciences, Shenzhen University of Advanced Technology, Shenzhen, Guangdong, PR China;*

[3]*Center for Cancer Immunotherapy of Institute of Biomedicine and Biotechnology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, PR China.*

Corresponding authors:
Zhibin Wang

Department of Biopharmaceutical Sciences, Faculty of Pharmaceutical Sciences, Shenzhen University of Advanced Technology, Shenzhen, Guangdong, PR China.

Center for Cancer Immunotherapy of Institute of Biomedicine and Biotechnology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, PR China.

Email: wangzhibin@suat-sz.edu.cn

Other author's email:

Haiwei Quan, 17582928763@163.com

Zhiguang Xu, zg.xu2@siat.ac.cn

KEYWORDS: ANLN, lung cancer, metastasis, single cell, epithelial cells

ABSTRACT

Lung cancer recurrence and metastasis remain major causes of cancer-related mortality, yet the molecular determinants driving these processes are not fully defined. Herein we did integrated analyses of bulk transcriptomic data and single-cell RNA sequencing data from multi-cohorts and identified key regulators of lung cancer progression. By integrating gene features related to survival, recurrence, and metastasis from independent cohorts and further analyzing the composition of the tumor microenvironment, we identified ANLN as a core progression-related gene with poor prognosis. Its expression was elevated in recurrent and metastatic tumors and correlated with reduced overall survival. xCell analysis revealed epithelial enrichment and relative immune and stromal depletion in ANLN-high tumors. Single-cell analyses of lymph node metastases demonstrated that ANLN is predominantly expressed in epithelial and proliferating tumor cells and is associated with extensive transcriptional and remodeling of the tumor microenvironment. Functional scoring and enrichment analyses revealed that ANLN-high epithelial cells exhibit coordinated activation of proliferative and migratory programs alongside suppression of immune-associated features. Pseudotime trajectory analysis further positioned ANLN enrichment at a critical intermediate state during AT2-derived epithelial evolution toward an invasive phenotype. These findings were further validated in an independent single-cell dataset capturing the transition from in situ to invasive lung cancer. In summary, our results identify ANLN as a marker of a conserved, invasion-prone epithelial state underlying lung cancer recurrence and metastasis, providing mechanistic insights and potential therapeutic implications.

1.INTRODUCTION

Lung cancer remains the most diagnosed malignancy and the leading cause of cancer-related mortality globally[5]. Clinically, Lung cancer is broadly classified into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) with NSCLC accounting for approximately 85% of all cases[6]. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two predominant histological subtypes of NSCLC. Although therapeutic strategies such as surgery, chemotherapy, radiotherapy, and combination regimens have achieved progress in controlling lung cancer progression, clinical outcomes for NSCLC patients remain suboptimal. Lung cancer is characterized by a low five-year survival rate and patients without recurrence beyond five years are typically considered cured[7]. Notably, approximately 20-40% of patients with stage I NSCLC still experience tumor recurrence after surgical resection[8-10], late recurrence still occurs in approximately 8-11% of cases[11-12]. Highlighting the urgent need to elucidate the underlying mechanisms driving disease progression and recurrence.

Cancer metastasis is one of the leading causes of cancer-related death[13-14]. It represents a highly complex, multistep process involving the coordinated activation of multiple biological pathways. This process includes the acquisition of invasive and migratory capabilities by tumor cells, detachment from the primary tumor site, survival during dissemination, and colonization of distant organs to form secondary tumors[15-16]. Cancer metastasis is governed by intricate biological processes, in which tumor cells undergo epithelial–mesenchymal transition, metabolic reprogramming, and escape from immune surveillance[17-19]. Despite substantial advances in elucidating metastasis-associated molecular mechanisms, the key regulatory factors that drive the

transition of tumor cells from localized growth to metastatic dissemination remain incompletely understood.

Tumor recurrence and metastasis are complex events driven by multiple biological processes, including enhanced tumor cell proliferation, increased migratory capacity, and immune evasion[20]. Accumulating evidence has implicated genomic alterations and epigenetic dysregulation in the malignant progression of lung cancer[21-23]. Additionally, the tumor microenvironment (TME) plays a critical role in driving metastasis: for instance, in particular, enrichment of alveolar type II epithelial cells and tumor microenvironment remodeling characterized by macrophage polarization have been associated with malignant progression[24-26]. However, most previous studies have focused on individual molecular events or cell populations, an integrated analysis that comprehensively links survival, recurrence, metastasis, and dynamic cell-state transitions at single-cell resolution is still lacking. Such an integrated approach is essential to identify key regulators that govern the metastatic cascade and could serve as potential therapeutic targets.

Herein we used an integrative analytical strategy combining large-scale transcriptomic datasets with co-expression network analysis, intercellular communication inference, and pseudotime trajectory analysis. Through this approach, we identified ANLN as a key determinant of a highly plastic and invasion-prone epithelial state and demonstrated its involvement in transcriptional reprogramming and microenvironmental remodeling during lung cancer metastasis. These

findings provide novel insights into the molecular mechanisms of NSCLC metastasis and highlight ANLN as a potential therapeutic target for improving clinical outcomes.

2.MATERIALS AND METHODS

Data Collection

To investigate key genes associated with lung cancer recurrence and metastasis, we retrieved publicly available datasets from the Gene Expression Omnibus (GEO): GSE41271, GSE166720, GSE277742, and GSE189357. Additionally, we collected two datasets related to overall survival (OS) in lung cancer patients: GSE157009 and GSE42127. After filtering, samples included those with recurrence (n = 119) or metastasis (n = 28), and non-recurrence (n = 152) or non-metastasis (n = 45) for downstream analyses. All datasets were processed and analyzed using R (v.4.5.2). Differential gene expression was assessed using the limma package (v.3.58.1), applying thresholds of $|\log_2 FC| \geqslant 0.3$ with $p<0.05$, and $|\log_2 FC| \geqslant 0.58$ with $p<0.01$ for the identification of recurrence or metastasis associated genes.

Survival Analysis

Survival analysis was performed using the R packages survival (v3.5.7) and survminer (v0.4.9). A survival object was constructed via Surv, followed by fitting a Cox proportional-hazards model with coxph to estimate hazard ratio (HR) and p-value. Kaplan–Meier survival curves were then generated using survfit and visualized by ggsurvplot.

Immune Infiltration Analysis

Patients were stratified according to the expression of the key gene signature associated with lung cancer recurrence and metastasis, and their tumor transcriptome matrices were processed by xCell(v1.1.0) to estimate the relative abundance of diverse immune and stromal cell populations (cell-type enrichment scores) within each sample.

Enrichment Analysis

To explore the biological pathways involved in genes associated with lung cancer metastasis and recurrence, gene set enrichment analysis (GSEA) was performed on transcriptomic data. An ordered gene list was generated by ranking all differentially expressed genes from high to low. This ranked list was then compared against the Hallmark, Gene Ontology (GO), and KEGG gene sets. The results of this analysis are used to identify key genetic functions and signaling pathways associated with lung cancer recurrence and metastasis.

CellChat Analysis

To compare intercellular communication between the ANLN-high and ANLN-low expression groups, the CellChat pipeline (v.1.6.1) was used to analyze cell–cell signaling interactions. A separate CellChat object was constructed for each group from the corresponding Seurat objects[1]. The human CellChatDB database was used as the reference. Communication probabilities were computed using the truncated mean method. Intercellular communication networks for each group were then inferred and aggregated using default parameters. To assess differences in cell–cell communication between the high and low ANLN expression groups, the total number of inferred interactions and the overall interaction strength were calculated and compared.

Differential signaling pathways between the two groups were analyzed, and the relative contribution of specific signaling pathways was further evaluated.

scRNA-seq Analysis

Single-cell data from metastatic and non-metastatic lymph nodes were processed using Seurat v5.3.1[2]. Raw matrices from 22 samples were loaded individually, converted into Seurat objects, and merged while retaining sample-level metadata. Quality filtering removed cells with <200 or >10,000 detected genes, UMI counts <500, or mitochondrial、hemoglobin gene fractions exceeding 10%. After normalization, identification of 3,000 highly variable genes, scaling, and PCA, batch correction was performed using Harmony[3], followed by neighbor graph construction, clustering (resolution = 0.8), and UMAP visualization based on the Harmony. We identified the gene expression markers for each cluster using FindAllMarkers function. Cell clusters were annotated as known cell types utilizing canonical marker genes.

Gene-signature Scoring

To evaluate malignant programs within epithelial cells at single-cell resolution, we defined four gene sets representing distinct biological processes: Proliferation, Motility, Immunity, and Metastasis. The Proliferation set was based on the term “GOBP mitotic nuclear division”, “GOBP cell division”, “KEGG cell cycle”, “GOBP epithelial cell proliferation” and “GOBP mitotic cytokinesis”, “GOBP DNA replication”; the Motility set was based on “GOBP extracellular matrix organization” and “GOBP cortical actin cytoskeleton organization”; the Immunity set utilized the “GOBP adaptive immune response” and “KEGG natural killer cell

mediated cytotoxicity"; the Metastasis set aggregated these three gene sets. AddModuleScore package was used to calculate scores for each cell and applied in downstream pseudotime, correlation, and functional analyses.

