Spatial dynamics of the tumor microenvironment in emerging resistance to targeted therapy in *EGFR*-mutated NSCLC

Satoshi Nakamura[1] Daisuke Shibahara[1,*], Kentaro Tanaka[1,2], Yasuyuki Kishikawa[1], Mikiko Hashisako[3], Keita Nakatomi[4], Noriaki Nakagaki[5], Mikihiro Kohno[6], Koichi Azuma[7], Ritsu Ibusuki[1], Kohei Otsubo[1], Yasuto Yoneshima[1], Eiji Iwama[1], Yoshinao Oda[3], Isamu Okamoto[1]

[1]Department of Respiratory Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

[2]Department of Pulmonary Medicine, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan

[3]Department of Anatomic Pathology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

[4]Department of Respiratory Medicine, Kyushu Central Hospital, Fukuoka, Japan

[5]Department of Respiratory Medicine, Fukuoka Red Cross Hospital, Fukuoka, Japan

[6]Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

[7]Division of Respirology, Neurology, and Rheumatology, Department of Internal Medicine, Kurume University School of Medicine, Fukuoka, Japan

*Corresponding Author: Daisuke Shibahara

Email address: shibahara.daisuke.805@m.kyushu-u.ac.jp

Postal address: 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

Disclosure

Daisuke Shibahara has received research funding from Eli Lilly Japan K.K. Koichi Azuma has received honoraria from AstraZeneca, MSD, Chugai Pharmaceutical, Ono Pharmaceutical, Bristol-Myers Squibb, Takeda Pharmaceutical, Taiho Pharmaceutical, and Amgen. Yasuto Yoneshima has received honoraria from AstraZeneca. Kentaro Tanaka has received honoraria from Chugai Pharmaceutical, AstraZeneca, Ono Pharmaceutical, Bristol-Myers Squibb, Eli Lilly, Takeda Pharmaceutical, Novartis Pharma K.K., Merck Biopharma, Kyowa Kirin, Daiichi-Sankyo, Pfizer, Amgen, Janssen Pharmaceutical K.K., and MSD, and is an advisory board member for Pfizer and Novartis Pharma K.K. Eiji Iwama has received honoraria from AstraZeneca and Chugai Pharmaceutical. Isamu Okamoto has received honoraria and research funding from Daiichi Sankyo, Chugai Pharmaceutical, Eli Lilly Japan K.K., AstraZeneca, Taiho Pharmaceutical, Boehringer Ingelheim, and Ono Pharmaceutical; honoraria from Takeda Pharmaceutical and Novartis Pharma K.K.; and research funding from Bristol-Meyers Squibb and MSD Oncology. Other authors declare no potential conflicts of interest.

__Abstract__

Epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) provide therapeutic benefit in *EGFR*-mutated non-small cell lung cancer (NSCLC); however, early resistance develops in some patients. To explore the underlying mechanisms, we performed spatial transcriptomics using Visium HD and StarDist processing on paired pre- and post-treatment samples from the same patients. In early resistance samples, fibroblasts increased in proportion and exhibited enhanced activation following EGFR-TKI treatment, while a distinct tumor cell cluster showed TNF-α/NF-kB pathway activation even prior to treatment. Furthermore, in the pre-treatment sample with early resistance, specific tumor clusters interacted with immune and stromal cells through inflammatory pathways, including MIF/CD74 and PDGF signaling. Immature tertiary lymphoid structures (TLSs) were enriched in early resistance, whereas mature TLSs were observed in long-term responders. These findings highlight the role of tumor heterogeneity and an inflammatory tumor microenvironment in early EGFR-TKI resistance, providing insights into potential therapeutic strategies to improve treatment outcomes.

__Introduction__

Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (EGFR-TKIs) demonstrate durable therapeutic efficacy in the majority of patients with *EGFR*-mutated non-small cell lung cancer (NSCLC)[1]. Nevertheless, resistance to EGFR-TKIs inevitably emerges over time after treatment initiation[2,3]. While several resistance mechanisms to EGFR-TKIs have been identified, the underlying causes of resistance remain unknown in 40-50% of cases[4–6]. Furthermore, some patients develop early resistance during treatment, whereas others achieve long-term responses, yet the reasons for these divergent outcomes remain unclear. Therefore, there is a critical need to elucidate the precise mechanisms of resistance not only to develop effective therapies for overcoming resistance but also to extend the duration of response to EGFR-TKIs. Increasing evidence suggests that the tumor microenvironment (TME) plays a pivotal role in shaping the response to EGFR-TKIs[7–12]; however, spatial analyses of TME using paired pre- and post-treatment samples from the same patients with *EGFR*-mutated NSCLC have been highly limited[13].

Spatial transcriptomics has emerged as a powerful approach to unravel the TME[13–18]. As an array-based method, Visium HD allows comprehensive mRNA capture without limiting the analysis to specific target genes, as well as allowing analysis of the entire sample, compared to in situ hybridization and sequencing methods. In this study, we conducted a detailed analysis of the TME in paired pre- and post-treatment samples from patients with *EGFR*-mutated NSCLC using Visium HD. Furthermore, we converted Visium HD spot data into single cell data using StarDist, a deep learning-based method for nuclei/cell detection and segmentation. By integrating the spatial coordinates derived from StarDist into the single-cell data, we conducted spatial

transcriptomic analysis at the single-cell level while preserving the actual spatial relationships between cells. Our findings elucidate the spatial dynamics of TME, highlighting the contribution of an inflammatory TME and tumor heterogeneity to the early onset of resistance mechanisms in *EGFR*-mutated NSCLC.

__Result__

Integrative analysis of spatial transcriptomics and nuclear segmentation

To investigate differences in the duration of response to osimertinib, a third-generation EGFR-TKI, we collected and analyzed paired pre- and post-treatment samples from four patients with *EGFR*-mutated NSCLC (Fig. 1a and Extended Data Fig. 1). Clinical characteristics of patients are listed in Supplementary Table 1. Based on each patient's progression-free survival duration[2], the samples were categorized into one early resistance (ER), one standard response (SR), and two long-term response (LR) groups. All samples were processed using Visium HD platform, which provides a spatial resolution of 2 µm per spot. To reconstruct single cells from the data, we employed StarDist[19], a deep learning-based tool, which enables the detection and segmentation of nuclei from hematoxylin and eosin (H&E) stained images generated alongside the spatial transcriptomics data (Fig 1b and Extended Data Fig. 2). Using StarDist, we aggregated 2 µm spots that overlapped with the polygonal coordinates corresponding to the detected nuclei, thereby reconstructing spatial transcriptomics data mapped to individual cell positions (Fig. 1c). Additionally, we incorporated the centroid coordinates of each polygon as new spot coordinates into the data structure to ensure compatibility with existing tools for spatial transcriptomics analysis (Extended Data Fig. 3a-c).

A total of 814,636,020 reads was obtained from Slide 1 (ER, SR), and 730,477,318 reads were obtained from Slide 2 (LR1, LR2). StarDist identified 141,324 nuclei in Slide 1 and 99,658 nuclei in Slide 2. Samples were partitioned based on coordinates from their location on the image, and the sample names were linked to the data (Extended Data Fig. 3d). The integrated data for these eight samples were then processed for quality control, filtering, normalization, dimensionality reduction, batch correction, nearest neighbor graph building, and Leiden clustering, and visualized using Uniform Manifold Approximation and Projection (UMAP). The cell population was broadly categorized into three main lineages: epithelial cells, stromal cells, and immune cells, with a small number of blood cells (Fig. 1d and Extended Data Fig. 3e). Within the epithelial cell population, tumor cells, basal cells, club cells, alveolar cell type 2 and goblet/ciliated cells were further divided (Fig. 1e and Extended Data Fig. 4a). Immune and stromal cells were clustered into lymphocytes, plasmablasts, macrophages, fibroblasts, endothelial cells, and smooth muscle cells (Fig. 1f,g and Extended Data Fig. 4b-c). Additionally, lymphocytes were subdivided, plasmablasts were classified by the immunoglobin they produced, and macrophages were distinguished from mixed cells (Extended Data Fig. 4d-i). The results of these clustering analyses were integrated with the polygonal coordinate data obtained from StarDist, enabling the visualization of distinct cell populations on spatial coordinates (Fig. 1h-k and Extended Data Fig. 5,6).

Changes of TME between pre- and post-treatment samples

To investigate changes in tumor cell characteristics before and after treatment, we performed differentially gene expression (DGE) analysis following pseudo-bulk processing, comparing tumor cells from pre- and post-treatment samples (Fig. 2a).

Genes with significantly higher expression in the pre-treatment group demonstrated a tendency toward favorable prognosis (Extended Data Fig. 7a-d), whereas genes with higher expression in the post-treatment group were associated with poorer prognosis (Extended Data Fig. 7e-h). Gene Set Enrichment Analysis (GSEA) revealed significant enrichment in gene sets associated with cell cycle and proliferation in post-treatment tumor cells (Fig. 2b,c and Extended Data Fig. 7i,j). These results indicate that tumor cells exhibit enhanced proliferative activity after EGFR-TKI resistance.

We next compared the proportion of stromal and immune cell clusters between pre- and post-treatment samples. The proportion of stromal cells, including fibroblasts and endothelial cells, was elevated in the ER_post sample compared to the ER_pre sample (Fig. 2d). DGE analysis of fibroblasts from ER samples revealed increased expression of collagen-related genes following treatment. (Fig. 2e). GSEA further demonstrated that a gene set related to extracellular matrix production was significantly enriched in fibroblasts from the ER_post sample (Fig. 2f). Additionally, collagen gene expressions were significantly elevated in ER_post sample compared to the other samples (Fig 2g and Extended Data Fig. 7k). These results suggest that fibroblasts exhibit increased activity following EGFR-TKI treatment, contributing to changes in the TME associated with early resistance.

In the T cell cluster, CD4-expressing T cells were present in slightly higher numbers, while the low expression levels of CD4, CD8A, and FOXP3 impeded their classification (Fig. 2h and Extended Data Fig. 7l). This could be attributed to the filtering effect of the analytical methods employed. In the macrophage cluster, tumor-associated macrophage markers (SPP1, C1QC) were highly expressed in accordance with the distribution of M2-like markers (CD163, MSR1)[20–22], suggesting an immune-

suppressive phenotype that was consistently observed across all samples. (Fig. 2i and Extended Data Fig. 7m).

Spatial distribution of tumor cell heterogeneity between early resistance and long-term response samples

To explore factors contributing to early resistance, tumor cells from all pre-treatment samples were categorized into two groups: ER_pre and SLR_pre (comprising SR_pre, LR1_pre, and LR2_pre) for DGE analysis (Fig. 3a). GSEA revealed significant enrichment in the inflammatory response, TNF-alpha signaling via NF-κB pathway, IL-6 pathway, and chemokine signaling pathways in ER_pre tumor cells (Fig. 3b,c and Extended Data Fig. 8a,b), indicating that the TME of ER_pre was characterized by active inflammation.

To better understand the factors that contribute the differences between samples, we performed a comparative analysis of tumor cell clusters. Based on differentially expressed genes, we classified and labeled each tumor cell cluster using cancer-related genes[23–28] as discriminatory markers (Fig. 3d,e and Extended Data Fig.8c,d). Applying inferCNV, copy number variation (CNV) analysis revealed a tendency for consistent CNV alterations in each tumor cell cluster when compared to epithelial cell clustering as a reference group (Extended Data Fig. 8e). To quantify the distribution, we visualized the tumor cluster across samples and mapped it spatially for each sample (Fig. 3f,g). The tc_S100A6 and tc_AQP1 clusters were present in all pre- and post-treatment samples, accounting for a high proportion in nearly all samples. The tc_S100A6 cluster exhibited significantly higher expressions of *CCND1* and *ERBB2* compared to the other clusters (Extended Data Fig. 8f,g), while *MET* and *CD74* expressions were markedly

elevated in the tc_AQP1 and tc_TMSB4X cluster, respectively (Extended Data Fig. 8h,i). GSEA based on DGE analysis of the tc_S100A6 cluster revealed significant enrichment of gene sets related to cell proliferation, including E2F targets and G2-M checkpoint, compared to the others (Fig. 4a,b). In contrast, gene sets linked to inflammatory response, epithelial mesenchymal transition (EMT), and the KRAS signaling were not enriched (Fig. 4c). DGE analysis and GSEA of the tc_AQP1 cluster demonstrated upregulation of pathways associated with mTORC1 and IL6/JAK/STAT3 signaling, both of which have been implicated in EGFR-TKI resistance[4,29–31], in addition to EMT and inflammatory responses (Fig. 4d-f).

