

<https://doi.org/10.1038/s41698-025-01027-z>

# Integrative multi-omics and machine learning reveal critical functions of proliferating cells in prognosis and personalized treatment of lung adenocarcinoma

Check for updates

Shun Wang<sup>1,5</sup>, Dingtao Hu<sup>2,5</sup>, Ruohuang Wang<sup>3,5</sup>, Jie Huang<sup>1</sup> ✉ & Baoqing Wang<sup>1,4</sup> ✉

Lung adenocarcinoma (LUAD) is a major cause of cancer-related mortality globally. Proliferating cells, crucial components of the tumor immune microenvironment (TIME), play a significant role in cancer progression and immunotherapy response. Herein, we utilized multi-omics data and employed a multifaceted approach to delineate the proliferating cell landscape in LUAD. The Scissor algorithm was applied to identify Scissor+ proliferating cell genes associated with prognosis. An integrative machine learning program, comprising 111 algorithms, was developed to construct a Scissor+ proliferating cell risk score (SPRS). The SPRS model demonstrated superior performance in predicting prognosis and clinical outcomes compared to 30 previously published models. The role of SPRS and five pivotal genes in immunotherapy response was evaluated, and their expression was experimentally verified. Multifactorial analysis confirmed SPRS as an independent prognostic factor affecting LUAD patient survival. High- and low-SPRS groups exhibited different biological functions and immune cell infiltration in the TIME. High SPRS patients showed resistance to immunotherapy but increased sensitivity to chemotherapeutic and targeted therapeutic agents. Our study elucidates the dynamics of proliferating cells in LUAD, enhancing prognostic accuracy and highlighting the potential of SPRS and its constituent genes for personalized therapeutic interventions.

Globally, lung cancer remains the leading cause of cancer-related death and the second most commonly diagnosed cancer<sup>1</sup>, with lung adenocarcinoma (LUAD) being the most prevalent histological subtype<sup>2</sup>. Despite significant advancements in diagnostic techniques and treatment strategies of LUAD, the patient outcomes remain frustratingly poor, with five-year overall survival still stagnating below 20%. Chemotherapy is the basis of LUAD treatment. However, its toxic side effects make it intolerable for many patients who cannot benefit from it<sup>4,5</sup>. The advent of molecular targeted therapy and immunotherapy has reshaped the therapeutic landscape for LUAD patients; however, only less than 20% of patients harbor drug-

sensitive mutations<sup>6</sup>. Therefore, there is an urgent need for further research into biomarkers that can predict the efficacy of chemotherapy, targeted and immunotherapy for LUAD to improve the long-term clinical outcomes of LUAD patients<sup>7</sup>.

LUAD is characterized by strong immunogenicity, and the imbalance of tumor immune microenvironment (TIME) is one of the important manifestations of LUAD<sup>8</sup>. The cross-talk and interaction between cancer cells and the immune microenvironment play essential roles in cancer progression, metastatic spread, immune evasion, as well as drug resistance<sup>9</sup>. Within this milieu, proliferating cells, including lymphocytes, myeloid cells,

<sup>1</sup>Department of Respiratory Medicine, Shanghai Xuhui Central Hospital, Zhongshan-Xuhui Hospital, Fudan University, Shanghai, China. <sup>2</sup>Department of Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei, China. <sup>3</sup>Department of Otolaryngology, The Second Affiliated Hospital of the Naval Military Medical University (Shanghai Changzheng Hospital), Shanghai, China. <sup>4</sup>Department of Respiratory Medicine, Zhongshan Hospital, Fudan University, Shanghai, China. <sup>5</sup>These authors contributed equally: Shun Wang, Dingtao Hu, Ruohuang Wang. ✉ e-mail: [syarhuang@163.com](mailto:syarhuang@163.com); [Wang.baoqing@zs-hospital.sh.cn](mailto:Wang.baoqing@zs-hospital.sh.cn)

cancer cells, and stromal cells, collectively drive critical pathological processes such as tumor growth, immune evasion, and therapy resistance<sup>10</sup>. While the heterogeneity of these proliferating cell populations is undeniable, converging evidence suggests that their dynamic and coordinated behavior exerts a profound, integrated influence on cancer progression and patient outcomes<sup>11</sup>. Importantly, it is now recognized that the proliferative program is a shared feature among diverse cell types under stress or activation within the TME, linking cell cycle states to not only malignant expansion but also immune modulation and stromal reprogramming<sup>12,13</sup>.

Despite significant progress in cancer biology, the detailed profiling of proliferating cells in lung diseases remains underexplored. The interactions between these proliferating cells and the immune system are intricate, with profound implications for disease prognosis and treatment outcomes that are only beginning to be fully recognized. Understanding these dynamics can enlighten us about their prognostic significance and ability to guide targeted therapeutic approaches. Single-cell RNA sequencing (scRNA-seq) serves as a potent technique for probing cellular heterogeneity, discerning distinct cell states, identifying marker genes, and elucidating associated functions, thereby facilitating the development of personalized therapeutic strategies<sup>10,14</sup>. Additionally, scRNA-seq analysis of immune cells within the tumor microenvironment (TME) has elucidated the molecular characteristics of these cells, offering new insights into cancer immunity<sup>15,16</sup>. Multi-omics methodologies, including bulk and sc-RNA sequencing coupled with spatial transcriptomics (ST), offer a promising avenue for dissecting the complexity of the TME and the roles of proliferating cells<sup>14,17,18</sup>. These advanced methodologies offer an unprecedented opportunity to map the single-cell and spatial resolution of lung diseases, facilitating the understanding of how the cellular heterogeneity of proliferating cells contributes to tumor progression and response to therapy.

In this study, we analyzed scRNA-seq data from multiple samples spanning normal lung tissue to LUAD and observed that proliferating cells were significantly enriched in idiopathic pulmonary fibrosis (IPF) and LUAD tissues compared to chronic obstructive pulmonary disease (COPD) and normal tissues. Using the Scissor algorithm, we identified 22 Scissor+ proliferating cell genes with significant prognostic implications for LUAD patients. We then employed 111 machine learning combinations to develop the Scissor+ proliferating cell risk score (SPRS) to elucidate these model genes and construct the predictive model. Furthermore, we investigated the intricate interactions between model genes, SPRS levels, and the immunological features within LUAD. Additionally, we examined and validated the therapeutic response of these model genes and SPRS to immunotherapy and chemotherapy, aiming to provide further insights into personalized medicine for LUAD.

## Results

### Deciphering lung disease progression through scRNA profiling

To investigate the microenvironmental landscape across the spectrum of human lung disease stages, we performed a comprehensive scRNA-seq analysis on a diverse set of 93 samples including healthy lung tissue from 28 individuals, COPD from 18, IPF from 32, and LUAD from 15 subjects (Fig. 1a). After rigorous quality control and the meticulous exclusion of doublets, 368,904 cells remained for detailed scrutiny. To mitigate and correct for potential batch effects among the samples (Fig. S1a), we employed harmony analysis, a sophisticated computational method. This was followed by a series of analytical techniques including principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) to reduce dimensions and cluster the data. Our unsupervised clustering strategy successfully identified 24 distinct cell clusters (Fig. S1b), with no significant batch effects discernible across the diverse samples (Fig. S1a).

Further analysis annotating these clusters into various cell types, including T cells, natural killer (NK) cells, neutrophils, macrophages, monocytes, dendritic cells (DCs), mast cells, epithelial cells, proliferating cells, endothelial cells, lung-associated fibroblasts (LAFs), B and plasma B cells. This classification was based on the characteristic expression profiles of canonical marker genes (Fig. 1c–e, Fig. S1b, c, and Table S3). Our findings

were corroborated by cell ratio and Ro/e algorithm analyses, which highlighted a notable enrichment pattern of proliferating cells specifically within the IPF and LUAD groups (Fig. 1f, g and Fig. S1d). This observation provides critical insights into the cellular dynamics of proliferating cells during lung disease progression.

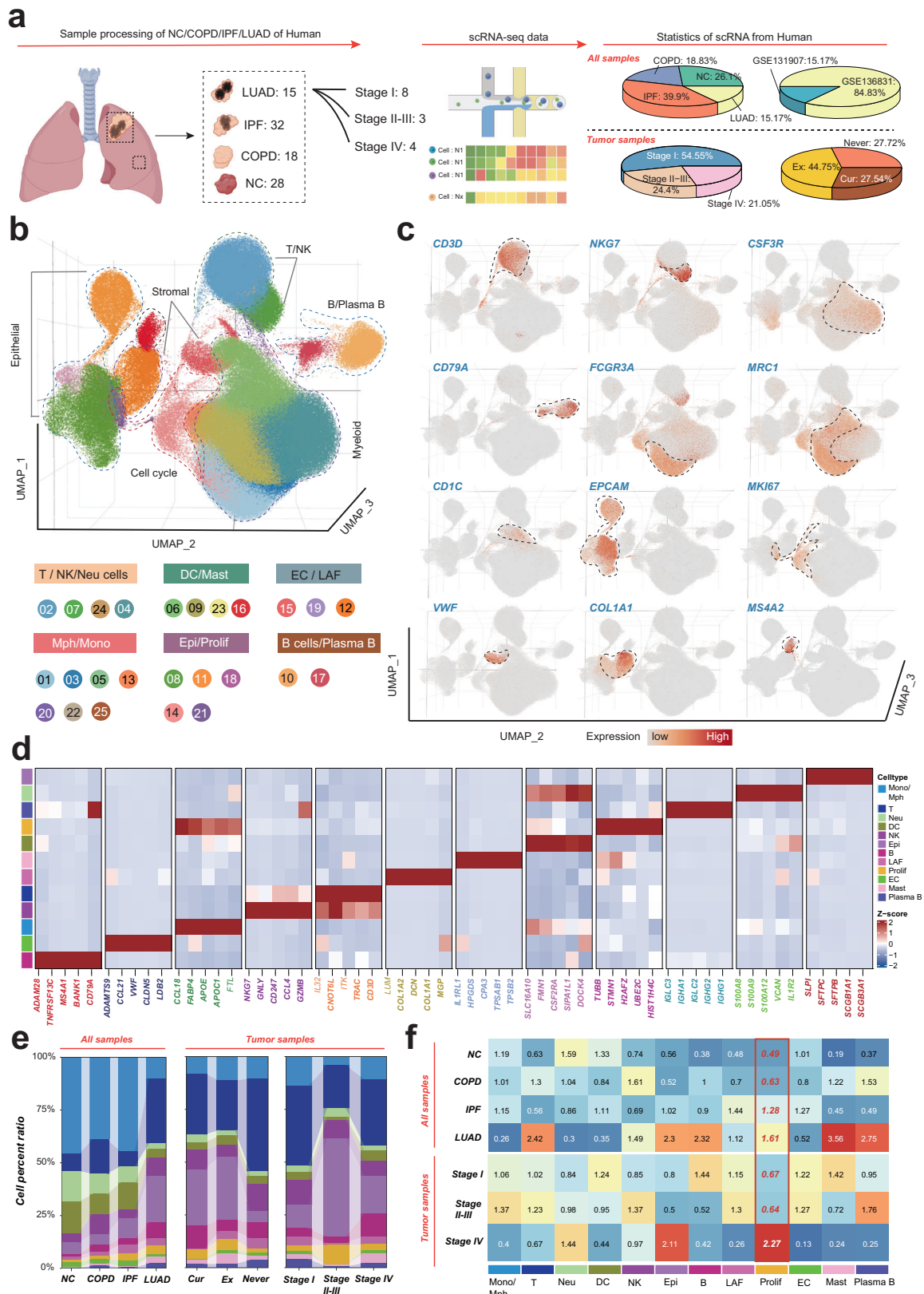
### Delineating proliferating cell heterogeneity and molecular signatures in human lung disease progression

We meticulously sorted 9353 proliferating cells from our scRNA-seq analysis and divided them into six distinct subpopulations (Fig. 2a and Table S4). The proliferating subpopulations were characterized by their unique surface markers and subset-specific markers, as delineated in Fig. 2b and Table S4. To gain deeper insights into the functional attributes of these subpopulations, we identified the top five biological processes significantly enriched among the marker genes for each cell cluster, effectively generating a molecular fingerprint for each cluster (Fig. 2c). To delineate the developmental trajectories of proliferating cells, we employed the SCTOUR algorithm. These analyses unveiled a complex, intersecting pattern of differentiation pathways among the proliferating cell subsets, with cluster C3\_KRT8 emerging as a central node (Fig. 2d, e). Given that proliferating cells represent a heterogeneous cell subset characterized by specific proliferative features, the inferred pseudotemporal differentiation trajectories should not be interpreted as definitive biological differentiation pathways. Rather, they highlight potential synergistic interactions among spatially proximate proliferating cell subsets, inferred from their UMAP-based trajectory directions. For example, the close spatial proximity and directional trajectory from C2\_MMP9 to C1\_FABP4 on the UMAP plot suggest a potential functional interplay between these two subpopulations in lung cancer progression. To test this hypothesis, we conducted intercellular communication analysis using the CellChat tool. As shown in Fig. 2f, C3\_KRT8 serves as a major sender, with C2\_MMP9 and C1\_FABP4 being the primary receivers of cellular signaling. The MIF-CD74 + CD44 signaling pathway was identified as a key mediator of communication among these subpopulations (Fig. 2g, h). Notably, spatial transcriptomics analysis revealed spatial colocalization among C1\_FABP4, C2\_MMP9, and C3\_KRT8 at the spatial resolution, further supporting the notion of their potential synergistic role in LUAD progression (Fig. 2i).