Pseudotime Analysis

To investigate the dynamic transitions of epithelial and proliferating cells, pseudotime trajectory analysis was performed using the Monocle3 package[4]. Normalized count matrices and cell metadata exported from Seurat were converted into a cell_data_set object, followed by PCA preprocessing and Uniform manifold approximation and projection (UMAP) based dimensionality reduction. The trajectory graph was learned and cells were ordered in pseudotime, using the epithelial population as the root state. Module scores representing metastasis, EMT, stemness, and proliferation related programs were calculated from curated gene sets and mapped along the pseudotime continuum to assess coordinated transcriptional changes.

3.RESULTS

An integrated analysis of survival, recurrence, and metastasis identifies a core gene set critical for lung cancer progression.

Survival analysis of the GSE157009 and GSE42127 datasets identified 2,499 and 2,213 genes associated with overall survival (OS), with 165 genes overlapping between the two cohorts (Fig. 1A), which represents a cross-dataset consensus survival-associated signature. Differential expression analysis comparing the Recurrence (Rec) group with the Non-Recurrence (Non-Rec)

group revealed 572 genes that were significantly upregulated in the Rec group (Figure 1B). In contrast, 323 genes were upregulated in the Metastasis (Met) group relative to the corresponding control group (Figure 1C), indicating substantial transcriptomic reprogramming associated with both progression phenotypes. The Venn diagram integrating 165 survival-related genes with the 572 recurrence-upregulated and 323 metastasis-upregulated genes identified four shared genes—ANLN, TPX2, CDCA8, and RAD54L (Fig. 1D), suggesting that these genes may act as key regulators simultaneously linked to recurrence, metastasis and reduced overall survival.

Further analysis of the upregulated genes using GSEA revealed significant enrichment of the EMT and Angiogenesis pathways in both the Rec (Fig. 1E) and Met groups (Fig. 1F). These pathways are known biological drivers of tumor invasion, facilitating cancer cell migration and dissemination. In addition, GO analyses indicated extracellular matrix remodeling, collagen assembly, regulation of cell adhesion and migration, and vasculature development were all upregulated in the Rec and Met groups (Fig. 1G，1H). These processes form the essential microenvironmental basis that promotes tumor invasion, which is consistent with the malignant biological characteristics of recurrence and metastasis observed in these samples.

WGCNA and xCell analyses identify ANLN as a key biomarker associated with tumor metastasis and recurrence.

To identify core genes underlying lung cancer recurrence and metastasis, weighted gene co-expression network analysis (WGCNA) was constructed for the Rec cohort (Fig. 2A) and the Met cohort (Fig. 2B), respectively. Module–trait correlation analysis revealed that in the

metastasis cohort (Fig. 2C), the turquoise module showed the strongest association with metastatic status. Among the four candidate genes, ANLN, TPX2, and CDCA8, were clustered within this module, suggesting that these three genes represent a metastasis-related gene program. In the recurrence cohort (Fig. 2D), the brown module displayed a significant positive correlation with recurrence, and similarly, ANLN, TPX2 and CDCA8 were assigned to this module. These findings indicate that the three genes belong to the co-expression network module strongly associated with lung cancer recurrence and metastasis, and show synergistic changes.

To further explore the relationship between these genes and the tumor microenvironment (TME), xCell was applied to infer immune and stromal components from the recurrence cohort. Samples were then stratified into high and low-expression groups based on the expression levels of ANLN, TPX2, and CDCA8, respectively. Notably, the ANLN-high group showed the most significant increase in epithelial cell proportion, whereas the Immune score, Stroma score, and Microenvironment score were all reduced to varying degrees (Fig. 2I), indicating that ANLN high expression tumors are more inclined towards the microenvironment characteristics of epithelial enrichment, immune and matrix depletion. Comparative evaluation of the differences in xCell-derived indices between high and low expression groups of the three candidate genes (Fig. 2J), further demonstrated that ANLN-high tumors exhibited the largest increase in epithelial cell. Although reductions in immune, stromal, and microenvironment scores were observed across all three high-expression gene groups, these reductions were not the most prominent in the ANLN-high group. Considering the established role of epithelial-related programs in tumor

recurrence and metastasis, ANLN was ultimately identified as the most representative key progression gene among the three candidates.

ANLN expression was significantly higher in recurrent compared with non-recurrent tumors ($p < 0.05$; Fig. 2E). Similarly, ANLN expression was markedly elevated in metastatic samples compared with non-metastatic samples ($p < 0.01$; Fig. 2F), suggesting a strong link between ANLN and both recurrence and metastasis in lung cancer. To further evaluate the prognostic relevance of ANLN, patients in the two survival cohorts (GSE157009 and GSE42127) were divided into ANLN-high and ANLN-low groups, and quartile-based Kaplan–Meier survival curves were generated (Fig. 2G–H). In both cohorts, patients with high ANLN expression exhibited significantly worse overall survival (hazard ratio [HR] = 1.7 and HR = 3.1, both $p < 0.05$), indicating that high ANLN expression is closely associated with poor prognosis in lung cancer.

ANLN-high epithelial cells are markedly enriched in lymph-node metastatic lesions and exhibit extensive cell type-specific transcriptional reprogramming.

To further dissect the role of ANLN in lung cancer metastasis, we analyzed the GSE277742 dataset, which contains single-cell RNA sequencing data (scRNA-seq) from lymph nodes of 18 lung cancer patients with lymph node metastasis (Met group) and 4 patients without metastasis (Non-Met group). After quality control, normalization, dataset integration, and principal component analysis (PCA), a total of 145,570 cells were retained for downstream analyses

Annotation of the integrated single-cell dataset identified eight major cell populations, including

Epithelial cells (EPCAM, KRT8, KRT18), Proliferation cells (MKI67, TOP2A), T cells (CD3D, CD3E), B cells (MS4A1, CD79A), Macrophages (C1QA), Monocytes (FCGR3B, CSF3R, CXCR2), Neutrophils (S100A8, S100A9), and Plasma cells (MZB1, JCHAIN) (Fig. 3A, 3F).

Examination of ANLN expression across different clinical groups (Fig. 3C) revealed that ANLN signals were predominantly localized to epithelial and proliferation-related clusters, with markedly stronger and broader expression in the Met group. This suggests that ANLN-high epithelial cells constitute the major tumor cell population within metastatic lymph nodes. Quantitative comparison of cellular composition between the Met and Non-Met groups (Fig. 3B, 3D) further showed that the Met group exhibited a substantial increase in epithelial and proliferating cells, whereas immune populations, including T cells and myeloid cells, were relatively enriched in the Non-Met group. These findings indicate a transition toward an epithelial-dominated cellular architecture in metastatic lymph nodes. We next performed differential expression analysis within each of the eight cell populations (Fig. 3E). The proliferation and epithelial compartment showed the largest and second largest number of differentially expressed genes, underscoring the pivotal role of epithelial cells in the metastatic progression of lung cancer.

Collectively, these results demonstrate that lung cancer lymph node metastasis is characterized by the accumulation of ANLN-high epithelial tumor cells and extensive transcriptional and compositional remodeling across multiple cell lineages, providing single-cell evidence for the central role of ANLN in metastatic dissemination.

High ANLN expression enhances metastatic pathways in epithelial cells during lymph node metastasis of lung cancer.

To further elucidate the functional role of ANLN within metastatic lymph nodes, we stratified the 18 metastatic samples into ANLN-high (n = 9) and ANLN-low (n = 9) groups based on ANLN expression levels and compared their single-cell characteristics. UMAP visualization (Fig. 4A) showed markedly elevated ANLN expression in the high expression group, with signals predominantly localized to epithelial and proliferation-associated cell populations.

Building on these observations, we next examined differential gene expression across the major single-cell compartments between the ANLN-high and ANLN-low groups (Fig. 4B). The largest numbers of DEGs were detected in the epithelial, proliferation, and macrophage clusters, indicating that high ANLN expression not only alters transcriptional programs within tumor epithelial cells but is also associated with substantial transcriptional changes in immune cells—particularly macrophages. Notable DEG alterations were also observed in T-cell and monocyte populations, suggesting that ANLN-associated remodeling exerts broad, multi-lineage effects across the metastatic tumor ecosystem.

To investigate ANLN-related regulatory pathways, KEGG and GO enrichment analyses were performed using DEGs derived from the comparison between ANLN-high and ANLN-low groups (Fig. 4C, 4D). KEGG analysis revealed significant enrichment in canonical oncogenic processes, including Cell cycle progression, Chromosome segregation, DNA replication, p53

signaling, Glycolysis/Gluconeogenesis, and Ubiquitin mediated proteolysis. GO analysis further highlighted enrichment in local adhesion, cell migration, double-strand break repair, epithelial cell migration, epithelial mesenchymal transition (EMT), and extracellular matrix organization, which are key biological programs implicated in tumor invasion and dissemination.

Collectively, these findings demonstrate that in lung cancer lymph node metastasis, high ANLN expression defines a highly proliferative and strongly migratory epithelial cell population, accompanied by widespread transcriptional alterations across multiple cell types. These results suggest that ANLN may promote metastatic progression by reprogramming tumor epithelial cells and reshaping their surrounding microenvironment.

CellChat analysis reveals enhanced intercellular communication strength and selective activation of metastasis associated signaling pathways in ANLN-high tumors.