Next, we analyzed distinct clusters that showed differences compared to the other samples. The tc_TMSB4X cluster was predominantly observed in ER_pre, with a lower frequency in other samples, suggesting its potential involvement in early resistance (Fig. 3f). GSEA followed by DGE analysis revealed significant enrichment of genes associated with metabolic pathways, while genes related to cell proliferation exhibited relatively low activity (Fig. 4g-i). The tc_MMP7 cluster was detected in both pre- and post-treatment ER samples, yet was nearly absent in other samples, supporting its potential role in early resistance (Fig. 3f). Additionally, DGE analysis and GSEA of the tc_MMP7 demonstrated significant enrichment of pathways related to EMT, inflammatory response, IL-6/JAK/STAT3 signaling, KRAS signaling, and TNF-alpha signaling via NF-kB (Fig. 4j-l). The tc_SCGB3A1 cluster was the most abundant in LR1_pre (Fig. 3f), possibly indicating a role in the long-term response. DGE analysis and GSEA showed a decrease in E2F targets and G2-M checkpoint, suggesting a relatively slow cell cycle (Extended Data Fig. 9a-c). The tc_AZGP1 cluster was characteristic of SR_pre (Fig. 3f), although no significant enrichment was observed

(Extended Data Fig. 9d-f). This cluster may have been affected by the extremely low cell number.

To investigate differences in the TME across tumor clusters, we quantified the spatial distribution of surrounding cell types within the ER_pre samples. For each tumor cluster, we calculated the proportion of cell types within concentric annular regions extending up to 1000 μm from each tumor cell (Extended Data Fig. 9g). This analysis revealed that the tc_S100A6 and tc_AQP1 clusters were diffusely distributed, whereas the tc_TMSB4X and tc_MMP7 clusters were each localized in close proximity, a pattern also observed in the spatial mapping of ER_pre (Fig. 3g). Notably, the tc_MMP7 cluster displayed the highest proportion of immune cells within its surrounding microenvironment. As shown in Figure 4l, GSEA of the tc_MMP7 cluster revealed an enrichment of genes associated with several inflammatory responses. Taken together, these findings suggest that this cluster is characterized by an "immune hot" state.

Maturity of tertiary lymphoid structures in TME

To examine the spatial relationships of each cluster within the TME, we performed a neighborhood enrichment analysis using the spatial map coordinates annotated with cell types. This analysis revealed a close spatial relationship between T cells, B cells, and follicular dendritic cells (FDCs), particularly in the ER_pre and LR1_pre samples (Fig. 5a). By visualizing these cell populations on the spatial map, they were distributed across several regions in ER_pre sample, whereas only one region was observed in LR1_pre sample (Extended Data Fig. 10a,b). In ER_pre sample, the clustering regions exhibited a structure where B cells and FDCs were intermixed, sparsely surrounded by

T cells (Fig. 5b and Extended Data Fig. 10c). In contrast, in the LR1_pre sample, the FDCs were positioned around the B cells, with T cells surrounding the outer perimeter (Fig. 5c and Extended Data Fig. 10d). Moreover, CXCL13 expression was elevated within these structures, indicating the presence of tertiary lymphoid structures (TLS). According to previous reports, the poorly organized clustering region in the ER_pre sample corresponded to an immature TLS, whereas the well-structured regions in the LR1_pre sample were identified as a mature TLS[32].

Enrichment of MIF- and PDGF- mediated cell-cell interactions in the early resistant TME

To further elucidate the characteristics of TME, we analyzed intercellular interactions using COMMOT. This Python library calculates cell-cell interactions based on ligand-receptor competition and spatial proximity. We extracted ligand-receptor gene pairs with sufficient expression levels from the CellChat database of known ligand-receptor interactions (Supplementary Table 2). Pathway comparisons across samples revealed elevated expression of genes associated with the macrophage migration inhibitory factor (MIF) pathway in ER and SR (Fig. 6a,b). MIF is a ligand of the CD74/CD44 or CD74/CXCR4 complex and regulates the inflammatory response[33]. To quantify ligand-receptor interactions, we calculated signal strength based on the spatial proximity of ligand- and receptor-expressing cells and their expression levels. Comparing the ER_pre and LR1_pre spatial signal vectors, the signals on the spatial map were directed to regions containing TLS and inflammatory cell infiltration (Fig. 6c,d). Signal intensity and direction was further analyzed by cell type (Fig. 6e-g). In both ER_pre and LR1_pre, signals showed bidirectional signaling, predominantly among T cells, follicular

dendritic cells, and B cells. Tumor cells exhibited limited signals in the LR1_pre sample, whereas tumor cells in the ER_pre sample displayed signals directed toward macrophages and TLS-related cell types. Some tumor cell clusters activated this pathway via an autocrine mechanism. These findings suggest that tumor cells actively stimulated macrophages and TLS through the MIF pathway in the ER_pre sample.

As shown in Figure 2d, the proportion of fibroblasts and endothelial cells increased post-treatment in the ER sample. Therefore, we focused on the platelet-derived growth factor (PDGF) pathway, which is associated with fibroblast proliferation[34]. The signal was not localized to specific regions within the ER_pre and LR1_pre samples (Fig. 6h,i). Bidirectional signaling was observed between fibroblasts and endothelial cells in both samples (Fig. 6j-l). In the ER_pre sample, tumor cells, including tc_TMSB4x and tc_MMP7 clusters, exhibited signals directed toward stromal cells. In contrast, tumor cell clusters in LR1_pre sample showed limited signal directed at stromal cells. These results suggest that tumor cells interact with stromal cells via the PDGF pathway, possibly contributing to the formation of an early-resistant TME.

__Discussion__

Although various mechanisms of resistance to EGFR-TKIs in *EGFR*-mutated NSCLC have been reported, there are limited studies comparing early resistance cases with long-term responders[3,35,36]. In this study, we showed that the cell proliferation of residual tumors was enhanced following osimertinib treatment. Genes significantly upregulated after treatment were associated with poor prognosis[37–39], emphasizing the need to target the remaining proliferative tumor cells. Tumor cell clusters, including the tc_S100A6 and tc_AQP1, persisted before and after osimertinib treatment, suggesting that

resistance has developed through the activation of downstream or alternative pathways. In fact, genes involved in resistance mechanisms such as CCND1, HER2, or MET were significantly high in these clusters, with their amplifications being associated with EGFR-TKI resistance [4]. Compared to tc_S100A6 and tc_AQP1 clusters, the tc_TMSB4X cluster, predominantly observed in ER_pre, demonstrated a relatively slow cell cycle, suggesting a quiescent state associated with EGFR-TKIs resistance[40]. We showed that the MIF-CD74 pathway was enhanced in the early resistance sample, with the tc_TMSB4X cluster activating this pathway via an autocrine mechanism. Notably, this cluster exhibited significantly higher CD74 expression compared to the other clusters. CD74 has been previously associated with resistance to EGFR-TKIs[41], suggesting that the tc_TMSB4X cluster acts as a reservoir for resistant cells. Furthermore, this cluster showed elevated bile acid metabolism, potentially reflecting suppression of immune cell responses[42]. In contrast, the tc_SCGB3A1 cluster, which also exhibited a slow cell cycle, was mainly present in LR1_pre and disappeared after osimertinib treatment, implying characteristics of low proliferation and low malignancy. The tc_MMP7 cluster had already exhibited an abundant TNF-alpha signaling via NF-kB even before the initiation of osimertinib treatment. The activation of TNF-alpha signaling induced by osimertinib has been shown to contribute EGFR-TKI resistance[43–45], indicating it as an early resistance mechanism. Furthermore, in spatial distribution analysis, the tc_MMP7 cluster was in an immunologically hot state due to a large amount of plasmablasts infiltrating the surrounding area. A recent report has shown that $CXCR1^{+}$ neutrophils contribute to early resistance to EGFR-TKI through TNF signaling via NF-kB[13]. However, in our analysis, the number of neutrophils was extremely low

and thus not included in the assessment. Taken together, our findings reveal that tumor heterogeneity contributes to resistance to EGFR-TKIs.

Increasing evidence suggests that interactions between tumor cells and surrounding stromal or immune cells involved in mechanisms of EGFR-TKI resistance[9,10,46]. Cancer associated fibroblasts (CAFs) have been reported as key regulators of the TME[47]. Our comparison analysis revealed an increase in the proportion of stromal cells in early resistance samples after osimertinib treatment, while it was no change in the SR samples, and it decreased in the LR samples. Notably, fibroblasts exhibited a marked increase, along with enhanced extracellular matrix production and upregulation of collagens such as COL1A1 and COL3A1, suggesting a potential correlation with early resistance[31]. In addition, PDGFRA and PDGFRB, which are receptors for PDGF signaling, are known markers of tumor-associated fibroblasts and regulate angiogenesis[47]. In the ER_pre sample, this PDGF signaling was active between tumor cells and stomal cells, suggesting that tumor cells interact with stromal cells via the PDGF pathway, thereby inducing an early resistant TME.

We demonstrated that inflammatory response was significantly upregulated in early resistance tissues compared to long-term response tissues before osimertinib treatment. We further showed that TME in early resistance was characterized by immature TLSs and infiltration of B cells and plasmablasts, whereas TME in long-term response exhibited mature TLSs with reduced immune cell infiltration. Previous reports have shown that immature TLSs were implicated in tumor-promoting effects, while mature TLSs were tumor-suppressive[32,48,49]. Our findings suggest that the degree of TLS maturation is associated with early resistance mechanisms. *EGFR*-mutated NSCLC is generally characterized by low immunogenicity, often described as an “immune cold”

state, which contributes to resistance to immunotherapy[8,9]. In contrast, an “immune hot” state, defined by TLS formation and high immune cell infiltration, has been associated with enhanced responsiveness to immunotherapy[50–54]. Notably, patients with a short response to EGFR-TKIs have shown better responses to immunotherapy[55], suggesting that their TME exhibits an “immune hot” state. Consistently, we showed that inflammatory responses were enriched in ER_pre samples. Taken together, these results indicate that an “immune hot” TME serves as a predictor of early resistance in *EGFR*-mutated NSCLC (Fig. 7).

Visium HD allows for high-resolution analysis, but it is difficult to distinguish between noise genes and differentially expressed genes. We addressed this by detecting and segmenting cell nuclei, reconstructing read count data and cell coordinate data to improve data reliability. While this approach minimized false positives, some clusters still contained mixed cell types, which we subsequently excluded. As a result of these processes, comprehensive gene expression analysis and spatial analysis has become possible while improving the accuracy of single-cell data and spatial coordinates. For further improvement in precision, multi-omics approaches are required.

This study has several limitations. First, the size of the bronchoscopic lung biopsy specimen was small compared to the size of a Visium HD slide, limiting the amount of data that could be obtained. The number of tumor cells may have been insufficient to analyze the changes pre- and post-treatment. Second, enhancing single-cell resolution resulted in reduced count data, potentially affecting the analysis of differentially expressed genes. Third, we cannot rule out the possibility that our findings were affected by sample variability. Future studies should collect additional clinical data to validate these findings.