### Prognostic relevance of proliferating cell subtypes in LUAD

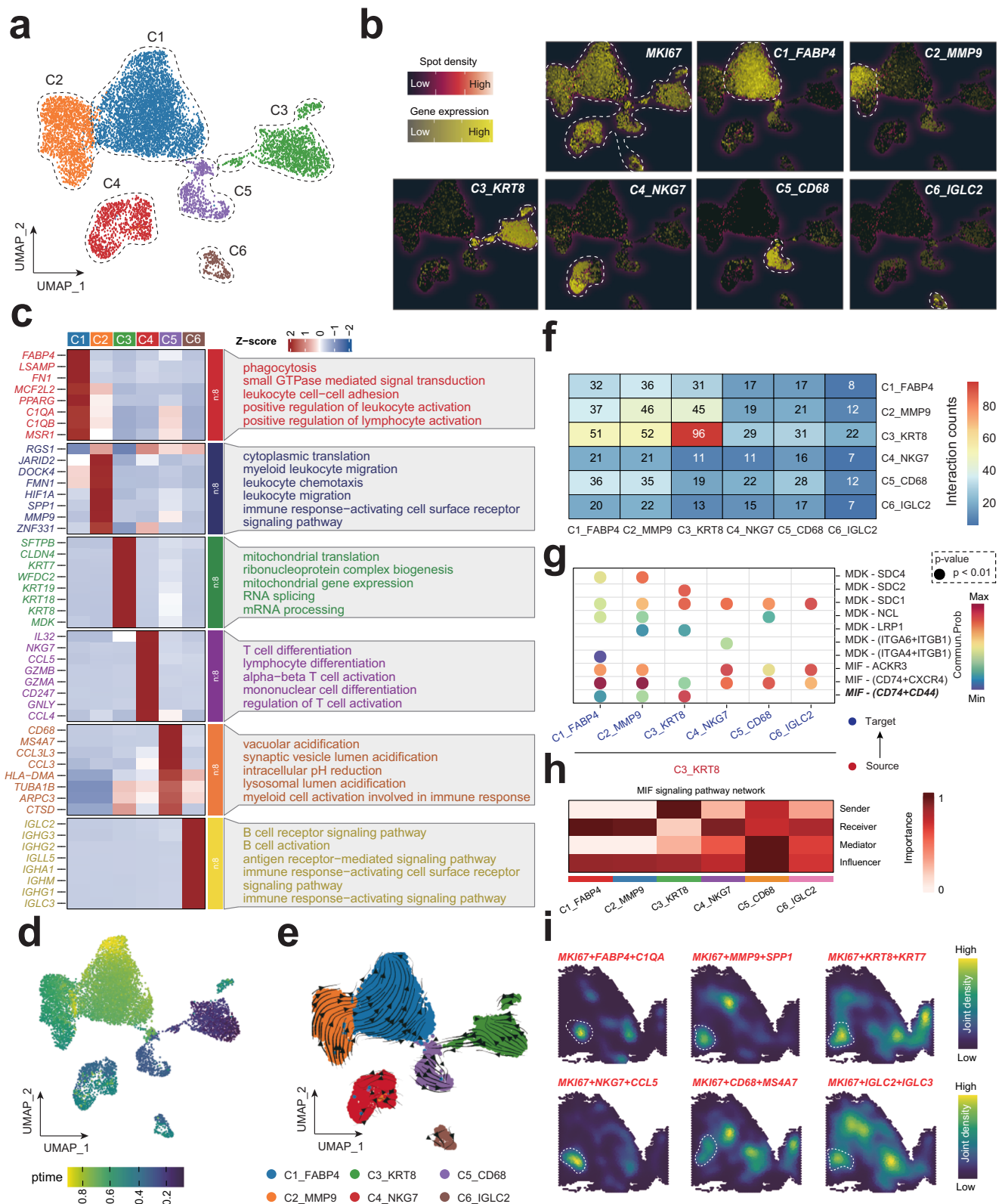
To unravel the clinical implications of various proliferating cell subtypes in LUAD, we leveraged the “Scissor” algorithm, a cutting-edge approach in single-cell data analysis, to pinpoint cell subgroups closely associated with distinct disease phenotypes within scRNA data. Our analysis revealed a predominance of Scissor+ labeled cells in clusters C2\_MMP9 and C3\_KRT8 (Fig. 3a, b). Of note, a total of 663 Scissor+ proliferating cell genes were identified (Fig. 3c and Table S5). Functional enrichment analysis hinted at an upregulation of cell-cycling and oncogenic pathways within the Scissor+ group, including the G2m Checkpoint and Epithelial-Mesenchymal Transition pathways, suggesting a potential role in the carcinogenesis and development of LUAD (Fig. 3d, e and Table S6).

Employing the Ro/e algorithm, we observed a distinct distribution of proliferating cell subsets across Scissor groups. Notably, clusters C2\_MMP9 and C3\_KRT8 were significantly enriched in the Scissor+ group, while C1\_FABP4, C5\_CD68, and C6\_IGLC2 subtypes were particularly enriched in the Scissor- group. This divergence in enrichment patterns may indicate varying prognostic impacts in LUAD, with the presence of subtypes like C2\_MMP9 and C3\_KRT8 in the Scissor+ group potentially indicating a more aggressive phenotype. To further identify the prognostic role of various proliferating subsets in LUAD, single-sample gene set enrichment analysis (ssGSEA) was applied to LUAD samples in the TCGA-LUAD dataset to calculate scores for various proliferating cell subtypes based on the top 200 significant genes (Table S7). Patients were categorized into score-low and score-high groups, with their prognosis correlation subsequently analyzed. Interestingly, only the Scissor+ proliferating cells group was associated with an unfavorable prognosis in LUAD. Importantly, NicheNet



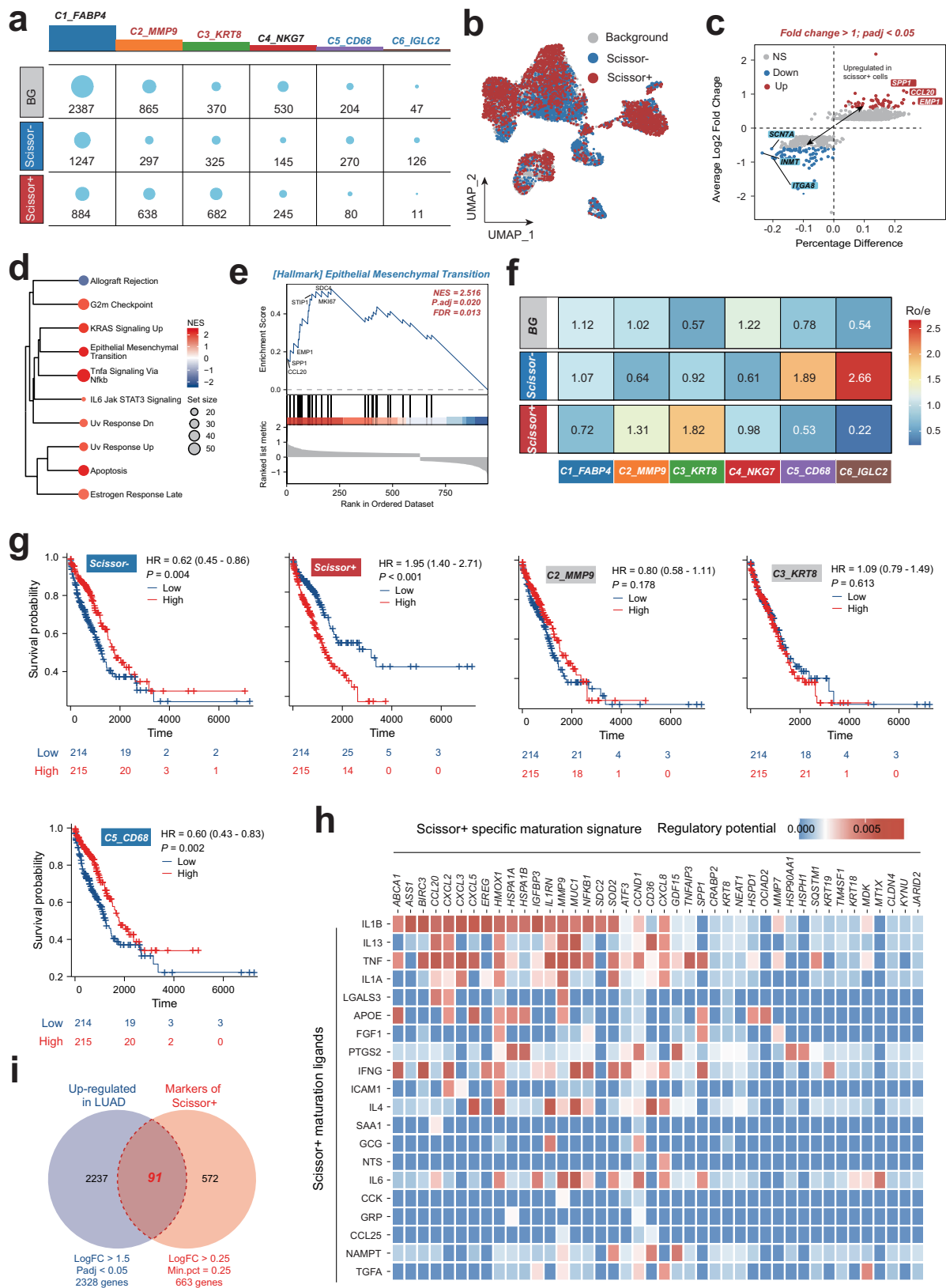
**Fig. 1 | Cell populations of 93 samples across the progression of lung disease.**  
**a** Schematic of the data generation, study design and the statistics of single-cell data from different datasets. **b** Uniform manifold approximation and projection (UMAP) analysis of all cells, colored by cell types. **c** UMAP plots showing expression

of canonical marker genes of major cell populations. **d** The expression levels of representative signature genes. **e** Stacked barplot showing proportions of major cell populations across groups. Colors represent major cell populations. **f** Heatmap showing tissue preferences of major cell population in each groups revealed by Ro/e.



**Fig. 2 | Characterization of proliferating cell subsets and their interactions in lung disease progression.** **a** UMAP plot of single-cell RNA sequencing data revealing distinct clusters of proliferating cells from various lung disease stages. Each point represents a cell, colored according to its cluster identity. **b** Heatmap displaying the expression of select marker genes characteristic of the annotated cell types within the identified clusters from the UMAP analysis. **c** Expression of significant markers and the enrichment of specific biological processes within each

proliferating cell cluster. **d**, **e** SCTOUR algorithm depicts the development trajectory of various proliferating subtypes. **f** Intercellular communication analysis using CellChat, depicting signaling interactions among proliferating cell subsets. **g** Dot plot showing pair-ligands between C3\_KRT8 and other subsets. **h** MIF signaling pathway network in proliferating cell subsets. **i** Spatial transcript analysis showing the joint density of various proliferating cell subsets.



**Fig. 3 | Identification and prognostic significance of Scissor+ proliferating cell genes in LUAD.** **a** Distribution of Scissor-, Scissor+, and background (BG) proliferating cell numbers in each cell type. **b** UMAP plots showing the distribution of proliferating cells in various Scissor groups. **c** The differentially expressed genes in Scissor- and Scissor+ proliferating cells. **d, e** Hallmark enrichment analysis of the upregulated genes in Scissor+ group. **f** Heatmap showing tissue preferences of proliferating cell subtypes in each Scissor groups revealed by Ro/e. **g** Kaplan-Meier

survival curves for LUAD patients stratified by the enriched scores of various proliferating cell subtypes calculate by the single-sample gene set enrichment analysis (ssGSEA) based on their significant markers. **h** Heatmap showing potential ligands driving the phenotype of Scissor+ proliferating cells inferred by NichNet analysis. **i** Venn plot showing the intersection of the LUAD-upregulated genes and the Scissor+ proliferating cell-associated markers. LUAD lung adenocarcinoma.

analysis predicted that IL1B ligands may drive the specific phenotype of Scissor+ proliferating cells, offering a potential therapeutic target for this aggressive subtype (Fig. 3h).

### Regulation mechanism of scissor+ proliferating cells and its cellular communication networks in LUAD

Previously, our analysis identified IL1B as a potential key ligand regulating the specific phenotype of Scissor+ proliferating cells (Fig. 3h) using NicheNet analysis. To validate the scientific rationale of this hypothesis, we examined the expression of IL1B and its receptor (IL1RL1, ADRB2, IL1R2 and IL1R1), as well as the IL1B ligand-receptor score (IL1BRCscore) calculated by AddModuleScore, along with well-known cell cycle and proliferation-related genes such as CCND1 and MKI67, across various proliferating cell subsets. Intriguingly, we found that these factors exhibited the highest and most specific expression in the primary subsets of the Scissor+ group, C2\_MMP9 and C3\_KRT9 (Fig. 4a), suggesting that IL1B may play a critical role in regulating the phenotype of Scissor+ proliferating cells.

We further assessed the prognostic value of IL1B in LUAD through a meta-analysis of 16 LUAD cohorts using a forest plot (Fig. 4b). The results indicated that high expression of IL1B was significantly associated with poor survival in LUAD patients (HR = 1.167, 95% CI: 1.071–1.273,  $p < 0.01$ ). Despite some degree of heterogeneity among cohorts, the overall findings to some extent, at least suggest that IL1B could serve as a negative prognostic biomarker for LUAD patients. Notably, we observed a substantial spatial colocalization between IL1B and Scissor score in LUAD samples, further supporting that IL1B may regulate the specific phenotype of Scissor+ proliferating cells. However, these findings warrant further validation in the context of clinical settings and future research.

To elucidate the mechanisms by which proliferating cells influence LUAD prognosis, we investigated the cellular communication between proliferating cells and other cell types within LUAD. We found that the Scissor+ subsets predominantly acted as signal senders, while immune cells (Mono/Mph), epithelial cells, and stromal cells (LAF and EC) served as signal receivers, with frequent intercellular communication occurring between them (Fig. 4d, e). Further analysis revealed that the FN1-CD44 axis may play a crucial role in the interactions between Scissor+ subsets and these cells, with bidirectional signaling occurring (Fig. 4f, g). Notably, LAF was identified as the primary signal sender, while Scissor+ subsets predominantly functioned as receivers (Fig. 4h, i). The expression specificity of FN1 and CD44 across different cell subpopulations was consistent with this hypothesis (Fig. 4j). Previous studies have shown that the FN1-CD44 ligand-receptor pair is involved in cell adhesion, migration, and invasion<sup>19</sup>. Therefore, LAF may interact with Scissor+ subsets via the FN1-CD44 axis, thereby enhancing their proliferative and invasive capabilities and promoting LUAD progression.