To systematically investigate intercellular communication patterns associated with ANLN expression, we systematically compared the intercellular communication patterns between the ANLN-high and ANLN-low groups using CellChat analysis. Figures 5A–5D illustrate the global communication networks among the eight major cell populations in the two groups. The ANLN-high group exhibited a weaker overall interaction strength but markedly greater number of intercellular interactions than the ANLN-low group, indicating that ANLN-high metastatic lesions establish a more extensive but less intense communication network.

At the cell-type level, we assessed the differences in interaction patterns between the two groups (Fig. 5E, 5F). Figure 5E depicts the heatmap of differential interaction counts, and Figure 5F shows the corresponding differences in interaction strength. The results revealed that the incoming and outgoing signaling of epithelial, proliferation, and macrophage populations was substantially elevated in the ANLN-high group, suggesting that ANLN mediates coordinated remodeling across multiple cell types within the metastatic microenvironment.

We further compared the relative information flow of signaling pathways between ANLN-high and ANLN-low groups (Fig. 5G). Pathways including SPP1, VISFATIN, PARs, CCL, GDF, IL-1, GRN, TWEAK, CXCL, ANNEXIN, IL-16, BAFF, and GALECTIN were markedly enhanced in the ANLN-high group. These ligand–receptor axes are broadly implicated in inflammatory amplification, cellular migration, immune modulation, and extracellular matrix (ECM) remodeling, consistent with highly invasive biological programs associated with metastasis.

High ANLN expression is associated with enhanced epithelial EMT and migration related programs.

To further elucidate the functional programs associated with ANLN in metastatic epithelial cells, we first stratified epithelial cells from the lymph node single-cell dataset into ANLN expressing ($ANLN^+$) and ANLN non expressing ($ANLN^-$) populations. Differential expression analysis was performed between these two groups, followed by gene set enrichment analysis (GSEA). The results revealed that $ANLN^+$ epithelial cells were significantly enriched for pathways related to

cell proliferation, cell motility, and immune regulation compared with ANLN^{-} epithelial cells (Fig. 6A).

Based on the GSEA results, three functional module scores were constructed. Proliferation score was calculated using genes derived from the top six proliferation- and cell cycle–related pathways, Motility score was computed from genes associated with migration and cytoskeletal remodeling, and Immunity score was derived from immune-related pathways. Using these scores, metastatic samples were stratified into ANLN-high and ANLN-low groups at the sample level, and functional differences between the two groups were systematically evaluated. Compared with ANLN-low tumors, epithelial cells from ANLN-high tumors exhibited significantly elevated proliferation scores (Fig. 6B) and motility scores (Fig. 6C), whereas immunity scores were markedly reduced (Fig. 6D). To capture the integrated metastatic potential, a composite metastasis score was calculated by combining the proliferation and motility scores with the immunity score. Notably, the composite metastasis score was significantly higher in ANLN-high tumors (Fig. 6E), indicating a globally enhanced pro-metastatic transcriptional state.

To further assess these associations at a continuous single-cell level, correlation analyses were performed across the epithelial cells from metastatic samples. ANLN expression showed significant positive correlations with the proliferation score (Fig. 6F) and motility score (Fig. 6G), while displaying a negative correlation with the immunity score (Fig. 6H). In addition, ANLN expression was positively correlated with the composite metastasis score (Fig. 6I).

Collectively, these findings demonstrate that high ANLN expression is tightly linked to coordinated activation of proliferation- and migration-related programs and suppression of immune-associated transcriptional activity in epithelial cells, supporting a central role for ANLN in promoting epithelial plasticity and metastatic progression in lung cancer.

Pseudotime trajectory analysis of AT2-derived epithelial cells revealed that ANLN defines a critical transitional state.

To further elucidate the dynamic role of ANLN during metastatic progression, we performed subclustering and pseudotime trajectory analysis on epithelial cells from metastatic lymph nodes. UMAP visualization identified five major epithelial subpopulations, including AT1, AT2, basal, ciliated, and club cells (Fig. 7A). Expression of the AT2 marker SFTPA1 was predominantly enriched in the AT2 cluster (Fig. 7E). AT2 is the origin of lung cancer cells[27-29], so it is crucial to analyze the development of this subgroup.

Monocle-based trajectory inference revealed a continuous temporal progression of AT2 cells state along the principal lineage (Fig. 7B). Mapping ANLN expression onto this trajectory demonstrated that ANLN was not maximally expressed at the initial or terminal stages, but instead showed prominent enrichment at an intermediate region of the trajectory (Fig. 7C–D). Quantitative analysis further confirmed that ANLN expression increased in a stage-dependent manner and peaked during the intermediate pseudotime window (Fig. 7F), suggesting that ANLN marks a critical transitional phase during epithelial progression rather than a static lineage endpoint.

To characterize the functional features of this ANLN-high transitional state, we examined the dynamic changes of multiple programmatic scores along pseudotime. The proliferation score increased markedly at the intermediate stage (Fig. 7G), indicating activation of cell-cycle–related programs. In parallel, the motility score also peaked in the mid-trajectory region (Fig. 7H), reflecting enhanced cytoskeletal remodeling and migratory capacity. In contrast, the immunity score exhibited a decreasing trend during the intermediate phase (Fig. 7I), suggesting attenuation of immune-associated programs. Integration of these features into the composite metastasis score revealed a pronounced peak at the same intermediate pseudotime window (Fig. 7J), closely mirroring the ANLN expression pattern and defining a state characterized by high proliferative and migratory potential coupled with reduced immune-related features.

To exclude the possibility that this pattern merely reflects a generic proliferative phenotype, we further constructed a separate pseudotime trajectory within the proliferation cells belonging to the eight subtypes identified in cancer metastasis components (Fig. 7K, 3A). Notably, ANLN expression remained selectively enriched at specific mid-to-late stages of this trajectory (Fig. 7L), rather than being uniformly elevated across all proliferating cells, indicating that ANLN marks a more advanced, invasion-prone functional state.

Collectively, these pseudotime analyses demonstrate that ANLN is involved in regulating a critical intermediate state during the transition of AT2-derived epithelial cells toward a highly invasive phenotype. This transitional state is characterized by coordinated activation of

proliferation, motility, and metastasis-associated programs, accompanied by attenuation of immune-related features, providing dynamic trajectory-level evidence for the role of ANLN in promoting lymph node metastasis.

Independent single-cell validation confirms enrichment of ANLN-high invasive epithelial programs during lung cancer progression.

To independently validate the association between ANLN expression and invasive epithelial programs, we analyzed the single-cell RNA-seq dataset GSE266330, which profiles the transition from in situ lung cancer to invasive lung cancer. After standard preprocessing and clustering, cells were annotated into eight major populations based on canonical markers, including epithelial cells, T/NK cells, B cells, monocytes, endothelial cells, fibroblasts, and others (Fig. 8A,8K).

We focused on epithelial cells and compared their distribution between in situ and invasive tumors (Fig. 8B). Feature plots show sparse expression of ANLN in situ lesions, whereas invasive tumors exhibit stronger and more widespread ANLN expression within epithelial clusters (Fig. 8C–D). Quantitative analysis further demonstrated that the fraction of ANLN positive epithelial cells was significantly higher in invasive tumors than in situ tumors ($p < 0.01$; Fig. 8E), indicating enrichment of ANLN high epithelial cells during invasive progression.

To validate the functional scoring framework we defined, we applied the same Proliferation, Motility, Immunity, and Metastasis scores to epithelial cells in the GSE266330 dataset. Using these predefined gene sets, epithelial cells from invasive tumors exhibited significantly higher

Proliferation and Motility scores with reduced Immunity scores, compared with those from in situ tumors (Fig. 8F-H). Consistent with the coordinated behavior of these gene sets observed previously, the Metastasis score—integrating Proliferative, Migratory, and Immune scores—was markedly elevated in the invasive group (Fig. 8I). Moreover, correlation analysis across all epithelial cells demonstrated a significant positive association between ANLN expression and the metastasis score (Fig. 8J), confirming that the ANLN linked metastatic program we identified is reproducibly captured in an independent single-cell dataset, reinforcing the role of ANLN as a marker and potential driver of invasive epithelial states during lung cancer progression.

4. DISCUSSION

In recent years, the rapid development of multi-omics analyses has greatly advanced our understanding of tumor progression mechanisms and prognostic evaluation. In this study, by integrating multi-cohort transcriptomic data with single-cell analyses, we revealed the critical role of ANLN in lung cancer recurrence and metastasis. ANLN encodes an actin-binding protein that plays an essential role in cell growth, migration, and cytoskeletal dynamics[30]. Although ANLN has been reported to be associated with cell proliferation in multiple tumor types[31-32], its precise biological function and role in lung cancer metastasis remain insufficiently characterized. Herein we performed an integrated analysis combining survival, recurrence, and metastasis phenotypes to explore the potential regulatory role of ANLN in malignant lung cancer progression.

We found that ANLN expression was consistently elevated in both recurrent and metastatic samples. In addition, high ANLN expression was significantly associated with poor overall

survival. This poor overall survival indicates that ANLN may contribute to the formation of an aggressive tumor phenotype.