In conclusion, our study suggests that an inflammatory TME and tumor heterogeneity contribute to early resistance mechanisms in *EGFR*-mutated NSCLC. Notably, even before treatment, certain tumor cell clusters already exhibit upregulated resistance-related pathways.

__Method__

Sample preparation

We collected formalin fixed paraffin embedded (FFPE) block specimens from four *EGFR*-mutated NSCLC patients pre- and post-treatment with osimertinib from multiple institutions. The biopsy methods for the specimens were surgical biopsy, bronchoscopic biopsy, surgical pleural biopsy, and single hole thoracoscopic pleural biopsy. Patient clinical data are listed in Supplementary Table 1. Patients were classified into early resistance (ER, n=1), standard response (SR, n=1), and long-term response (LR, n=2) based on the duration of progression free survival. These specimens were originally performed to determine treatment. We obtained written consent from all patients for specimen collection and research use. This study was carried out in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of Kyushu University Hospital (approval no. 23297-01)

10X Genomics Visium HD data generation

In this study, two capture areas on a Visium HD slide were utilized, with four samples assigned to each capture area. A total of eight samples consisting of paired pre- and post-treatment samples from four patients (ER, SR, LR1 and LR2) were analyzed, and each sample was differentiated from (ER_pre, ER_post, SR_pre, SR_post, LR1_pre,

LR1_post, LR2_pre, LR2_post). Specifically, ER and SR samples were assigned to one capture area, while LR1 and LR2 samples were assigned to the other capture area. Initially, RNA was extracted from FFPE tissue sections prepared for RNA quality assessment, and the DV200 value was measured using the 4200 Tape Station. The DV200 value for all samples was greater than or equal to 30%. The FFPE tissue sections were then subjected to deparaffinization and H&E staining following the Visium HD FFPE Tissue Preparation protocol provided by 10x Genomics (Visium HD FFPE Tissue Preparation Handbook, CG000684 Rev A). After staining, tissue sections were imaged using the EVIDENT VS200 microscope in accordance with the 10X Genomics protocol (Visium HD Spatial Applications Image Guidelines CG000688 Rev A). Gene expression libraries were prepared from the H&E-stained sections following the 10X Genomics protocol (Visium HD Spatial Gene Expression Reagent Kits User Guide CG000685 Rev A). The quality of the prepared libraries was evaluated using the 4200 Tape Station, and all libraries were confirmed to have sufficient concentrations. The Visium libraries were sequenced using the NovaSeq X Plus platform with a paired-end configuration (read1: 150 bp, read2: 150 bp). Sequencing data (FASTQ files) were processed and analyzed using Space Ranger (version 3.0.0) software to assess data quality and generate spatial gene expression profiles.

Analysis Environment

The analysis was conducted using Python (version 3.10) on either Ubuntu (22.04.5 LTS) with JupyterLab (4.2.5) or Windows (Windows 11 Home edition) with PyCharm (Community Edition 2024.1.1). The primary Python libraries and their versions included: AnnData (0.10.8), AnnData2ri (1.3.2), Commot (0.0.3), GeoPandas (1.0.1),

GSEApy (1.1.3), HarmonyPy (0.0.10), inferCNVpy (0.5.0), Matplotlib (3.8.4), NumPy (1.26.4), Pandas (2.2.2), Rpy2 (3.5.16), Scanpy (1.10.3), SciPy (1.11.4), Seaborn (0.13.2), Squidpy (1.6.1) and StarDist (0.9.1). In some analyses, R Script was executed in Python. The primary R packages and their versions used were: tradeSeq (1.8.0), edgeR (3.36.0), and clusterExperiment (2.14.0). While the libraries were generally applied as described in their documentation, modifications were made to Commot to improve visual clarity in plots.

Nuclei detection and segmentation

Although Visium HD allows for the use of 8 µm square spots as single cells, these spots do not accurately reflect the true cellular coordinates. To preserve cellular coordinates as much as possible, nuclei detection and segmentation were performed using StarDist[19]. Specifically, Python code was developed based on Analysis Guide provided by 10x Genomics (https://www.10xgenomics.com/jp/analysis-guides/segmentation-visium-hd). Running this code generated coordinate data for individual nuclei represented as polygons. Each 2 µm spot count data within the polygon area was relabeled as a single cell. Additionally, the centroid coordinates of each polygon were added to the count data, along with metadata content, creating a data structure compatible with libraries commonly used in spatial transcriptomics analysis.

Sample labeling on spatial coordinates

In typical spatial transcriptomics analyses, all samples within a single capture area are assumed to be the same. However, in this study, different four samples were present within the same capture area. Since this situation is not typically supported by existing

tools, it was necessary to manually label samples based on spatial coordinates. Using the spatial coordinate data within the data structure, samples were divided into eight groups by referencing images and assigning coordinates manually. Since there was an overlapping area between LR1_pre and LR1_post, a blank area was created to avoid sample contamination.

Major clustering

The newly generated data from nuclei segmentation were analyzed using a pipeline similar to single-cell RNA-seq analysis with Scanpy package. Cells with fewer than 20 counts per cell or containing more than 10% mitochondrial gene, genes with fewer than 20 total counts were excluded from the analysis. Cells with a cross-sectional area of 900 µm2 or more on the image were excluded in consideration of the possibility that they were composed of multiple cells. After filtering, the data were calculated for highly variable genes (HVGs) and normalized using Pearson residuals, which adjusts for library size and overdispersion. Principal component analysis (PCA) using the top 4000 HVGs calculated the principal components up to the top 35. To account for batch effects across samples, Harmony integration was applied using the HarmonyPy library through the Scanpy with the sample labels. Subsequently, the nearest neighbors distance matrix and the neighborhood graph of the observations were calculated with a local neighborhood size of 100 and 35 principal components. A UMAP embedding was generated with a minimum distance parameter set to 0.5. Leiden clustering was performed at a resolution of 0.4 and identified 18 clusters. However, due to the high noise level in cells, it was possible that the differences between cells and noise could not be distinguished. Instead, we analyzed HVGs in the expression of each cluster and

compared the expression levels of known cell markers to classify the clusters into three major cell types: epithelial, stromal, and immune cell types (Supplementary Table 3).

Sub-cluster analysis

Sub-clustering was performed to further explore the epithelial, stromal, and immune cell populations identified in the initial clustering. For each cell type cluster, the analysis was start again from the raw count data. The data were filtered to the same threshold as the major clustering. HVGs were calculated within some sub-cluster, PCA was calculated, followed by the construction of neighborhood graphs, UMAP embedding, and clustering. The distinction between tumor cells and normal cells was evaluated using inferCNVpy, which is based on inferCNV, in addition to the pathological images. Normal airway epithelial cells obtained from the clustering results were set as non-malignant cells and analyzed. Further, for tumor cells, additional clustering was performed. For immune cells, three cell types—lymphocytes, plasmablasts, and macrophages—were identified, and each type was subsequently analyzed with individual clustering steps. Detailed parameters for each clustering step and cell annotation markers are provided in the supplementary materials (Supplementary Table 3,4).

Differential Gene Expression Analysis

We performed differential gene expression (DGE) analysis using created pseudo-bulk samples combined with the edgeR package[56,57]. Single-cell RNA-seq data were aggregated at the sample level for each cell type of interest. Samples with fewer than 30 cells of the target cell type were excluded. For comparisons between ER and SLR

groups, as well as pre- and post-treatment comparisons of the tc_AZGP1 cluster, two random replicates were generated for each group because the edgeR library requires biological replicates for statistical modeling. In the others, pseudo-bulk aggregation was performed without creating additional replicates.

After pseudo-bulk aggregation, DGE analysis was performed using the edgeR library[57]. Low-expressed genes were first filtered out, retaining only those with sufficient counts across experimental groups to improve statistical power. Library size normalization was carried out using the trimmed mean of M-values method, which adjusts for sequencing depth and compositional biases across samples. A design matrix was constructed to account for sample conditions and randomly generated replicates, enabling the estimation of both biological variation and group-specific effects. The quasi-likelihood variance estimation method was employed to generate a gene list for DGE. Adjusted p-values were calculated using the Benjamini-Hochberg method to control the false discovery rate. Results for each cell type were integrated back into the annotated data object to facilitate further interpretation. The aggregated gene expression values were subsequently used for downstream analyses, including the creation of volcano plots and GSEA.

The Kaplan-Meier Plotter database

We used data from lung cancer patients in the Kaplan-Meier Plotter database (http://kmplot.com/analysis/) to evaluate differentially expressed genes in tumor cells compared before and after treatment, respectively[58]. Patients on the database were divided into high and low expression groups by median expression values, and survival curves were plotted with no follow-up restrictions.

Integration of cell type and spatial coordinate data

The dataset used in this analysis was generated by integrating single-cell analysis for cell type identification with spatial coordinate data derived from StarDist. The StarDist-based coordinate data, which represents new spot coordinates, was transformed based on the ratio of the actual distance to the pixel coordinates, as described in the metadata. This integration resulted in a unified data structure compatible with analysis tools such as Scanpy and Squidpy.

Neighborhood enrichment analysis

Neighborhood enrichment analysis was performed on the spatial transcriptomics data to assess the enrichment of different cell clusters within their spatial neighborhoods. For each of the 8 sample datasets, spatial neighbors function within Squidpy was used to calculate the spatial neighbors of each spot. This was followed by the application of the neighborhood enrichment function to calculate the z-scores, using predefined cell cluster labels. To standardize the enrichment scores across all samples, the minimum and maximum values were first calculated across all datasets. Each z-score matrix was then normalized by subtracting the global mean and dividing by the standard deviation, ensuring that enrichment scores were comparable across samples. This normalization allowed for the unified analysis of neighborhood enrichment across different experimental conditions.

Spatial distribution and neighborhood analysis of tumor cells

The spatial distribution and neighborhood relationships of tumor cells were analyzed by integrating single-cell cluster information with spatial coordinate data. For each of the 8 sample datasets, the spatial positions of tumor cells were identified. To quantify the surrounding cell composition, we calculated the pairwise distances between each tumor cell and all other cell types within the tissue. To analyze the spatial distribution, we defined concentric annular regions extending up to 1000 μm from each tumor cell, with each region increasing in 20 μm increments. These regions served as distance bins, allowing us to categorize surrounding cells based on their proximity to tumor cells. We then calculated the proportion of each cell type within these bins to assess their spatial distribution relative to tumor clusters. This method enabled a quantitative evaluation of neighborhood enrichment patterns, providing insights into the TME and its cellular composition.

Cell-cell communication analysis

Cell-cell interactions were analyzed using COMMOT, a tool that enables comprehensive modeling of both direct and indirect interactions between different cell types while incorporating the spatial organization of cells within tissues[59]. For this analysis, smooth muscle cells, normal airway epithelial cells, and blood cells were excluded to focus on the TME. Additionally, genes expressed in more than 0.5% of all cells were first filtered. This gene list was further refined by extracting known ligand-receptor pairs from the CellChat database. Only ligand-receptor pairs co-occurring within a distance of 500 μm were considered for subsequent analysis. For the identified ligand-receptor pairs and their associated pathways, signal strength and vectors were

calculated based on the distance between sending and receiving cells as well as gene expression intensity.

Gene expression analysis and statistics

Gene expression scores were calculated using marker gene-based scoring method implemented in Scanpy. The “MIF Score” was derived from the MIF pathway-related genes: *MIF*, *CD74*, *CD44* and *CXCR4*. Background correction was performed using control gene sets of 5000 randomly selected genes to ensure robust scoring. Pairwise comparisons of each score between the Comparisons of MIF scores between samples and gene expression in tumor cells were performed using a pairwise comparison using a unilateral Mann-Whitney U test. Visualization was conducted with Seaborn for violin and strip plots annotated with the value of the p-value. The significance level of GSEA after DGE analysis was set as p value less than 0.05 and false discovery rate as 0.25.