### Prognostic significance of Scissor+ proliferating cell genes in LUAD

To gain insights into the prognostic implications of diverse proliferating cell subtypes in LUAD, we intersected a set of 2328 tumor-regulated genes with 663 Scissor+ proliferating cell genes, yielding 91 potential prognostic genes that were carried forward into subsequent analyses (Fig. 3i). A focused Cox regression analysis within the TCGA-LUAD cohort narrowed down these to 22 candidate model genes. We further advanced our analysis by employing machine-learning techniques to pinpoint model genes with predictive potential. Within the TCGA-LUAD training cohort, we developed a robust predictive model, integrating 111 machine learning algorithms through the MIME package. The predictive prowess of each model was gauged by calculating the average C-index across all cohorts (Fig. 5a). Our research identified the Lasso + SuperPC combination as superior in TCGA, GSE31210, GSE50081, GSE72094, and META cohorts, indicating its enhanced prognostic capacity. This combination was thus selected for deeper analysis. The partial probability deviation in the LASSO regression was minimized by optimizing the penalty parameter lambda using 10-fold cross-validation (Fig. 5b). This process identified five model genes, namely

FAM83A, ANLN, HMGA1, ECT2, and PRC1 that demonstrated decisive values across all LUAD cohorts by Cox regression analysis (Fig. S2). Moreover, the SPRS calculated using the Lasso + SuperPC combination exhibited a high C-index, showcasing great prognostic discrimination across each cohort (Fig. 5c, d). Notably, the identified combination also demonstrated great 1-year area under the curve (AUC) among various LUAD cohorts (Fig. 5e). To ascertain the prognostic impact of SPRS, we performed a meta-analysis of univariate COX regression via Mime, which revealed that the score calculated by SPRS is a potent prognostic indicator for LUAD patients (Fig. 5f, g). Moreover, we conducted a meta-analysis of the univariate cox survival analysis among the SPRS in 20 LUAD cohorts to extend and validate the prognostic values of SPRS in LUAD (Fig. S3).

Interestingly, the five SPRS model genes exhibited significantly higher protein expression in LUAD tissues compared to normal controls (Fig. S4a), consistent with their mRNA upregulation. Furthermore, elevated protein levels of FAM83A, ANLN, and ECT2 were significantly associated with poorer prognosis in LUAD patients (Fig. S4b). Strikingly, all model genes showed strong positive correlations with cell cycle progression and EMT-related pathways (Fig. S4c), suggesting that these genes may collectively drive tumor aggressiveness by coordinately regulating proliferative and metastatic programs in LUAD.

### Spatial profiling of SPRS activity and its association with LUAD malignancy and immune microenvironment

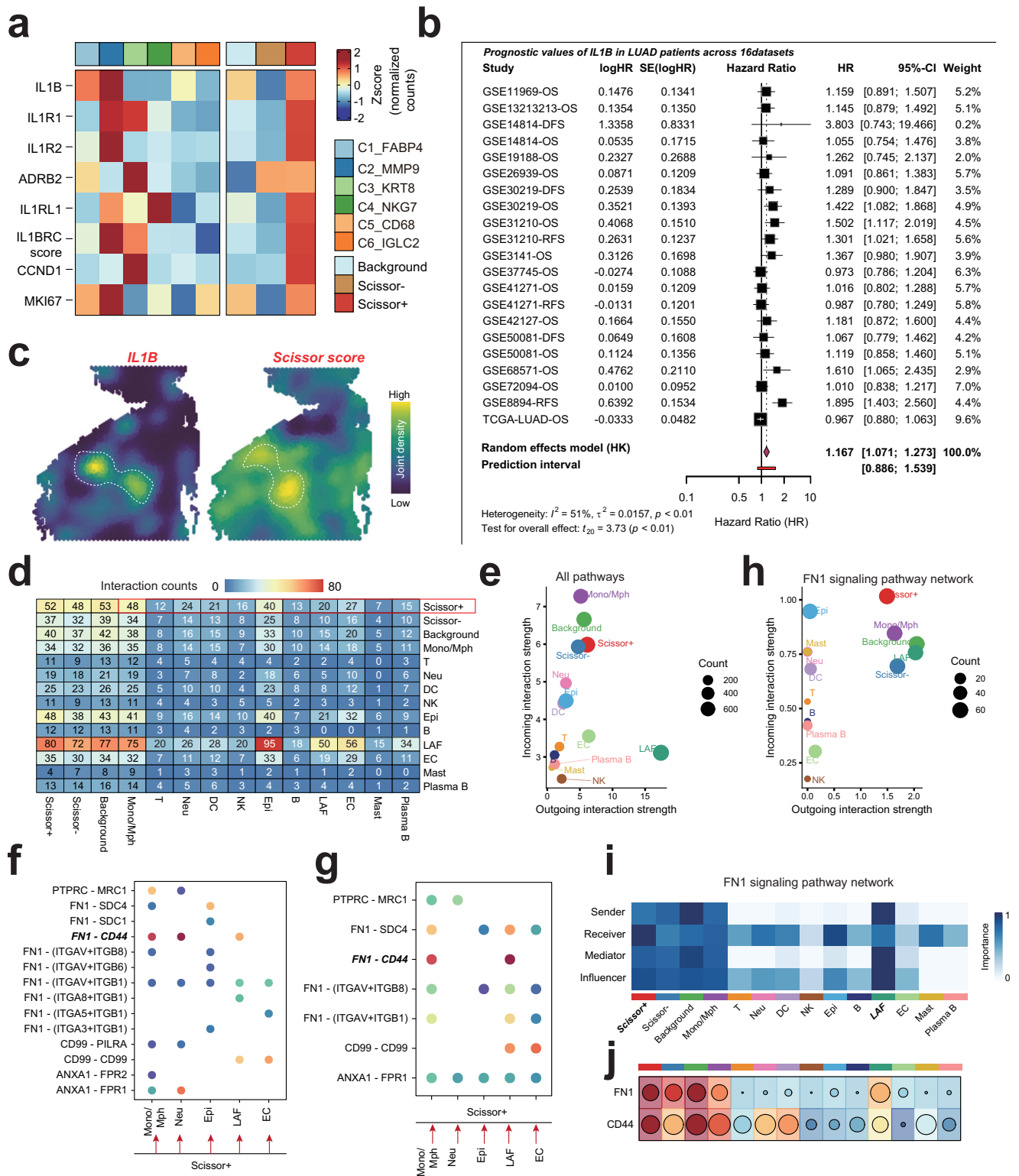
To elucidate the spatial distribution of SPRS activity within LUAD tissues, we employed the AUCell algorithm to determine AUC scores for SPRS using spatial transcriptomics data from GSE179572<sup>20</sup> (Fig. S5a). Deconvolution analysis was subsequently conducted to delineate the cellular composition within each spatially resolved tissue spot (Fig. S5b). Intriguingly, elevated SPRS AUC scores were predominantly observed in malignant and mixed malignant regions (Fig. S5c). Spearman correlation analysis further revealed a robust correlation between SPRS AUC scores and the abundance of tumor cells, macrophages, dendritic cells (DCs), and neutrophils within the LUAD microenvironment (Fig. S5d). This suggests that SPRS activity is preferentially localized to tumor cell populations and may contribute to the establishment of an immune-suppressive microenvironment.

Moreover, individual analysis of LPPG model genes demonstrated significantly higher expression levels in malignant and mixed malignant regions compared to normal tissue spots (Fig. S5e, f), with peak expression observed in purely malignant regions (Fig. S5g). Correlation analyses of all five SPRS components consistently echoed these findings, confirming their heightened activity within tumor cells (Fig. S5h). Collectively, these results position SPRS as a promising biomarker for evaluating LUAD malignancy and underscore its intricate interplay with the tumor microenvironment.

### Validation of the clinical significance of SPRS

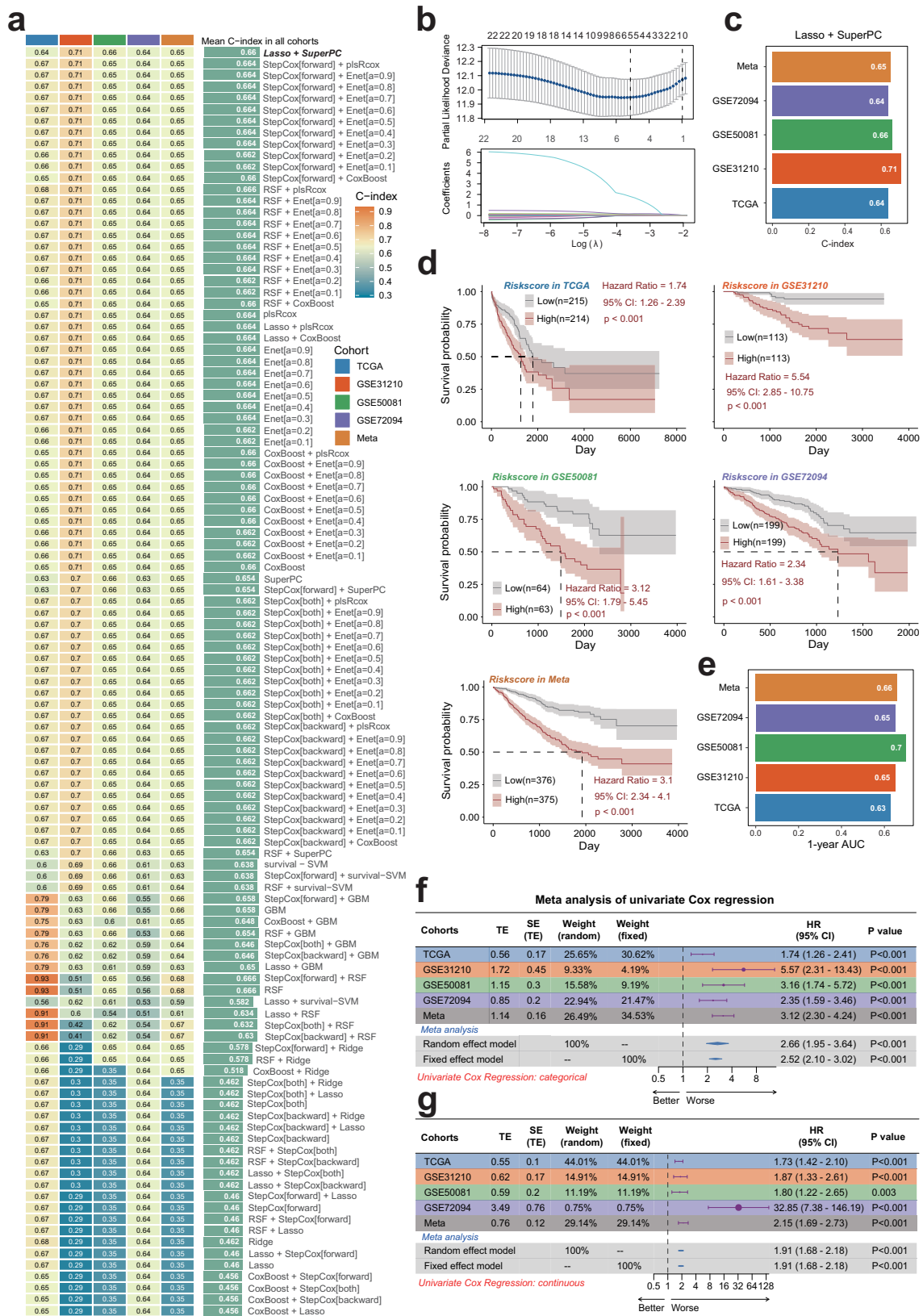
To substantiate the prognostic accuracy of the SPRS model for LUAD, an extensive review of the pertinent literature within the past half-decade was conducted. In this analysis, the SPRS's C-index was compared against 30 established mRNA-related prognostic profiles in LUAD. Remarkably, the SPRS surpassed these existing models in predicting LUAD outcomes across all datasets evaluated (Fig. 6a, b). Univariate and multivariate Cox regression analyses confirmed SPRS as an independent risk factor, distinct from traditional clinical parameters (Table S8).

Moreover, we constructed a nomogram model within the TCGA cohort, utilizing multivariate Cox and stepwise regression analysis to forecast 1-year, 3-year, and 5-year survival of LUAD patients. This model incorporated T stage, N stage, M stage, and SPRS as risk factors (Fig. S6a). The calibration curve validated the model's accuracy against actual outcomes (Fig. S6b). K-M curves underscored the model's clinical relevance (Fig. S6c). An assessment of the AUC values for the TCGA-LUAD cohort revealed that the nomogram model significantly enhanced the precision of OS predictions for LUAD patients (Fig. 6c). The predictive power of SPRS was also corroborated across additional LUAD cohorts, including GSE31210, GSE50081, GSE72094, and META, suggesting that SPRS holds value in prognostication across diverse



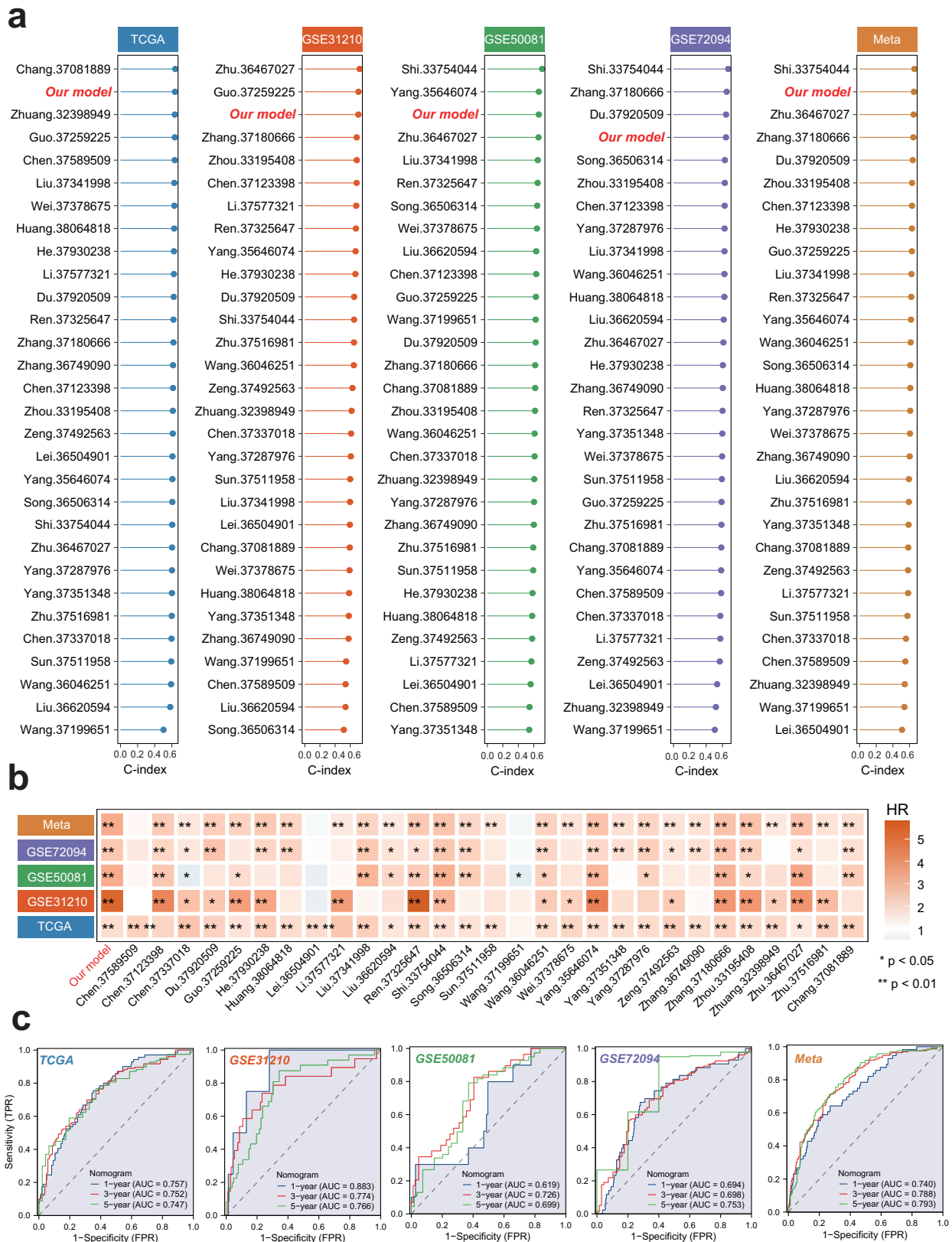
**Fig. 4 | Regulation mechanism of scissor+ proliferating cells and their cellular communication networks in LUAD. a** Heatmap displays the expression of IL1B and its receptors (IL1RL1, ADRB2, IL1R2, IL1R1) along with the IL1B ligand-receptor score (IL1BRCscore) and proliferation-related genes (CCND1, MKI67) across different proliferating cell subsets. **b** Meta-analysis of IL1B expression in LUAD cohorts. A forest plot summarizing the hazard ratios (HR) and 95% confidence intervals (CI) for IL1B expression across 16 LUAD cohorts. **c** Spatial colocalization analysis of IL1B and Scissor score in LUAD samples, supporting the hypothesis that IL1B regulates the phenotype of Scissor+ proliferating cells. **d** Schematic representation of cellular communication networks in LUAD. Scissor+