Based on our results, it seems that the profound impact of ANLN on the TME is through increased epithelial cell proportions and relative depletion of immune and stromal components. Our data further indicate the influence of ANLN on the TME is through predominant enrichment in epithelial and proliferating cell populations, which are accompanied by marked alterations in cellular composition and transcriptional programs. CellChat analysis further showed that ANLN-high tumors exhibited more complex intercellular communication networks and selectively activated multiple signaling pathways related to inflammatory regulation, cell migration, and extracellular matrix remodeling, suggesting that ANLN may promote metastatic progression by reshaping intercellular communication patterns. Consistently, subsequent GSEA and GO functional enrichment analyses revealed that genes associated with high ANLN expression were significantly enriched in pathways related to cell cycle regulation, epithelial–mesenchymal transition (EMT), cell migration, and extracellular matrix remodeling. These processes are widely recognized as fundamental molecular mechanisms underlying tumor metastasis, further supporting the hypothesis that ANLN plays a promotive role in lung cancer progression.

By integrating functional scoring analyses with pseudotime trajectory inference, we further demonstrated that ANLN is not merely a proliferation marker, but a critical intermediate state. Such state is the transition of AT2-derived epithelial cells toward a highly invasive phenotype.

This transitional state is characterized by concurrently enhanced proliferative and migratory capacities, accompanied by suppression of immune-related programs.

In summary, this study reveals the pivotal role of ANLN in lung cancer recurrence and lymph node metastasis from a multi-level and multi-scale perspective and proposes that ANLN may drive lung cancer progression toward a highly invasive state by coordinating tumor cell–intrinsic transcriptional programs with microenvironmental remodeling. Although further functional experiments are required to establish causal relationships, these findings provide new insights into the mechanisms of lung cancer metastasis and suggest that ANLN holds potential value as a prognostic biomarker and therapeutic target.

Limitations

A major limitation of this study is that all analyses were conducted using publicly available bulk and single-cell transcriptomic datasets. Although multiple independent bulk cohorts and two independent single-cell datasets were included for survival analysis, differential expression analysis, and reproducibility validation, the overall sample size was constrained by the availability of eligible datasets in public repositories. Specifically, 425 tumor samples were used for survival analysis, 344 samples were included for recurrence and metastasis related differential expression analysis, and two independent single-cell RNA-seq datasets comprising 31 samples and approximately 230,000 cells in total were analyzed.Therefore, potential selection bias and residual heterogeneity across datasets cannot be fully excluded.

In parallel, the conclusions of this study are primarily based on computational and statistical analyses. While an independent single-cell dataset was used for reproducibility validation, no additional experimental validation was performed. As a result, the findings should be interpreted as hypothesis-generating rather than definitive evidence of causal biological mechanisms. Furthermore, although single-cell transcriptomic data enabled cell-type-resolved analyses, the number of samples with detailed clinical annotations was limited, which may restrict the generalizability of the results across diverse patient populations. Future studies incorporating larger cohorts, and experimental validation will be necessary to further substantiate and extend the findings reported in this work.

FIGURES AND FIGURE LEGENDS

Figure 1

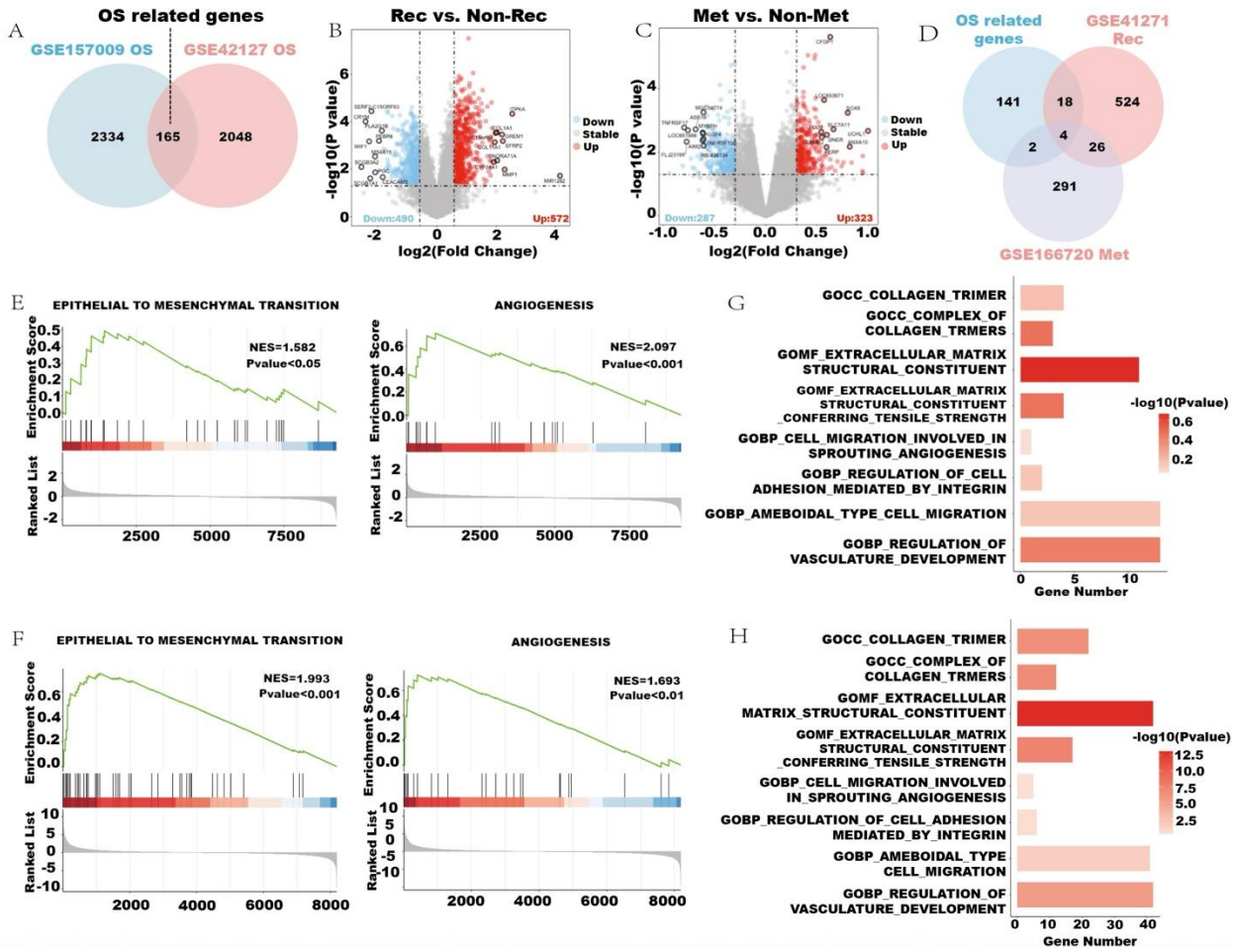
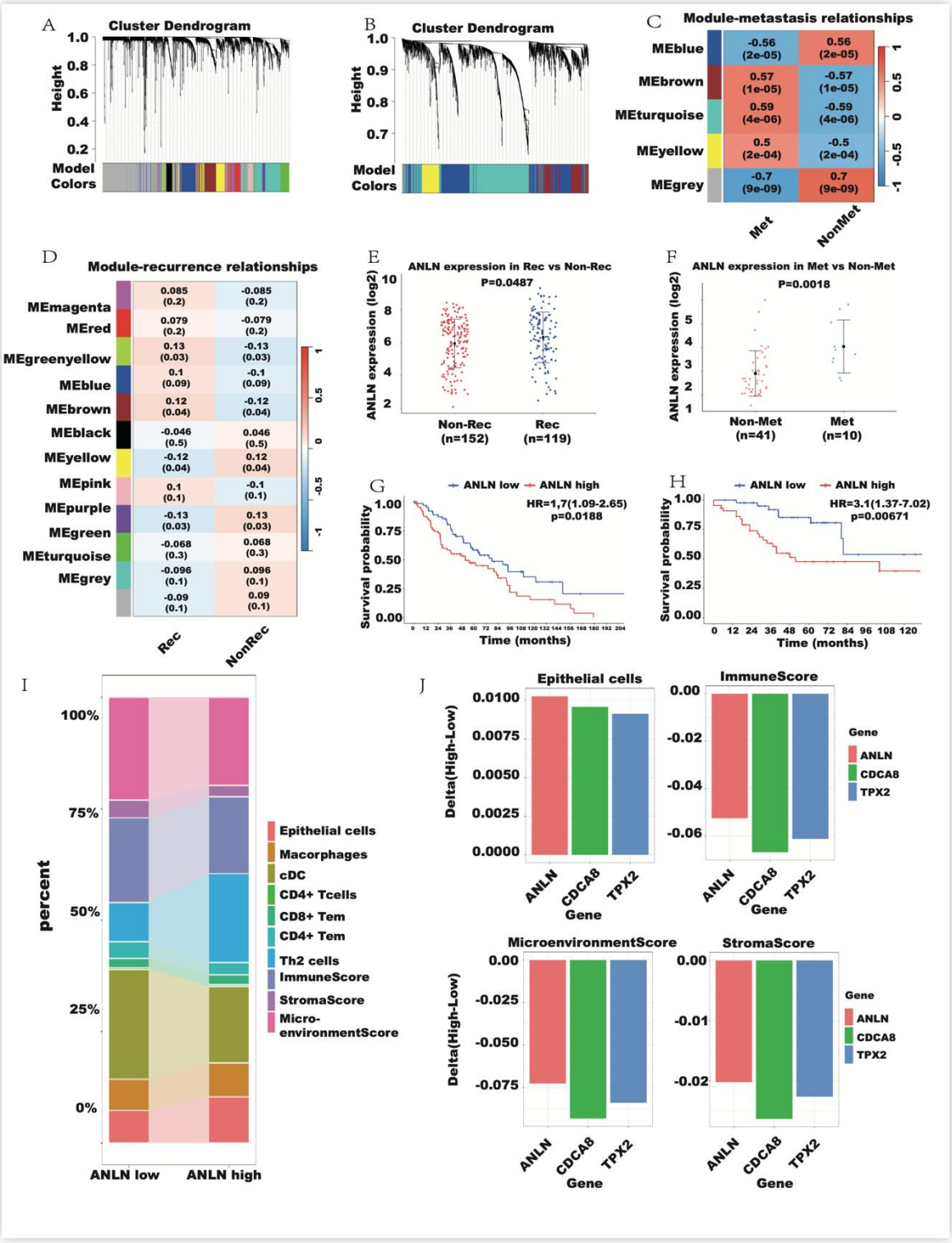