__Data availability__

The Visium HD dataset from this study is available on the GEO repository under accession numbers GSE288758.

_

_Reference__

1. Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *New England Journal of Medicine*. 2020;382(1):41-50. doi:10.1056/NEJMoa1913662

2. Soria JC, Ohe Y, Vansteenkiste J, et al. Osimertinib in Untreated EGFR-Mutated Advanced Non–Small-Cell Lung Cancer. *New England Journal of Medicine*. 2018;378(2):113-125. doi:10.1056/NEJMoa1713137

3. Oxnard GR, Hu Y, Mileham KF, et al. Assessment of Resistance Mechanisms and Clinical Implications in Patients With EGFR T790M–Positive Lung Cancer and Acquired Resistance to Osimertinib. *JAMA Oncol*. 2018;4(11). doi:10.1001/jamaoncol.2018.2969

4. Leonetti A, Sharma S, Minari R, Perego P, Giovannetti E, Tiseo M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer*. 2019;121(9):725-737. doi:10.1038/s41416-019-0573-8

5. Maynard A, McCoach CE, Rotow JK, et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell*. 2020;182(5):1232-1251.e22. doi:10.1016/j.cell.2020.07.017

6. Oh S, Koh J, Kim TM, et al. Transcriptomic Heterogeneity of EGFR-Mutant Non–Small Cell Lung Cancer Evolution Toward Small-Cell Lung Cancer. *Clinical Cancer Research*. 2024;30(20):4729-4742. doi:10.1158/1078-0432.CCR-24-0160

7. Madeddu C, Donisi C, Liscia N, Lai E, Scartozzi M, Macciò A. EGFR-Mutated Non-Small Cell Lung Cancer and Resistance to Immunotherapy: Role of the Tumor Microenvironment. *Int J Mol Sci*. 2022;23(12):6489. doi:10.3390/ijms23126489

8. Chen S, Tang J, Liu F, et al. Changes of tumor microenvironment in non-small cell lung cancer after TKI treatments. *Front Immunol*. 2023;14. doi:10.3389/fimmu.2023.1094764

9. Yang L, He YT, Dong S, et al. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer*. 2022;10(2):e003534. doi:10.1136/jitc-2021-003534

10. Matsumoto Y, Sawa K, Fukui M, et al. Impact of tumor microenvironment on the efficacy of epidermal growth factor receptor-tyrosine kinase inhibitors in patients with -mutant non-small cell lung cancer. *Cancer Science*. 2019;110(10):3244-3254. doi:10.1111/cas.14156

11. Yang CY, Liao WY, Ho CC, et al. Association between programmed death-ligand 1 expression, immune microenvironments, and clinical outcomes in epidermal growth factor receptor mutant lung adenocarcinoma patients treated with tyrosine

kinase inhibitors. *European Journal of Cancer*. 2020;124:110-122. doi:10.1016/j.ejca.2019.10.019

12. Hsu KH, Tseng JS, Yang TY, et al. PD-L1 strong expressions affect the clinical outcomes of osimertinib in treatment naïve advanced EGFR-mutant non-small cell lung cancer patients. *Sci Rep*. 2022;12(1):9753. doi:10.1038/s41598-022-13102-7

13. Wang H, Xiong A, Chen X, et al. CXCR1+ neutrophil infiltration orchestrates response to third-generation EGFR-TKI in EGFR mutant non-small-cell lung cancer. *Sig Transduct Target Ther*. 2024;9(1):1-15. doi:10.1038/s41392-024-02045-2

14. Zhang H, AbdulJabbar K, Moore DA, et al. Spatial Positioning of Immune Hotspots Reflects the Interplay between B and T Cells in Lung Squamous Cell Carcinoma. *Cancer Res*. 2023;83(9):1410-1425. doi:10.1158/0008-5472.CAN-22-2589

15. De Zuani M, Xue H, Park JS, et al. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat Commun*. 2024;15(1):4388. doi:10.1038/s41467-024-48700-8

16. Peng H, Wu X, Liu S, et al. Multiplex immunofluorescence and single□cell transcriptomic profiling reveal the spatial cell interaction networks in the non□small cell lung cancer microenvironment. *Clin Transl Med*. 2023;13(1):e1155. doi:10.1002/ctm2.1155

17. Wang S, Rong R, Yang DM, et al. Features of tumor-microenvironment images predict targeted therapy survival benefit in patients with EGFR-mutant lung cancer. *J Clin Invest*. 133(2):e160330. doi:10.1172/JCI160330

18. Dagogo-Jack I, Valiev I, Kotlov N, et al. B-Cell Infiltrate in the Tumor Microenvironment Is Associated With Improved Survival in Resected Lung Adenocarcinoma. *JTO Clinical and Research Reports*. 2023;4(7). doi:10.1016/j.jtocrr.2023.100527

19. Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-Convex Polygons. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Springer International Publishing; 2018:265-273. doi:10.1007/978-3-030-00934-2_30

20. Mantovani A, Sica A. Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Current Opinion in Immunology*. 2010;22(2):231-237. doi:10.1016/j.coi.2010.01.009

21. Matsubara E, Yano H, Pan C, et al. The Significance of SPP1 in Lung Cancers and Its Impact as a Marker for Protumor Tumor-Associated Macrophages. *Cancers (Basel)*. 2023;15(8):2250. doi:10.3390/cancers15082250

22. Song Y, Fu Y, Wang J, et al. Complement C1q induces the M2-polarization of tumor-associated macrophages in lung adenocarcinoma. *Genes Dis*. 2023;11(4):101093. doi:10.1016/j.gendis.2023.101093

23. Wei X, Dong J. Aquaporin 1 promotes the proliferation and migration of lung cancer cell in vitro. *Oncology Reports*. 2015;34(3):1440-1448. doi:10.3892/or.2015.4107

24. Li P, Lv X, Zhang Z, Xie S. S100A6/miR193a regulates the proliferation, invasion, migration and angiogenesis of lung cancer cells through the P53 acetylation. *Am J Transl Res*. 2019;11(8):4634-4649.

25. Yang Z, Luo J, Zhang M, et al. TMSB4X: A novel prognostic marker for non-small cell lung cancer. *Heliyon*. 2023;9(11):e21505. doi:10.1016/j.heliyon.2023.e21505

26. Yamada K, Kadota K, Fujimoto S, et al. MMP-7 expression is associated with a higher rate of tumor spread through air spaces in resected lung adenocarcinomas. *Lung Cancer*. 2023;175:125-130. doi:10.1016/j.lungcan.2022.12.001

27. Mazumdar J, Hickey MM, Pant DK, et al. HIF-2α deletion promotes Kras-driven lung tumor development. *Proc Natl Acad Sci U S A*. 2010;107(32):14182-14187. doi:10.1073/pnas.1001296107

28. Albertus DL, Seder CW, Chen G, et al. AZGP1 Autoantibody Predicts Survival and Histone Deacetylase Inhibitors Increase Expression in Lung Adenocarcinoma. *Journal of Thoracic Oncology*. 2008;3(11):1236-1244. doi:10.1097/JTO.0b013e318189f5ec

29. Li L, Li Z, Lu C, et al. Ibrutinib reverses IL-6-induced osimertinib resistance through inhibition of Laminin α5/FAK signaling. *Commun Biol*. 2022;5(1):1-14. doi:10.1038/s42003-022-03111-7

30. Fang W, Huang Y, Gu W, et al. PI3K-AKT-mTOR pathway alterations in advanced NSCLC patients after progression on EGFR-TKI and clinical response to EGFR-TKI plus everolimus combination therapy. *Transl Lung Cancer Res*. 2020;9(4):1258-1267. doi:10.21037/tlcr-20-141

31. Yamazaki S, Higuchi Y, Ishibashi M, et al. Collagen type I induces EGFR□TKI resistance in EGFR□mutated cancer cells by mTOR activation through Akt□independent pathway. *Cancer Sci*. 2018;109(6):2063-2073. doi:10.1111/cas.13624

32. Bruno TC. B cells to the forefront of immunotherapy. *Nature*. 2020;577(7791):474-476. doi:10.1038/d41586-019-03943-0

33. Calandra T, Roger T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol*. 2003;3(10):791-800. doi:10.1038/nri1200

34. Yao L, Rathnakar BH, Kwon HR, et al. Temporal control of PDGFRα regulates the fibroblast-to-myofibroblast transition in wound healing. *Cell Rep*. 2022;40(7):111192. doi:10.1016/j.celrep.2022.111192

35. Thress KS, Jacobs V, Angell HK, et al. Modulation of Biomarker Expression by Osimertinib: Results of the Paired Tumor Biopsy Cohorts of the AURA Phase I Trial. *Journal of Thoracic Oncology*. 2017;12(10):1588-1594. doi:10.1016/j.jtho.2017.07.011

36. Isomoto K, Haratani K, Hayashi H, et al. Impact of EGFR-TKI Treatment on the Tumor Immune Microenvironment in *EGFR* Mutation–Positive Non–Small Cell Lung Cancer. *Clinical Cancer Research*. 2020;26(8):2037-2046. doi:10.1158/1078-0432.CCR-19-2027

37. Jiang C, Zhang Y, Li Y, et al. High CEP55 expression is associated with poor prognosis in non-small-cell lung cancer. *Onco Targets Ther*. 2018;11:4979-4990. doi:10.2147/OTT.S165750

38. Luo A, Yin Y, Li X, Xu H, Mei Q, Feng D. The clinical significance of FSCN1 in non-small cell lung cancer. *Biomedicine & Pharmacotherapy*. 2015;73:75-79. doi:10.1016/j.biopha.2015.05.014

39. Li Y, Leng Y, Dong Y, et al. Cyclin B1 expression as an independent prognostic factor for lung adenocarcinoma and its potential pathways. *Oncol Lett*. 2022;24(6):441. doi:10.3892/ol.2022.13561

40. Lindell E, Zhong L, Zhang X. Quiescent Cancer Cells—A Potential Therapeutic Target to Overcome Tumor Resistance and Relapse. *Int J Mol Sci*. 2023;24(4):3762. doi:10.3390/ijms24043762

41. Kashima Y, Shibahara D, Suzuki A, et al. Single-cell analyses reveal diverse mechanisms of resistance to EGFR tyrosine kinase inhibitors in lung cancer. *Cancer Res*. 2021;81(18):4835-4848. doi:10.1158/0008-5472.CAN-20-2811

42. Varanasi SK, Chen D, Liu Y, et al. Bile acid synthesis impedes tumor-specific T cell responses during liver cancer. *Science*. 2025;387(6730):192-201. doi:10.1126/science.adl4100

43. Gong K, Guo G, Gerber DE, et al. TNF-driven adaptive response mediates resistance to EGFR inhibition in lung cancer. *J Clin Invest*. 128(6):2500-2518. doi:10.1172/JCI96148

44. Nam YW, Shin JH, Kim S, et al. EGFR inhibits TNF-α-mediated pathway by phosphorylating TNFR1 at tyrosine 360 and 401. *Cell Death Differ*. 2024;31(10):1318-1332. doi:10.1038/s41418-024-01316-3

45. Ibusuki R, Iwama E, Shimauchi A, et al. TP53 gain-of-function mutations promote osimertinib resistance via TNF-α–NF-κB signaling in EGFR-mutated lung cancer. *NPJ Precis Oncol*. 2024;8:60. doi:10.1038/s41698-024-00557-2

46. Alsaed B, Lin L, Son J, et al. Intratumor heterogeneity of EGFR expression mediates targeted therapy resistance and formation of drug tolerant microenvironment. *Nat Commun*. 2025;16(1):28. doi:10.1038/s41467-024-55378-5