subsets predominantly function as signal senders, while immune cells (Mono/Mph), epithelial cells, and stromal cells (LAF and EC) act as signal receivers. **e** Interaction network highlighting frequent intercellular communication between Scissor+ subsets and other cell types within LUAD. **f, g** Analysis of the FN1-CD44 axis in cellular interactions. The FN1-CD44 axis was identified as a key mediator of interactions between Scissor+ subsets and other cells, with bidirectional signaling observed. **h, i** FN1 signaling pathway network illustrating the primary signal senders and receivers. LAF was identified as the main signal sender, while Scissor+ subsets predominantly functioned as receivers. **j** Heatmap showing the expression of FN1 and CD44 in various cell types of lung diseases.



**Fig. 5 | Development and validation of SPRS for LUAD. a** Through a comprehensive computational framework, a combination of 111 machine learning algorithms was generated. The C-index of each model was calculated through the TCGA-LUAD, GSE31210, GSE50081, GSE72094, and META-LUAD (batch effects-removed combination of GSE31210, GSE50081 and GSE72094) cohorts and sorted by the average C-index of the validation set. **b** The hub gene selected through the

LASSO regression. **c** The C-index of SPRS calculated by Lasso+SuperPC algorithm across each datasets. **d** The relation between SPRS calculated by Lasso+SuperPC combined model and outcome of patients in different cohorts. **e** 1-year AUC of Lasso + SuperPC combined model among different cohorts. **f, g** Meta-analysis of univariate Cox result of Lasso+SuperPC combined model among different cohorts. LUAD lung adenocarcinoma, SPRS Scissor+ proliferating cell risk score.



**Fig. 6 | Comparison of SPRS with previously established models in LUAD.**  
**a** C-index of SPRS and 30 published models across the training and 4 testing datasets.  
**b** HR of SPRS and 30 published models across the training and 4 testing datasets.  
 SPRS, Scissor+ proliferating cell risk score. **c** ROC curve of the established

nomogram model based on the SPRS and relevant clinical parameters in the training and 4 testing datasets. LUAD, Lung Adenocarcinoma; ROC, Receiver operating characteristic; SPRS, Scissor+ proliferating cell risk score.

patient groups (Fig. S6d–o and Fig. 6c). Therefore, the SPRS not only outperforms existing models but also provides a robust tool for differentiating LUAD prognosis, offering a potentially transformative approach to patient risk stratification and therapeutic planning.

### Immune landscape of SPRS in LUAD

To explore the prognostic implications of the SPRS in LUAD, we conducted a thorough examination of the TIME using the “IOBR” R package, which integrates multiple algorithms for a comprehensive analysis of immune cell infiltration. Our findings revealed a significant increase in the levels of immune cells, including CAFs and MDSCs in patients with higher SPRS, pointing to an immunosuppressive microenvironment. This association suggests that a high SPRS may be linked to advanced immunosuppression mechanisms, which could be a critical factor in LUAD progression (Fig. 7a–d). Further analysis of immunotherapy-related biomarkers within the high SPRS group indicated an upregulation of these markers, potentially due to immune evasion strategies employed by the immunosuppressive TIME (Fig. 7b). Intriguingly, our data also demonstrated a positive correlation between high SPRS and elevated TMB and TNB, implying that an inflamed TIME with increased TMB might drive a more active adaptive immune response, which in turn could lead to enhanced immune evasion (Fig. 7e, f). This paradox highlights the complexity of the tumor microenvironment and suggests that high TMB alone may not be a sufficient predictor of ICB efficacy<sup>21</sup>. Further research is needed to explore the intricate relationships between TMB, SPRS, and immune evasion mechanisms to better understand their collective impact on treatment outcomes.

Survival analysis confirmed that SPRS could serve as a valuable complementary factor to TMB, TNB, CAFs, and MDSCs in predicting patient outcomes. LUAD patients with lower SPRS, along with reduced TMB or TNB, or decreased infiltration of CAFs or MDSCs, exhibited a more favorable survival prognosis (Fig. 7g). This suggests that SPRS could be a potent immunological factor for stratifying LUAD patients and informing treatment strategies.

### Alteration landscape of SPRS model genes in LUAD

Overall, the analysis of samples from the TCGA-LUAD project revealed multiple chromosomal copy number variations (Fig. 8a). Notably, FAM83A exhibited the highest mutation rate (7%) in LUAD, primarily through amplifications (Fig. 8b). The genomic alteration rate was higher in the top 25% of samples exhibiting elevated FAM83A expression, with increased proportions of both losses and gains (Fig. 8c). Furthermore, in LUAD, the Spearman rank correlation coefficient between the Gistic2-calculated FAM83A copy number score and FAM83A mRNA expression levels was 0.28, indicating a moderate positive correlation (Fig. 8d). notably, FAM83A expression showed an increase trend from homozygous deletions to high copy number amplifications (Fig. 8e). Additionally, numerous mutations associated with FAM83A expression were identified, with the top three being KRAS, MUC16, and CSMD1 (Fig. 8f). Multiple studies have indicated that mutations in KRAS<sup>22</sup>, MUC16<sup>23</sup>, and CSMD1<sup>24</sup> not only promote tumor progression but may also undermine the efficacy of immunotherapy. Moreover, a Wilcoxon rank-sum test revealed significant differences in FAM83A expression levels between KRAS wild-type and mutant patients, with FAM83A expression significantly higher in KRAS mutant patients. This finding suggests that KRAS mutations may regulate FAM83A expression, impacting patient prognosis (Fig. 8g). Thus, elevated FAM83A expression likely reflects enhanced proliferation activity, genomic instability, and heterogeneity within these tumor cells, all of which are associated with poor prognosis and resistance to treatment. Therefore, a more precise and personalized treatment approach may be necessary for these patients.

Moreover, it is noticeable that the analyses performed above did not differentiate between cases with and without driver gene mutations or alterations, such as KRAS. It is well established that the TIME varies between these groups, and patients with driver gene mutations generally exhibit lower responsiveness to immunotherapy compared to those without such mutations or alterations. In the current study, we found that KRAS

mutations significantly influenced the prognosis of patients in the GSE72094-LUAD cohort. To eliminate potential confounding factors related to KRAS mutations, we further validated the correlation between SPRS and TIME in patients without KRAS mutations. Our findings indicated a trend similar to that observed in the TCGA-LUAD cohort (Fig. 8h), where the group with high SPRS displayed an immunosuppressive microenvironment characterized by greater infiltration of tumor-associated macrophages (TAM) and MDSC, along with elevated expression of immune checkpoint molecules (Fig. 8i, j) and worse clinical outcomes (Fig. 8k). These findings suggest that the SPRS model remains a robust predictor of the tumor immune microenvironment in LUAD, even after excluding the influence of driver gene mutations such as KRAS, thereby highlighting its potential utility in guiding more tailored and personalized treatment strategies for LUAD patients.

### SPRS as a predictor for personalized LUAD therapy response

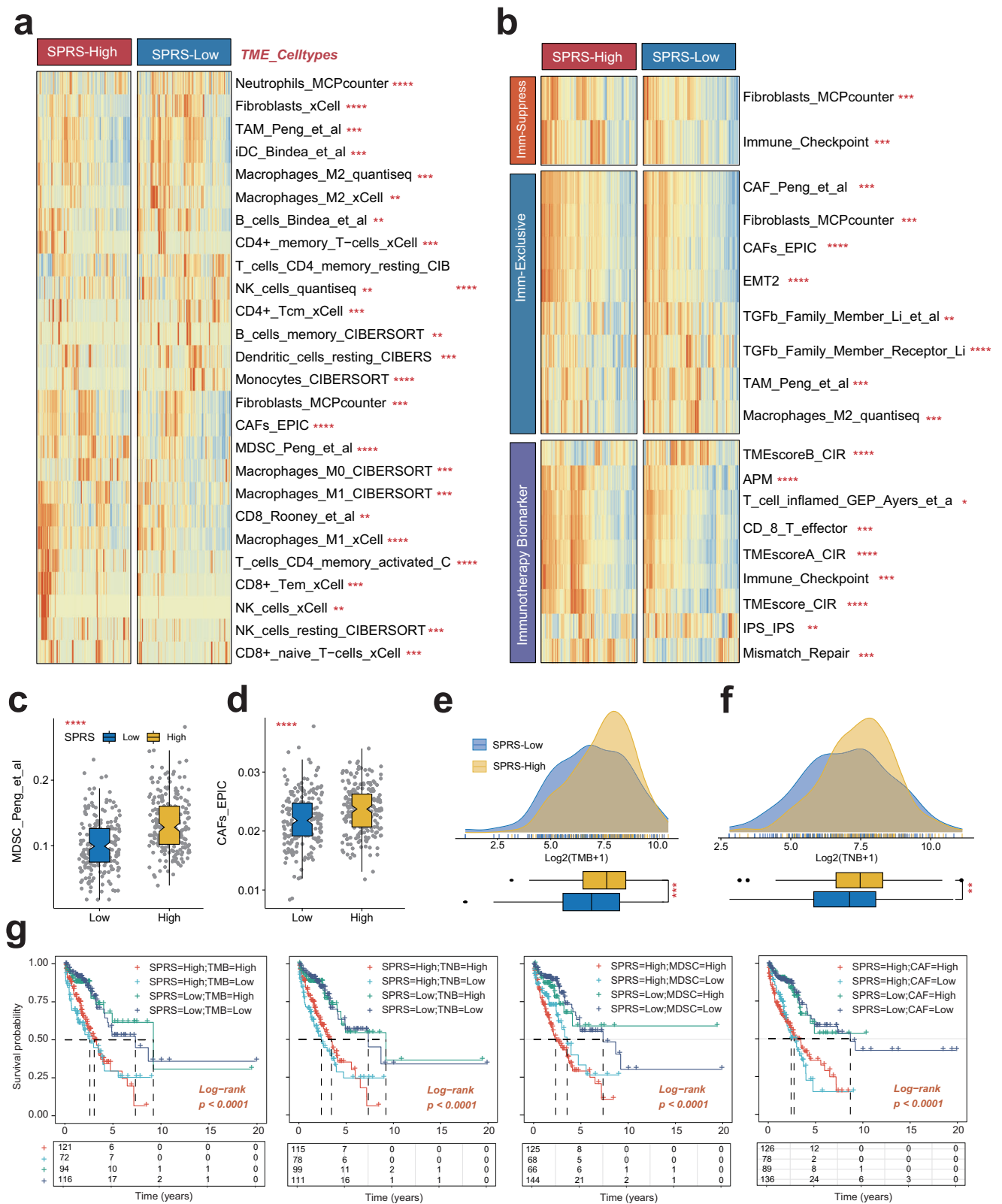
To evaluate the clinical relevance of the SPRS in LUAD immunotherapy, a meticulous analysis of the IMvigor210 cohort was undertaken. This analysis aimed to discern the prognostic and therapeutic implications of SPRS in LUAD patients. A comparative analysis of survival rates between patients with high and low SPRS revealed a significant difference, with the low SPRS group exhibiting more favorable outcomes (HR = 1.80,  $p < 0.001$ ; Fig. 9a). This correlation was further supported by the observation that a lower SPRS correlated with improved responses to immunotherapy in the IMvigor210 cohort, as indicated by the Wilcoxon rank-sum test ( $p = 0.0019$ , Fig. 9b). To reinforce these findings, the analysis was expanded to other independent additional immunotherapy cohorts, including GSE91061 and GSE78220. Consistent with our initial observations, patients with higher SPRS scores were associated with poorer clinical outcomes compared to those with lower SPRS (Fig. 9c, d). GSEA identified the enrichment of EMT and cell cycle-related pathways, including E2f targets, G2m checkpoints, and Myc target V1, in the high SPRS groups (Fig. 9e). These findings suggest that patients with elevated SPRS may exhibit a more aggressive disease phenotype, potentially due to enhanced immune evasion mechanisms.