Figure 1. An integrated analysis of survival, recurrence, and metastasis identifies a core gene set critical for lung cancer progression. (A) Venn diagram illustrating the overlap of OS-associated genes from GSE157009 and GSE42127, with 165 genes shared between two cohorts. (B) Volcano plot showing differentially expressed genes between recurrence and non-recurrence samples in GSE157009 (572 upregulated). (C) Volcano plot showing differentially expressed genes between metastatic and non-metastatic samples in GSE42127 (323 upregulated). (D) Venn

diagram integrating OS, recurrence, and metastasis-associated genes, revealing four overlapping core genes. (E) GSEA of recurrence-associated upregulated genes demonstrating enrichment of EMT and Angiogenesis pathways. (F) GSEA of metastasis-associated upregulated genes showing similar enrichment of EMT and Angiogenesis. (G–H) GO enrichment analyses showing ECM remodeling, adhesion, migration, and angiogenesis pathways enriched in both recurrence (G) and metastasis (H) gene sets.

510 **Figure 2**



511

512 **Figure 2.** (A, B) WGCNA for recurrence and metastasis cohorts showing co-expression module
513 structure. (C) Module–metastasis correlations identify the turquoise module as strongly
514 associated with metastasis, with ANLN, TPX2, and CDCA8 included. (D) Module–recurrence
515 correlations identify the brown module as significantly associated with recurrence, with ANLN,
516 TPX2, and CDCA8 included. (E, F) ANLN expression is increased in recurrence (E) and
517 metastasis samples (F). (G, H) Kaplan–Meier curves demonstrate significantly worse survival in
518 ANLN-high patients in both cohorts. (I) xCell shows epithelial enrichment and immune、
519 stromal depletion in ANLN-high tumors. (J) Comparative microenvironmental changes across
520 three candidate genes further support ANLN as the key representative progression gene.

521

522

Figure 3

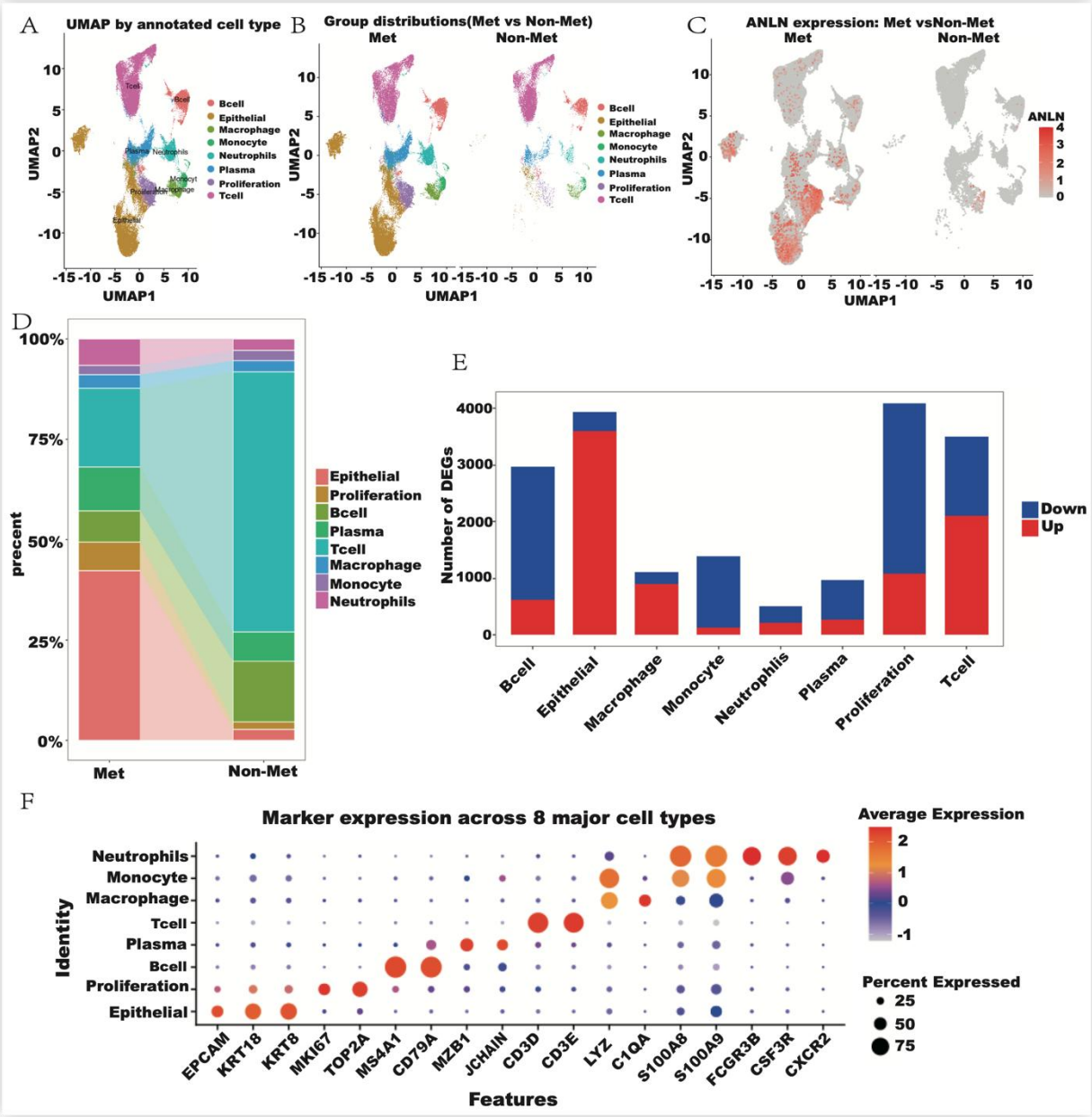


Figure 3. (A) UMAP visualization of 145,570 integrated lymph-node cells from metastatic (Met) and non-metastatic (Non-Met) lung cancer samples, identifying eight major cell types: Epithelial, Proliferation, T cells, B cells, Macrophages, Monocytes, Neutrophils, and Plasma cells. (B) Comparison of cell-type composition shows a marked increase in epithelial and proliferating

529 tumor cells in metastatic lymph nodes, whereas immune lineages (T cells, myeloid cells) are
530 proportionally enriched in non-metastatic samples. (C) ANLN expression map indicates that
531 ANLN is predominantly expressed in epithelial and proliferative compartments, with
532 substantially stronger and broader expression in metastatic lesions. (D) Quantitative analysis of
533 population shifts further demonstrates epithelial and proliferation expansion and immune
534 depletion in metastatic samples. (E) Differential expression analysis across eight cell types
535 reveals that epithelial cells harbor the largest number of DEGs between Met and Non-Met
536 samples, highlighting epithelial transcriptional rewiring as a major driver of lymph-node
537 metastasis. (F) Dot-plot of canonical markers validates the identities of the eight annotated cell
538 populations supporting downstream analyses.
539
540