47. Luo H, Xia X, Huang LB, et al. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat Commun*. 2022;13(1):6619. doi:10.1038/s41467-022-34395-2

48. Helmink BA, Reddy SM, Gao J, et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*. 2020;577(7791):549-555. doi:10.1038/s41586-019-1922-8

49. Petitprez F, de Reyniès A, Keung EZ, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020;577(7791):556-560. doi:10.1038/s41586-019-1906-8

50. Ukita M, Hamanishi J, Yoshitomi H, et al. CXCL13-producing CD4+ T cells accumulate in the early phase of tertiary lymphoid structures in ovarian cancer. *JCI Insight*. 7(12):e157215. doi:10.1172/jci.insight.157215

51. Germain C, Gnjatic S, Tamzalit F, et al. Presence of B Cells in Tertiary Lymphoid Structures Is Associated with a Protective Immunity in Patients with Lung Cancer. *Am J Respir Crit Care Med*. 2014;189(7):832-844. doi:10.1164/rccm.201309-1611OC

52. Cabrita R, Lauss M, Sanna A, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*. 2020;577(7791):561-565. doi:10.1038/s41586-019-1914-8

53. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat Rev Cancer*. 2022;22(7):414-430. doi:10.1038/s41568-022-00466-1

54. Leong TL, Bryant VL. B cells in lung cancer—not just a bystander cell: a literature review. *Transl Lung Cancer Res*. 2021;10(6):2830-2841. doi:10.21037/tlcr-20-788

55. Yoshida H, Kim YH, Ozasa H, et al. Nivolumab in non-small-cell lung cancer with EGFR mutation. *Annals of Oncology*. 2018;29(3):777-778. doi:10.1093/annonc/mdx745

56. Murphy AE, Skene NG. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat Commun*. 2022;13(1):7851. doi:10.1038/s41467-022-35519-4

57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140. doi:10.1093/bioinformatics/btp616

58. Győrffy B. Transcriptome-level discovery of survival-associated biomarkers and therapy targets in non-small-cell lung cancer. *British Journal of Pharmacology*. 2024;181(3):362-374. doi:10.1111/bph.16257

59. Cang Z, Zhao Y, Almet AA, et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods*. 2023;20(2):218-228. doi:10.1038/s41592-022-01728-4

__Figure legends__

Fig. 1 | Analysis overview and clustering, visualization of cell type and spatial location. a, Biopsy interval pre- and post-osimertinib treatment and patient classification. **b**, Detecting and segmenting cell nuclei using StarDist on high-resolution tissue images loaded with Visium HD protocol. **c**, Visualization of detection and segmentation by StarDist and 2 µm spot extraction of identified cell nuclei positions. **d**, UMAP of annotated cell types by single cell RNA-seq based on cell nuclei identified by StarDist. **e**, UMAP with epithelial cells re-clustered and labeled for cell type. **f**, UMAP with Immune cells re-clustered and labeled for cell type. **g**, UMAP with stromal cells re-clustered and labeled for cell type. **h**,**i**, H&E-stained pathological image of a certain area of ER_pre (h), and image of the same area with cell type identification reflected in polygon coordinate data obtained by StarDist (i). **j**,**k**, H&E-stained pathological images of magnified regions of LR2_pre (j), and images of cell type-identified same regions reflected in polygonal coordinate data obtained by StarDist (k). Normal epithelial cells and tumor cells were distinguished, and surrounding mesenchymal cells and immune cells were identified with high accuracy.

Fig. 2 | Comparative analysis before and after treatment. a, Volcano plot showing the DGE analysis results by extracting tumor cell clusters from all samples and dividing them into two groups before and after treatment (All_pre vs All_post). **b**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of All_pre vs All_post. **c**, GSEA plot with high absolute value of normalized enrichment score between All_pre and All_post. **d**, Stacked bar chart showing the proportion of cell types within each sample. **e**, Volcano plot showing DGE analysis results comparing fibroblasts extracted

from ER_pre and ER_post. **f**, GSEA plot of the only significant gene set in Hallmark gene sets. **g**, Violin plot and strip plot showing raw counts of *COL1A1* gene for each sample. One-sided Mann-Whitney U tests were performed to compare the other samples with ER_post, and the test results were described in the graph. ***$p < 0.0005$. **h**, Expression levels of *CD3D*, *CD4*, *CD8A*, and *FOXP3* related to T cell subtypes displayed on UMAP of T cell parts extracted from lymphocyte clusters. **i,** Expression levels of genes related to tumor-associated macrophage markers (SPP1, C1QC) and M2 macrophage markers (CD163, MSR1) on UMAP of Macrophage..

Fig. 3 | Analysis exploring mechanisms of tumor heterogeneity and early resistance. **a**, Volcano plot showing the DGE analysis results by extracting tumor cell clusters from pre-treatment samples and dividing them into two groups: ER_pre and SLR_pre. **b**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of ER_pre vs SLR_pre. **c**, GSEA plot with high absolute value of normalized enrichment score between ER_pre and SLR_pre. **d**, Matrix plot showing clustering of tumor cells based on differentially expressed genes. **e**, UMAP with Tumor cells re-clustered and labeled for differentially expressed genes. **f**, Violin plot showing the distribution of tumor cell clusters per sample. **g**, Image of tumor cell cluster positions plotted on HE-stained tissue image of Slide 1 and Slide 2.

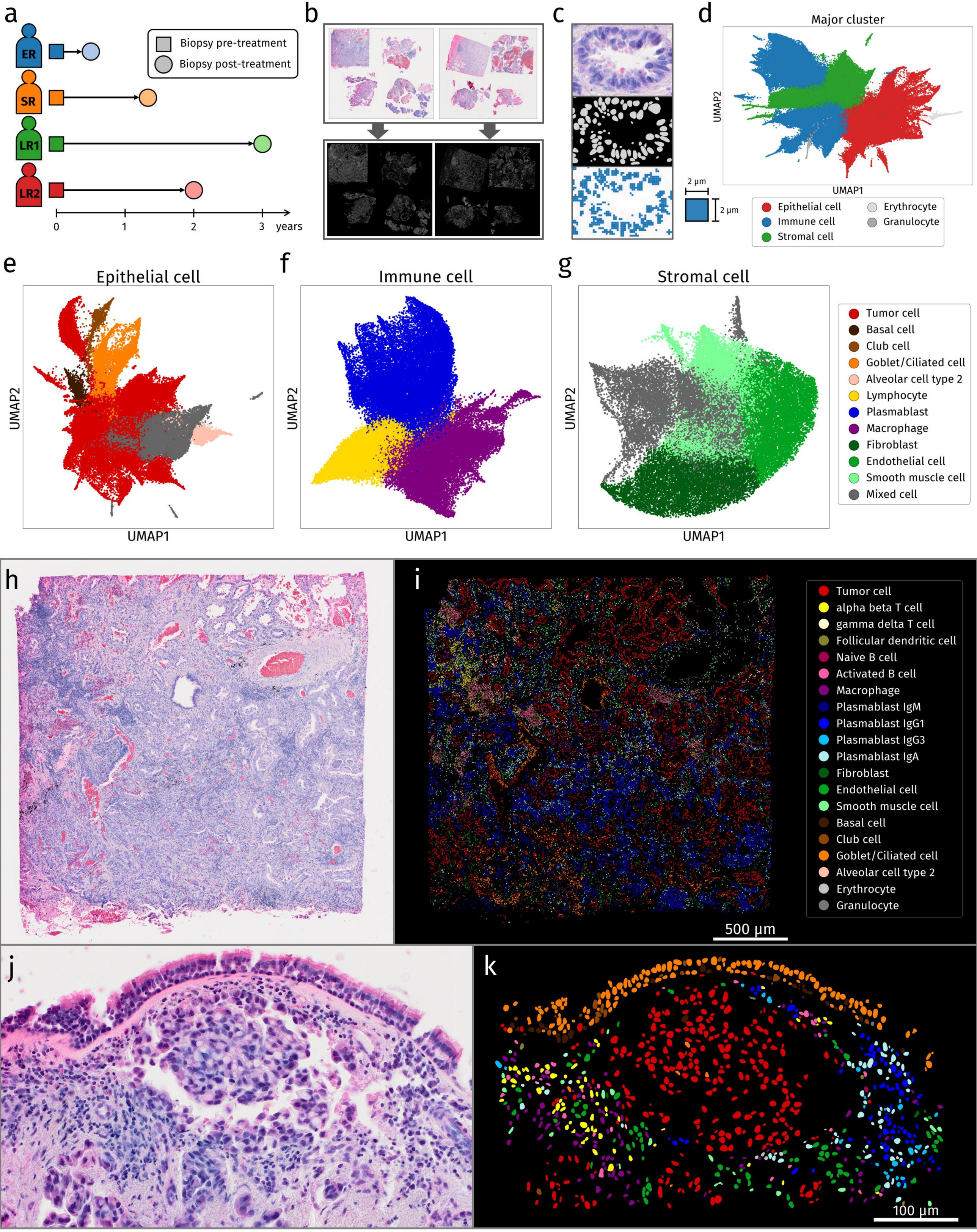
Fig. 4 | Individual analysis of tumor cell clusters. a, Volcano plot showing DGE analysis results between the tc_S100A6 cluster and the rest. **b**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of the tc_S100A6 cluster vs the rest. **c**, GSEA plot with high absolute value of normalized enrichment score between the

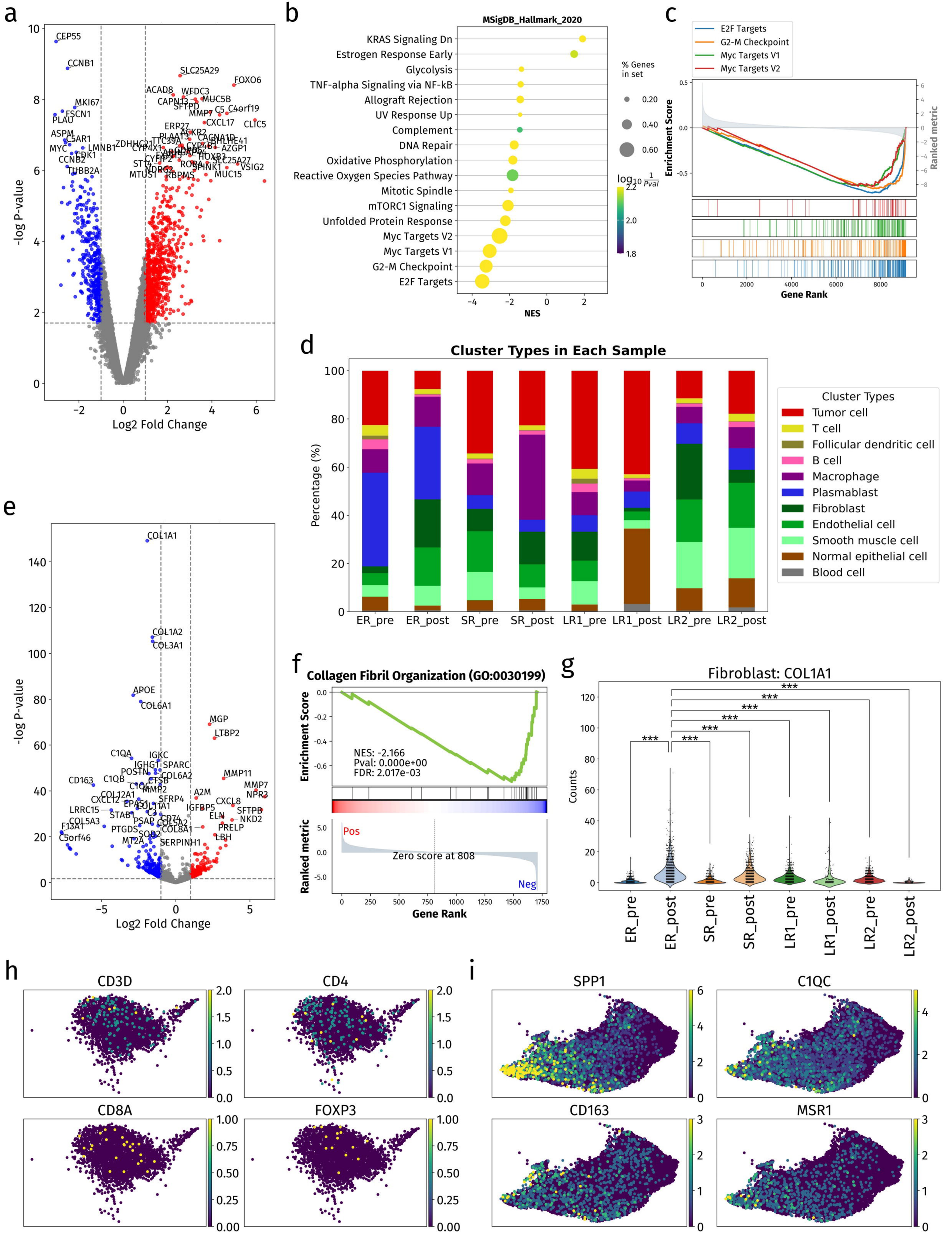
tc_S100A6 cluster and the rest**.** **d**, Volcano plot showing DGE analysis results between the tc_AQP1 cluster and the rest. **e**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of the tc_AQP1 cluster vs the rest. **f**, GSEA plot with high absolute value of normalized enrichment score between the tc_AQP1 cluster and the rest**. g**, Volcano plot showing DGE analysis results between the tc_TMSB4X cluster and the rest. **h**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of the tc_TMSB4X cluster vs the rest. **i**, GSEA plot with high absolute value of normalized enrichment score between the tc_TMSB4X cluster and the rest**.** **j**, Volcano plot showing DGE analysis results between the tc_MMP7 cluster and the rest. **k**, Dot plot of GSEA with Hallmarker gene sets based on DGE analysis of the tc_MMP7 cluster vs the rest. **l**, GSEA plot with high absolute value of normalized enrichment score between the tc_MMP7 cluster and the rest**.**