To identify potential therapeutic strategies for LUAD patients with high SPRS, the GDSC database was utilized to screen for drugs that may be sensitive to this patient subgroup. The analysis revealed a negative association between SPRS and several drugs, including cisplatin, docetaxel, paclitaxel, and gefitinib, suggesting a potential increased sensitivity to these chemotherapeutic and targeted therapies in LUAD patients with high SPRS (Fig. 9f, g and Table S9). Furthermore, given that PRC1 demonstrated the highest HR value based on univariate Cox regression analysis, its subcellular localization and protein level were examined using the HPA database. PRC1 was found to be primarily localized to the plasma membrane, microtubules, cytokinetic bridge, and midbody. Notably, PRC1 protein levels were elevated in LUAD tissues compared to normal tissues (Fig. 9h, i), implicating PRC1 as a potentially significant factor in LUAD progression and aggressiveness.

To strengthen the validity of our study, we used RT-qPCR to validate the expression of the identified model genes (Table S10). The expression levels of five SPRS model genes were evaluated in the clinical LUAD samples, including three adjacent cancer tissues and three cancerous tissues. All model genes showed an increasing trend in the LUAD tissues (Fig. 9j), consistent with our bioinformatic analysis, thereby reinforcing the credibility of our study.

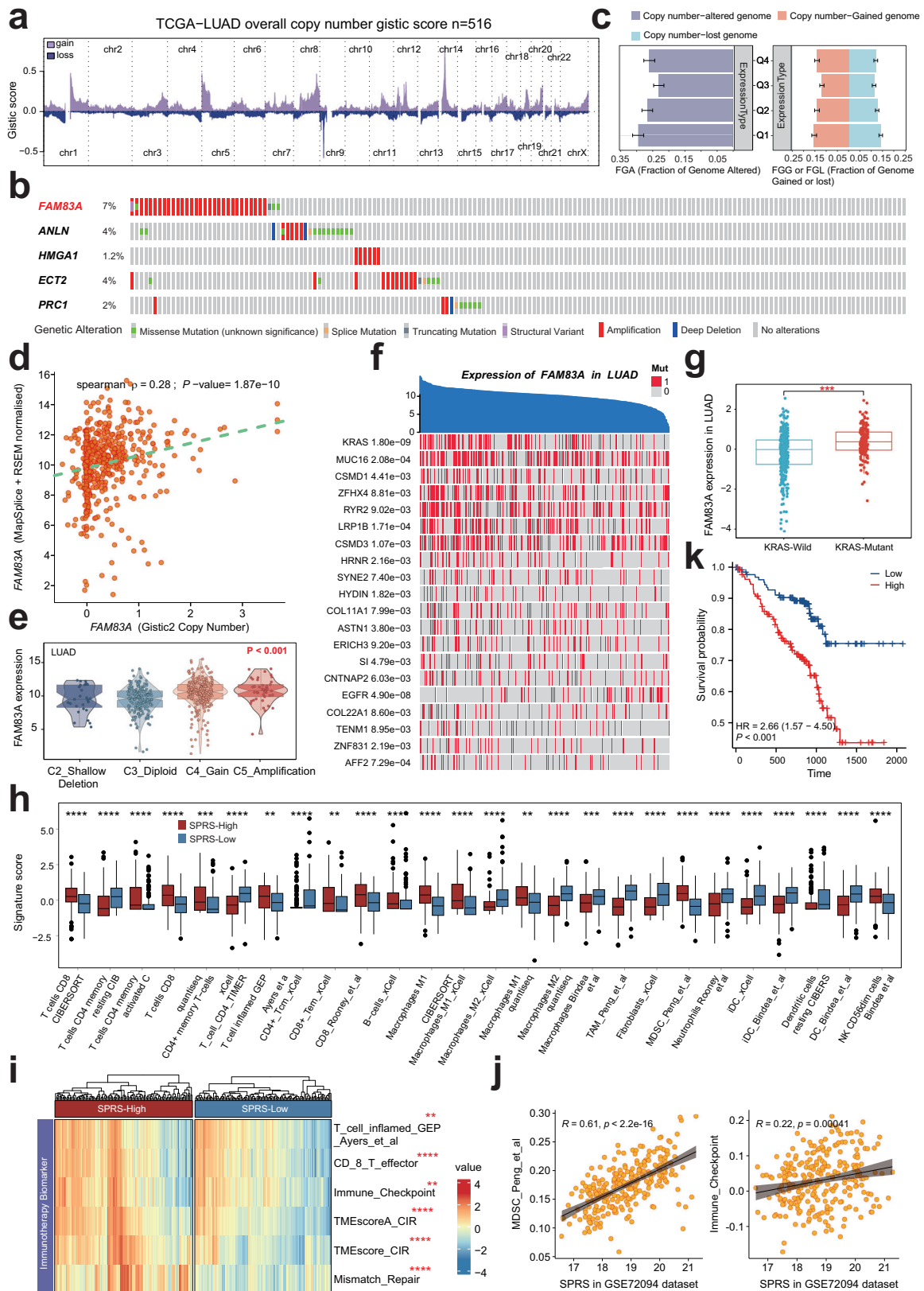
### Spatial transcriptomics unveils elevated SPRS in LUAD tissues

To elucidate the spatial distribution of SPRS in lung diseases, we conducted ST analysis on a diverse cohort of tissues, including three healthy lung samples, two adjacent non-cancerous tissues, and five LUAD tissues (Fig. 10a). Given the complexity inherent in each spatial spot, which may encompass multiple cell types, we employed a signature-based analytical approach to quantify the enrichment of various cellular components within each spot (Fig. 10b). Our analysis revealed significant heterogeneity among malignant cells, contrasting with the co-localization of fibroblasts and proliferating cells (Fig. 10b). Upon calculating the SPRS for each ST sample,



**Fig. 7 | Immune landscape and survival analysis based on SPRS in LUAD. a** The distribution of TME immune cell type signatures between high- and low-SPRS patients in LUAD. **b** The distribution of immune suppression, immune exclusion, and immunotherapy-related signatures between high- and low-SPRS patients in LUAD. **c, d** The distribution of TMB and TNB between high- and low-SPRS patients in LUAD. **e, f** The distribution of MDSC and CAF between high- and low-SPRS

patients in LUAD. **g** Survival analysis combined SPRS with TMB, TNB, MDSC, and CAFs in LUAD patients. LUAD lung adenocarcinoma, SPRS Scissor+ proliferating cell risk score, CAFs cancer-associated fibroblasts, MDSCs myeloid-derived suppressor cells, TMB tumor mutational burden, TNB tumor neoantigen burden; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , ns not significant.



we observed a notable elevation in both SPRS scores and the proportion of SPRS-positive cells in LUAD tissues compared to non-tumorous controls (Fig. 10c, d and Fig. S7a–f). These findings underscore the potential of SPRS as a spatial biomarker for LUAD, reflecting the intricate interplay between gene expression patterns and tumorigenesis.

### Multi-omics characterization of proliferating cell-driven immunosuppression across LUAD progression stages

The above results explored the role of proliferating cells in different stages of lung disease from a multi-omics perspective. However, the cohort of LUAD patients included only 15 individuals. We further extracted the LUAD

**Fig. 8 | Alteration landscape of SPRS model genes and their impact on LUAD progression.** **a** This diagram shows the copy number variation in the genomes of 516 samples from the TCGA-LUAD project. The *x*-axis in the figure marks the different chromosome numbers, and the *y*-axis represents the gistic score values. Red bars indicate a higher gistic score (increase in copy number), and cyan bars indicate a lower gistic score (decrease in copy number). **b** Mutation profile and frequencies of SPRS model genes in LUAD from the cBioPortal database. **c** Percentage of FGL, FGG, and FGA in FAM83A expression subsets for LUAD. **d** Correlation analysis of Gistic2 gene copy number variation score and FAM83A gene expression in LUAD scatter plot. Each point in the graph represents a sample, the *x*-axis represents the gene copy number score calculated by Gistic2, and the *y*-axis represents the corresponding gene expression. **e** Differential expression of FAM83A between various copy number type. **f** Correlation of FAM83A expression levels and specific genetic

mutations in LUAD. The blue area at the top represents individual patient samples, and the *y*-axis represents the FAM83A expression level of the samples, arranged in reverse order. The red vertical line below represents the mutated gene, otherwise the wild type (gray). **g** The distribution of median, quartile and data is presented to explore the difference in the distribution of FAM83A gene expression in wild-type (blue) and mutant (red) KRAS cells. \*\*\*,  $p < 0.001$ . **h** Validation of the correlation between SPRS and TIME in LUAD patients without KRAS mutation in GSE72094 cohort. **i** The distribution of immunotherapy-related signatures between high- and low-SPRS patients in LUAD patients without KRAS mutation in GSE72094 cohort. **j** Scatter plot shows the correlation between SPRS levers and immune checkpoint molecules as well as the MDSC infiltration levers in LUAD patients without KRAS mutation in GSE72094 cohort. **k** Kaplan–Meier survival curves for patients with low and high SPRS in LUAD patients without KRAS mutation in GSE72094 cohort.

samples from the test cohort and integrated them with a larger single-cell atlas of LUAD, reconstructing a dataset that includes different stages of LUAD, comprising 32 patients with Stage I, 5 with Stage II, 8 with Stage III, and 28 with Stage IV disease (Fig. 11a). Following a rigorous quality control process, we included a total of 259,254 high-quality single-cell profiles from LUAD patients, categorizing them into 15 distinct cell subpopulations based on their unique gene expression characteristics (Fig. 11b). Using the AddModuleScore algorithm, we calculated the SPRS score for each cell and categorized them into SPRS\_high and SPRS\_low groups based on median expression levels (Fig. 11c). Employing the Ro/e algorithm, we found that proliferating cells exhibited a significant tissue preference in later stages of LUAD (Stages II–IV), consistent with observations from the testing cohort. Additionally, patients categorized as SPRS\_high had a higher enrichment of macrophages, CAFs, and proliferating cells (Fig. 11d), further revealing an immunosuppressive microenvironment associated with higher SPRS levels and resistance to immunotherapy.

We extracted the top 50 markers of proliferating cells from the validation cohort's and discovered that the combined high expression of proliferation-related genes was significantly associated with poorer prognosis across 2,491 LUAD patients in 14 datasets (combined HR = 3.277, 95% CI: 2.536–4.235,  $p < 0.01$ ), with low heterogeneity observed among different cohorts ( $I^2 = 25\%$ ). This indicates that comprehensive markers of proliferating cells can effectively distinguish high-risk patients and provide robust prognostic stratification (Fig. 11e).

To further analyze and verify the spatial organization of immune cell types in high- and low-SPRS tumors as well as to explore interactions between proliferating cells and the TIME in LUAD, we examined the spatial localization of SPRS-positive cells in tumor tissues and their associations with immune cell infiltration and tumor heterogeneity across the 10 LUAD ST samples. Similarly, we categorized each ST spot based on the median SPRS score into SPRS\_high and SPRS\_low groups. We identified samples #P24\_T2 and #P10\_T2 as high-SPRS tumors, while #P16\_T1 and #P25\_T1 were categorized as low-SPRS tumors, correlating with the observed differences in the proportions of SPRS-positive cells across the various ST samples (Fig. 9d and Fig. S5). Furthermore, using the CellTrek algorithm, we deconvoluted the cellular composition of the high and low-SPRS tumors (Fig. 11g) and found that high-SPRS samples exhibited a higher proportion of CAFs and a lower proportion of CD8 + T cells (Fig. 11h), corroborating the results observed at the single-cell level (Fig. 11d).

## Discussion

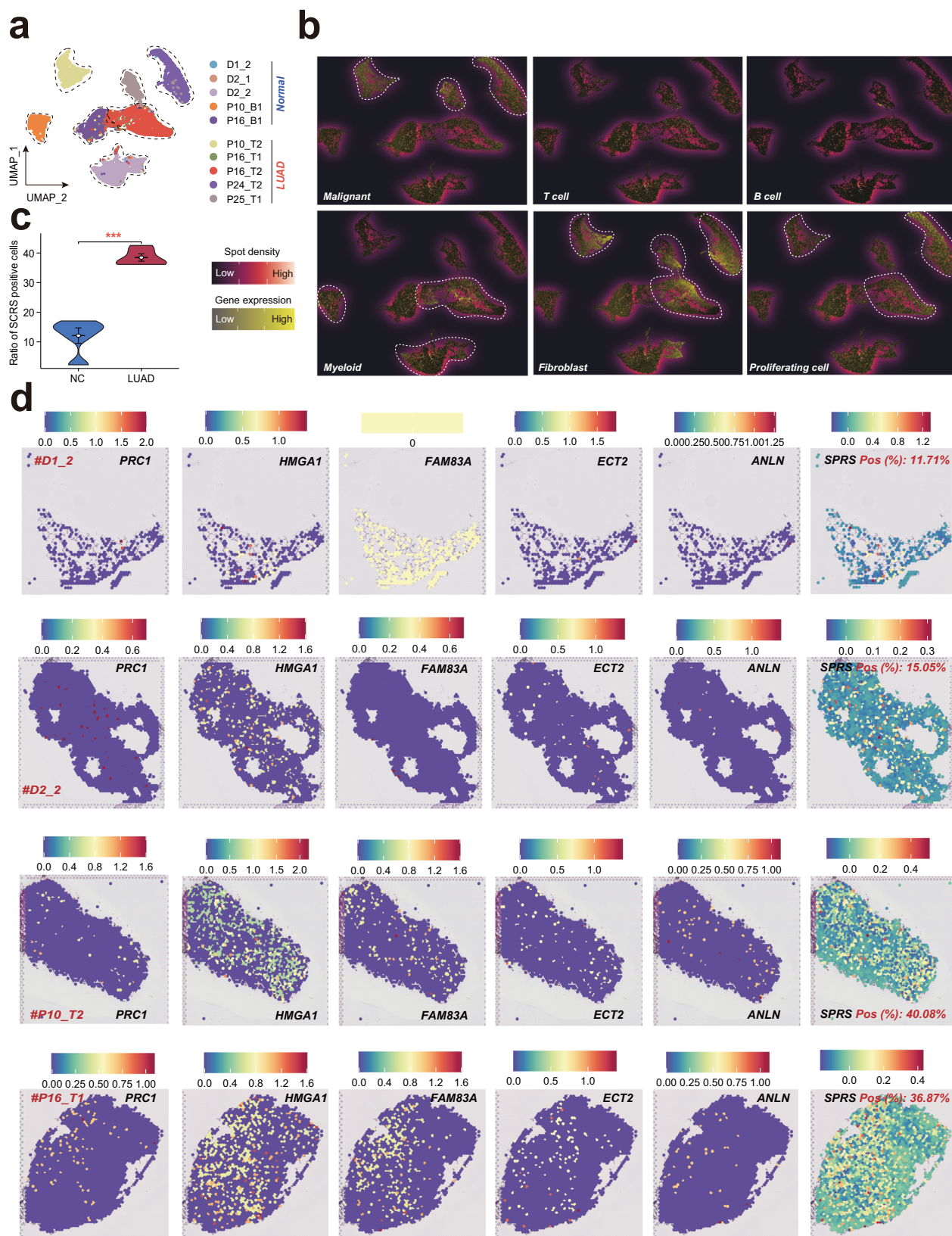
Within the TIME of cancers, proliferating cells are identifiable throughout the entire spectrum of carcinogenesis and development of cancers<sup>10</sup>. Recent single-cell multi-omics studies have highlighted the critical role of proliferating immune cell populations in tumor-immune interactions. Specifically, cycling T cells, myeloid precursors, and proliferative stromal cells are significantly enriched in areas of active tumor-immune engagement. These cells not only respond rapidly to environmental cues but also play a crucial role in local immunosuppression and can predict responses to immunotherapy. Moreover, high-dimensional transcriptomic profiling has revealed that transcriptional signatures associated with proliferation often

transcend traditional lineage boundaries, capturing the functional convergence of various cell subsets involved in cell cycle progression and DNA replication. This finding provides new insights into the immune dynamics within the tumor microenvironment and their implications for therapeutic responses. However, the genetic landscape, multimolecular mechanisms, and immune functions of proliferating-associated genes in LUAD are still not fully understood.