Figure 4

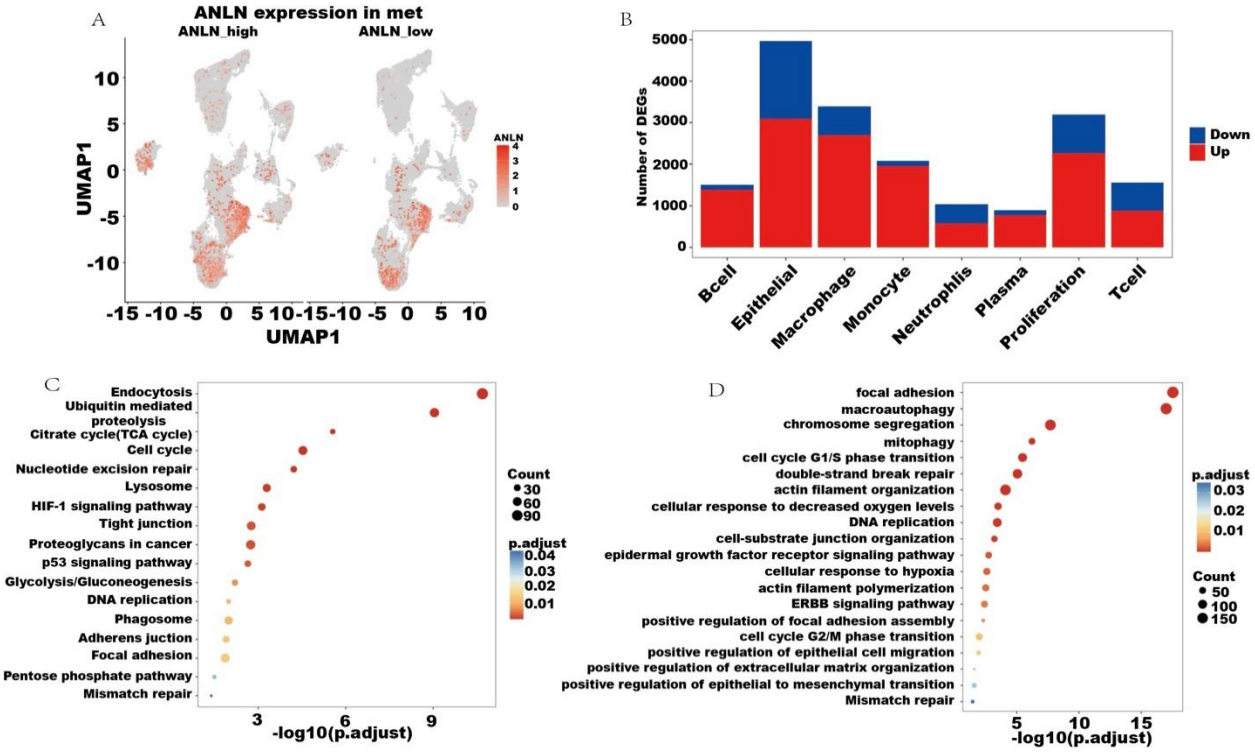
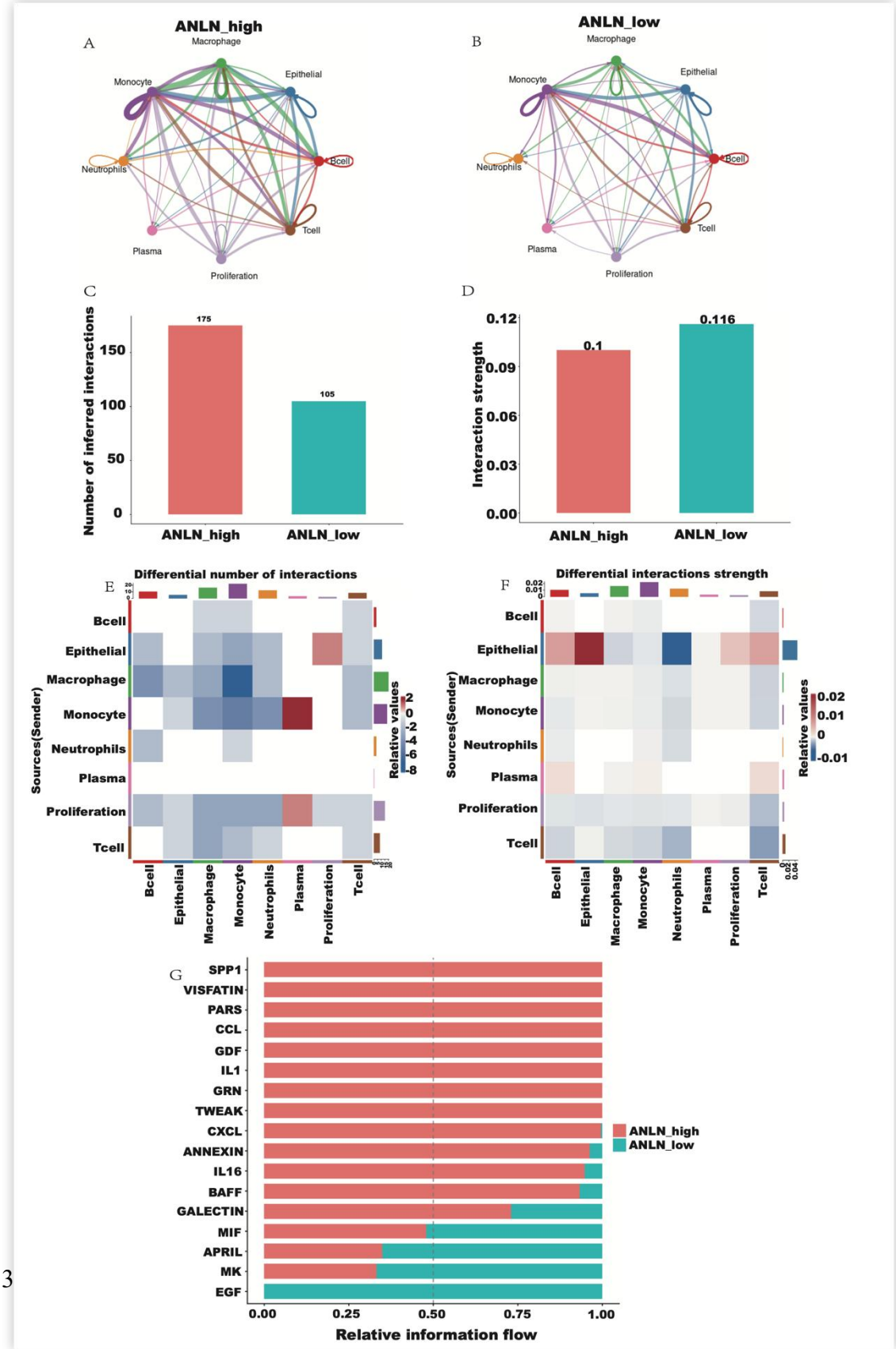


Figure 4. (A) UMAP visualization shows markedly elevated ANLN expression in epithelial and proliferating cells in the ANLN-high group. (B) Differential expression analysis reveals the largest DEG changes in epithelial, proliferative, and macrophage populations, with additional remodeling in T cells and monocytes. (C) KEGG enrichment highlights activation of cell cycle, DNA replication, p53 signaling, and metabolic pathways in ANLN-high tumors. (D) GO analysis identifies enrichment of adhesion, migration, cytoskeletal remodeling, DNA repair, and EMT-related processes.