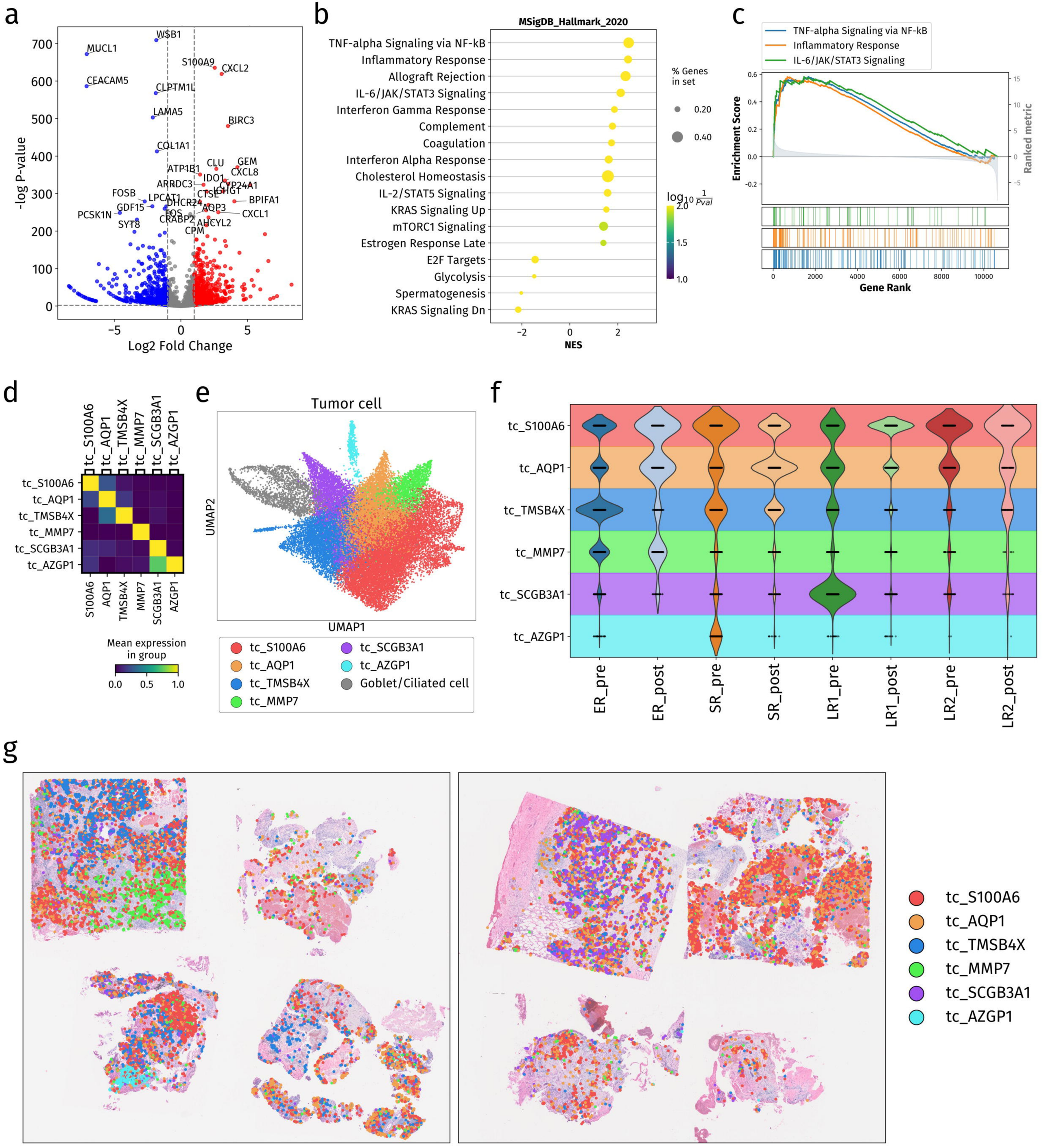
Fig. 5 | Spatial neighborhood enrichment analysis and TLS identification. a, Heatmap representing spatial neighborhood enrichment analysis results. 5000 surrounding cells were counted from each cell, and the proportion of that cell type was expressed as z-scores normalized to all samples. **b**, Enlarged images of a TLS in the ER_pre sample. The area surrounded by the white dotted line showed images of HE-stained tissue, cell types, and the expression level of CXCL13. **c**, Enlarged images of a TLS in the LR1_pre sample. The area surrounded by the white dotted line showed images of HE-stained tissue, cell types, and the expression level of CXCL13.

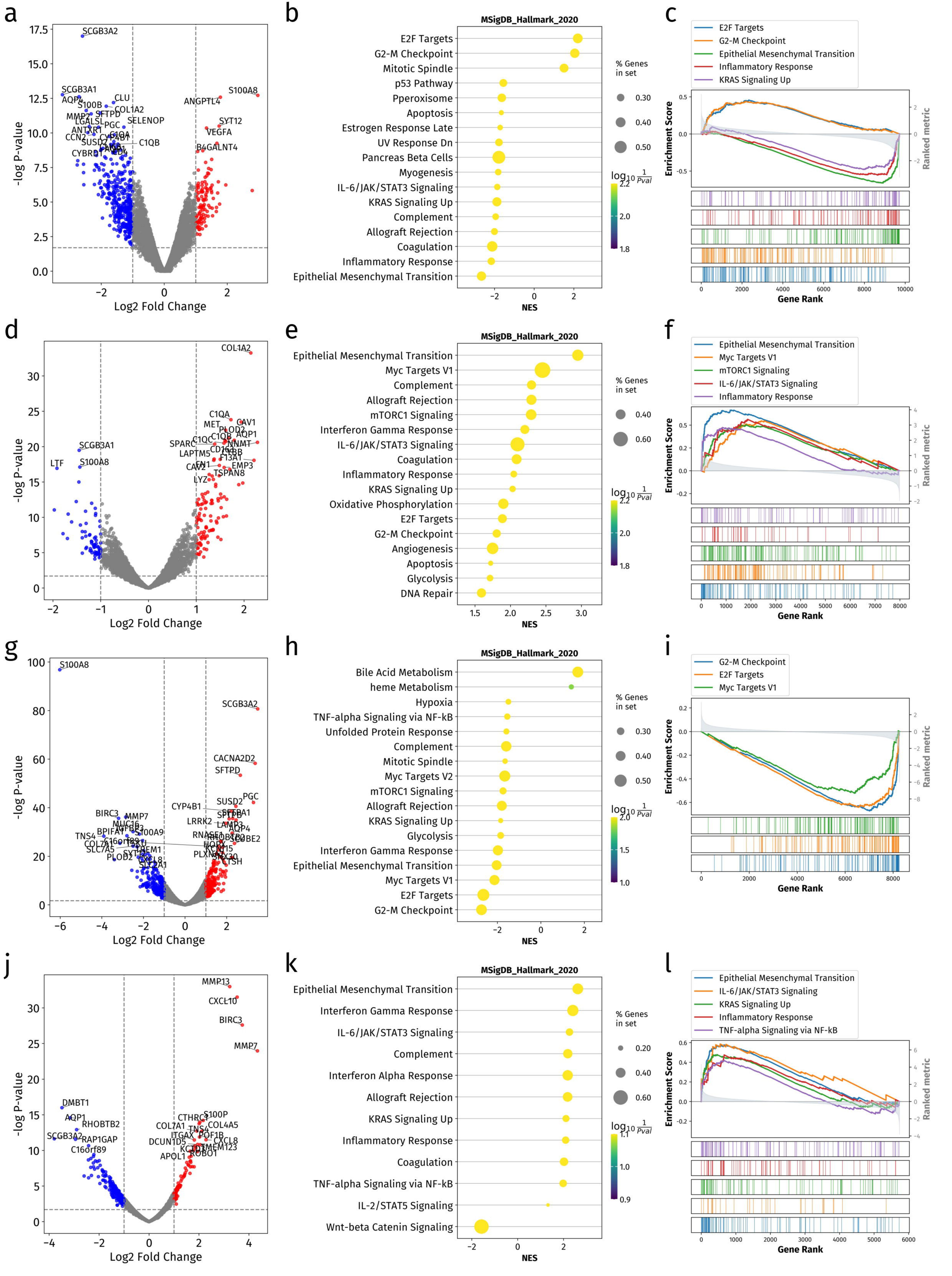
Fig. 6 | Cell-cell interaction analysis using COMMOT. a, Violin plot and strip plot showing MIF scores for each sample. One-sided Mann-Whitney U tests were performed to compare the other samples with the ER_pre sample, and the test results were described in the graph. ns $p \geq 0.05$; ***$p < 0.0005$. **b**, Images of the expression levels of *MIF*, *CD74*, *CD44*, and *CXCR4*, which constitute the MIF pathway, and were expressed in 0.5% or more of all cells in each sample. **c**, An image of spatial directions of cell-cell communication of MIF pathway in the ER_pre sample. **d**, An image of spatial directions of cell-cell communication of MIF pathway in the LR1_pre sample. **e**, Visualization of significant MIF pathway signal networks summarized by cell type in ER_pre and LR1_pre samples. **f**, Dot plot of significant MIF pathway signal networks summarized by cell type in the ER_pre sample. **g**, Dot plot of significant MIF pathway signal networks summarized by cell type in the LR1_pre sample. **h**, An image of spatial directions of cell-cell communication of PDGF pathway in the ER_pre sample. **i**, An image of spatial directions of cell-cell communication of PDGF pathway in the LR1_pre sample. **j**, Visualization of significant PDGF pathway signal networks summarized by cell type in ER_pre and LR1_pre samples. **k**, Dot plot of significant PDGF pathway signal networks summarized by cell type in the ER_pre sample. **l**, Dot plot of significant PDGF pathway signal networks summarized by cell type in the LR1_pre sample.