In this study, we delve into the multifaceted role of proliferating cells during the progression of lung diseases, leveraging cutting-edge multi-omics profiling and machine learning techniques to uncover novel insights. Our approach integrates bulk RNA sequencing, scRNA-seq, and stRNA-seq data to provide a comprehensive view of the proliferating cell landscape within lung diseases. By applying the Scissor algorithm, we have identified a subset of proliferating cell genes that are associated with prognostic implications, a discovery that could revolutionize our understanding of LUAD pathogenesis. The development of a SPRS, constructed through an integrative machine learning program comprising 111 algorithms, represents a significant step forward in the personalized medicine paradigm. This score not only predicts patient prognosis with remarkable accuracy but also informs tailored therapeutic strategies, aligning with the growing recognition of the importance of precision medicine in cancer treatment.

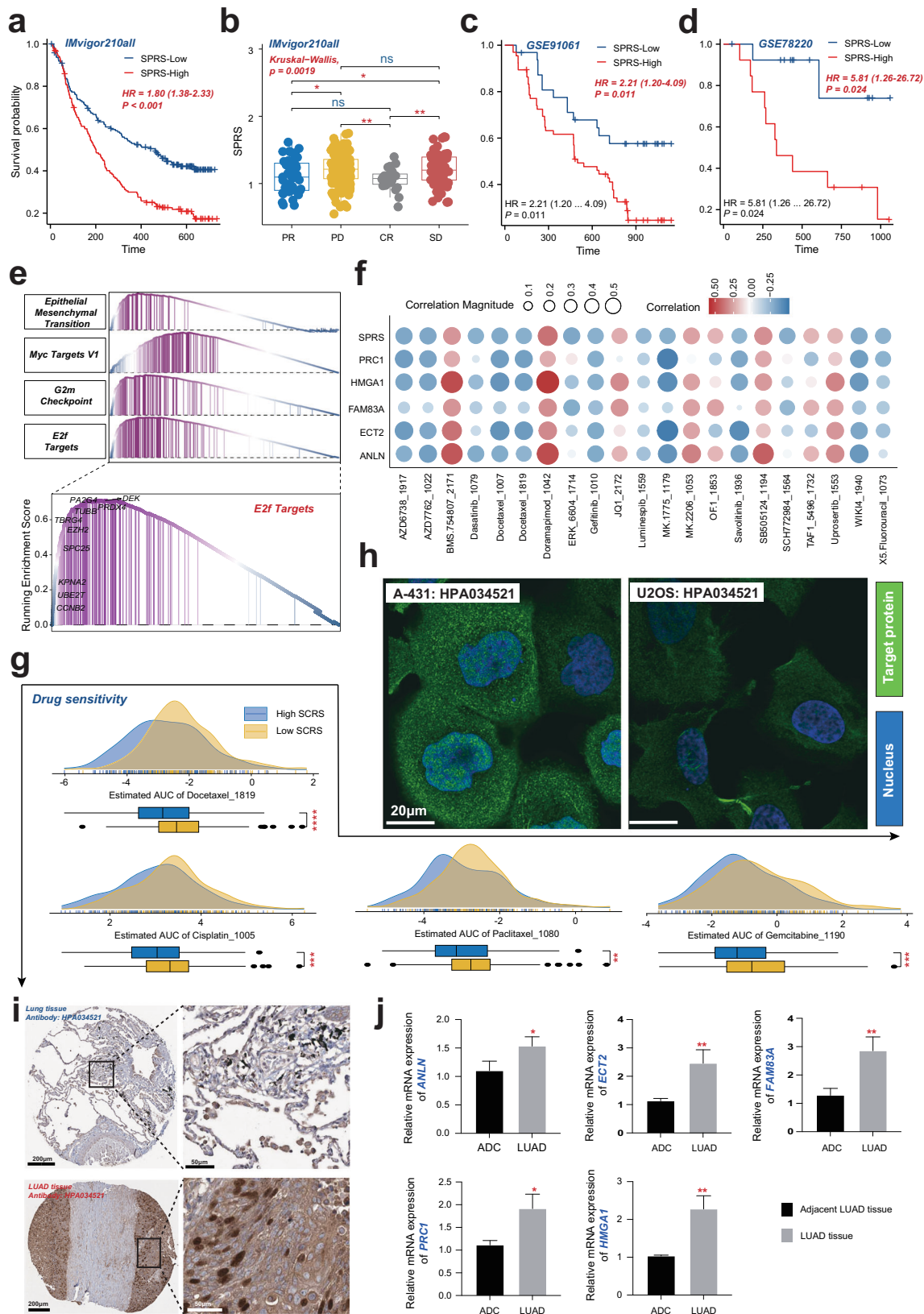
Scissor+ proliferating cells may regulate tumor progression through several mechanisms. They can drive cell cycle progression via the G2M checkpoint<sup>25</sup>, promoting genomic instability and tumor growth. These cells may also express genes associated with EMT, such as COL5A2<sup>26</sup> and ITGAV<sup>27</sup>, which facilitate tumor invasion and metastasis. Our analysis implicates IL1B as a key ligand in regulating the specific phenotype of Scissor+ proliferating cells. The identified IL1B receptors (IL1RL1, ADRB2, IL1R2, and IL1R1) alongside an IL1B ligand-receptor score exhibit significant expression patterns correlating with cell cycle and proliferation-associated genes like CCND1 and MKI67, specifically within C2\_MMP9 and C3\_KRT9 subgroups. This suggests a crucial role for IL1B in maintaining the phenotypic characteristics of Scissor+ proliferating cells. Furthermore, the FN1-CD44 axis emerges as a significant pathway through which Scissor+ proliferating cells interact with other cell types in the TIME, which potentially enhances their proliferative and invasive capabilities. The bidirectional signaling mediated by FN1 and CD44 underscores the complex intercellular dynamics that contribute to tumor progression and the establishment of an immunosuppressive microenvironment. To substantiate our mechanistic hypotheses regarding IL1B's role in regulating Scissor+ proliferating cells, future studies should engage experimental validation techniques. In vitro functional assays could involve the knock-down of IL1B to assess changes in the proliferation and apoptosis of Scissor+ cell subpopulations, while in vivo models utilizing IL1B receptor inhibitors could provide insights into tumor growth and immune microenvironment remodeling.

The current study constructed a SPRS signature to predict the prognosis and individualized therapeutic response of patients with LUAD. This model deepens our comprehension of the genetic intricacies surrounding proliferating cell-associated genes in LUAD and provides a basis for developing targeted prevention and treatment strategies. Among the



**Fig. 9 | Spatial transcriptomics analysis of SPRS in LUAD.** **a** UMAP plots showing spots from all sections, color-coded according to their sample source. **b** Signature-based strategy to assess the enrichment of various cell types within each spot. **c** Comparisons of SPRS positive cell ratios between cancer and noncancerous tissues.

**d** Expression of the five SPRS model genes and SPRS score in each spot across samples. LUAD lung adenocarcinoma, SPRS scissor+ proliferating cell risk score, UMAP uniform manifold approximation and projection.



**Fig. 10 | Analysis of SPRS in LUAD and its implications for immunotherapy and chemotherapy response.** **a** Kaplan–Meier survival curves for patients with low and high SPRS from the IMvigor210 cohort, demonstrating differential survival probabilities. **b** Graphical representation of patient responses to immunotherapy, categorized by SPRS levels, with statistical significance indicated. **c**, **d** Comparative survival analysis from two independent cohorts (GSE91061 and GSE78220) showing the association between SPRS levels and survival outcomes. **e** GSEA plot showing the enrichment of cell cycle-related pathways in high SPRS LUAD samples. **f** Correlations

between SPRS/model genes and IC50 of specific drugs from GDSC database. Red and blue dots indicate positive and negative correlations, respectively. **g** Ridge plot depicting the distribution of AUC values for various drugs, suggesting a link between SPRS and drug sensitivity. **h** Cellular localization of PRC1 protein in A-431 and U2OS from the HPA database. **i** Relative protein levels of PRC1 in LUAD tissue compared to a normal tissue, as determined by the HPA database. **j** Relative mRNA expression levels of five SPRS model genes in LUAD tissues compared to normal tissues, as determined by RT-qPCR. LUAD, Lung Adenocarcinoma; SPRS, Scissor+ proliferating cell risk score.



**Fig. 11 | Multi-omics characterization of proliferating cell dynamics and their clinical implications in LUAD.** **a** Cohort composition and staging distribution of the integrated LUAD dataset ( $n = 73$  patients: 32 Stage I, 5 Stage II, 8 Stage III, 28 Stage IV). **b** Uniform Manifold Approximation and Projection (UMAP) of 259,254 high-quality single-cell profiles from LUAD patients, annotated into 15 distinct cell subpopulations based on canonical marker genes. **c** Stratification of cells into SPRS\_high and SPRS\_low groups based on median AddModuleScore values. **d** Tissue preference analysis (Ro/e algorithm) revealing stage-dependent enrichment of proliferating cells in advanced LUAD (Stages II–IV) as well as relative abundance

of macrophages, cancer-associated fibroblasts (CAFs), and proliferating cells in SPRS\_high tumors. **e** Forest plot of meta-analysis (14 datasets,  $n = 2491$  patients) demonstrating consistent association between proliferating cell marker expression and poor prognosis. **f** Spatial transcriptomics (ST) tissue preference of various samples in SPRS\_high and SPRS\_low tumors highlighted. **g** CellTrek-based deconvolution of ST spots. **h** Quantification of cell-type proportions in SPRS\_high versus SPRS\_low tumors, confirming immunosuppressive microenvironment features (elevated CAFs, reduced CD8 + T cells) observed in single-cell data.

Importantly, multiple studies have demonstrated that key genes in the SPRS model significantly contribute to immune evasion beyond their roles in cell proliferation. FAM83A promotes PD-L1 expression by regulating the EGFR/MAPK signaling pathway and downregulating antitumor immune pathways such as MHC-I and interferon signaling, thereby impairing the ability of immune cells to recognize tumors<sup>35,36</sup>; ANLN indirectly enhances the expression of immune checkpoint molecules and reduces tumor antigen exposure by affecting the cytoskeleton and related signaling pathways, such as PI3K/AKT and Wnt, which in turn facilitates immune evasion<sup>37,38</sup>; HMGA1 reinforces immunosuppressive conditions by remodeling chromatin and regulating the transcription of immune/inflammatory-related genes, thereby increasing the recruitment of immunosuppressive cells such as tumor-associated macrophages and Tregs, resulting in a “cold” tumor microenvironment that suppresses CD8 + T cell and NK cell activity<sup>39,40</sup>; The upregulation of ECT2 is associated with increased Treg cell infiltration and the enhanced expression of immunosuppressive factors, leading to a diminished immune response<sup>33,41</sup>; The aberrant high expression of PRC1 affects the infiltration of tumor-associated immune cells and antigen presentation, promoting the formation of an immunosuppressive microenvironment and reducing T cell activity<sup>42,43</sup>. In summary, these five genes collaborate through various mechanisms to synergistically enhance the ability of lung cancer to evade immune surveillance.

As highlighted in our study, PRC1 is overexpressed in LUAD tissues. The elevated protein levels in cancerous tissues compared to normal tissues position it as a compelling candidate for further inquiry. Future studies should aim to assess the clinical validity of PRC1 in larger, more diverse patient cohorts, evaluating its prognostic utility across different stages of LUAD. Moreover, mechanistic studies should be undertaken to elucidate the precise pathways through which PRC1 exerts its effects within the tumor microenvironment. This could involve examining the interaction of PRC1 with other cellular components within the TIME of LUAD, its regulation at the transcriptional and post-transcriptional levels, and its potential as a predictive biomarker for response to various treatment modalities. In addition, the development of targeted therapies against PRC1 could be explored, with a focus on inhibiting its function to impede tumor growth and enhance treatment responses.

To substantiate our bioinformatic discoveries, we conducted RT-qPCR analyses to assess the relative expression of key model genes across both adjacent non-neoplastic and neoplastic tissues. These analyses confirmed our initial findings with consistency. However, to enhance our comprehension, further exploration at the protein level is imperative. In evaluating the clinical prognostic significance of SPRS in LUAD, we compared the C-index of SPRS against 30 other LUAD prognostic signatures. Notably, the SPRS model demonstrated superior predictive accuracy across all cohorts. Additionally, we developed several nomograms incorporating the SPRS scores and pertinent clinical variables. Both the prognostic nomogram and subsequent ROC analyses revealed that SPRS possesses robust predictive capabilities for 1-year, 3-year, and 5-year overall survival rates in LUAD patients. Yet, given the intricacies of LUAD, more extensive research is essential to elucidate the SPRS model’s function across various LUAD tissue subtypes.