552 **Figure 5**



3

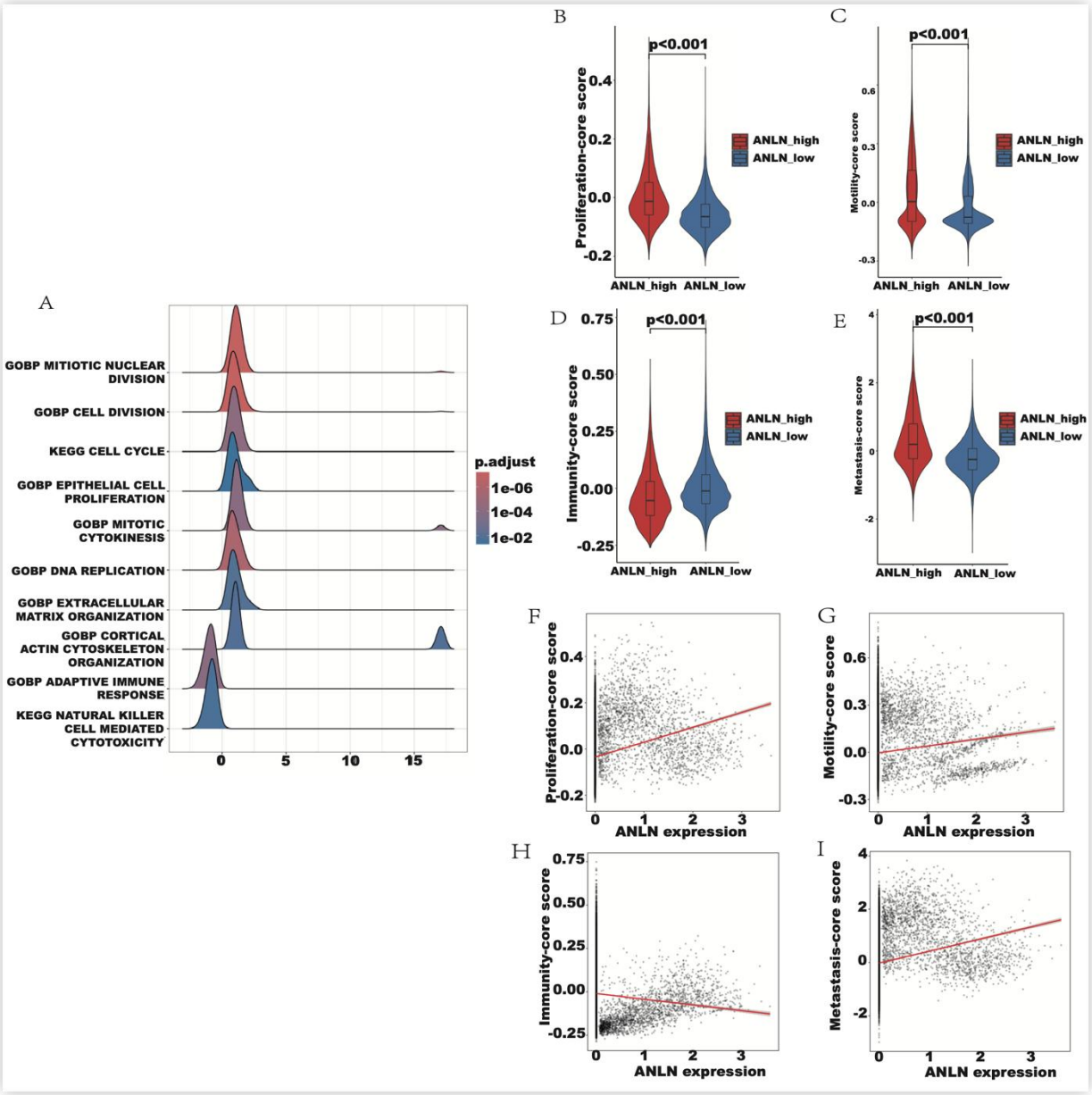
553

554 **Figure 5.** ANLN-high metastatic tumors exhibit expanded intercellular communication networks
555 and enhanced pro-metastatic signaling programs. (A–D) CellChat analysis shows that ANLN-
556 high samples possess markedly increased numbers of cell–cell interactions compared with
557 ANLN-low tumors, forming a weaker but more extensive communication network across the
558 eight major cell populations. (E, F) Differential heatmaps interaction number (E) and interaction
559 strength (F) indicate that epithelial, proliferating, and macrophage lineages. (G) Relative
560 information flow analysis reveals those pro-metastatic pathways—including SPP1, VISFATIN,
561 PARs, CCL, GDF, IL-1, GRN, TWEAK, CXCL, ANNEXIN, IL-16, BAFF, and GALECTIN—
562 are strongly elevated in ANLN-high tumors,reflecting enhanced inflammation, migration,
563 immune modulation, and ECM remodeling.

564

565

566 **Figure 6**



567

568

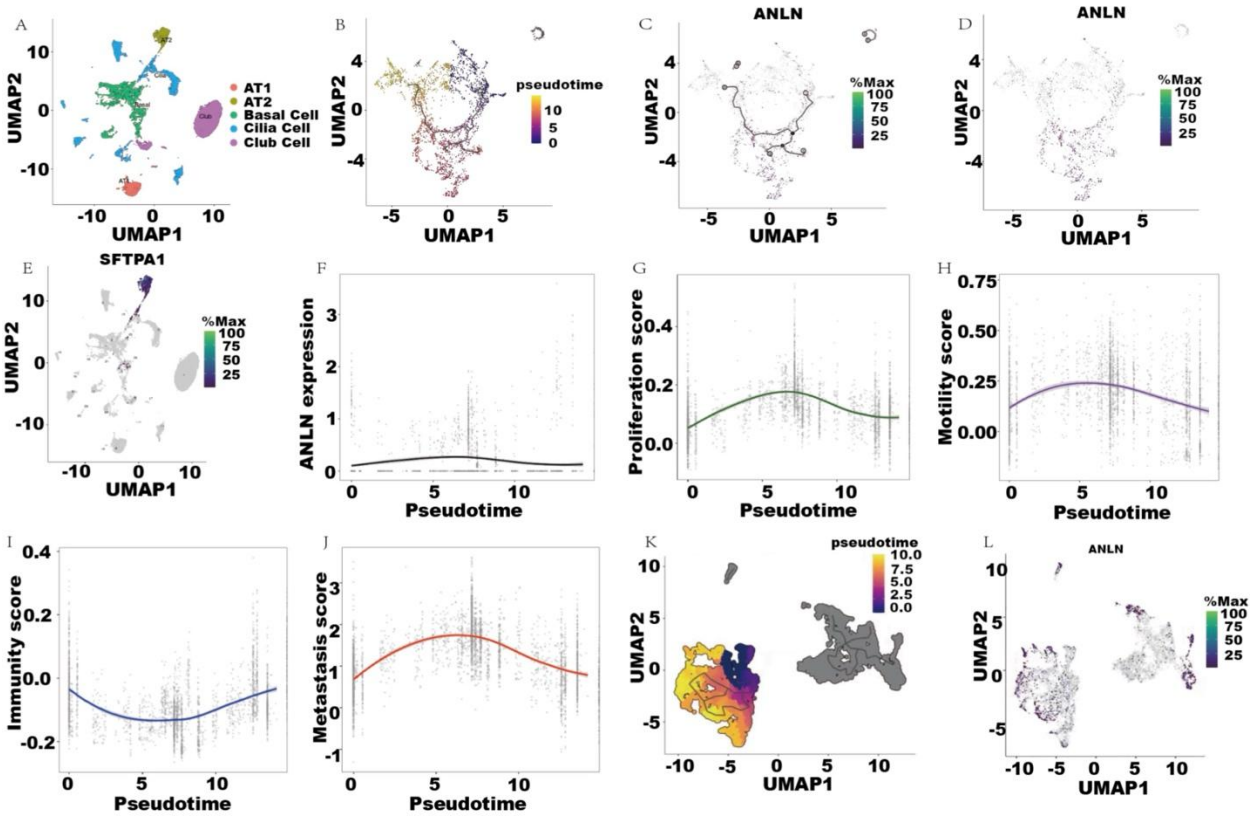
569 **Figure 6.** (A) GSEA plot comparing ANLN+ vs ANLN- epithelial cells, showing enrichment of

570 Proliferation, Motility, and Immune related pathways used to construct three module scores. (B)

571 Proliferation score is significantly higher in ANLN high samples. (C) Motility score is

572 significantly elevated in ANLN high samples. (D) Immunity score is significantly lower in

573 ANLN high samples. (E) Metastasis score is markedly increased in ANLN high tumors. (F)
574 Single-cell correlation showing ANLN expression positively correlates with proliferation score.
575 (G) ANLN expression positively correlates with motility score. (H) ANLN expression negatively
576 correlates with immunity score. (I) ANLN expression positively correlates with the composite
577 metastasis score.
578
579 **Figure 7**



580

581 **Figure 7**. (A) UMAP of epithelial subclusters (AT1, AT2, basal, cilia, club) from metastatic
582 lymph nodes. (B) Monocle-inferred pseudotime trajectory of epithelial cells. (C–D) ANLN
583 expression overlaid on the trajectory and UMAP, showing enrichment at an intermediate
584 pseudotime region. (E) Feature plot of the AT2 marker SFTPA1. (F) ANLN expression
585 dynamics along pseudotime. (G–J) Program score changes along pseudotime: Proliferation (G),

586 Motility (H), Immunity (I), and Metastasis score (J). (K) Pseudotime trajectory reconstructed
587 within proliferation cells. (L) UMAP of ANLN expression overlaid on the proliferating-lineage
588 trajectory.

589
590

Figure 8

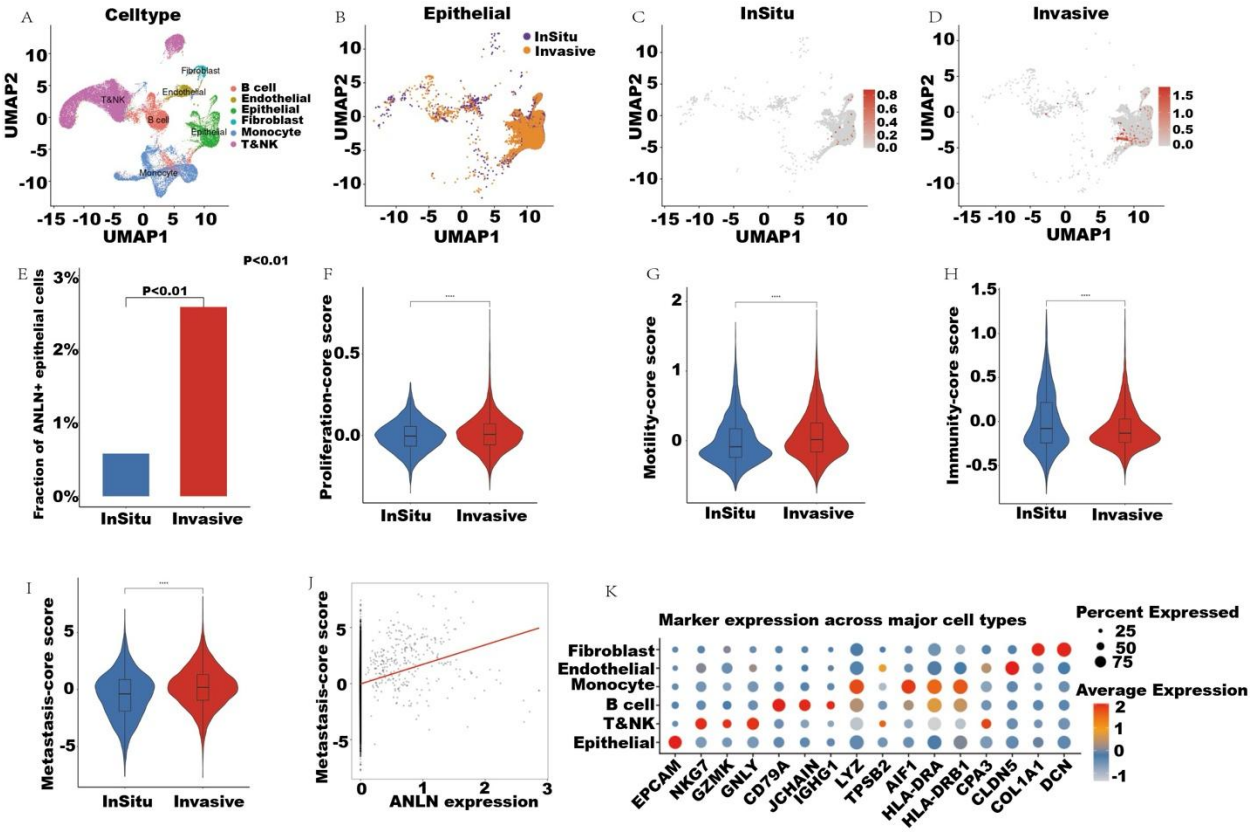


Figure 8. (A) UMAP of GSE266330 showing eight major cell types annotated by canonical markers. (B) UMAP of epithelial cells colored by stage (in situ vs invasive). (C–D) Feature plots of ANLN expression in epithelial cells from in situ (C) and invasive (D) tumors. (E) Fraction of ANLN-positive epithelial cells in in situ and invasive tumors. (F–H) Violin plots comparing Proliferation (F), Motility (G), and Immunity (H) scores in epithelial cells between groups. (I) Violin plot of the composite metastasis score in epithelial cells. (J) Correlation between ANLN expression and metastasis score across epithelial cells. (K) Dotplot of canonical marker expression validating cell-type annotation.