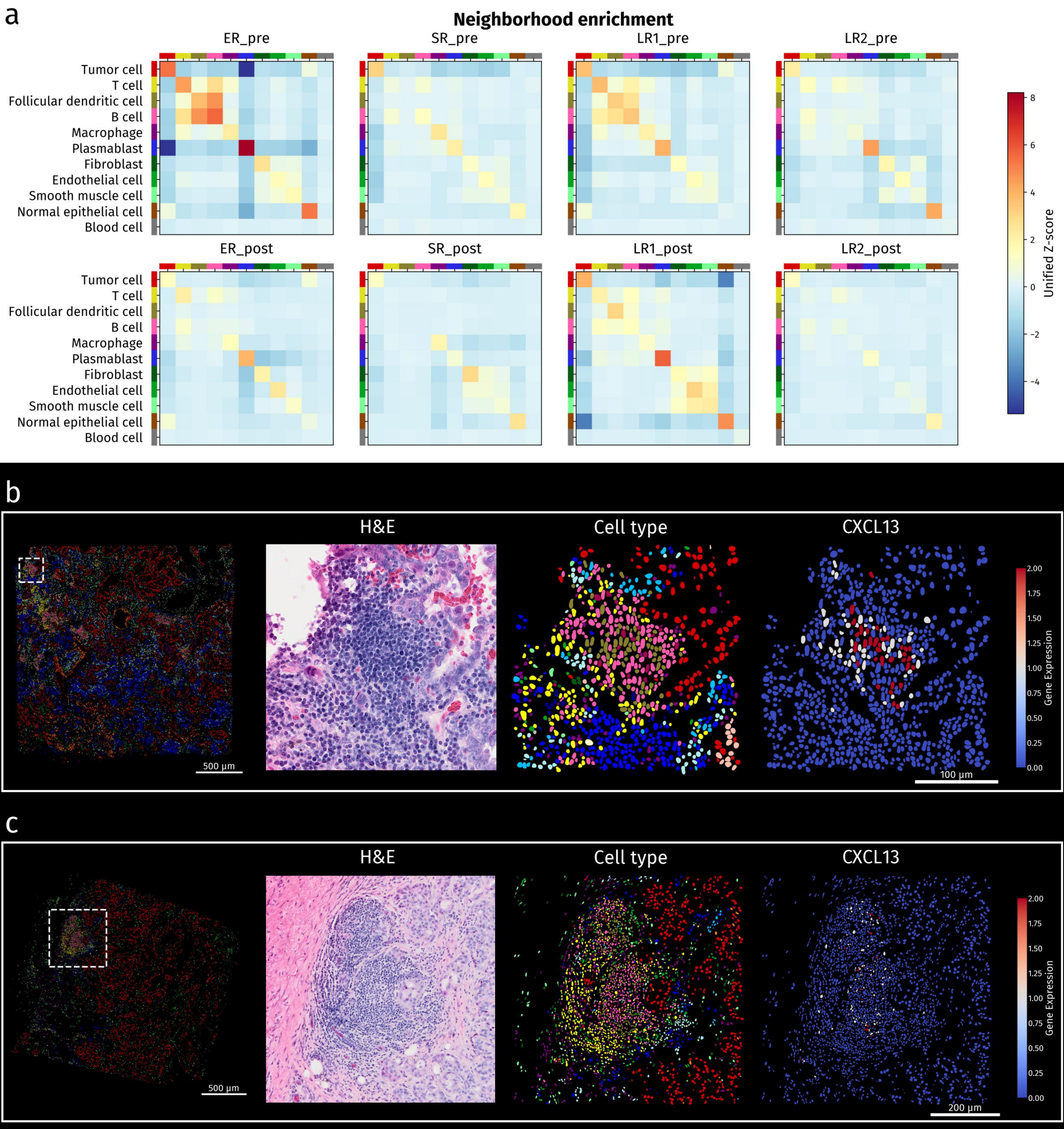
Fig. 7 | Proposed model of early resistance to EGFR-TKIs in the tumor microenvironment. Schematic diagram summarizing findings from the analysis of this study

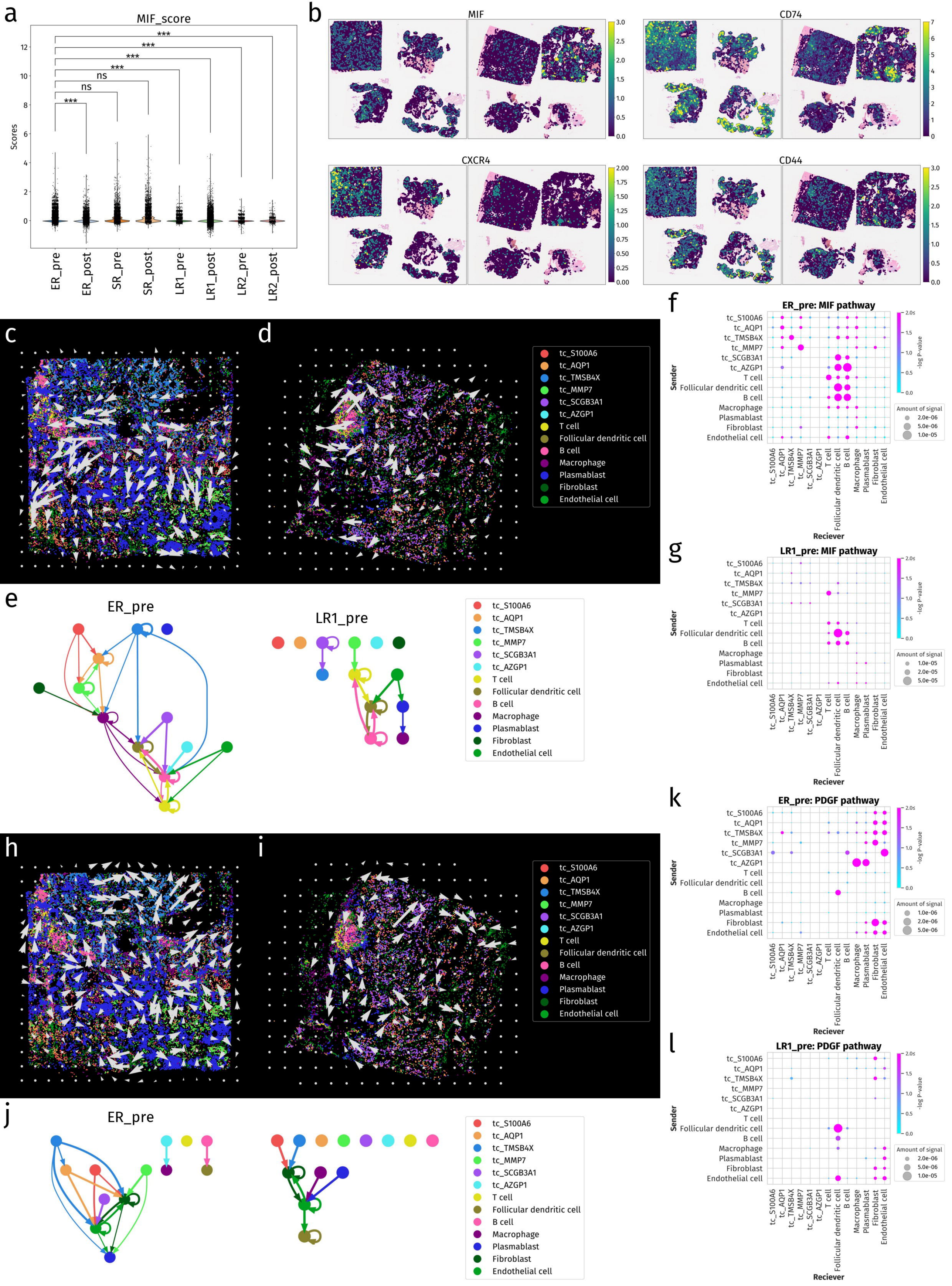
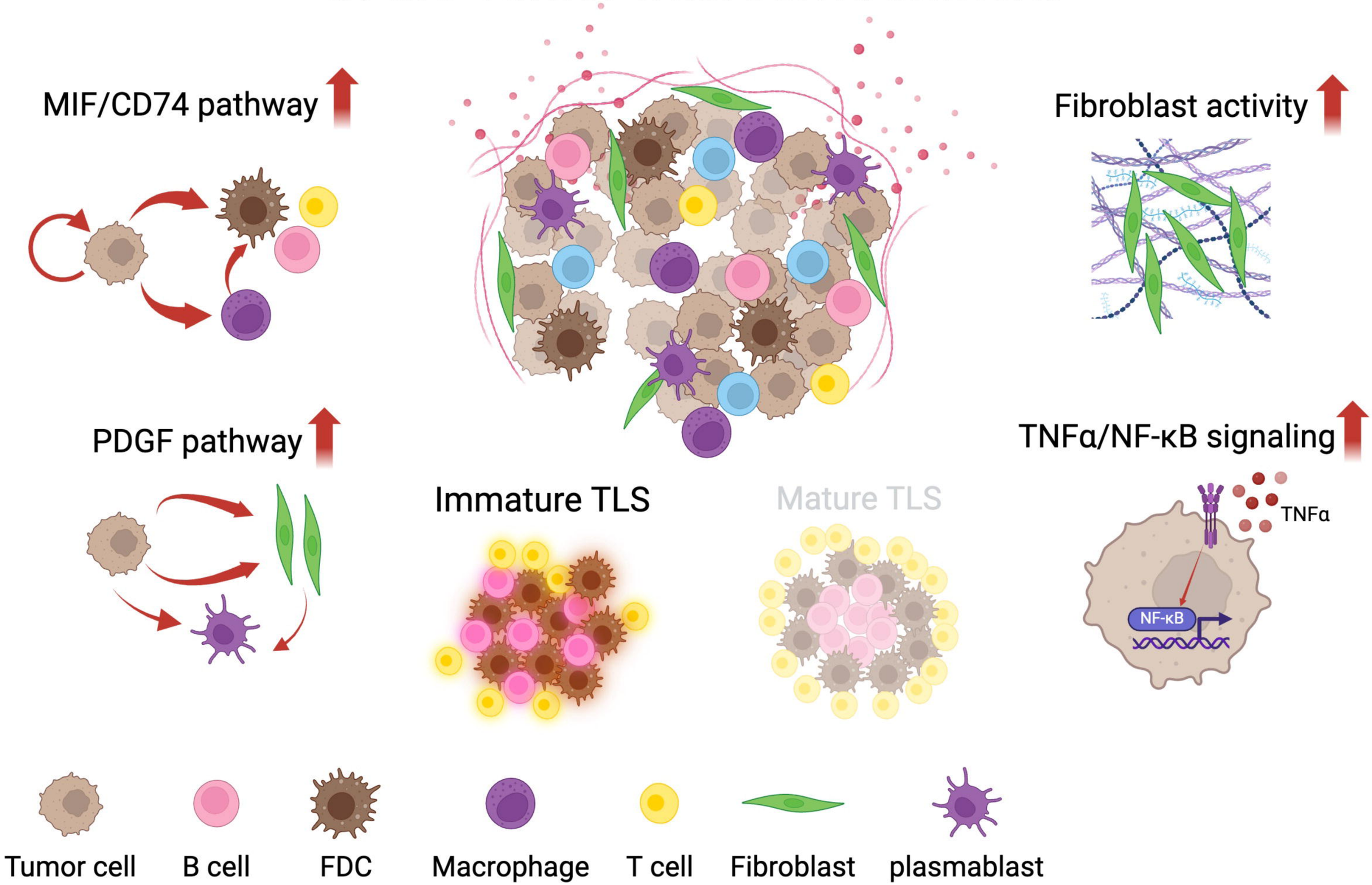