Tumors are characterized by immune evasion, a trait that permits malignant cells to sidestep the host’s immune surveillance, fostering tumor progression and metastasis<sup>2,8,10,16,44–46</sup>. Intriguingly, tumors with elevated

SPRS levels showed diminished infiltration of anti-tumor immune cells, such as CD4/CD8 + T lymphocytes and NK cells, compared to those with lower SPRS levels. Conversely, immunosuppressive cell types, including macrophages, CAFs and MDSCs, were upregulated in tumors with high SPRS. This immunosuppressive cell distribution profile, marked by an abundance of fibroblasts and MDSCs in patients with high SPRS, suggests that these tumors exhibit a more immunosuppressive phenotype.

Immunotherapy presents LUAD patients with the potential to prolong survival<sup>8,18</sup>. Our validation across various immunotherapy cohorts revealed that patients with low SPRS levels responded more favorably to immunotherapy. This observation aligns with our analysis and indicates that SPRS could be instrumental in identifying populations likely to benefit from immunotherapy. Given the poor prognosis associated with high SPRS levels in immunotherapy, we utilized the GDSC database to systematically identify potential therapeutic agents. Our findings suggest that LUAD patients with high SPRS may respond to chemotherapeutic and targeted agents, including cisplatin, paclitaxel, docetaxel, and gemcitabine.

Our study offers novel perspectives on the role of proliferating cells in pulmonary pathologies and sheds light on the clinical implications of the SPRS signature. However, several limitations warrant acknowledgment. Firstly, our analysis was primarily based on bioinformatics, focusing on five proliferating-cell-associated genes within the signature model and SPRS. These genes were further examined using RT-qPCR in the clinical samples. To substantiate these findings, additional immunohistochemistry and western blotting experiments using clinical specimens are necessary. Validation of both mRNA and protein levels in larger clinical samples is crucial to firmly establish their clinical significance. Secondly, the findings from our bioinformatics analysis require further experimental validation through both *in vitro* and *in vivo* studies. Future research should include patients undergoing immunotherapy, along with comprehensive demographic data to confirm the prognostic accuracy of the SPRS signature and elucidate its relationship with ICB responsiveness and survival outcomes in LUAD patients. Moreover, the specific roles of IL1B and the FN1-CD44 signaling pathway in the regulation of Scissor+ proliferating cells must be experimentally validated to clarify their contributions to tumor dynamics and the immune microenvironment. Finally, there may be biases arising from data heterogeneity across different datasets (e.g. TCGA and GEO), which could affect the generalizability of our findings. Larger, multicentric, prospective studies are needed to strengthen the clinical utility of the SPRS model.

In conclusion, our study provides a robust framework for understanding the complex dynamics of proliferating cells in the development of lung diseases, particularly for LUAD. The integration of multi-omics data and machine learning not only enhances our prognostic capabilities but also opens new avenues for targeted therapeutic interventions. These findings underscore the potential of SPRS and its constituent genes as critical tools in the advancement of personalized treatment strategies for LUAD patients. Future research should focus on validating these findings in larger cohorts and exploring the mechanistic pathways through which these genes influence LUAD progression and treatment response.

## Methods

### Data acquisition and preprocessing

This study leverages scRNA-seq data extracted from a curated collection of 93 human lung samples, providing a detailed examination of cellular

diversity and temporal changes within the pulmonary microenvironment across a continuum of pathologies, from normal to LUAD. Our dataset includes 28 healthy lung samples of transplant donors from the GSE131907 dataset<sup>44</sup>, and an additional set of 18 COPD, 32 IPF, and 15 LUAD samples, all sourced from the GSE136831 dataset<sup>47</sup>.

RNA-sequencing (RNA-seq) expression profiles along with detailed clinical and pathological data of LUAD patients were retrieved from The Cancer Genome Atlas (TCGA). To expand our analysis, we included gene expression data and comprehensive clinical and pathological information from three additional LUAD cohorts (GSE31210, GSE50081, and GSE72094)<sup>48–50</sup>, and two immunotherapy-focused cohorts, GSE91061 and GSE78220<sup>51,52</sup>. The IMvigor210 dataset was also utilized to validate the immunotherapy responsiveness of our prognostic model<sup>53</sup>. The detailed depiction of the clinical characteristics of patients encompassed in our study is presented in Table S1–2. Furthermore, we have incorporated ST data from ten samples, corresponding to both normal lung tissues and LUAD tissues, retrieved from the E-MTAB-13530 cohort<sup>45</sup>. To further support our investigation, drug sensitivity data were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database<sup>54</sup>.

To validate the impact of proliferating cells on LUAD progression and the TIME in a larger cohort, we combined 54 samples from the GSE131907 cohort with 15 LUAD samples from the testing set of GSE136831 to form a validation set. To verify the differential protein levels of SPRS model genes between normal and LUAD samples, we utilized the CPTAC-LUAD dataset (<https://cptac-data-portal.georgetown.edu/datasets>). Reverse-phase protein array (RPPA) data were downloaded from the TCGA database<sup>55</sup>. Based on previously published studies<sup>55</sup>, we calculated the pathway activity scores of the target genes for 10 cancer-related pathways (TSC/mTOR, RTK, RAS/MAPK, PI3K/AKT, hormone ER, hormone AR, EMT, DNA damage response, cell cycle, and apoptosis pathways) using the sparkle (<https://grswsci.top>) database to further explore the functional roles of these genes. The Spearman correlation and *p*-values between the target genes and pathway activity scores were calculated using the `cor.test` function.

### scRNA-seq data processing

For the processing of scRNA-seq data, we predominantly utilized the ‘Seurat’ R package (version 4.3.1)<sup>46</sup>. Our approach encompassed a thorough quality control workflow, integrating a series of function calls including “NormalizeData,” “FindVariableFeatures,” “ScaleData,” “RunPCA,” “FindNeighbors,” “FindClusters,” “RunUMAP,” and “FindAllMarkers.” This standardized methodology allowed us to curate a dataset comprising over 360,000 cells for subsequent analyses. Cell type annotation was achieved using the “SingleR” package, combining its capabilities with established literature markers<sup>56</sup>. To systematically quantify the enrichment of various cell clusters across distinct pathological groups, we computed Ro/e values using the “epitools” R package<sup>16</sup>. The biological functionality of marker genes for each cell type was evaluated utilizing the “ClusterGVis” and “org.Hs.eg.db” R packages alongside the “erichCluster” function<sup>57</sup>. Additionally, we employed the “AddModuleScore” function to assess expression patterns across different proliferating cell subsets. Pseudotime analysis, crucial for inferring developmental trajectories of identified cell populations, was conducted with the “SCTOUR”<sup>58</sup> and the slingshot algorithms<sup>59</sup>. Moreover, we implemented the “Scissor” R package<sup>60</sup> to identify proliferating subsets with significant associations to poor prognosis in LUAD. Scissor integrates phenotype-linked bulk RNA-seq with scRNA-seq data to pinpoint subpopulations most correlated with the clinical phenotype of interest, in our study, patient death. The approach quantifies similarity between each single cell and each bulk sample via Pearson correlation, then fits a regularized regression model that links these similarities to the phenotype using the Cox regression for survival data. Cells with nonzero regression coefficients are classified as Scissor+ (positively associated with the phenotype) or Scissor- (negatively associated). By leveraging bulk clinical data, Scissor efficiently identifies clinically relevant cellular subpopulations even when single-cell data alone may lack power, making it a robust tool for dissecting cellular heterogeneity in complex diseases<sup>7</sup>.

To further dissect the interactions between different proliferating cell subsets, we used the R package ‘CellChat’ (version 1.1.1)<sup>61</sup>. Using the “Secreted Signaling” category, we calculated communication probabilities and networks with the `computeCommunProb`, `filterCommunication`, and `computeCommunProbPathway` functions. The overall cell-cell communication network was aggregated using `aggregateNet`, and specific signaling pathways were visualized in heatmaps with the `netVisual_heatmap` function.

For these prognosis-linked proliferating subsets, a more in-depth NicheNet (version 1.0.0)<sup>62</sup> analysis was performed to predict potential ligands driving their transcriptomic changes and distinctive phenotypes, following previously established procedures<sup>63</sup>. The top 20 ligands, receptors, and target genes, ranked by `aupr_corrected`, were visualized in heatmaps.

### Spatial transcriptomics analysis

Employing the “Seurat” package (version 4.4.1), we meticulously reanalyzed ST data from three normal lung tissues, two non-tumorigenic tissues, and five LUAD tissues. The quality control process was executed with stringent precision, closely following established methodologies from the literature<sup>14,64</sup>. After stringent quality control to filter low-quality spots, we applied the SCTransform method for normalization and variance stabilization. Subsequently, the Harmony algorithm was used to integrate data across samples. Subsequently, the SPRS signature scores were integrated into the metadata of the ST dataset using the “AddModuleScore” function, with default parameters specified by the package. To visualize the spatial distribution of gene expression, we employed the “SpatialFeaturePlot” function from the Seurat package. Spots with an SPRS score greater than zero were designated as positive. Additionally, we utilized the Nebulosa function to visualize the joint density of genes specific to different cell subpopulations, thereby inferring their spatial localization within the tissue.

### Cellular components deconvolution using CellTrek

To acquire the spatial coordinates of the cell types from the scRNA cohort, we initially integrated the ST and scRNA-seq expression data using the CellTrek R package<sup>65</sup>, employing the standard settings for co-dimension reduction. We subsequently generated a sparse graph leveraging a random forest algorithm and established a spot-cell similarity matrix for individual cells, incorporating spatial coordinate data to enhance the analysis.

### Constructing a prognostic signature using multiple machine-learning algorithms

To achieve data standardization and improve comparability across different cohorts, we initially normalized all datasets using Z-scores. The TCGA cohort, with its comprehensive treatment information, served as the training set to explore the relationship between Scissor+ proliferating cell genes, treatment strategies, and prognostic outcomes. Other cohorts were used as validation sets. Due to the limited sample sizes of the GSE31210, GSE50081, and GSE72094 cohorts, they were combined into a META-LUAD cohort, with batch effects addressed using the “ComBat” algorithm from the “SVA” package.

In developing our model, we identified 2328 genes significantly upregulated in the TCGA-LUAD cohort. These genes were intersected with markers specific to the Scissor+ group to refine the gene set for modeling. We employed the MIME R package to integrate these genes into 111 different machine learning algorithms for model construction<sup>66</sup>. For individual algorithms, model genes were selected based on their significance, and the predictive models were constructed directly using the inherent methodologies of each algorithm. For combined algorithm approaches, such as the LASSO+SuperPC combination, a two-step process was employed: initially, the LASSO algorithm was used to filter the model genes based on their significance, followed by the application of SuperPC to build the predictive model. Each algorithmic configuration was rigorously trained on the TCGA cohort and subsequently tested on the GEO-LUAD and META-LUAD cohorts to ensure robustness and prevent overfitting. Our objective was to

comprehensively evaluate the effectiveness of various machine learning models across multiple datasets in order to identify the most reliable models for predicting outcomes in LUAD. For each LUAD patient, the SPRS was calculated based on the established model. Patients were classified into high or low SPRS groups using the median SPRS as a threshold. To evaluate the prognostic significance of SPRS, we employed the Kaplan-Meier (KM) survival analysis using the “survminer” R software package, with statistical significance determined using the log-rank test at a threshold of  $p < 0.05$ <sup>14</sup>. Additionally, we compared the prognostic ability of the SPRS model and other 30 established LUAD-related prognostic signatures using the MIME package.

Subsequently, we endeavored to construct prognostic nomogram models<sup>67</sup> by integrating the SPRS with relevant survival-related clinical parameters in both the training TCGA-LUAD cohort and testing cohorts. To identify the constructed models' predictive fidelity, calibration curve analysis and receiver operating characteristic (ROC) curve analysis were applied using the “rms”, “survival”, and “timeROC” R packages<sup>14,68</sup>. To further examine the consistency of findings across diverse cohorts, we employed the  $I^2$  statistic, a measure that estimates the proportion of total variability in study outcomes attributable to heterogeneity rather than random error. An  $I^2$  value exceeding 50% was interpreted as evidence of substantial heterogeneity. To scrutinize the robustness of our results, we conducted sensitivity analyses by sequentially omitting individual datasets and re-estimating the pooled hazard ratios (HRs) and corresponding confidence intervals (CIs). This approach facilitated the detection of any dataset that may unduly sway the aggregate findings. Additionally, we assessed potential publication bias through visual inspection of funnel plots and by conducting Egger's regression test<sup>69</sup>. A  $p$ -value below 0.05 in Egger's test was deemed to signify significant publication bias. For these analyses, the  $I^2$  statistic and Egger's regression test were computed using the R package “metafor” (version 2.4.0), while Cox regression analysis was performed with the “survival” package (version 3.3-1).

### Alteration analysis of the SPRS model genes

To elucidate the mutational profiles of SPRS model genes, we initially examined the genomic copy number alterations in 451 samples from the TCGA-COAD project, utilizing GISTIC scores for analysis. Additionally, we assessed the distribution of FGL, FGG, and FGA across different expression subgroups of FAM83A in LUAD, employing the TukeyHSD function in R (version 4.3.1). To further investigate whether FAM83A expression levels are associated with specific mutational genotypes, we applied a permutation test-based approach, leveraging the `independence_test` function from the `coin` package to evaluate the independence between gene expression and mutational status. For comprehensive analysis, RNA sequencing data, mutation data, and copy number alteration (CNA) data for TCGA-LUAD patients were retrieved from the cBioPortal database (<http://www.cbioportal.org/>)<sup>70</sup>, which enabled us to construct a detailed mutational landscape of SPRS model genes in LUAD.

### In-depth exploring of the immunological landscape and immunotherapy responses of SPRS

To comprehensively analyze the immunological differences between patients with high- and low-SRBS, we employed the Immuno-Oncology Biology Research (IOBR) R package<sup>71</sup> to calculate enrichment scores for each sample in TCGA-LUAD cohorts. We then compared the distribution variances of tumor mutational burden (TMB), tumor neoantigen burden (TNB), and immunological components such as cancer-associated fibroblasts (CAFs) and myeloid-derived suppressor cells (MDSCs) between these cohorts. Using SPRS, we regrouped patients for survival analysis. The effectiveness of the immunotherapy response was assessed by scrutinizing the survival delay in patients responding to immunotherapy. This analysis was validated and further enriched with data from additional cohorts, including GSE78220 and GSE91061<sup>51,52</sup>, thereby ensuring a robust and multi-dimensional understanding of the relationship between SPRS and immunotherapy outcomes.