AUTHOR CONTRIBUTIONS

H.Q and Z.X performed the research and drafting the manuscript. Z.W conceived and mentored this project, as well as prepared the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

All data used in this study are publicly available. Bulk RNA sequencing datasets used for survival and differential expression analyses were obtained from the Gene Expression Omnibus (GEO) database under the accession numbers GSE157009(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157009), GSE42127(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42127), GSE41271(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41271), and GSE166720 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166720). Single-cell RNA sequencing datasets used for primary analysis and independent validation were also retrieved from GEO under the accession numbers GSE277742(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE277742) and GSE189357(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE189357).Gene set enrichment analyses were performed using gene sets from the Molecular Signatures Database (MSigDB, version 7.5.1), including the Hallmark, Gene Ontology (GO), and KEGG collections (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp).Cell–cell communication analyses were conducted using curated ligand–receptor interaction

databases implemented in CellChat (https://github.com/sqjin/CellChat).No newly generated datasets were produced in this study.

FUNDING INFORMATION

This work was supported by the National Key R&D Program of China (2023YFA0915700) to ZW.

Ethics, Consent to Participate, and Consent to Publish

Ethics approval and consent to participate were not required for this study, as all data analysed were obtained from publicly available databases and de-identified prior to use. Consent for publication is not applicable.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

REFERENCE

1. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using cell chat. Nat Commun. 2021;12(1):1088. https://pubmed.ncbi.nlm.nih.gov/33597522/.

2. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304.https://pubmed.ncbi.nlm.nih.gov/37231261/.

3. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods. 2019;16(12):1289–96. https://pubmed.ncbi.nlm.nih.gov/31740819/.

4. Chen Z, Wang X, Wu H, Fan Y, Yan Z, Lu C, et al. X-box binding protein 1 as a key modulator in healing endothelial cells, a novel EC phenotype promoting angiogenesis after MCAO. Cell Mol Biol Lett. 2022;27(1):97.https://pubmed.ncbi.nlm.nih.gov/36348288/.

5. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin.74, 229–263 (2024). https://pubmed.ncbi.nlm.nih.gov/38572751/.

6. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin 2022;72:7-33. http://www.ncbi.nlm.nih.gov/pubmed/35020204.

7. Noroxe, D. S. & Sorensen, J. B. Ultra-late relapse with a single cerebellar metastasis 10 years after complete surgery for stage IIA non-small cell lung cancer (bronchioalveolar carcinoma). J. Thorac. Oncol. 7, 764–765 (2012). http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=22425928.

8. Altorki, N. et al. Lobar or sublobar resection for peripheral stage IA non-small-cell lung cancer. N. Engl. J. Med.388, 489–498 (2023). https://pubmed.ncbi.nlm.nih.gov/36780674/.

9. Mahvi, D. A., Liu, R., Grinstaff, M. W., Colson, Y. L. & Raut, C. P. Local cancer recurrence: the realities, challenges, and opportunities for new therapies. CA Cancer J. Clin.68, 488–505 (2018). https://pubmed.ncbi.nlm.nih.gov/30328620/.

10. Martini, N. et al. Factors influencing ten-year survival in resected stages I to IIIa non-small cell lung cancer. J. Thorac. Cardiovasc. Surg. 117, 32–36 (1999).http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=9869756.

11. Maeda, R. et al. Long-term outcome and late recurrence in patients with completely resected stage IA non-small cell lung cancer. J. Thorac. Oncol. 5, 1246–1250 (2010).http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=20559152.

12. Chen, K. et al. Spatiotemporal genomic analysis reveals distinct molecular features in recurrent stage I non-small cell lung cancers. Cell Rep.40, 111047 (2022). https://pubmed.ncbi.nlm.nih.gov/35830809/.

13. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. Nat Med. (2006) 12:895–904. https://pubmed.ncbi.nlm.nih.gov/16892035/.

14. Marx V. Tracking metastasis and tricking cancer. Nature. 2013;494(7435):133–136. doi: 10.1038/494131a. https://pubmed.ncbi.nlm.nih.gov/23389545/.

15. Gerstberger S, Jiang Q, Ganesh K. Metastasis. Cell. (2023) 186:1564–79. https://pubmed.ncbi.nlm.nih.gov/37059065/.

16. Husemann, Y. et al. Systemic spread is an early step in breast cancer. Cancer Cell 13, 58–68 (2008).http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=18167340.

17. Galassi C, Chan TA, Vitale I, Galluzzi L. The hallmarks of cancer immune evasion. Cancer Cell. 2024 Nov 11; https://pubmed.ncbi.nlm.nih.gov/39393356/.

18. Spano D, Zollo M. Tumor microenvironment: a main actor in the metastasis process. Clin Exp Metastasis. 2012;29(4):381–395. doi: 10.1007/s10585-012-9457-5. [DOI] [PubMed] [Google Scholar] [Ref list] https://pubmed.ncbi.nlm.nih.gov/22322279/.

19. Rhim, A. D. et al. EMT and dissemination precede pancreatic tumor formation. Cell 148, 349–361 (2012). http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=22265420.

20. Li, Y., Liu, F., Cai, Q. et al. Invasion and metastasis in cancer: molecular insights and therapeutic targets. Sig Transduct Target Ther 10, 57 (2025). https://doi.org/10.1038/s41392-025-02148-4.

21. Lau SCM, Pan Y, Velcheti V, Wong KK. Squamous cell lung cancer: Current landscape and future therapeutic options. Cancer Cell. 2022 Nov 14;40(11):1279-1293. doi: 10.1016/j.ccell.2022.09.018. Epub 2022 Oct 20. PMID: 36270277. https://pubmed.ncbi.nlm.nih.gov/36270277/.

22. Alexandrov LB et al. Signatures of mutational processes in human cancer. Nature 500, 415–21 (2013). https://pubmed.ncbi.nlm.nih.gov/23945592/

23. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012 Sep 27;489(7417):519-25. doi: 10.1038/nature11404. Epub 2012 Sep 9. Erratum in: Nature. 2012 Nov 8;491(7423):288. Rogers, Kristen. PMID: 22960745; PMCID: PMC3466113.

24. Mainardi, S. et al. Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. Proc. Natl Acad. Sci. USA111, 255–260 (2014). https://pubmed.ncbi.nlm.nih.gov/24367082/.

25. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature587, 619–625 (2020). https://pubmed.ncbi.nlm.nih.gov/33208946/.

26. Han, G. et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. Nature627, 656–663 (2024). https://pubmed.ncbi.nlm.nih.gov/38418883/.

27. Mainardi, S. et al. Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. Proc. Natl Acad. Sci. USA111, 255–260 (2014). https://pubmed.ncbi.nlm.nih.gov/24367082/.

28. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature587, 619–625 (2020). https://pubmed.ncbi.nlm.nih.gov/33208946/

29. Han, G. et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. Nature627, 656–663 (2024). https://pubmed.ncbi.nlm.nih.gov/38418883/.

30. Cao YF, Wang H, Sun Y, Tong BB, Shi WQ, Peng L, Zhang YM, Wu YQ, Fu T, Zou HY, Zhang K, Xu LY, Li EM. Nuclear ANLN regulates transcription initiation related Pol II clustering and target gene expression. Nat Commun. 2025 Feb 2;16(1):1271. doi: 10.1038/s41467-025-56645-9. PMID: 39894879; PMCID: PMC11788435. https://pmc.ncbi.nlm.nih.gov/articles/PMC11788435/.

31. Wang A, Dai H, Gong Y, Zhang C, Shu J, Luo Y, Jiang Y, Liu W, Bie P. ANLN-induced EZH2 upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis. J Exp Clin Cancer Res. 2019 Aug 8;38(1):347. doi: 10.1186/s13046-019-1340-7. PMID: 31395079; PMCID: PMC6686567. https://pmc.ncbi.nlm.nih.gov/articles/PMC6686567/.

32. Piekny AJ, Maddox AS. The myriad roles of Anillin during cytokinesis. Semin Cell Dev Biol. 2010 Dec;21(9):881-91. doi: 10.1016/j.semcdb.2010.08.002. Epub 2010 Aug 21. PMID: 20732437. https://pubmed.ncbi.nlm.nih.gov/20732437/.