a
ER
SR
LR1
LR2
Biopsy pre-treatment
Biopsy post-treatment
0
1
2
3
years
b
c
2 µm
2 µm
d
Major cluster
UMAP2
UMAP1
Epithelial cell
Immune cell
Stromal cell
Erythrocyte
Granulocyte
e
Epithelial cell
UMAP2
UMAP1
f
Immune cell
UMAP2
UMAP1
g
Stromal cell
UMAP2
UMAP1
Tumor cell
Basal cell
Club cell
Goblet/Ciliated cell
Alveolar cell type 2
Lymphocyte
Plasmablast
Macrophage
Fibroblast
Endothelial cell
Smooth muscle cell
Mixed cell
h
i
Tumor cell
alpha beta T cell
gamma delta T cell
Follicular dendritic cell
Naive B cell
Activated B cell
Macrophage
Plasmablast IgM
Plasmablast IgG1
Plasmablast IgG3
Plasmablast IgA
Fibroblast
Endothelial cell
Smooth muscle cell
Basal cell
Club cell
Goblet/Ciliated cell
Alveolar cell type 2
Erythrocyte
Granulocyte
500 µm
j
k
100 µm


a
-log P-value
Log2 Fold Change
CEP55
CCNB1
MKI67
FSCN1
PLAU
ASPM
C5AR1
MYC
LMNB1
CDK1
CCNB2
TUBB2A
SLC25A29
FOXO6
ACAD8
WFDC3
MUC5B
SFTPD
C5
C4orf19
CXCL17
CLIC5
ERP27
ACKR2
CAGNA1D
BHLHE41
AZGP1
HOXB3
SLC25A27
VSIG2
MUC15
b
MSigDB_Hallmark_2020
KRAS Signaling Dn
Estrogen Response Early
Glycolysis
TNF-alpha Signaling via NF-kB
Allograft Rejection
UV Response Up
Complement
DNA Repair
Oxidative Phosphorylation
Reactive Oxygen Species Pathway
Mitotic Spindle
mTORC1 Signaling
Unfolded Protein Response
Myc Targets V2
Myc Targets V1
G2-M Checkpoint
E2F Targets
NES
% Genes in set
0.20
0.40
0.60
c
E2F Targets
G2-M Checkpoint
Myc Targets V1
Myc Targets V2
Enrichment Score
Ranked metric
Gene Rank
d
Cluster Types in Each Sample
Percentage (%)
ER_pre
ER_post
SR_pre
SR_post
LR1_pre
LR1_post
LR2_pre
LR2_post
Cluster Types
Tumor cell
T cell
Follicular dendritic cell
B cell
Macrophage
Plasmablast
Fibroblast
Endothelial cell
Smooth muscle cell
Normal epithelial cell
Blood cell
e
-log P-value
Log2 Fold Change
COL1A1
COL1A2
COL3A1
APOE
COL6A1
MGP
LTBP2
C1QA
IGKC
IGHG1
SPARC
POSTN
COL6A2
CTSB
CD163
C1QB
C1QC
MMP2
MMP11
MMP7
NPR3
COL12A1
CXCL12
SFRP4
A2M
CXCL8
LRRC15
STAB1
COL11A1
IGFBP5
SFTPB
COL5A3
PSAP
CD74
ELN
NKD2
F13A1
PTGDS
COL5A2
COL8A1
PRELP
C5orf46
SOD2
MT2A
SERPINH1
LBH
f
Collagen Fibril Organization (GO:0030199)
Enrichment Score
NES: -2.166
Pval: 0.000e+00
FDR: 2.017e-03
Ranked metric
Pos
Zero score at 808
Neg
Gene Rank
g
Fibroblast: COL1A1
Counts
ER_pre
ER_post
SR_pre
SR_post
LR1_pre
LR1_post
LR2_pre
LR2_post
h
CD3D
CD4
CD8A
FOXP3
i
SPP1
C1QC
CD163
MSR1


a
-log P-value
Log2 Fold Change
MUCL1
WSB1
S100A9
CXCL2
CEACAM5
CLPTM1L
LAMA5
BIRC3
COL1A1
CLU
GEM
ATP1B1
CXCL8
IDO1
ARRDC3
CYP24A1
IGHG1
FOSB
CTSE
LPCAT1
BPIFA1
DHCR24
GDF15
AQP3
PCSK1N
CXCL1
FOS
CRABP2
SYT8
AHCYL2
CPM
b
MSigDB_Hallmark_2020
TNF-alpha Signaling via NF-kB
Inflammatory Response
Allograft Rejection
IL-6/JAK/STAT3 Signaling
Interferon Gamma Response
Complement
Coagulation
Interferon Alpha Response
Cholesterol Homeostasis
IL-2/STAT5 Signaling
KRAS Signaling Up
mTORC1 Signaling
Estrogen Response Late
E2F Targets
Glycolysis
Spermatogenesis
KRAS Signaling Dn
NES
% Genes in set
0.20
0.40
$\log_{10}\frac{1}{Pval}$
c
TNF-alpha Signaling via NF-kB
Inflammatory Response
IL-6/JAK/STAT3 Signaling
Enrichment Score
Ranked metric
Gene Rank
d
tc_S100A6
tc_AQP1
tc_TMSB4X
tc_MMP7
tc_SCGB3A1
tc_AZGP1
Mean expression in group
e
Tumor cell
UMAP1
UMAP2
tc_S100A6
tc_AQP1
tc_TMSB4X
tc_MMP7
tc_SCGB3A1
tc_AZGP1
Goblet/Ciliated cell
f
ER_pre
ER_post
SR_pre
SR_post
LR1_pre
LR1_post
LR2_pre
LR2_post
g
tc_S100A6
tc_AQP1
tc_TMSB4X
tc_MMP7
tc_SCGB3A1
tc_AZGP1


a
-log P-value
Log2 Fold Change
SCGB3A2
SCGB3A1
AQP4
S100B
CLU
COL1A2
MMP2
SFTPD
LGALSL
SELENOP
PGC
ANTXR1
CCN2
SUSD2
C1QB
CYBRD1
S100A8
ANGPTL4
SYT12
VEGFA
B4GALNT4
b
MSigDB_Hallmark_2020
E2F Targets
G2-M Checkpoint
Mitotic Spindle
p53 Pathway
Pperoxisome
Apoptosis
Estrogen Response Late
UV Response Dn
Pancreas Beta Cells
Myogenesis
IL-6/JAK/STAT3 Signaling
KRAS Signaling Up
Complement
Allograft Rejection
Coagulation
Inflammatory Response
Epithelial Mesenchymal Transition
NES
% Genes in set
0.30
0.40
0.50
c
E2F Targets
G2-M Checkpoint
Epithelial Mesenchymal Transition
Inflammatory Response
KRAS Signaling Up
Enrichment Score
Ranked metric
Gene Rank
d
-log P-value
Log2 Fold Change
COL1A2
C1QA
MET
CAV1
PLOD2
C1QB
AQP1
C1QC
SPARC
NNMT
CD163
LAPTM5
F13A1
FN1
CAV2
EMP3
TSPAN8
LYZ
SCGB3A1
LTF
S100A8
e
MSigDB_Hallmark_2020
Epithelial Mesenchymal Transition
Myc Targets V1
Complement
Allograft Rejection
mTORC1 Signaling
Interferon Gamma Response
IL-6/JAK/STAT3 Signaling
Coagulation
Inflammatory Response
KRAS Signaling Up
Oxidative Phosphorylation
E2F Targets
G2-M Checkpoint
Angiogenesis
Apoptosis
Glycolysis
DNA Repair
NES
% Genes in set
0.40
0.60
f
Epithelial Mesenchymal Transition
Myc Targets V1
mTORC1 Signaling
IL-6/JAK/STAT3 Signaling
Inflammatory Response
Enrichment Score
Ranked metric
Gene Rank
g
-log P-value
Log2 Fold Change
S100A8
SCGB3A2
CACNA2D2
SFTPD
PGC
SUSD2
CYP4B1
BIRC3
MMP7
MUC16
IGFBP3
BPIFA1
S100A9
TNS4
COL7A1
SLC7A5
PLOD2
TREM1
LRRK2
LAMP3
AQP4
RNASE1
HOPX
KCNJ15
PLXNA2
CTSH
h
MSigDB_Hallmark_2020
Bile Acid Metabolism
heme Metabolism
Hypoxia
TNF-alpha Signaling via NF-kB
Unfolded Protein Response
Complement
Mitotic Spindle
Myc Targets V2
mTORC1 Signaling
Allograft Rejection
KRAS Signaling Up
Glycolysis
Interferon Gamma Response
Epithelial Mesenchymal Transition
Myc Targets V1
E2F Targets
G2-M Checkpoint
NES
% Genes in set
0.30
0.40
0.50
i
G2-M Checkpoint
E2F Targets
Myc Targets V1
Enrichment Score
Ranked metric
Gene Rank
j
-log P-value
Log2 Fold Change
MMP13
CXCL10
BIRC3
MMP7
DMBT1
AQP1
RHOBTB2
SCGB3A2
RAP1GAP
C16orf89
CTHRC1
S100P
COL7A1
COL4A5
TNS4
ITGAX
POF1B
DCUN1D5
CXCL8
APOL1
ROBO1
k
MSigDB_Hallmark_2020
Epithelial Mesenchymal Transition
Interferon Gamma Response
IL-6/JAK/STAT3 Signaling
Complement
Interferon Alpha Response
Allograft Rejection
KRAS Signaling Up
Inflammatory Response
Coagulation
TNF-alpha Signaling via NF-kB
IL-2/STAT5 Signaling
Wnt-beta Catenin Signaling
NES
% Genes in set
0.20
0.40
0.60
l
Epithelial Mesenchymal Transition
IL-6/JAK/STAT3 Signaling
KRAS Signaling Up
Inflammatory Response
TNF-alpha Signaling via NF-kB
Enrichment Score
Ranked metric
Gene Rank


a
Neighborhood enrichment
ER_pre
SR_pre
LR1_pre
LR2_pre
ER_post
SR_post
LR1_post
LR2_post
Tumor cell
T cell
Follicular dendritic cell
B cell
Macrophage
Plasmablast
Fibroblast
Endothelial cell
Smooth muscle cell
Normal epithelial cell
Blood cell
Unified Z-score
8
6
4
2
0
−2
−4
b
H&E
Cell type
CXCL13
500 µm
100 µm
Gene Expression
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
c
H&E
Cell type
CXCL13
500 µm
200 µm
Gene Expression
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Tumor cell
alpha beta T cell
gamma delta T cell
Follicular dendritic cell
Naive B cell
Activated B cell
Macrophage
Plasmablast IgM
Plasmablast IgG1
Plasmablast IgG3
Plasmablast IgA
Fibroblast
Endothelial cell
Smooth muscle cell
Basal cell
Club cell
Goblet/Ciliated cell
Alveolar cell type 2
Erythrocyte
Granulocyte


a
MIF_score
Scores
ER_pre
ER_post
SR_pre
SR_post
LR1_pre
LR1_post
LR2_pre
LR2_post
b
MIF
CD74
CXCR4
CD44
c
d
tc_S100A6
tc_AQP1
tc_TMSB4X
tc_MMP7
tc_SCGB3A1
tc_AZGP1
T cell
Follicular dendritic cell
B cell
Macrophage
Plasmablast
Fibroblast
Endothelial cell
e
ER_pre
LR1_pre
f
ER_pre: MIF pathway
Sender
Reciever
-log P-value
Amount of signal
g
LR1_pre: MIF pathway
h
i
j
ER_pre
k
ER_pre: PDGF pathway
l
LR1_pre: PDGF pathway

Early Resistance Mechanisms to EGFR-TKIs in the Tumor Microenvironment