### Significance of the SPRS in drug sensitivity

We conducted a Gene Set Enrichment Analysis (GSEA) to identify pathways that were significantly enriched in tumor samples stratified by low and high SPRS scores<sup>72</sup>. Utilizing the GDSC dataset, we predicted drug sensitivity in LUAD patients based on the SPRS and model genes. The half-maximal inhibitory concentration (IC50) values for a variety of common drugs were determined using the “oncoPredict” package. To evaluate the differences in drug IC50 values between the high and low SPRS groups, we employed the Wilcoxon test, with statistical significance set at  $p < 0.05$ . A lower IC50 value is indicative of enhanced drug sensitivity.

### Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

Total RNA was extracted from LUAD clinical samples. This extraction process employed a cellular RNA-quick purification kit (10606ES60; EScience Biotech). Following RNA isolation, the cDNA was synthesized using Hifair® II 1st Strand cDNA Synthesis Kit (11119ES60, EScience Biotech). The RT-qPCR analysis was conducted on the LightCycler® 480 system (Roche) with an RT-qPCR kit (11203ES03, EScience Biotech). The relative mRNA expression levels were calculated using the comparative Ct method by normalizing gene expression with the corresponding  $\beta$ -actin, and the  $\Delta$ Ct value was calculated as follows:  $\Delta$ Ct = Ct ( $\beta$ -actin) – Ct (target gene). All data were presented as the mean  $\pm$  SEM from three independent experiments. Primer sequences used in this study are listed in Table S10.

Additionally, we also investigated the cellular location and the protein level of PRC1 between the normal and LUAD tissues using the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>).

### Statistical analysis

All statistical analyses were conducted using R software, version 4.2.1. We employed the Student's t-test to assess differences in normally distributed continuous variables between two groups. For variables that did not adhere to a normal distribution, the Wilcoxon test was utilized. When comparing more than two groups, we applied a one-way ANOVA test for parametric data or a Kruskal–Wallis test for non-parametric data. Spearman's correlation analysis was used to evaluate the relationships between continuous variables. Kaplan-Meier survival analysis was employed to plot survival curves, and the log-rank test was used to analyze the data. A two-tailed  $p$ -value threshold of 0.05 was established to determine statistical significance.

### Data availability

The datasets used in this paper are available online, as described in the “Methods” section. No new algorithms were developed for this article. All code generated for analysis is available from the authors upon request.

### Code availability

No new algorithms were developed for this article. All code generated for analysis is available from the authors upon request.

Received: 9 October 2024; Accepted: 23 June 2025;

Published online: 18 July 2025

### References

1. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *Ca Cancer J. Clin.* **73**, 17–48 (2023).
2. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).
3. Lee, J. W. et al. The combination of MEK inhibitor with immunomodulatory antibodies targeting programmed death 1 and programmed death ligand 1 results in prolonged survival in KRAS/p53-driven lung cancer. *J. Thorac. Oncol.* **14**, 1046–1060 (2019).
4. Shiraiishi, Y., Sekino, Y., Horinouchi, H., Ohe, Y. & Okamoto, I. High incidence of cytokine release syndrome in patients with advanced NSCLC treated with nivolumab plus ipilimumab. *Ann. Oncol.* **34**, 1064–1065 (2023).

5. Topalian, S. L. et al. Five-year survival and correlates among patients with advanced melanoma, renal cell carcinoma, or non-small cell lung cancer treated with nivolumab. *Jama Oncol.* **5**, 1411–1420 (2019).
6. Sarode, P. et al. Reprogramming of tumor-associated macrophages by targeting beta-catenin/FOSL2/ARID5a signaling: a potential treatment of lung cancer. *Sci. Adv.* **6**, eaaz6105 (2020).
7. Zhang, J., Hu, D., Fang, P., Qi, M. & Sun, G. Deciphering key roles of B cells in prognostication and tailored therapeutic strategies for lung adenocarcinoma: a multi-omics and machine learning approach towards predictive, preventive, and personalized treatment strategies. *EPMA J.* **16**, 127–163 (2025).
8. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
9. Eissmann, M. F. et al. Il-33-mediated mast cell activation promotes gastric cancer through macrophage mobilization. *Nat. Commun.* **10**, 2735 (2019).
10. Chen, S. et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat. Cell Biol.* **23**, 87–98 (2021).
11. Pelka, K. et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752 (2021).
12. Liu, Y. et al. Metabolic reprogramming in the tumor microenvironment: unleashing T cell stemness for enhanced cancer immunotherapy. *Front. Pharmacol.* **14**, 1327717 (2023).
13. Marelli-Berg, F. M., Fu, H. & Mauro, C. Molecular mechanisms of metabolic reprogramming in proliferating cells: implications for T-cell-mediated immunity. *Immunology* **136**, 363–369 (2012).
14. Hu, D. et al. Multi-omic profiling reveals potential biomarkers of hepatocellular carcinoma prognosis and therapy response among mitochondria-associated cell death genes in the context of 3p medicine. *EPMA J.* **15**, 321–343 (2024).
15. Yang, Y. et al. Pan-cancer single-cell dissection reveals phenotypically distinct B-cell subtypes. *Cell* **187**, 4790–4811 (2024).
16. Xue, R. et al. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. *Nature* **612**, 141–147 (2022).
17. Qin, Y., Pu, X., Hu, D. & Yang, M. Machine learning-based biomarker screening for acute myeloid leukemia prognosis and therapy from diverse cell-death patterns. *Sci. Rep.* **14**, 17874 (2024).
18. Wang, S. et al. Machine learning reveals diverse cell death patterns in lung adenocarcinoma prognosis and therapy. *Npj Precis. Oncol.* **8**, 49 (2024).
19. Bulle, A. & Lim, K. H. Beyond just a tight fortress: contribution of stroma to epithelial-mesenchymal transition in pancreatic cancer. *Signal Transduct. Target. Ther.* **5**, 249 (2020).
20. Sudmeier, L. J. et al. Distinct phenotypic states and spatial distribution of CD8 (+) T cell clonotypes in human brain metastases. *Cell Rep. Med.* **3**, 100620 (2022).
21. Conway, J. R., Kofman, E., Mo, S. S., Elmarakeby, H. & Van Allen, E. Genomics of response to immune checkpoint therapies for cancer: implications for precision medicine. *Genome Med.* **10**, 93 (2018).
22. Araujo, H. A. et al. Mechanisms of response and tolerance to active ras inhibition in KRAS-mutant non-small cell lung cancer. *Cancer Discov.* **14**, 2183–2208 (2024).
23. Saad, H. M. et al. The potential role of MUC16 (CA125) biomarker in lung cancer: a magic biomarker but with adversity. *Diagnostics* **12**, 2985 (2022).
24. Gialeli, C. et al. Complement inhibitor CSMD1 modulates epidermal growth factor receptor oncogenic signaling and sensitizes breast cancer cells to chemotherapy. *J. Exp. Clin. Cancer Res.* **40**, 258 (2021).
25. Xie, Q. et al. Multi-omics analysis identifies glioblastoma dependency on H3K9ME3 methyltransferase activity. *Npj Precis. Oncol.* **9**, 78 (2025).
26. Wu, L. et al. Prognostic value of EMT gene signature in malignant mesothelioma. *Int. J. Mol. Sci.* **24**, 4264 (2023).
27. Kemper, M. et al. Integrin alpha-v is an important driver in pancreatic adenocarcinoma progression. *J. Exp. Clin. Cancer Res.* **40**, 214 (2021).
28. Yuan, W. et al. Screening and identification of miRNAs negatively regulating FAM83a/WNT/beta-catenin signaling pathway in non-small cell lung cancer. *Sci. Rep.* **14**, 17394 (2024).
29. Wang, T., Zhu, X. & Wang, K. CircMIIP contributes to non-small cell lung cancer progression by binding miR-766-5p to upregulate FAM83a expression. *Lung* **200**, 107–117 (2022).
30. Cao, Q. et al. LncRNA cytor promotes lung adenocarcinoma gemcitabine resistance and epithelial-mesenchymal transition by sponging mir-125a-5p and upregulating anln and RRM2. *Acta Biochim. Biophys. Sin.* **56**, 210–222 (2024).
31. Suzuki, C. et al. ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. *Cancer Res.* **65**, 11314–11325 (2005).
32. Saed, L., Jelen, A., Mirowski, M. & Salagacka-Kubiak, A. Prognostic significance of HMGA1 expression in lung cancer based on bioinformatics analysis. *Int. J. Mol. Sci.* **23**, 6933 (2022).
33. Kosibaty, Z., Murata, Y., Minami, Y., Noguchi, M. & Sakamoto, N. Ect2 promotes lung adenocarcinoma progression through extracellular matrix dynamics and focal adhesion signaling. *Cancer Sci.* **112**, 703–714 (2021).
34. Zhan, P. et al. Prc1 contributes to tumorigenesis of lung adenocarcinoma in association with the Wnt/beta-catenin signaling pathway. *Mol. Cancer* **16**, 108 (2017).
35. Yuan, Y. et al. Promotion of stem cell-like phenotype of lung adenocarcinoma by FAM83A via stabilization of ErbB2. *Cell Death Dis.* **15**, 460 (2024).
36. Fei, W. et al. High-risk histological subtype-related FAM83a hijacked FOXM1 transcriptional regulation to promote malignant progression in lung adenocarcinoma. *PeerJ* **11**, e16306 (2023).
37. Saito, Y. et al. A therapeutically targetable taz-tead2 pathway drives the growth of hepatocellular carcinoma via ANLN and KIF23. *Gastroenterology* **164**, 1279–1292 (2023).
38. Guo, E. et al. Alternatively spliced anln isoforms synergistically contribute to the progression of head and neck squamous cell carcinoma. *Cell Death Dis.* **12**, 764 (2021).
39. Zhong, C., Zhang, Q., Bao, H., Li, Y. & Nie, C. Hsa\_circ\_0054220 upregulates HMGA1 by the competitive RNA pattern to promote neural impairment in MPTP model of Parkinson's disease. *Appl. Biochem. Biotechnol.* **196**, 4008–4023 (2024).
40. Li, K. J. et al. Nat10 promotes prostate cancer growth and metastasis by acetylating mRNAs of HMGA1 and KRT8. *Adv. Sci.* **11**, e2310131 (2024).
41. Lv, E., Sheng, J., Yu, C., Rao, D. & Huang, W. Long noncoding RNA mapkap5-as1 promotes metastasis through regulation mir-376b-5p/ect2 axis in hepatocellular carcinoma. *Dig. Liver Dis.* **55**, 945–954 (2023).
42. Uribe, M. L. et al. TSHZ2 is an EGF-regulated tumor suppressor that binds to the cytokinesis regulator PRC1 and inhibits metastasis. *Sci. Signal.* **14**, ea6e156 (2021).
43. Ciapponi, M., Karlukova, E., Schkolziger, S., Benda, C. & Muller, J. Structural basis of the histone ubiquitination read-write mechanism of rybp-prc1. *Nat. Struct. Mol. Biol.* **31**, 1023–1027 (2024).
44. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).
45. De Zuani, M. et al. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat. Commun.* **15**, 4388 (2024).
46. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
47. Adams, T. S. et al. Single-cell rna-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, ea61983 (2020).
48. Der, S. D. et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage Ia patients. *J. Thorac. Oncol.* **9**, 59–64 (2014).

49. Okayama, H. et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* **72**, 100–111 (2012).
50. Schabath, M. B. et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* **35**, 3209–3216 (2016).
51. Hugo, W. et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).
52. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949 (2017).
53. Mariathasan, S. et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
54. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
55. Liu, C. J. et al. GSCA: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic and immunogenomic levels. *Brief. Bioinform.* **24**, bbac558 (2023).
56. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
57. Zhang, J. Clustergvis: one-step to cluster and visualize gene expression matrix. <https://github.com/junjunlab/ClusterGVis> (2022).
58. Li, Q. Sctour: a deep learning architecture for robust inference and accurate prediction of cellular dynamics. *Genome Biol.* **24**, 149 (2023).
59. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
60. Sun, D. et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat. Biotechnol.* **40**, 527–538 (2022).
61. Jin, S., Pliikus, M. V. & Nie, Q. Cellchat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat. Protoc.* **20**, 180–219 (2025).
62. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
63. Cheng, S. et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809 (2021).
64. Wu, R. et al. Comprehensive analysis of spatial architecture in primary liver cancer. *Sci. Adv.* **7**, eabg3750 (2021).
65. Wei, R. et al. Spatial charting of single-cell transcriptomes in tissues. *Nat. Biotechnol.* **40**, 1190–1199 (2022).
66. Liu, H. et al. Mime: a flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection. *Comp. Struct. Biotechnol. J.* **23**, 2798–2810 (2024).
67. Balachandran, V. P., Gonen, M., Smith, J. J. & Dematteo, R. P. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* **16**, e173–e180 (2015).
68. Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).
69. Cheng, J., He, Z., Jing, J., Liu, Y. & Zhang, H. *Integrating Machine Learning and Multi-omics to Identify Key SUMOylation Molecular Signature in Sarcoma*. (Life Conflux, 2024).
70. de Bruijn, I. et al. Analysis and visualization of longitudinal genomic and clinical data from the AACR project genie biopharma collaborative in CBioPortal. *Cancer Res.* **83**, 3861–3867 (2023).
71. Zeng, D. et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* **12**, 687975 (2021).
72. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

## Acknowledgements

This study was supported by a grant from the Clinical Research Project of Shanghai Municipal Health Commission (202140473) and a grant from the Health System Peak Discipline Construction Project of Xuhui District, Shanghai (SHXHZDXK202312).

## Author contributions

All authors searched the literature, designed the study, interpreted the findings, and revised the manuscript. Shun Wang, Ruohuang Wang, and Dingtao Hu carried out data management and statistical analysis and drafted the manuscript. Ruohuang Wang and Dingtao Hu helped with cohort identification and data management. Shun Wang contributed to the critical revision of the manuscript. Jie Huang and Baoqing Wang were responsible for conceptualization, data collection, curation, supervision, writing review, and editing. Shun Wang, Ruohuang Wang, and Dingtao Hu contributed equally to this manuscript and should be considered as co-first authors.

## Competing interests

The authors declare no competing interests.

## Ethical statement

This study was reviewed and approved by the Ethics Committee of Shanghai Xuhui Central Hospital (Approval No.: 2022-021). Written informed consent was obtained from all participants prior to enrollment. All procedures strictly adhered to the principles of the Declaration of Helsinki and relevant national/international ethical guidelines.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-025-01027-z>.

**Correspondence** and requests for materials should be addressed to Jie Huang or Baoqing Wang.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025